Mining the proteome: the application of tandem mass spectrometry to endocrine cancer research

Neil Sharma1, Ashley Martin2 and Christopher J McCabe1

1School of Clinical and Experimental Medicine, Institute for Biomedical Research and 2School of Cancer Sciences, University of Birmingham, Birmingham B15 2TT, UK

(Correspondence should be addressed to C J McCabe; Email: mccabcjz@bham.ac.uk)

Abstract

Tandem mass spectrometry (MS/MS) permits the detection of femtomolar quantities of protein from a wide variety of tissue sources. As endocrine cancers are frequently aetiologically complex, they are particularly amenable to mass spectrometry. The most widely studied aspect is the search for novel reliable biomarkers that would allow cancers to be diagnosed earlier and distinguished from benign tumours. MS/MS allows for the rapid analysis of blood and urine in addition to tumour tissue, and in this regard it has been applied on research involving thyroid, pancreatic, adrenal and ovarian cancers with varying degrees of success, as well as additional organ sites including breast and lung. The description of an individual cancer proteome potentially allows for personalised management of each patient, avoiding unnecessary therapies and targeting treatments to those which will have the most effect. The application of MS/MS to interaction proteomics is a field that has generated recent novel targets for chemotherapy. However, the technology involved in MS/MS has a number of drawbacks that at present prevent its widespread use in translational cancer research, including a poor reproducibility of results, in part due to the large amount of data generated and the inability to accurately differentiate true from false-positive results. Further, the current cost of running MS/MS restricts the number of times the experiments can be repeated, contributing to the lack of significance and concordance between studies. Despite these problems, however, MS/MS is emerging as a front line tool in endocrine cancer research and it is likely that this will continue over the next decade.

Endocrine-Related Cancer (2012) 19 R149–R161

Introduction

Cellular function is determined by the interaction of proteins within and around the cell, and mapping the range of these interactions (the ‘interactome’ (Sanchez et al. 1999)) is increasingly being recognised as a way of identifying novel therapeutic targets. It is for this reason that proteomics has become one of the most rapidly advancing fields in cancer research. Studying the expression and interactions of proteins in both normal and malignant tissues significantly enhances the understanding of the mechanisms of disease progression and pathogenesis. In addition, the difference in expression of proteins within cancer tissues and in the circulation can lead to the identification of biomarkers that may help in the diagnosis of malignancy. Following from this, if an individual patient’s cancer proteome is determined, treatment specific to his/her cancer type may be feasible, potentially reducing unnecessary treatments with the incumbent side effects and costs. Finally, by mapping the human interactome, new therapeutic targets may be identified, which can lead to the development of novel cancer treatments.

In order to progress with proteomic research, it is critical that accurate methods exist to determine protein expression per se, as well as the presence and quantitation of post-translational modifications (PTM) and protein interactions with other proteins, DNA, lipids and carbohydrates. These requirements have led to the development of specific sample preparation...
protocols, whilst the actual analytical process is often (but not exclusively) based on the same approach, digestion of the protein sample into peptides using a protease followed by a mass spectrometric analysis of HPLC-separated peptides. Sequencing the human genome (Lander et al. 2001, Venter et al. 2001) meant that for each gene, the exact amino acid chain could be calculated, and from this, all possible peptide chains caused by cleavage at specific sites could be generated. Coupling this with the improved computer power over the last decade has, what was once a laborious and time-consuming task, become increasingly quick and straightforward, opening up enormous possibilities within the field of endocrine cancer research (Hanash & Taguchi 2010). Overall, the development of these modern proteomic approaches in combination with sophisticated database searches now permits the reliable identification of proteins in a range of samples types and complexity levels.

**General principles**

The principle behind protein identification by mass spectrometry lies in the fact that every protein has a predictable fragmentation pattern (peptide fingerprint) when cleaved by a digestive enzyme (Steen & Mann 2004). The most commonly used enzyme is trypsin, which cleaves peptide chains at the carboxyl side of lysine and arginine residues. An efficient, reproducible digestion is essential to generate worthwhile data, and currently, the most widely used method is the FASP protocol from Mann’s group (Wisniewski et al. 2009). In this approach, proteins are solubilised using 4% SDS, a very effective solubilising agent, which is removed by centrifuging the sample through a 35 kDa cut-off filter where the unfolded proteins are retained and the detergent removed in the filtrate. After concentration and purification, samples are digested using endoproteinase Lys C and/or trypsin and the peptides are harvested for analysis. The use of strong denaturants ensures that the proteins are unfolded allowing the proteases to gain access to cleavage sites producing a relatively complete and therefore reproducible digest. Methods that use different detergent removal strategies after, rather than before, proteolytic digestion have recently demonstrated improved peptide coverage (Bereman et al. 2011). Another very common approach is the digestion of proteins in gel slices following separation of the proteins using SDS–PAGE. This process is also reliable, as once again the protein substrates are denatured and unfolded.

Tryptic digestion will often generate very complex peptide mixtures that are usually analysed by liquid chromatography (LC)–MS/MS, where the peptides are separated using reverse phase HPLC before mass spectrometry. Typically, the peptides are loaded onto a ‘trap-column’ in a solvent that contains 2–4% acetonitrile and an acid that acts as an ion pair. The peptides are eluted using a gradient of acetonitrile where the rate of the gradient is matched to the sample complexity. There has been considerable improvement in the design and performance of HPLC instrumentation and columns, allowing routine implementation of nanobore (sub μm/min flow rate) methods using columns with internal diameters of 50–100 μm. Therefore, peptides elute from these columns in a very small volume, meaning that the actual concentration of peptides in the HPLC eluate is relatively high (Steen & Mann 2004).

The mass spectrometric analysis of the HPLC-resolved peptides typically uses two basic approaches: electrospray ionisation (ESI) and matrix-assisted laser desorption/ionisation with a time of flight mass spectrometer (MALDI ToF/ToF). In the case of a MALDI instrument, a defined period of HPLC column outflow is automatically spotted onto a MALDI target plate. Typically, 700 or more spots are generated for a complex sample. In ESI, the column eluate is sprayed into the mass spectrometer source via an emitter (a fine needle that can conduct an electrical current), which is spatially isolated from the actual mass spectrometer. A high charge (1–5 kV) is applied to the emitter that ionises the peptides in the acidified HPLC solvent. As this liquid exits the fine tip, small droplets are generated and manipulated by electric fields according to their mass to charge ratio (m/z). The ions reach a detector and a mass spectrum is generated. By examining the separation pattern of the peaks produced, the specific m/z of each ion can be calculated. Tandem mass spectrometry (MS/MS) refers to the further fragmentation of the most abundant precursor ions using an inert gas. These are again passed through the spectrometer and the individual product ions assigned further peaks based on their m/z. Each product ion is the result of the loss of one or more amino acid residues from the precursor ion; by examining the population of peaks and their m/z (the peptide ‘tandem mass spectrum’), it is therefore possible to determine the amino acid sequence of the ion, and therefore the structure of the precursor ion. Once the mass spectrum has been matched to a peptide sequence, computer databases use the predictable cleavage sites of trypsin to match the peptide to a parent protein.
Sample generation protocols

The varied design and capabilities of the commercially available mass spectrometers and the associated HPLCs and columns mean that a wide range of sample types and complexities can be analysed, although care has to be taken to ensure that the protocols do not bias against a specific group of proteins. For example, large, small, acidic, basic or hydrophobic proteins can be lost using inappropriate methods.

Protein complexes

The isolation of proteins using affinity purification of a target or bait protein has allowed the detection of novel protein–protein interactions in many studies. Antibodies to native or tagged proteins have been used extensively in this work, and efforts to decrease the detection of non-specific binding partners have used tandem affinity tags (TAP-TAG; Rigaut et al. 1999). However the protein complex is isolated, it is usual that the proteins are identified following resolution of the proteins using SDS–PAGE, band excision and trypic digestion. It is important to perform suitable controls to detect non-specific interactions, and it is particularly useful if some interactions with the target protein are already known as these can act as positive controls, as does the detection of the target; as well as individual protein–protein interactions, the duration of these interactions and their relation to cellular function can also be determined, mapping out cell signalling pathways (Bisson et al. 2011).

Post-translational modifications

The development of group-specific peptide/protein purification protocols for the study of PTM has increased the utility of proteomics in many areas of research.

Glycosylation, which regulates a range of protein functions (localisation, secretion and enzyme activity) and plays critical roles in cell signalling and adhesion, can be quantified following chemical enrichment or isotopic labelling. Assessment of phosphorylation, the most common PTM, has improved considerably since the development of reliable phosphopeptide enrichment protocols that use binding to metal ions (Fe$^{3+}$ and Ga$^{3+}$, termed immobilised metal ion affinity chromatography) (Jensen & Larsen 2007) or metal dioxides (titanium or zirconium, termed metal oxide affinity chromatography) (Larsen et al. 2005). Phosphopeptide purification using these protocols is not complete, with acidic peptides (containing glutamic and/or aspartic acid) being a particular problem. Approaches to increase the specificity of the purification include the use of agents (2,5-dihydroxybenzoic acid or glycolic acid) that compete with the low affinity acidic peptide binding. Tyrosine-phosphorylated peptides/proteins are considerably rarer, but as reliable antibodies exist for this species (but not phospho-serine or threonine), immunofluorescence enrichment has been used (Blagoev et al. 2004, Schumacher et al. 2007). The development of various fragmentation approaches (electron-capture dissociation on Fourier transform ion cyclotron resonance instruments and electron-transfer dissociation on ion traps, orbitraps and ESI ToF instruments, has helped to improve the sensitivity of phosphoproteomics (Boersema et al. 2009).

Ubiquitin and the small ubiquitin-like modifier (SUMO) are highly conserved PTM proteins that regulate the half-life, intracellular location and activity of modified proteins. Ubiquitinated and SUMOylated proteins can also be identified and quantified by MS/MS, although the variation in SUMO isoform and ubiquitin number can make detection complex. This complexity and the fact that conventional database search engines can only allow for modifications with a fixed mass means that a new data analysis strategy was required. Here, the presence of predicted characteristic fragment ion patterns in MS/MS spectra based on the sequence of the SUMO isoform under investigation is used to detect SUMOylated proteins (SUMmOn; Jeram et al. 2010).

Protein quantitation

Initially, the analysis of complex samples (often termed ‘shot-gun proteomics’) focused on identifying as many peptides/proteins as possible in complex samples such as blood serum/plasma, urine, cerebrospinal fluid (CSF) or tissue/cell lysates. However, it soon became clear that it is more relevant to determine either the absolute or the relative abundance of proteins, particularly in biomarker discovery.

Several approaches have been developed, and the actual protocol used is usually limited by the sample type, cost, mass spectrometer available and the quality of data required. All the methods have advantages and disadvantages compared with other protocols, and the choice of approach can be affected by the experience and personal preferences of the group performing the analyses.

Currently, the most common approaches use stable isotope-labelling strategies such that two (or more) isotopically labelled versions of each peptide are produced (Elliott et al. 2009). These molecules are chemically identical and behave exactly in the same
way during purification steps but vary by a known mass. Therefore, as they are analysed under identical conditions, differences in the peak intensities of the isotopic versions represent an accurate estimation of the relative abundance of the isotopic species.

The vast majority of quantitative proteomic methods attempt to determine the relative amount of large numbers of different peptides/proteins in complex samples without any prior knowledge of the proteins under investigation. However, a focused analysis of specific proteins can be performed if the protein of interest is known and if preliminary analysis can determine which unique tryptic peptides are consistently observed in LC–MS/MS as digests of the target protein. Then, an isotopically labelled version of the peptide is synthesised and a known amount is spiked into the samples. When analysed by LC–MS/MS, the natural and isotopic versions co-elute, meaning that their relative MS spectra intensities can be used to calculate the concentration of endogenous peptide in the original sample (Gerber et al. 2003, Elliott et al. 2009). More recently, the approach has been improved to a large extent because of improvements in the performance of ‘triple quad’ mass spectrometers. These instruments are capable of selecting a specific narrow m/z range of ions in their first quadrupole, which is then fragmented by CID in the second. The third quadrupole selects for a specific m/z (of a dominant fragment ion) to analyse, which generates a ‘cleaner’ spectrum since the parent ion was isolated before fragmentation. This process is termed selected reaction monitoring (SRM). Due to the potential to select for specific m/z values, the sensitivity of SRM is considerably greater than typical LC–MS/MS. This facet can be used not only to quantify potential biomarkers but also to identify the gene products of specific mutations that are known to be tumorigenic (Wang et al. 2011). The major difficulty with this approach is the amount of work required to determine which peptide and fragment ions (often termed an SRM transition) are the most appropriate to use. Many peptides can be quantified in a single run (multiple reaction monitoring, or MRM) as it is possible to programme the mass spectrometer to detect/select several m/z values and to change the ions under study with time; the simultaneous quantification of 45 different proteins in plasma using MRM has been reported (Kuzyk et al. 2009). These techniques are summarised in Fig. 1.

Stable isotope labelling by amino acid in culture (SILAC) is a method of isotopically labelling proteins metabolically, where two (or more) cell types, or the same cell type under different conditions, are grown in either ‘heavy’ or ‘light’ media (Ong et al. 2002). In the ‘heavy’ media, either the naturally occurring 13C-arginine or the 12C-lysine is replaced by 13C versions. When the cells grow and use the amino acids from the media to make new proteins, the cells grown in the ‘heavy’ media isotopically label their own proteins, and if sufficient rounds of cell division are allowed, the labelling will approach equilibrium. An attempt to extend the SILAC approach for the quantitation of proteins in tissue samples has been reported (Geiger et al. 2010). This ‘super SILAC’ approach is relatively expensive as several cell lines have to be grown in 13C-labelled amino acid-containing media.

In order to quantify proteins in tissue/clinical samples (including body fluids) using stable isotopes, it is much more usual to chemically modify proteins/peptides using ‘light’ or ‘heavy’ reagents. The reagents target reactive groups such as thiols or amines, of which the latter is the most widely used as every peptide contains an N-terminus. The SILAC and peptide derivatisation methods have the significant disadvantage that the sample complexity doubles when a different isotopic version of each peptide is produced. The alternative strategy of using isobaric labelling reagents that fragment during MS/MS analysis to produce isotopically labelled reporter groups has been developed and is now the most commonly used peptide derivatisation strategy in use (Ross et al. 2004). This method, termed isobaric tags for relative and absolute quantification (iTRAQ), was originally developed as a four-plex approach and has now been expanded to an eight-plex system. As the tags are isobaric, any given derivatised peptide from each sample has the same mass and is therefore isolated for MS/MS analysis together. Upon dissociation, the reporter groups (with different numbers of ‘heavy’ atoms) are released and detected in the MS/MS spectrum. There is no increase in the sample complexity with this approach, and the proteins are quantified by comparing the relative intensity of the labelled reporter ion fragments. Disadvantages with iTRAQ are that the reagents are relatively expensive, considerable sample manipulation takes place before labelling/mixing (making reproducibility difficult) and the mass of reporter ions is below the mass range of some commonly used mass spectrometers (e.g. 150 Da for ion traps).

### Tandem mass spectrometry and endocrine cancer

As MS/MS is more widely used in the study of the molecular basis of cancer, techniques to improve accuracy and combat technical difficulties are
There is an increasing body of evidence demonstrating that MS/MS has important and wide range of applications in endocrine cancer, including identifying biomarkers, facilitating personalised medicine and also in developing novel therapies. Endocrine cancers are often difficult to distinguish from benign tumours of endocrine organs, and in some cases present late that results in a poor prognosis. In addition, there is often a complex interplay between the ongoing normal physiology of the cell, aberrant production of secreted hormones and disordered cellular function. Together, this results in endocrine cancer research being particularly amenable to MS/MS.

**Identification of biomarkers**

One of the most significant areas of cancer biology in which MS/MS has had an impact is the identification of biomarkers. These are used in pre-diagnosis screening as well as prognostication of cancers and in identification of recurrence. Standard molecular biological techniques such as immunohistochemistry and quantitative PCR remain in routine use to identify genes that may serve to stratify a tumour into a particular sub-group. However, using MS/MS, plasma and urine can be analysed as well as tumour tissue, enabling the detection of circulating proteins that have been either secreted by the cancer or have ‘leaked’ out.

---

![Figure 1](image-url)

**Figure 1** The process of triple quadrupole mass spectrometry. Two quadrupole mass spectrometers are in series (Q1 and Q3), with a third non-mass filtering quadrupole in-between (Q2). Q1 selects an ion of a specific mass, which is then fragmented in Q2 by collision with inert gases (e.g. argon, helium and nitrogen). In standard MS/MS, all fragments are analysed. In selected and multiple reaction monitoring, only one, or a specific number, of the fragment ions is analysed, allowing for much quicker sample processing.
BRAF and development. Genetic mutations such as 

tages, and there is a continued need for further 

calcitonin for medullary cancer, have their disadvan-

thyroglobulin for differentiated thyroid cancer and 

The two most commonly used serum markers, 

available that predict cancer within a thyroid nodule. 

malignancy, there remains a lack of reliable markers 

Despite thyroid cancer being the commonest endocrine 

Mass spectrometry in thyroid cancer 

Despite thyroid cancer being the commonest endocrine 
malignancy, there remains a lack of reliable markers 
available that predict cancer within a thyroid nodule. 
The two most commonly used serum markers, 
thyroglobulin for differentiated thyroid cancer and 
calcitonin for medullary cancer, have their disadvantages, 
and there is a continued need for further development. Genetic mutations such as RET/PTC, 
BRAF and RAS have been identified in thyroid cancer 
specimens (Namba et al. 2003, Carta et al. 2006, 
Stanojevic et al. 2011), but these again require a tissue 
 sample in order to identify them. While fine needle 
aspiration cytology (FNAC) can be useful for this, the 
amount of material obtained is often minimal and, even 
with ultrasound guidance, not necessarily fully representative of the tumour. As an alternative 
approach, a number of studies over the past decade 
have examined the proteome of thyroid samples. Initially, 2D electrophoresis was used to identify 
protein spots from normal, goitrous, hyperplastic, adenomatous and malignant thyroid specimens 
followed by sequencing from PVDF membranes and 
MS/MS to determine which proteins were most abundant (Srisomsap et al. 2002). Cathepsin B was 
reportedly over-expressed in thyroid cancer, and 
ATP synthase and prohibitin were specifically over-
expressed in the papillary variant, proteins that had not 
been previously linked to thyroid cancer. Although this 
first study used both MS/MS and traditional methods, 
as MS/MS technology has advanced, the need to run 
synchronous techniques has diminished. Later, 
proteins over-expressed in 29 thyroid tumour samples 
when compared with normal thyroid were identified 
using SELDI-TOF MS/MS (Sofiadis et al. 2010). 

When the spectra produced from nuclear and cytosolic 
fractons of cell lysates were analysed, the 
Ca$^{2+}$ binding ion S100A6 was over-expressed in 
PTC compared with FTC and normal thyroid, which 
was confirmed by western blotting and immuno-

histochemistry (Sofiadis et al. 2010). Additionally, 
2-hydroxyglutarate levels have been shown to be 
elevated in PTC compared with both normal thyroid 
and hyperplastic nodules (without the IDH1 and IDH2 
mutations usually seen in gliomas and myeloid 
leukaemias), although these results are yet to be 
validated (Rakheja et al. 2011). 

Characterisation of a thyroid tumour, then, may 
prove useful in further stratifying tumour tissue post-
biopsy/surgery. However, it is the high-throughput 
analysis of serum and urine for secreted proteins that 
may serve as biomarkers that highlight the potential of 
MS/MS as a vital tool in cancer proteomics. Two 
recent studies have taken different approaches in the 
identification of the thyroid cancer secretome. The first 
used an in vitro model of TPC-1 (PTC) and CAL62 
(anaplastic thyroid cancer (ATC)) cell lines (Kashat 
et al. 2010). After culturing cells in serum-free 
conditions, media was collected, trypsinised and run 
through LC–MS/MS. Proteins were classed as high 
confidence if identified by two or more peptides with 
≥95% confidence. Forty-six proteins were shortlisted; 
the mean number of peptides identified was 4.82 with 
average sequence coverage of 12.5%. After further 
validation, six proteins were determined to be present 
in the thyroid cancer secretome: biotinidase, nucleolin, 
enolase1, PTMA, CYR61 and clusterin. The expression of these proteins in the serum of patients 
with thyroid cancer was then examined, but unfortunately, none of the proteins were proven to be over-
expressed in this setting. This inability to translate 
results from an in vitro to an in vivo setting is a 
shortcoming observed in a number of MS/MS studies. 
However, identifying those proteins selectively 
secreted by thyroid tumours remains a fundamental 
 imperative. Using human serum directly to identify 
secreted proteins has the advantage of being applicable 
to a large number of patients and also being 
representative of the in vivo setting. SELDI-TOF 
MS/MS was used to study the serum protein profile 
of 224 patients, 108 with PTC and 116 normal controls 
(Fan et al. 2009). Haptoglobin alpha-1 was found to be 
up-regulated in PTC while apolipoproteins CI and CII 
were down-regulated. Further, as the stage of tumour 
increased, there was a corresponding rise in haptoglo-
bin alpha-1 and fall in the apolipoproteins. Further 
work is now needed to consolidate these observations 
and to determine whether these proteins may be of use
in the initial investigation of patients with suspected thyroid cancer. To differentiate PTCs from both nodular goitres and normal thyroids, a metabolomic approach comparing the three groups found that the greatest difference occurred in lipid metabolism, with 3-hydroxybutyric acid expression much higher in the PTC group compared with the others (Yao et al. 2011). The group suggests that this may therefore be a potential marker to distinguish between PTC and benign nodular lesions.

**Mass spectrometry and pancreatic tumours**

While thyroid cancer generally has a good prognosis, pancreatic ductal adenocarcinoma has a very low 5-year survival rate, which is in part due to the fact that most tumours are discovered late (Borgida et al. 2011). As such, a serum biomarker may be a useful tool in screening for the disease, with candidate proteins being identified by MS/MS. The serum of patients with pancreatic cancer compared with control patients without pancreatic cancer has been assessed (Rong et al. 2010). Using MALDI-TOF MS/MS and linear ion trap-coupled MS, two proteins were isolated: mannose binding lectin 2 and myosin light chain kinase 2, both were up-regulated in the serum from patients with pancreatic adenocarcinoma. These results were confirmed by western blotting, suggesting that these markers may be useful adjuncts to the existing biomarkers, CA19-9 and CEA. Despite using similar methods, another study reported the presence of a peptide chain from sulphhydryl oxidase 1 as their only significant finding and linked this to the expression of protein in pancreatic tumour cell lines (Antwi et al. 2009). These studies highlight the fact that proteomic research using MS/MS can yield very variable results even though the methodology may be similar. To reduce the error rate and the potential for missing low concentration biomarkers in *in vitro* studies, the use of SILAC with subsequent multidimensional separation and MS/MS has been suggested (Yu et al. 2009). MMP7, BCAM and ICAM1 were identified, although the former was not validated by examining the difference between pooled cancer and control serum samples and was therefore discounted. Both BCAM and ICAM1 were then examined in an independent set of ten cancer and ten control serum samples, both were significantly elevated in the cancer group compared with controls. This work suggested that SILAC is therefore useful for the detection of those proteins that would otherwise be over-shadowed by the more abundant serum proteins such as albumin and immunoglobulins, and this strategy has been adopted more generally (Brewis & Brennan 2010). When the proteome of pancreatic juice was investigated, again using 2D gel electrophoresis and MALDI-TOF MS/MS, MMP9, DJ1 and A1BG were elevated in cancer compared with control, suggesting that these may be helpful in differentiating between pancreatic ductal carcinoma and chronic pancreatitis (Tian et al. 2008). However, a similar study (Park et al. 2011) that compared the pancreatic juice following secretin stimulation in patients with pancreatic cancer, chronic pancreatitis and no disease found that BIG2, PRDX6 and REG1z were over-expressed in cancer and confirmed this with immunohistochemical analysis of cancer tissue. This highlights the potential variability between similar studies and, given the way in which proteins are identified, the potential inaccuracy in reporting. It is therefore critical that the published results of MS/MS analyses contain as much relevant information as possible for others to accurately compare and appraise their work.

**Mass spectrometry and ovarian cancer**

Previous studies did not report the confirmation of known biomarkers of pancreatic cancer in their results, which may indicate that a number of proteins and potentially strong candidates are being missed by these tests. However, examining the proteome of ascitic fluid in patients with ovarian cancer to identify potential new biomarkers has been partially successful in this regard (Kuk et al. 2009). When analysing the data, all known high-abundance serum proteins were disregarded as well as those identified by only one peptide or in one sample. Twenty-five proteins were found, which had previously been associated with ovarian cancer (from which they implied validation of their methodology), and a further 52 proteins will be the subject of validation experiments in the future. Unfortunately, CA125, one of the most widely used biomarkers of ovarian cancer (Schorge et al. 2010), was not identified. However, this may reflect the fact that CA125 is both a high-molecular weight molecule and highly glycosylated, and its mass would therefore fall outside the analysed range. These factors, coupled with the fact that CA125 is present in low concentrations even in ovarian cancer, may explain its omission.

**Mass spectrometry and adrenal cancer**

The above studies have attempted to find biomarkers that may be able to confirm either the presence of disease or the recurrence of cancer after treatment. None have specifically sought to develop a novel
screening test, either because at present it would not be practical to screen the population for a particular low-incidence cancer or because valid biomarkers already exist. Despite the wealth of potential biomarkers available, only a small number make it to regular clinical use in screening due to a lack of sensitivity or specificity. One area in which a novel screening approach would be of benefit is in phaeochromocytoma and paraganglioma. High-risk patients (e.g. MEN2, von Hippel–Lindau disease and neurofibromatosis type 1) currently undergo urinary/plasma testing for metanephrine and normetanephrine (Adler et al. 2008). MS/MS has been used over the past few years to quantify the urinary and plasma concentrations of these catecholamines with positive results (de Jong et al. 2007, Perry et al. 2007). More recently, a technique using LC–MS/MS, which does not suffer from drug interference, has been developed which is now in routine use (Marrington et al. 2010), demonstrating that besides being a sophisticated laboratory research tool, MS/MS also has a place in the diagnostics of cancer patients. A summary of potential biomarkers is shown in Table 1.

**Personalised medicine**

The search for biomarkers of endocrine cancer does not seek to find a way of tailoring treatment, rather identifying novel proteins that can then be studied in a more translational context. When treating patients with cancer, future therapies are likely to evolve away from protocol-driven management strategies to a more tailored, individual approach (Pirmohamed 2010). In this respect, MS/MS can be used to identify proteomic

<table>
<thead>
<tr>
<th>Source</th>
<th>Biomarker</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thyroid</td>
<td>2-Hydroxyglutarate</td>
<td>Rakheja et al. (2011)</td>
</tr>
<tr>
<td>Thyroid tissue (PTC/normal/hyperplastic nodule)</td>
<td>3-Hydroxybutyric acid</td>
<td>Yao et al. (2011)</td>
</tr>
<tr>
<td>Serum (PTC/nodular goitre/normal)</td>
<td>PTMA</td>
<td>Kashat et al. (2010)</td>
</tr>
<tr>
<td>TPC-1 cells (human PTC cell line)</td>
<td>Nucleolin</td>
<td></td>
</tr>
<tr>
<td>CAL62 cells (human anaplastic cell line)</td>
<td>Biotinidase</td>
<td></td>
</tr>
<tr>
<td>Thyroid tissue (PTC/FTC/normal)</td>
<td>Ca$^{2+}$-binding ion</td>
<td>Sofiadis et al. (2010)</td>
</tr>
<tr>
<td>Serum (PTC/normal)</td>
<td>Haptoglobin alpha-1</td>
<td>Fan et al. (2009)</td>
</tr>
<tr>
<td>Thyroid tissue (PTC/normal/nodular/hyperplastic)</td>
<td>Cathepsin B</td>
<td>Srisomsap et al. (2002)</td>
</tr>
<tr>
<td>Pancreas</td>
<td>ATP synthase</td>
<td></td>
</tr>
<tr>
<td>Pancreatic juice (ductal carcinoma/chronic pancreatitis/no disease)</td>
<td>Prohibitin</td>
<td></td>
</tr>
<tr>
<td>Serum (carcinoma/normal)</td>
<td>Mannose-binding lectin 2</td>
<td>Rong et al. (2010)</td>
</tr>
<tr>
<td>Serum (low-molecular weight fraction) (ductal carcinoma/normal)</td>
<td>Myosin light chain kinase 2</td>
<td></td>
</tr>
<tr>
<td>CAPAN2 cells (human pancreatic ductal adenocarcinoma cell line) with validation in serum from cancer patients</td>
<td>Sulphhydryl oxidase 1</td>
<td>Antwi et al. (2009)</td>
</tr>
<tr>
<td>Pancreatic juice (ductal carcinoma/chronic pancreatitis)</td>
<td>BCAM</td>
<td>Yu et al. (2009)</td>
</tr>
<tr>
<td>Adrenal</td>
<td>ICAM1</td>
<td></td>
</tr>
<tr>
<td>Urine (phaeochromocytoma/normal)</td>
<td>MMP9</td>
<td>Tian et al. (2008)</td>
</tr>
<tr>
<td></td>
<td>DJ1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A1BG</td>
<td></td>
</tr>
<tr>
<td>Ovarian</td>
<td>Metanephrine</td>
<td>Adler et al. (2008)</td>
</tr>
<tr>
<td>Ascitic fluid (advanced ovarian carcinoma)</td>
<td>Normetanephrine</td>
<td>Marrington et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>25 Previously identified biomarkers</td>
<td>Kuk et al. (2009)</td>
</tr>
<tr>
<td></td>
<td>52 Potential candidates</td>
<td></td>
</tr>
</tbody>
</table>
differences between different groups of cancer patients. While most ovarian cancers will be initially susceptible to platinum-based chemotherapy, many will eventually become resistant (Cho & Shih Ie 2009). Using ESI-MS/MS to examine the proteome of primary and post-chemotherapy ovarian cancers, 11 proteins were identified, which were expressed at higher levels in the recurrent group, all of which were confirmed by RT-PCR (Jinawath et al. 2010). By knocking down two of these (RELA (NF-κB p65) and STAT5), carboplatin-resistance in vitro was reversed. Most importantly, by treating the cells with either an NF-κB or a STAT5 inhibitor, their sensitivity to carboplatin was significantly enhanced. A similar study compared the proteome of paclitaxel-resistant and -sensitive cell lines, using both MALDI-TOF and LC–MS/MS (Di Michele et al. 2009). A number of proteins that showed significant differential expression between the two were demonstrated and grouped into those related to stress response, metabolism, protein biosynthesis and apoptosis. Collectively, these studies provide a possible way of determining whether a patient will benefit from a particular chemotherapeutic agent before its delivery, potentially avoiding both unnecessary side effects and cost.

Previous techniques have performed proteomic analyses on either serum or fresh tissue, but in some cases, the specimen may have already been paraffin embedded and examined by a pathologist. Several methods exist for carrying out HPLC–MS/MS on formalin-fixed and paraffin-embedded (FFPE) tissue, which are much more applicable to a hospital setting (Gustafsson et al. 2010, Paulo et al. 2011). In this way, archived samples could be revisited in certain cases or as new discoveries regarding treatment are made. This could then open up a new treatment for a patient, regardless of when the original diagnosis was made or surgery performed. As mentioned earlier, statistical analysis is often hampered by insufficient sample numbers for what are often rare or unusual malignancies and by the short follow-up time that the patients including in current studies have available. By using FFPE tissue, which has been archived and for which long-term patient outcomes are known, studies will be able to compare larger numbers of samples and correlate different proteomic measures to specific patient factors, such as recurrence and response to treatment.

Interaction proteomics
MS has enabled the large-scale identification of protein–protein interactions (Downard 2006, Volkel et al. 2010), and from these, large interaction networks have been generated. Tyrosine kinase inhibitors (TKIs) are promising agents in the treatment of a number of endocrine cancers (Raymond et al. 2011, Verbeek et al. 2011), and proteomic analysis of interactions can identify more accurately specific kinases to target (Gorla et al. 2009). Directly related to this, MS/MS allows the identification of post-translational protein modifications (Chung et al. 2011), including phosphorylation, which can lead to further therapeutic insight. However, as observed with CA125 earlier, if high-molecular weight proteins are examined, the presence of PTMs may obscure the data; analyses must therefore be tailored to specific events, such as phosphorylation or glycosylation. It has been demonstrated that resistance to the TKI lapatinib in breast cancer cell lines is linked to increased phosphorylation of Src family kinases (Rexer et al. 2011). Further, by treating these cells with Src inhibitors, the sensitivity to lapatinib is restored. Similarly, LC–MS/MS has been used to explore the phosphorylation of the epidermal growth factor receptor (EGFR; Zhang et al. 2011). Thirty phosphorylation sites were documented and were related to activating mutations in cancer cell lines as well as sensitivity to the EGFR inhibitor erlotinib. These strategies thus highlight the potential to target treatment for specific cancers based on the proteomic profile of the tumour tissue and to monitor and address disease progression in the light of the ongoing treatment regimen.

Conclusions
Proteomic studies on endocrine cancer are beginning to advance through the application of MS/MS. Improvements in information technology have enabled the processing of the vast amounts of data produced by the techniques into meaningful results. As with any emerging analytical tool, there are both advantages and disadvantages to its application, and these must be addressed before it fully finds its niche within the field.

One of the huge benefits MS/MS that realises is the ability to identify and quantitate peptides at femtomolar concentrations in an otherwise heterogeneous mixture. This allows the analysis of serum, urine and exocrine fluids for biomarkers. While MS/MS may not then be needed for the ultimate detection of these proteins in clinical practice, it is their identification that is facilitated and this in turn may have profound translational implications. However, the transition from traditional techniques has not been smooth and indeed is still ongoing. At present, reproducibility of results is poor, despite standard techniques being used.
In addition, the hunt for reliable biomarkers is hampered by the continued difficulty of identifying well-established biomarkers in large and heterogeneous datasets. The Clinical Proteomic Tumor Analysis Consortium (CPTAC), as part of the National Institutes of Health in the USA, is an example of a body set-up to address these problems (http://proteomics.cancer.gov/programs/cptacnetwork). They identified key barriers to the development of new biomarkers (either technological limitations or the very large numbers of candidate proteins that require clinical validation) and re-designed the process behind biomarker discovery. This involved the additional step of verifying candidate biomarkers after initial identification, using targeted quantitative assays capable of examining large numbers of samples, and therefore providing appropriate statistic power. Only those proteins that passed this step would progress on to clinical validation, thus ensuring that only those with the most potential would be examined. Using these methods, individual Protein Characterisation Centres feed data into a central hub, which allow the data to be used by scientific groups around the world, resulting in targeted accurate assays and statistical tests (Addona et al. 2009, Chang et al. 2011).

While technological advances and better availability have somewhat improved access to MS/MS, it remains an expensive and time-consuming technique. Unfortunately, the cost of processing a set of samples leads to many groups minimising the number. Given the problems mentioned earlier regarding large datasets and false-positive/negative results, this restriction inevitably contributes to the poor reproducibility of data. Concerns have been raised as to the accumulation of errors when datasets are combined, especially with respect to the ‘hidden cost’ of validating a large number of results (White 2011). However, others have described protocols for the targeting of the strongest candidates from these results, thus allowing verification of a smaller, more accurate set of proteins (Whiteaker et al. 2011).

Despite these shortcomings, the application of MS/MS on endocrine cancer research should be welcomed and encouraged. The publication of complete datasets, both online or as supplementary material, will enable cross-checking of results and stimulate collaboration between groups. Authors should ensure that a standard methodological approach is used, or where different that the full details are made available, as even small variations may result in differing endpoints. Greater care should be taken in the validation of results of MS/MS; at present, the technique does not generate a high enough degree of sensitivity or specificity to be used in isolation.

As the above difficulties are addressed and overcome, MS/MS will prove to be a valuable translational asset in the care of endocrine cancer patients. By analysing large numbers of tumour tissue samples, more biomarkers will be discovered that will help in screening for the disease. In addition, by elucidating the proteome of an individual cancer after diagnosis, correlations between treatment resistance and susceptibility can be determined, thereby reducing the burden of unnecessary therapies. In turn, new targets will become apparent as the pathogenesis of the disease is better understood. These aims are not realistic at present, given that it is not possible to analyse every patient with an endocrine cancer using MS/MS unless they are part of an ongoing research study. However, with continued funding and development, it is likely that the next decade will see the use of MS/MS becoming a routine investigative tool in translational research on endocrine cancer.

Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding
This work was supported by the Medical Research Council (Grant No. G1000359) and a Royal College of Surgeons of England/Get Ahead Charitable Trust Research Fellowship.

References


Brewis IA & Brennan P 2010 Proteomics technologies for the global identification and quantification of proteins. *Advances in Protein Chemistry and Structural Biology* 80 1–44. (doi:10.1016/B978-0-12-381264-3.00001-1)


Gustafsson JO, Oehler MK, McColl SR & Hoffmann P 2010 Citric acid antigen retrieval (CAAR) for tryptic peptide imaging directly on archived formalin-fixed paraffin-embedded tissue. *Journal of Proteome Research* 9 4315–4328. (doi:10.1021/pr9011766)


N Sharma et al.: Mining the proteome


Received in final form 26 April 2012
Accepted 2 May 2012
Made available online as an Accepted Preprint
3 May 2012