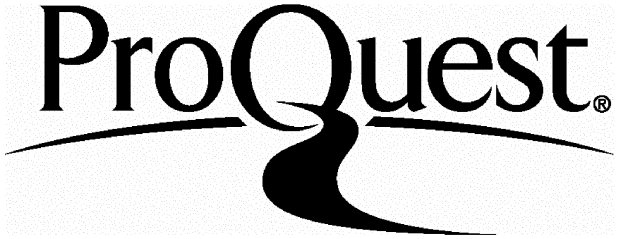
The Analysis of Cytochrome P450 Proteins by Mass Spectrometry

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

CATHERINE SUSANA LANE

Department of Pharmaceutical and Biological Chemistry
The School of Pharmacy
University of London
UK

July 2004



Abstract

The cytochromes P450 (P450s) are a superfamily of mixed function oxidases with a central role in the metabolism of a large number of drugs, xenobiotics and endogenous compounds. Mammalian P450s are membrane bound and hydrophobic in character. The completion of the draft of the human genome revealed the presence of about 106 different P450 genes, of which 57 are thought to be functional. Traditional methods for the analysis of P450s rely on indirect detection techniques such as immunoblotting, activity assays and the detection of P450 mRNA. Mass spectrometry (MS) provides an attractive alternative approach since it offers uniquely the ability to directly detect low levels of multiple proteins simultaneously and without pre-selection.

A method for the analysis of cytochrome P450 proteins by MS has been developed. Microsomal proteins are separated by 1-dimensional sodium dodecyl sulphate polyacrylamide gel electrophoresis, followed by in-gel tryptic digestion of the protein bands and peptide analysis by nano-liquid chromatography (nano-LC) coupled to nano-electrospray tandem mass spectrometry (ES-MS/MS), with identification by database searching. Work with biological samples culminated in a study in which the P450 expression profile was determined for six sets of colorectal liver metastases and corresponding liver samples from patients with metastatic colorectal cancer of the liver. Fourteen distinct P450 enzymes were positively identified (CYP1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1, 3A4, 4A11, 4F2, 4F11, 8B1 and 27A1), thirteen in liver and twelve in tumour tissue. The P450 profiles of the tumours demonstrate that metastatic cancers in liver potentially have extensive drug-metabolising capabilities, which are likely to be important in determining the outcome of treatment. Further studies have been initiated, which aim to identify and quantify functional P450s using mechanism based inactivators (MBIs). These are substrates that form metabolites that covalently adduct to the active site, leaving the enzyme permanently modified. The use of benzyl isothiocyanate (BITC) as a MBI of CYP2E1 led to the identification of a BITC-modified catalytic site peptide, ^{283}LYTMDGITVTVADLFFAGTETTSTTLR^{309}. The site of modification was determined by MS/MS to be one of six C-terminal threonine and serine residues. The mass of the modification was consistent with the adduction of one molecule of benzyl isocyanate and oxygen.

For my Parents

Acknowledgements

I would like to thank my supervisors Laurence Patterson and Bill Griffiths for all their help, advice and encouragement and for the huge amounts of time they have put in, especially over the last few months. Thanks must also go to Kevin Welham for guidance in my first and second years and in particular to Bill Griffiths for taking over the baton in the last few months of my third year. A big thank-you to Mike Cocksedge and Emmanuel Samuel for much needed technical support.

I am indebted to the many people at the School of Pharmacy and elsewhere, too numerous to mention individually, who have been more than happy to take the time to help me and to answer my (often foolish) questions on countless subjects. Special thanks to Amanda Wilgoss, Andy Wilderspin, Rosemary Smyth and Mike Munday for copious amounts of help with all things biological; to Roger Jee for assistance with statistics and to Dave Secrett and others at ThermoElectron for greatly appreciated LCQ advice.

Many thanks to Andy Wilderspin for the use of his laboratory.

Thank-you to Amanda Wilgoss for expressing CYP1B1 and for making the CYP1B1 Western Blots of the human tissue samples; to Rosemary Smyth for the SOD samples; to Brian Davidson and Barry Fuller for the human liver and tumour samples and James Hewes for the patient information; and to Anita Yakkundi and Rebecca Gallagher for the RIF-1 murine tumours.

I am indebted to Gary Woffendin for the original method for the in-gel tryptic digestion of proteins and to Yuqin Wang for the S-carbamidomethylation procedure.

Thanks finally to the people who have kept me sane: to Amanda and Rosemary, Jo and Himali, and to my Dad: I couldn't have done this without you.

Abbreviations

1DE	one-dimensional gel electrophoresis
2DE	two-dimensional gel electrophoresis
ac	alternating current
AMT	accurate mass and time
AQ4N	anthraquinone-di-N-oxide
APS	ammonium persulphate
ATP	adenosine triphosphate
BIC	benzyl isocyanate
BITC	benzyl isothiocyanate
BSA	bovine serum albumin
CAD	collision-activated dissociation
dc	direct current
DF 203	2-(4-amino-3-methylphenyl)benzothiazole
DLPC	1,2-didecanoyl-sn-glycero-3-phosphocholine
DNA	deoxyribonucleic acid
DTT	dithiothreitol
E. coli	*Escherichia coli*
ECD	electron capture dissociation
EDTA	ethylenediaminetetra-acetic acid
EGTA	ethylene glycol-bis(2-aminoethylether)-N,N,N',N'-tetraacetic acid
ELISA	enzyme-linked immunosorbent assay
EM	extensive metaboliser
ER	endoplasmic reticulum
ES	electrospray
FAD	flavin adenine dinucleotide
FMN	flavin mononucleotide
FT	Fourier transform
FTMS	Fourier transform mass spectrometer
g	acceleration due to gravity
g	grams
HFBA	heptafluorobutyric acid

HIV	human immunodeficiency virus
HRP	horseradish peroxidase
ICAT	isotope coded affinity tag
ICR	ion cyclotron resonance
ID	inner diameter
IM	intermediary metaboliser
LC	liquid chromatography
microsomes	small (~100 nm in diameter) closed vesicles formed by the ER after disruption by homogenisation
MALDI	matrix-assisted laser desorption ionisation
MBI	mechanism-based inactivator
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MudPIT	multi-dimensional protein identification technology
m/z	mass to charge ratio
NADH	nicotinamide adenine dinucleotide, reduced form
NADPH	nicotinamide adenine dinucleotide phosphate, reduced form
Nano-ES	nano-electrospray mass spectrometry
Nano-LC	nano-scale reversed phase liquid chromatography
Nano-LC-ES	nano-LC coupled to nano-ES
NCBI	National Center for Biotechnology Information
NP-40	nonidet P-40
ompC	outer membrane protein C
OD	outer diameter
P450	cytochrome P450
PM	poor metaboliser
PMF	peptide mass fingerprinting
PMSF	phenylmethanesulphonyl fluoride
PSD	post source decay
PTM	post-translational modification
QTOF	quadrupole-time of flight
rf	radio frequency
RP	reversed phase

S9	supernatant after centrifugation at 9000 *g*
SBTI	soya-bean trypsin inhibitor
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SILAC	stable isotope labelling with amino acids in cell culture
SOD	Cu/Zn superoxide dismutase
TAP	tandem affinity purification
TEMED	N,N,N',N'-tetramethylethylenediamine
TEV	tobacco etch virus
TIC	total ion count
TFA	trifluoroacetic acid
tnaA	tryptophanase
TOF	time of flight
Tris	tris(hydroxymethyl)methylamine
UDP	uridine diphosphate
UM	ultra-rapid metaboliser
UV	ultra-violet

Amino acid abbreviations

One letter code	Three letter code	Amino acid
A	Ala	Alanine
B	Asx	Asparagine or aspartic acid
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
Z	Glx	Gluatmine or glutamic acid

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 From genomics to proteomics

The genomic sequencing of numerous organisms has transformed biological and medical research, providing the foundation for the large scale interpretation of gene and cellular function. The term proteome was coined in 1994 to describe the set of proteins encoded by the genome [1]. Proteomics, the study of the proteome, has come to encompass the identification, characterisation and quantification of the complete set of proteins expressed by the entire genome in the lifetime of a given cell, tissue or organism, including isoforms, polymorphisms and modifications, protein-protein interactions and the structural description of proteins and their complexes. Proteins carry out most biological functions, and in order to understand how cells work, one must study what proteins are present, what they do and how they interact with one another. If the genome represents the words in a dictionary, then the proteome provides the definitions, with the interactions of the proteins with each other and the other molecules in their environment providing the grammar to form meaningful language.

Although genomics provided the "blueprint" for the potential gene products that are the focus of proteomic studies, the challenges of proteomics are larger and far more complex than the huge but basically straightforward task of mapping the genome. In contrast to the static nature of the genome, which is essentially identical in every cell of an organism, the proteome is dynamic, constantly changing and responding to internal and external stimuli. Whereas DNA sequencing has the enabling technologies of the polymerase chain reaction and automated sequencing, proteomics must cope with problems of limited and variable sample material, a protein abundance dynamic range of more than 10^6-fold, post-translational modifications and a plethora of perturbations due to development, environment, drugs and disease [2]. The total number of human genes is not known, but has been estimated at around 25,000 [3]; this number of genes could give rise to over two million protein components [4].

The sheer scale and complexity of proteomics research makes it very much a technology-driven enterprise, and one that is rapidly evolving with the invention and

development of methods and techniques. A recent *Nature* overview [2] cites five "central pillars" of proteomics research as being mass spectrometry-based proteomics, array-based proteomics (for example the yeast two-hybrid system), structural proteomics (encompassing the use of X-ray crystallography, nuclear magnetic resonance spectroscopy, electron microscopy and electron tomography), clinical proteomics (i.e. disease states and drug discovery) and proteome informatics. Yet it is difficult to section proteomics into different areas: the boundaries between the subjects are blurred, and much proteomics research involves the combination of technologies from several different fields.

1.2 Mass spectrometry-based proteomics

Mass spectrometry (MS) is a technique whose beginnings date back to the studies performed by J. J. Thomson [5] and his student F. W. Aston [6] in the early days of the last century. MS measures the mass to charge ratios (*m/z*) of gas-phase ions with extremely high sensitivity. In the last 20 years, MS has played an increasingly significant role in the biological sciences; today, MS is the most sensitive method for the structural characterisation of biomolecules [7]. The successes of MS in biology are largely due to the introduction of the "soft ionisation" techniques of electrospray (ES) [8-11] and matrix-assisted laser desorption ionisation (MALDI) [12,13], which allow the transfer of large, polar, thermally labile biomolecules into the gaseous phase for mass analysis, without prior derivatisation. These achievements were recognised by the awarding of the 2002 Nobel prize in chemistry to John Fenn and Koichi Tanaka [14], for their pioneering work on ES and MALDI.

At the same time as ES and MALDI were becoming commonplace in biological MS, the concept of proteomics was emerging. The origins of what has come to be called proteomics date back to the invention of two dimensional gels thirty years ago, which provided the first practical method for displaying thousands of proteins on a single gel [15,16]. Biological mass spectrometry allowed the large-scale identification of those proteins, and has now developed to such an extent as to supersede the two-dimensional gels that originally gave proteomics its impetus.

The range of existing and potential applications of MS to proteomics is extensive, and a detailed account would fill many pages. An overview is presented below.

1.2.1 Instrumentation

Mass spectrometry measures the *m/z* ratios of gaseous ions. Every mass spectrometer now consists of:

- an ion source, to produce ions from the sample
- one or more mass analysers, to separate the ions according to their *m/z* ratios
- a detector, to register the number of ions emerging from the last analyser
- a computer, to process the data, to produce the mass spectrum in a suitable form and to control the instrument through feedback

Each mass spectrometer also has an inlet device to introduce the analyte into the ion source, for example a liquid chromatograph or a direct insertion probe (Figure 1.1).

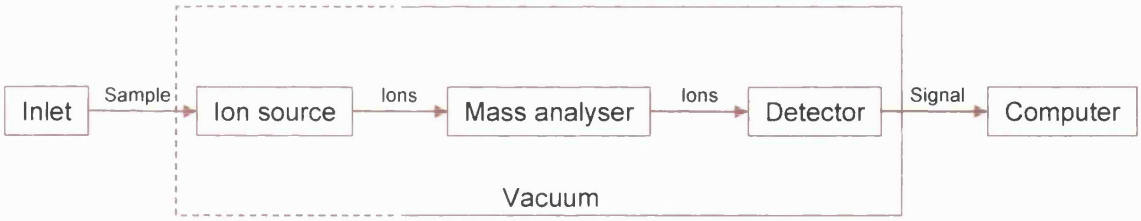


Figure 1.1 A basic diagram for a mass spectrometer.

1.2.1.1 Ion sources

In order to analyse a sample by MS, it must first be vaporised and ionised. The two ionisation techniques most commonly used for the mass spectrometric analysis of proteins and peptides are ES and MALDI. ES produces gaseous ions from solution phase samples, and can therefore be easily coupled to liquid-based separation technologies such as liquid chromatography (LC) and capillary electrophoresis. MALDI ionises samples out of a dry, crystalline matrix, and is generally used for the

analysis of simple peptide mixtures, whilst integrated ES-LC-MS systems are favoured for the analysis of more complex samples.

1.2.1.1.1 The electrospray process

The generation of macroions by electrospray was first demonstrated by Dole *et al* in 1968 [17], but it was Fenn's group at Yale University that first coupled ES with MS [8,9]. The electrospray process (Figure 1.2) transfers ions in solution into gaseous ions at atmospheric pressure, which are sampled into the vacuum system of the mass spectrometer through a series of sampling apertures separating successive vacuum stages.

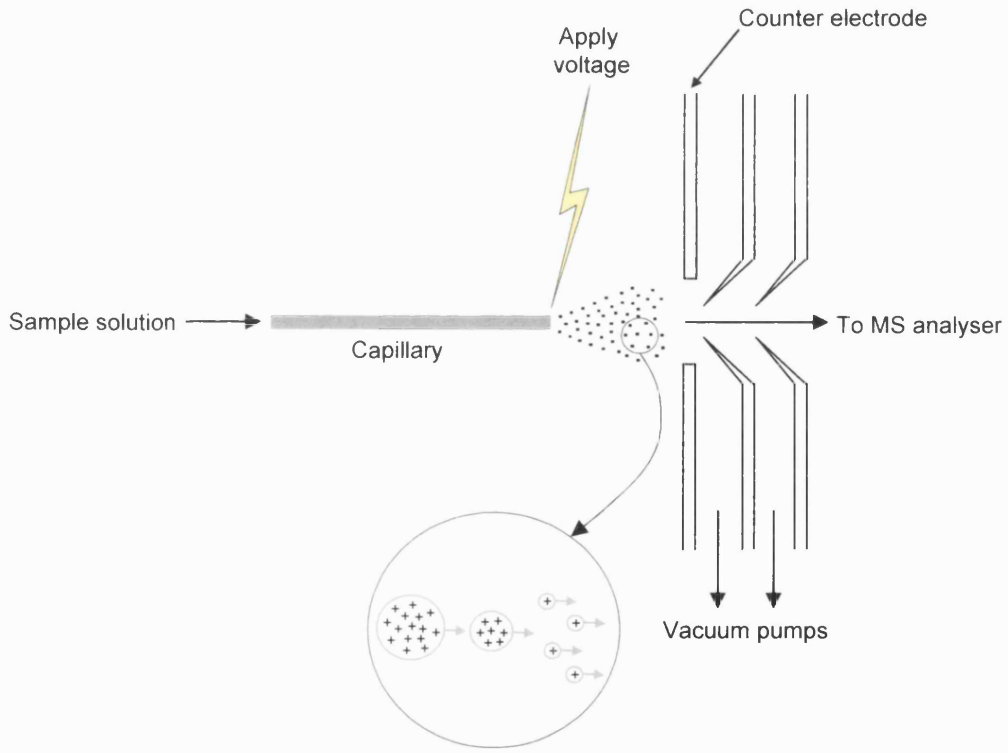


Figure 1.2 Schematic diagram of an electrospray interface and the ES process.

The mechanisms involved in the production of isolated gaseous ions by electrospray are not fully understood [18]. The sample solution flows at low flow rates (nl min^{-1} to µl min^{-1}) through a capillary tube to which a high voltage (1 to 6 kV) is applied. The solution flowing through the capillary experiences an electric field set up between

the capillary and a counter electrode and, assuming a positive potential is applied to the capillary, positive ions in solution will accumulate at the surface of the tip, which becomes drawn out, assuming a conical shape known as a "Taylor cone" (Figure 1.3). As the liquid is forced to hold more electric charge the cone is drawn out into a filament that, when the surface tension is exceeded by the applied electrostatic force, produces positively charged droplets (< 10 μm in diameter) via a "budding" process. Figure 1.3 shows a photograph of the spray produced. The droplets fly towards the counter electrode (or collector), which is opposite in charge to their own. As they fly towards the electrode they pass through either a heated capillary (180 to 270 °C) or a curtain of heated nitrogen to allow solvent to evaporate. The electrical charge density at the surface of the droplets increases as the droplet size decreases. The droplet deforms into a tear shape and, at the point where the electrostatic repulsion is greater than the surface tension (known as the "Raleigh limit"), it blows apart, emitting smaller particles (Figure 1.4). Depending on the initial size of the droplet, the particles leaving can either be smaller droplets that repeat the process, or discrete solvated surface ions. At atmospheric pressure collisions with the surrounding gases quickly desolvate the solvent-clustered ion, resulting in a quasi-molecular (or multi-charged) ion [19].

An alternative mechanism for gas phase ion production has been proposed in which ion emission is envisaged to occur from small, highly charged droplets, with the driving force for ion formation being the repulsion between the ion and the other charges on the droplet [20,21]. The relative importance of the two proposed mechanisms remains the subject of ongoing research and debate [22].

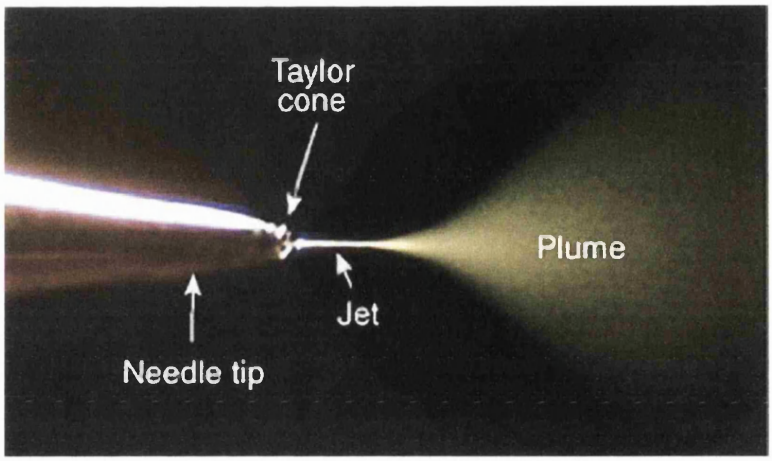
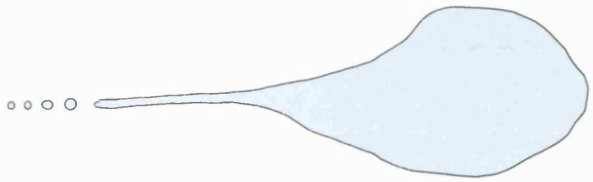


Figure 1.3 The electrospray process, viewed through a high-powered microscope. As the liquid begins to exit the needle it charges up and assumes a conical shape, known as the Taylor cone. At the tip of the cone, the liquid is drawn to a filament, which then becomes unstable, breaking up into a mist of charged droplets. Since the droplets are charged they repel each other strongly and fly apart, covering a wide surface area. Taken from http://www.newobjective.com/electrospray/.



Rayleigh: $q^2 = 8\,\pi^2\,\varepsilon_0\,\gamma\,D^3$

Figure 1.4 A drawing of a decomposing droplet in an ES source [23]. The critical point at which the surface tension of the droplet is overcome by the electrostatic repulsion of the surface charges is known as the Rayleigh stability limit: q = charge; ε_0 = permittivity of the environment; γ = surface tension and D = diameter of a supposed spherical droplet.

Significantly, the ES process occurs at relatively low temperatures (room temperature, or just above) and so large, thermally labile, polar molecules can be ionised without decomposition. Ionised molecules of the form $[M + H]^+$ or $[M + nH]^{n+}$ (or $[M - H]^-$ / $[M - nH]^{n-}$) are generally produced. The prerequisite for gaseous ion production with ES is that the analyte can be ionised in solution. If several ionisable sites are present then multiply charged ions will be produced, for example, denatured proteins typically carry one charge per 1000 Da [23]. By observing such multiply charged species, the effective mass range of the spectrometer can be extended to hundreds of thousands of daltons.

1.2.1.1.2 Matrix-assisted laser desorption ionisation

The technique of MALDI was introduced in 1988 when Karas and Hillenkamp described the analysis of proteins with molecular masses exceeding 10 kDa [12,13]. MALDI sublimates and ionises the analyte out of a dry, crystalline matrix using laser pulses.

The sample to be analysed is co-crystallised with a large excess of a matrix material that will strongly absorb the light from a laser. The laser is typically a nitrogen laser at 337 nm. Irradiation of the matrix causes rapid heating and localised sublimation of the matrix crystals. Since the matrix is in large excess and contains a chromophore for the laser light it will absorb essentially all of the laser radiation. As the matrix expands into the gas phase it takes with it intact analyte molecules; little internal energy is transferred to the analyte molecules, allowing ionisation without fragmentation. Ionisation can occur at any time during this process, but the exact origin of ions produced by the MALDI process is still not fully understood. The most widely accepted mechanism involves gas-phase proton transfer in the expanding matrix plume with photoionised matrix molecules [23]. This is illustrated in Figure 1.5.

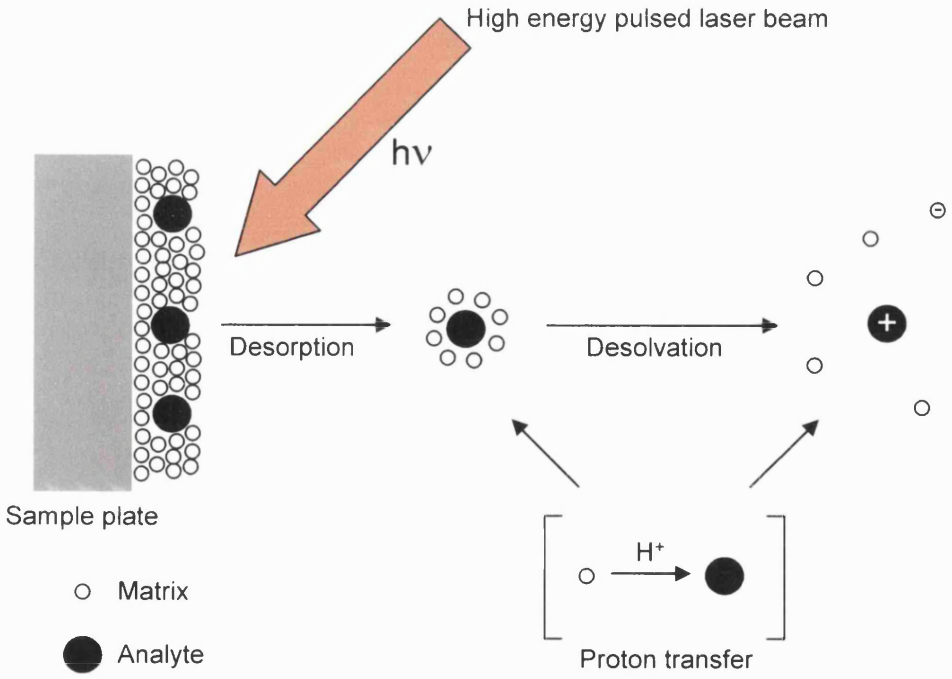


Figure 1.5 The MALDI process. Adapted from de Hoffmann and Stroobant [23].

The MALDI process is independent of the absorption properties and size of the compound to be analysed, therefore it allows the desorption and ionisation of analytes with very high molecular masses (greater than 100 000 Da).

1.2.1.2 Mass analysers

To a large extent, the information obtained from a proteomics experiment is determined by the operating conditions and performance specifications of the mass analyser. The mass analyser is the means by which the ions are separated and detected; instrumental parameters such as mass resolution, mass accuracy, mass range, sensitivity and tandem MS (MS/MS) capability are key in assessing performance and utility. There are four basic types of mass analyser currently in use for proteomics research: the ion trap, time-of-flight (TOF), quadrupole and Fourier transform (FT) ion cyclotron resonance (ICR) analysers. They are diverse in terms of design and performance, and can either be used as stand-alone analysers or, in some cases, put together in tandem to take advantage of their different strengths.

1.2.1.2.1 Quadrupoles

Quadrupole mass analysers consist of four precisely parallel rods equally spaced around a central axis. Opposing sets of rods have both a dc (direct current) and an ac (alternating current) or rf (radio frequency) voltage component, one set positive and the other set negative. Ions are introduced in a continuous beam along the central axis between the poles (Figure 1.6), and are filtered on the basis of their *m/z* ratios in the following manner: ions that pass between the two positive rods that are above a critical *m/z* ratio are transmitted through the centre of the quadrupole. This forms a high pass mass filter. Ions that pass between the two rods with a negative potential that are below a critical *m/z* ratio are transmitted through the centre of the quadrupole; this forms a low pass mass filter. Combining both sets of rods into a quadrupole arrangement overlaps the two mass filter regions, creating a "band pass" area of mutual stability (Figure 1.7) and allowing ions of a certain *m/z* ratio to pass through. Ions with *m/z* ratios outside this area of mutual stability cannot pass through

and run into the rods. The *m/z* ratio of the ions that are allowed to pass through the quadrupole is proportional to the voltage applied to the rods; the higher the voltage, the higher the *m/z* value that is allowed to pass. By altering the relative contributions of the dc and rf components, the width of the band pass area, and therefore the resolution, can be adjusted (wider band pass = wider peak = lower resolution; narrower band pass = narrower peak = higher resolution). Scanning a quadrupole mass analyser involves ramping the amplitude of the dc and rf voltages at a constant ratio, thus changing the position of the band pass region and allowing different masses to be transmitted.

For MS/MS analysis three quadrupoles can be configured together (to form a "triplequad"). The first and third quadrupoles are used for scanning, whilst the middle quadrupole is used as a collision cell. Ions in the second quadrupole are fragmented by low-energy collisions with a background gas such as nitrogen.

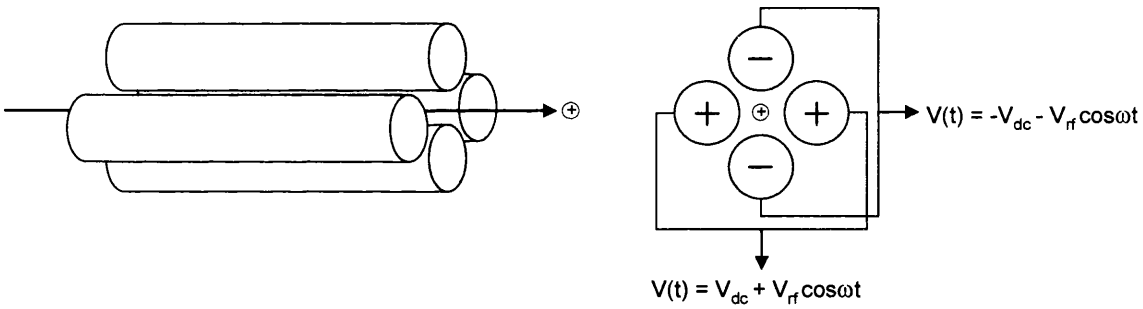


Figure 1.6 Schematic diagram of a quadrupole mass analyser, showing the direction of ion travel and the equations for the potentials applied to the rods. V(t) = voltage at time t; V_{dc} = direct potential; V_{rf} cos ωt = radio frequency component. Adapted from Willoughby *et al* [19].

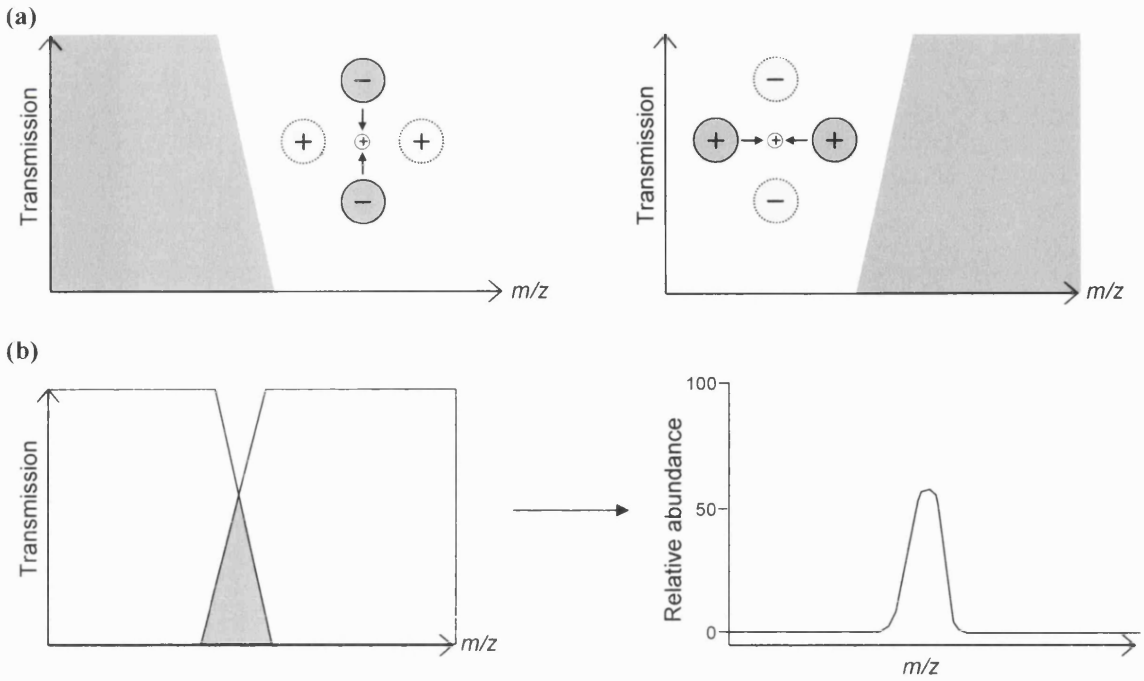


Figure 1.7 Mass filter created by a quadrupole mass analyser. **(a)** Negative rods create a low pass mass filter (left); positive rods create a high pass mass filter (right). **(b)** Combining negative and positive sets of rods into a quadrupole arrangement overlaps the two mass filter regions, creating an area of mutual stability which allows ions of a certain *m/z* to pass. Adapted from Willoughby *et al* [19].

1.2.1.2.2 Ion traps

Ion traps work by first trapping the ions and then detecting them based on their *m/z* ratios. Conceptually, an ion trap can be imagined as a quadrupole bent around on itself to form a closed loop. The inner rod is reduced to a point at the centre of the trap; the outer rod is a circular ring electrode and the top and bottom rods become two end-cap electrodes (Figure 1.8). Hence the ion trap is also referred to as the "quadrupole ion trap". Ions are held inside the trap and subjected to oscillating electric fields generated by a rf voltage applied to the ring electrode only. Unlike for quadrupoles, there is no dc component to the voltage, and the end caps are held near ground potential. For mass analysis, ions are guided up to the trap and a portion of the ions is sampled into the trap. Ions of all *m/z* values enter the trap at the same time. Once inside the trap, the ions are confined by the rf field and take on an oscillating frequency that is related to their *m/z* value (Figure 1.9). As the ions repel each other in the trap their trajectories expand as a function of time. To avoid ion losses from this expansion, helium gas is present inside the trap as a "buffer" gas, to

remove excess energy from the ions by collision; the helium "dampens" the ions into the centre of the ring electrode.

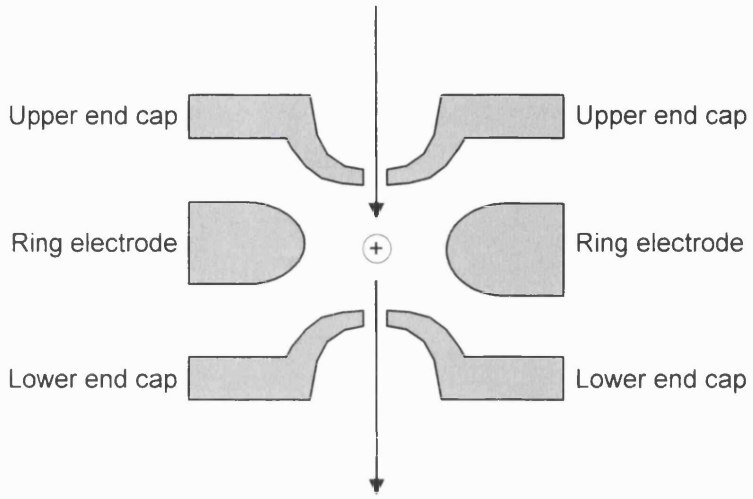


Figure 1.8 Schematic diagram of an ion trap mass analyser. Adapted from Willoughby *et al* [19].

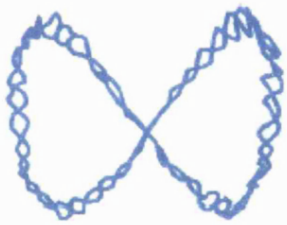


Figure 1.9 Oscillating figure of eight-shaped trajectory adopted by ions in an ion trap.

To scan the mass range, the rf voltage (i.e. amplitude) on the ring electrode is ramped. At the same time, a small rf voltage is applied to the end caps. As the amplitude on the ring electrode increases, the frequencies of the ion oscillations also increase. When the resonant frequency of an ion reaches the end cap frequency, the ion will become excited into an oscillating motion that is so large that it becomes destabilised and is ejected from the trap along the axis of the end-caps. Since the oscillating frequencies of the ions are a function of their mass, ions of different *m/z* values will exit the ion trap at different voltages and therefore at times.

MS^n can be carried out in ion traps. Ions are selected for MS/MS analysis by using an rf voltage applied to the end caps to selectively eject all ions in the trap except for a chosen precursor ion. A resonating frequency that corresponds to the resonant frequency of the isolated ion is then applied to the end caps at an amplitude of a few

percent of that required to eject the ion. The precursor ion starts to oscillate and collide with the helium buffer gas, which eventually induces fragmentation of the precursor ion. The fragment ions have different resonating frequencies from the parent ion and therefore are dampened into the centre of the trap by the helium gas. After a period of time, the rf voltage on the ring electrode is ramped, causing ejection of the ions in the manner described above. For MS^n analysis, this cycle is repeated to allow sequential stabilisation and subsequent fragmentation of successive product ions.

Ion traps are robust, sensitive and relatively inexpensive, and so are widely used for proteomics research. A disadvantage of the ion trap is its relatively low mass accuracy, due partly to the limited number of ions that can be accumulated in its centre before space-charging distorts their distribution, and therefore the accuracy of the measurement. Recently, the "linear" ion trap has been developed [24], which stores the ions in a cylindrical volume that is considerably larger than that of traditional ion traps, allowing increased sensitivity, resolution and mass accuracy.

1.2.1.2.3 Time-of-flight analysers

Analysis by TOF is based on the following principles:

- An accelerating potential (V) will give an ion of charge z an energy of zV, which can be equated to the kinetic energy of the ion:

$$zV = \frac{mv^2}{2}$$

where m = mass, v = velocity

- If all ions are accelerated with the same potential then ions of different mass with the same charge must be travelling at different velocities.
- But velocity (v) = distance (d) / time (t), therefore the equation can be re-written:

$$\frac{m}{z} = \frac{2Vt^2}{d^2}$$

and ions with different mass will take different amounts of time to travel the same distance.

Mass to charge ratios are determined by measuring the time that ions take to move through a field-free region between the source and the detector. Mass resolution is affected by slight variations in flight time, and factors that create a distribution in flight times among ions with the same *m/z* ratio will result in poor mass resolution. Two techniques are used to compensate for temporal (time of ion formation), spatial (location of ion formation) and kinetic (energy of ion formation) distribution. By introducing a time delay between ion formation and extraction of ions from the source, wide spatial and temporal distributions can be avoided. This is known as "delayed extraction". Ions are first allowed to expand into a field-free region in the source, then after a certain delay (nanoseconds to microseconds) a voltage pulse is applied to extract the ions outside the source. The second technique is the use of ion mirrors, or reflectrons, which compensate for variations in energy distribution. The reflectron creates a retarding field that deflects the ions, sending them back through the flight tube. The more energetic the ion, the deeper it penetrates the retarding field of the reflectron before being reflected. Thus a more energetic ion will travel a longer flight path and arrive at the detector at the same time as less energetic ions of the same mass. A schematic diagram of a reflectron TOF is shown in Figure 1.10.

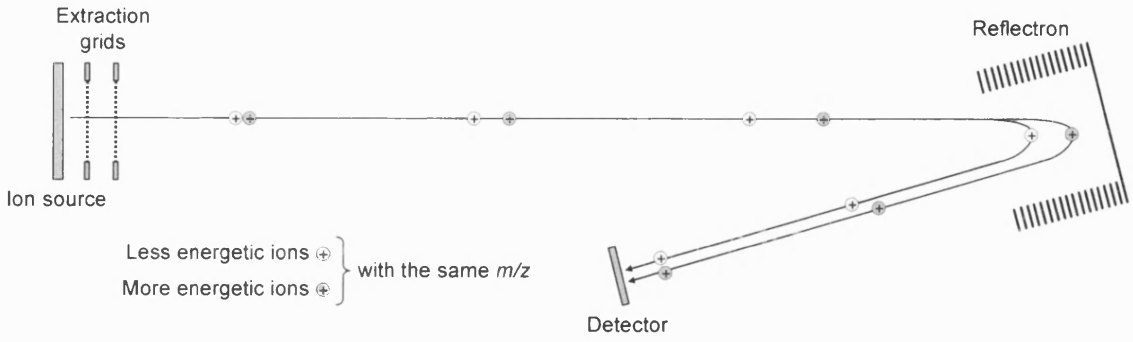


Figure 1.10 Basic components of a reflectron TOF analyser.

MS/MS can be achieved with reflectron TOF analysers by the observation of post source decay (PSD) fragments; however, this approach is protracted and arduous. MS/MS with TOF analysis is most commonly practiced by placing a collision cell between two TOF analysers, or by configuring a TOF analyser as the second stage in hybrid instruments (Section 1.2.1.3).

1.2.1.2.4 Fourier transform MS

A Fourier transform mass spectrometer (FTMS), also known as an ion-cyclotron resonance mass spectrometer (ICR-MS), is a type of ion trap consisting of a cubic cell inside a strong magnetic field. The cell has three sets of plates: trapping, transmitter and receiver plates (Figure 1.11). For mass analysis a continuous beam of ions is formed outside the cell and guided up to the trap, and a portion of the beam is pulsed into the cell. Inside the cell the ions are constrained by the strong magnetic field so that they move in cyclotron motion (circular orbit, see Figure 1.12) in a plane perpendicular to the magnetic field; they are also constrained by electric potentials applied to the trapping plates, which are also perpendicular to the magnetic field (Figure 1.11). Cyclotron motion is periodic and is characterised by its cyclotron frequency, the frequency with which an ion repeats its orbit. The cyclotron frequency, f_c, is determined by the strength of the magnetic field (B), the charge on the ion (z) and the mass of the ion (m):

$$f_c = \frac{zB}{2\pi m}$$

Thus for a constant magnetic field the m/z ratio of an ion is determined by measuring its cyclotron frequency. The radius of the cyclotron orbit depends on the kinetic energy of the ion and, since the cyclotron frequency is constant for an ion of given m/z, scales directly with an ion's velocity, or with the square root of the kinetic energy. When an ion is first trapped, the radius of its cyclotron orbit is usually small compared with the dimensions of the cell. In order to detect the ions, rf electric potentials are applied to the transmitter plates. An ion whose cyclotron frequency is in resonance with the frequency of the applied rf field will absorb energy and, as it

does so, the radius of its cyclotron orbit will increase. As the excited ions pass the receiver plates the frequency of their passage is detected as an induced current called the "image" current. To accomplish detection of ions with many different *m/z* ratios simultaneously, many frequencies are applied during the excitation event, most commonly by using a rapid frequency sweep in a short (1 ms) time period [25]. The image current that results is a composition of the frequencies and amplitudes of ions of many different *m/z* values. The signal is converted to a mass spectrum by applying a Fourier transform (Figure 1.13). Once the ions have been detected, a rf "quench" pulse is applied to eject the ions before the next lot of ions are introduced into the cell.

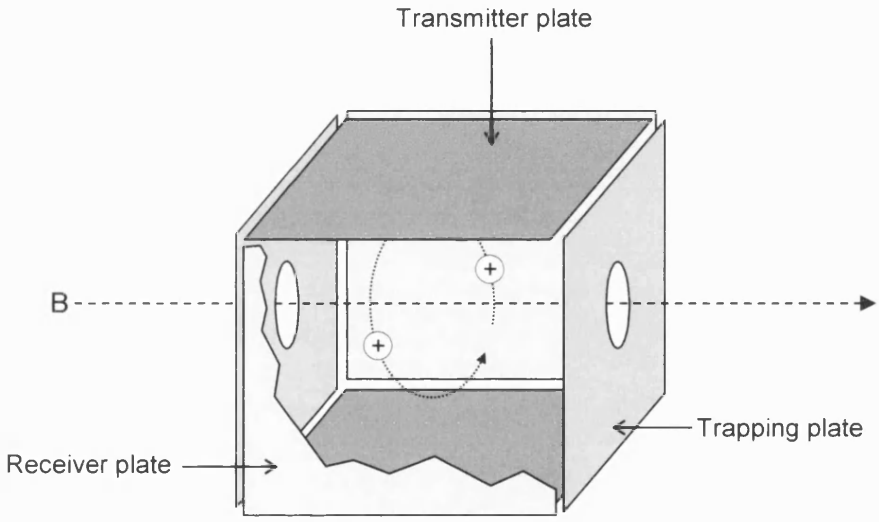


Figure 1.11 Schematic diagram of Fourier transform mass analyser (FTMS). Ions are injected into the trap along the same plane as the magnetic field, B, and trapped by a voltage applied to the trapping plates. The ions move in cyclotron motion in the plane perpendicular to the plane of the magnetic field. Adapted from Willoughby *et al* [19].

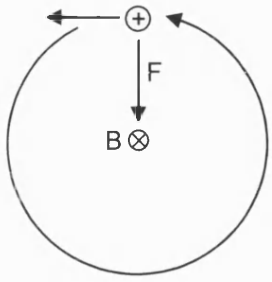


Figure 1.12 Cyclotron motion in the plane perpendicular to the plane of a magnetic field, B. The magnetic field is pointing into the plane of the page. An ion moving to the left experiences a downward force, F, that drives it into an anticlockwise orbit. Taken from Amster [25].

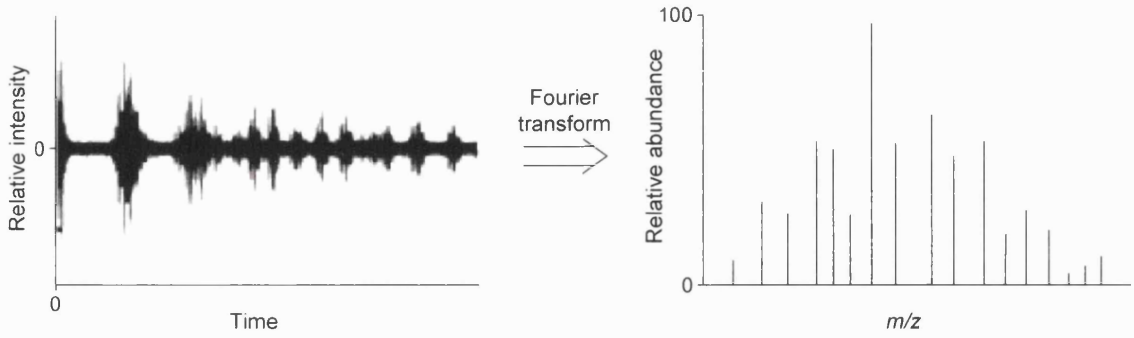


Figure 1.13 Time and mass domain for FTMS. Signal intensity as a function of time is transformed, through a Fourier transform, into signal intensity as a function of frequency and hence to *m/z*.

FTMS is unlike any other form of mass analyser because ion detection is non-destructive; signal to noise can be improved by averaging many cycles before transforming and storing the data.

MS/MS analysis with FTMS is similar to that in the ion trap: ions are first isolated by ejecting all other ions in the cell using resonance excitation, then a pulse of gas is introduced into the cell and a small voltage is applied to the transmitter plates. By varying the amplitude of the resonating frequency, the precursor ion starts to oscillate and collide with the background gas, inducing dissociation and production of fragment ions. After a period of time, the ions are excited into higher cyclotron orbits and detected.

FTMS provides high sensitivity, mass accuracy, resolution and dynamic range, but instruments are expensive and their operation complex; to date, these two factors have limited their routine use in proteomics research, in spite of their substantial potential [26].

1.2.1.3 Coupling ion sources to mass analysers

MALDI is traditionally coupled to TOF analysers, which measure the mass of intact peptides, whilst electrospray is mostly coupled to ion traps and triple quadrupoles, which allow the generation of fragment ion spectra from selected precursor ions. In the last few years, MALDI ion sources have been coupled to quadrupole ion-trap mass spectrometers [27], TOF-TOF analysers [28], in which two TOF sections are

separated by a collision cell, and quadrupole-TOF (QTOF) analysers [29], in which a collision cell is placed between a quadrupole mass filter and a TOF analyser. These instruments have good sensitivity, resolution and mass accuracy, and allow the fragmentation of MALDI-generated precursor ions. In addition, the QTOF can be used interchangeably with an ES ion source.

Although ion trap, TOF and hybrid TOF instruments are currently the most widely used for proteomics research, FTMS and linear ion traps are likely to become widespread in the future.

1.2.2 Peptide fragmentation theory

Protein identification and peptide sequencing by MS/MS necessitates knowledge of how peptides fragment in the gas phase. The transfer of ions into the gas phase by electrospray is not a highly energetic process. The desolvation process effectively cools the ions, and therefore gaseous ions entering into the mass spectrometer for analysis have low internal energy [22]. The energy required for the ions to undergo dissociation is introduced by collisional activation. Peptide ion fragmentation is promoted by a "mobile" proton, i.e. the fragmentation of most protonated peptides requires the involvement of a proton at the cleavage site [30]. The relative populations of the different protonated forms of a particular peptide depend on the internal energy content of the peptide and the gas-phase basicities of the potential sites of protonation. The addition of energy via collisional activation alters the initial population of protonated forms (i.e. "mobilises" the proton) and increases the population of protonated forms with energies higher than that of the most stable structure; for these excited molecules, the proton is located at various backbone heteroatoms. Protonation at backbone sites initiates cleavage of the backbone to produce b- and/or y-type fragments (Figure 1.14).

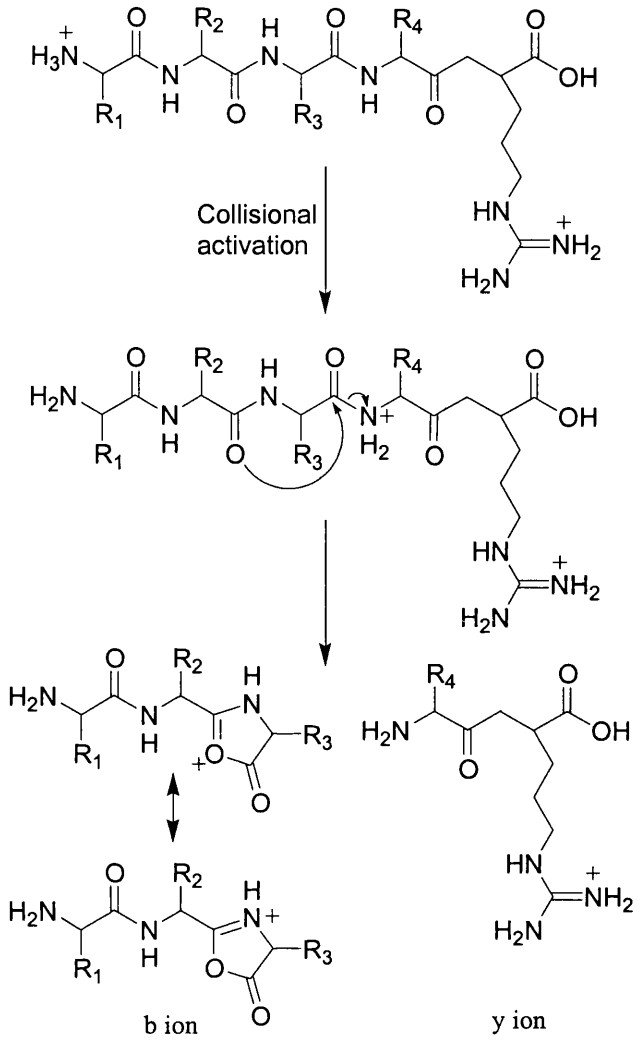


Figure 1.14 Fragmentation of a doubly charged tryptic peptide to produce b and y ions. Relocation of the N-terminus proton to promote charge-directed cleavage via collision-activated dissociation (CAD) may occur to either N or O backbone atoms [30].

For singly protonated peptides containing a strongly basic amino acid the energy required to relocate the proton to a position on the backbone and hence induce dissociation is considerably higher than for peptides containing no basic amino acids. Dissociation energy requirements are greatest for arginine-containing peptides and decrease in the order Arg-containing > Lys-containing > non-basic, i.e. in order of decreasing gas-phase basicity [31]. In doubly charged tryptic peptides the initial sites of protonation are generally the basic side chain of the C-terminal residue and the primary amine group at the N-terminus [22] (Figure 1.14). The energy barrier to transfer of the N-terminus proton to the peptide backbone and between different sites on the peptide backbone is low so that various different fragmentation pathways are promoted, resulting in MS/MS spectra with several different b and y ions. Tang and

Boyd [32] observed that fragmentations of tryptic peptide $[M + H]^{2+}$ ions resulted in product ion spectra in which the y ions appeared at appreciably higher relative abundance than their b ion counterparts. This was explained by the higher stability of the y ions due to the sequestering of the proton on the basic side chain of the C-terminal residue, whereas the b-ions have no strongly favoured site of charge, so are likely to undergo further intramolecular proton rearrangement and fragmentation. In addition, if high enough collisional activation is achieved then b ions can undergo further dissociation to form a ions, with concurrent loss of carbon monoxide (Figure 1.15).

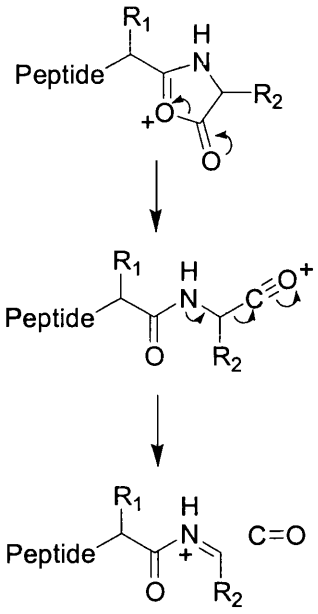


Figure 1.15 Dissociation of b ions to produce a ions and carbon monoxide.

1.2.2.1 Residue-specific cleavages

The presence of certain amino acid residues in a peptide will result in the selective cleavage of the peptide at those sites. Enhanced cleavage is observed N-terminal to proline residues because, as a tertiary amine, the backbone N atom of proline is more basic than the other backbone N atoms and the mobile proton is therefore more likely to reside at this position [33-35]. Conversely, fragmentation C-terminal to proline is rare because of steric hindrance [36] (Figure 1.16).

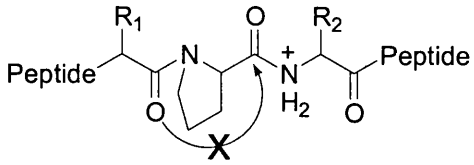


Figure 1.16 Cleavage C-terminal to proline residues is rare due to steric hindrance.

For peptides containing arginine and aspartic and/or glutamic acid, enhanced cleavage occurs C-terminal to the acid residue if the number of added protons is equal to or less than the number of arginines present [30]. In these cases, the arginine(s) tightly binds the proton(s) allowing the acidic hydrogen of the acid side chain to initiate cleavage (Figure 1.17).

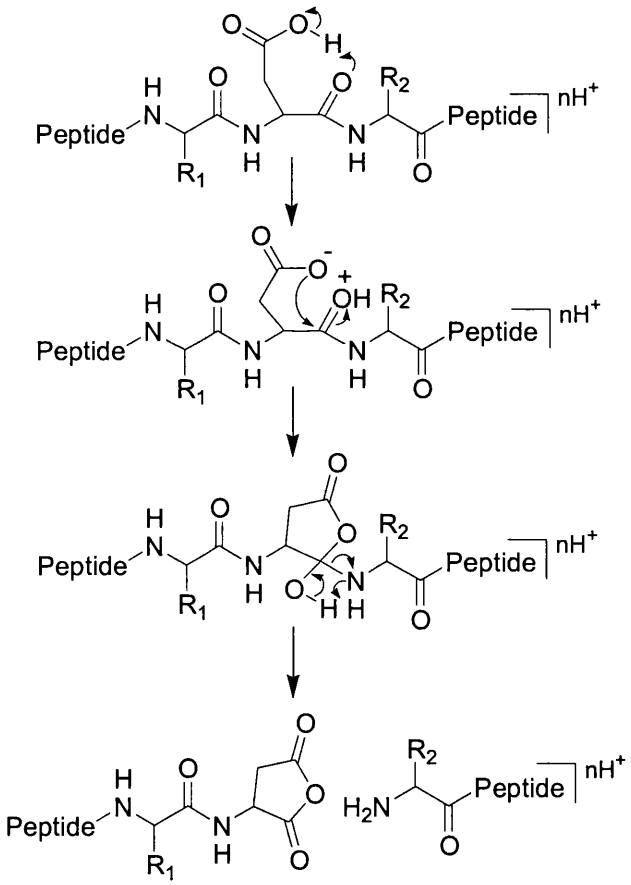


Figure 1.17 Selective cleavage occurs at positions C-terminal to acidic residues when the number of added protons (n) is equal to or less than the number of arginines present.

Where the number of added protons exceeds the number of arginines present, enhanced cleavage at acidic residues is not observed. For peptides containing lysine rather than arginine and aspartic acid and/or glutamic acid there is no enhancement of cleavage at acidic residues [30].

In contrast to the results for acidic residues, where selective cleavage occurs if there is no added proton available to catalyse fragmentation, enhanced cleavage occurs C-terminal to histidine when an added proton is present on the histidine side chain [30]. All singly protonated peptide ions containing both histidine and arginine fragment non-selectively, yet doubly protonated peptide ions with arginine and histidine, or singly protonated peptides containing histidine but not arginine, show selective cleavage. This is explained mechanistically in Figure 1.18.

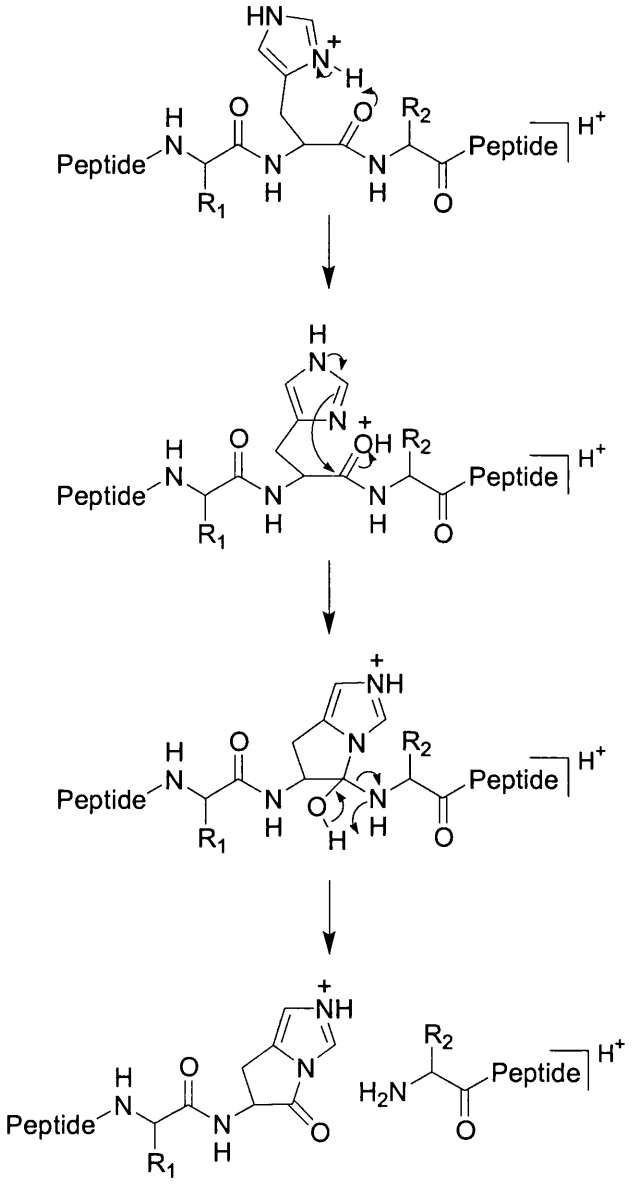


Figure 1.18 Enhanced cleavage is observed C-terminal to histidine when an added proton is present on the histidine side chain [30].

Depending on the identity of the amino acid residue C-terminal to histidine (R_2, Figure 1.18), proton transfer may occur from the histidine ring to the C-terminal leaving fragment.

1.2.3 Protein identification

There are two main routes by which proteins are identified using mass spectrometry. The classical proteomics approach involves the separation of the proteins in a mixture by two dimensional gel electrophoresis (2DE), followed by in-gel tryptic digestion and peptide mass fingerprinting (PMF) by MALDI-TOF MS. In this approach, proteins are identified by matching the list of observed peptide masses with a calculated list of all the expected peptide masses for each entry in a protein database. Observed 2DE-separated protein spots can be quantified by staining intensity. However, there are many problems with 2DE. It has only a very limited dynamic range: only the most abundant proteins are observed [37], and several classes of proteins are known to be excluded or underrepresented in 2DE patterns; these include very acidic or basic proteins, very large or small proteins and membrane proteins [38]. Improvements in 2DE technology such as sample prefractionation [39,40], large format, higher resolving gels [41], narrow pH-range gels [40] and more sensitive staining methods [42,43] have alleviated, but not eliminated these limitations.

PMF requires relatively pure samples, since mixtures of proteins will create mixtures of PMFs, making protein identification difficult [44]. The enabling of MALDI-MS/MS (Section 1.2.1.3) should facilitate the use of MALDI for the analysis of more complex samples, thus uncoupling MALDI-MS from 2DE. However, some prior separation of the sample will still be necessary; if liquid chromatography is used then the effluent from the chromatographic run must be deposited on a sample plate and mixed with MALDI matrix. Several systems have been designed to facilitate this [45-48], but only now is automation of the process being implemented [26].

The second major approach to the identification of proteins and peptides by mass spectrometry is the use of LC-ES-MS/MS. This method is at the centre of MS-based proteomics today; a generic LC-ES-MS/MS approach is illustrated in Figure 1.19.

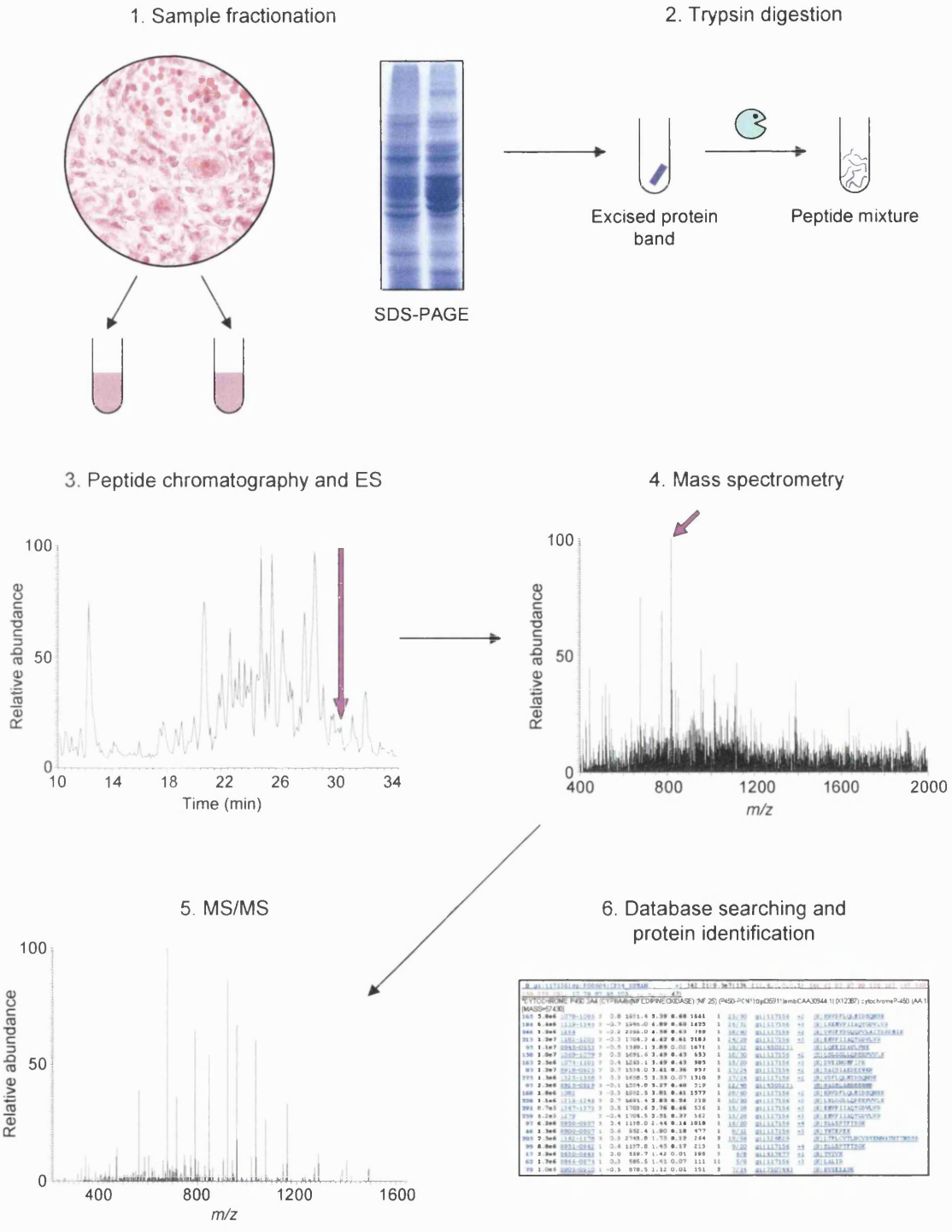


Figure 1.19 A generic mass spectrometry-based proteomics experiment, consisting of six stages. Proteins for analysis are first isolated from cell lysate or tissues by fractionation methods, often including a final step of SDS-PAGE. Proteins are then enzymatically digested and separated using one or more stages of liquid chromatography. The LC eluate is directed into an ES source, and the peptides are analysed first by full scan MS and then by MS/MS. Peptide MS/MS spectra are searched against a protein database using a search algorithm which assigns peptide identifications based on match criteria.

One dimension of peptide chromatography does not provide sufficient peak capacity to separate peptides generated from complex protein mixtures; to address this, various different combinations of protein and peptide separation schemes have been explored involving two or three dimensional chromatography and/or one-dimensional gel electrophoresis (1DE). A popular method is the "multidimensional protein identification technology", or MudPIT approach [49,50], which makes use of the orthogonal separation methods of strong cation exchange (SCX) followed by reversed-phase (RP) chromatography to separate peptides generated from proteolytic digestion of entire protein mixtures. The number of proteins identified using the MudPIT approach can far exceed the number identified using 2DE experiments, for example 1,484 unique proteins were identified from 5,540 unique peptides from a total cell lysate of *Saccharomyces cerevisiae* [50].

A major stumbling block in the use of LC-ES-MS/MS for the identification of peptides from biological matrices is that of informatics: the amount of data generated by the method is huge, and its analysis can be extremely daunting. This is discussed in more detail in Section 1.2.4.

1.2.4 LC-ES-MS/MS and the bioinformatics bottleneck

Protein identifications from peptide collision-activated dissociation (CAD) spectra are less ambiguous than those from PMF because the peak-pattern in the CAD spectrum provides additional information to the measured mass of the peptide. However, the amount of data generated is huge: many thousands of peptide CAD spectra can be acquired during an LC-ES-MS/MS run. In order to identify these peptides, spectra are scanned against protein sequence databases using a search algorithm. Several algorithms have been developed for this purpose [51]. The most commonly used algorithms are Sequest [52], Mascot [53] and MS-Tag [44,54]. Sequest (Section 2.14.2) adopts a cross-correlation approach, in which peptide amino acid sequences from a protein database are used to construct theoretical mass spectra, and the degree of overlap, or cross-correlation, between the theoretical and experimental mass spectra determines the best match. The Mascot method employs probability-based matching: the CAD fragment masses calculated from peptide sequences in the

database are compared with the experimentally observed peaks, and a score is calculated that reflects the statistical significance of the match between the experimental and theoretical spectra. The MS-Tag approach involves extracting a short, unambiguous section of the peptide's amino acid sequence from the experimental data, which is used along with the measured mass of the peptide to determine the protein of origin. For all of these approaches, identified peptides are compiled into a protein "hit-list". Currently available *de novo* sequencing algorithms are computationally intensive and require high quality data [51]. Therefore, for high-throughput proteomics studies, protein identifications are restricted to those proteins whose sequences appear in the searched database; many post-translationally modified or mutated proteins will be overlooked.

The algorithms used for searching MS/MS data are not infallible, and the main challenge in the interpretation of database search results is how to distinguish false positive identifications. If the best matches in all database searches were assumed to correspond to the correct peptides, then a large proportion of these assigned peptides would be wrong [55]. This situation can arise for several reasons: the scoring schemes used in current database search tools are based on a simplified representation of the peptide ion fragmentation process[36,56]; the charge states of the peptides selected for fragmentation are not always known; many MS/MS spectra are of poor quality or are mixtures of the fragmentations of more than one peptide; high quality MS/MS spectra will be incorrectly assigned if their true corresponding peptides are not in the database. Manual verification of peptide assignments is time consuming and is not feasible for the analysis of data sets that may contain tens of thousands of spectra.

The number of incorrect peptide identifications can be reduced by additional processing of the data before searching by, for example, removal of low quality spectra, clustering of redundant spectra and application of charge-state determination algorithms [57-62]. Additionally, the introduction of more advanced scoring systems that include additional knowledge of how peptides fragment [36,56,63] should generate further improvements. In order to attempt to separate correct from incorrect peptide assignments, filtering criteria, based on database search scores and other available data, can be applied. However, different researchers often use their own preferred filtering criteria, making it difficult to compare results between (or even within)

research groups. Software tools such as DTASelect [64], INTERACT [65] and CHOMPER [66], compatible with Sequest and Mascot, are available to facilitate the filtering of data. Yet the rates of false identifications resulting from the use of these filters are rarely estimated [51]. Consequently, the question of what constitutes protein identification in an LC-MS/MS experiment is difficult to answer.

In order to allow the comparison of proteomic data between groups, peptide assignments should be validated using statistical programs developed to be compatible with existing database search tools [51]. The development and application of robust, transparent tools for the statistical analysis of proteomic data is essential. Several computational methods have recently become available [67-72]. Only once these programs have become standardised and widely used can one of the long-term aims of proteomics begin to be realised: the creation of centralised, public databases of peptide and protein identifications with repositories for storing MS data. The combined results from many different research groups could then be merged and applied to the whole genome, validating expressed genes at the protein level, and enabling the elucidation of patterns of protein expression that would be missed in individual experiments.

1.2.5 Protein quantification

In a recent article in *Nature* [26], Aebersold and Mann argue that for the documentation of the expression of proteins as a function of cell or tissue state to be meaningful the data must be at least semi-quantitative, and that a list of the proteins detected in the different states is inadequate.

In both MALDI- and ES-MS the relationship between the concentration of an analyte in a sample and its measured signal intensity is dependent on multiple factors that are difficult to control and is, in fact, not completely understood [26]. Consequently, in the absence of internal standards, mass spectrometers are inherently poor quantitative devices. Quantitation in mass spectrometry-based proteomics is most commonly achieved by the use of stable isotope dilution. This makes use of the fact that two analytes differing only in stable isotope composition will be chemically almost

identical, eluting at very similar retention times, yet can be differentiated in a mass spectrometer due to their mass difference. The ratio of the ion currents for the analyte pair will be equal to the abundance ratio for the two analytes. Therefore relative abundances of proteins in different cell or tissue states can be calculated, and absolute abundances can be measured with the use of isotopically labelled standards. The three main ways of introducing stable isotopes to proteins are via chemical reactions, via enzymatic incorporation of ^{18}O from $H_2{}^{18}O$ during proteolysis, and by metabolic labelling using heavy isotope-labelled salts or amino acids.

The most widely used chemical labelling method is the isotope-coded affinity tag (ICAT) approach [73]. The ICAT reagent contains a biotin affinity tag, a linker that can incorporate stable isotopes and a thiol-specific reactive group. The reagent exists in heavy and light forms. Proteins from two different cell states are harvested, denatured, reduced and labelled at cysteine residues using the light or heavy ICAT reagents. The samples are then combined and digested with trypsin. The ICAT labelled peptides are isolated using the biotin affinity tag by affinity chromatography and analysed by LC-ES-MS/MS. The ratio of the ion intensities for ICAT-labelled pairs of peptides allows the calculation of the relative abundance of the parent protein in the two cell states, whilst MS/MS confirms the sequence of the peptide and identifies the protein. The ICAT strategy is shown in Figure 1.20. However, the original ICAT reagents were relatively large, and the presence of their fragments in the MS/MS patterns of peptides was found to complicate the analysis. To address this, a cleavable ICAT reagent was introduced [74], which contains an acid-cleavable linker that allows the removal of the affinity tag before MS analysis of the peptides. In addition, ^{2}H-labelling was replaced by ^{13}C-labelling, in order to eliminate chromatographic displacement effects. The disadvantage of the ICAT approach is that it is limited to the analysis of cysteine-containing proteins; however, this feature can be utilised to simplify complex mixtures of proteins.

One of the features of the chemical tagging approach is that isotope-tagging reactions can be chosen to direct the isotopes and attached affinity tags to specific functional groups or protein classes, allowing selective isolation and analysis of the tagged peptides. Tagging reactions that are specific for sulphydryl groups [73,75], amino groups [76], the active sites of serine [77] and cysteine hydroxylases [78], phosphate ester groups [79,80] and N-linked carbohydrates [81] have been described.

Enzyme catalysed incorporation of ^{18}O from $H_2{}^{18}O$ during proteolysis [82-84] results in each peptide generated being labelled at its carboxy terminal. The mass difference generated by ^{18}O incorporation is 4 Da, which makes quantitation difficult because isotope patterns for analyte pairs are likely to overlap. However, Heller *et al* [85] have applied this methodology to complex biological samples with some success.

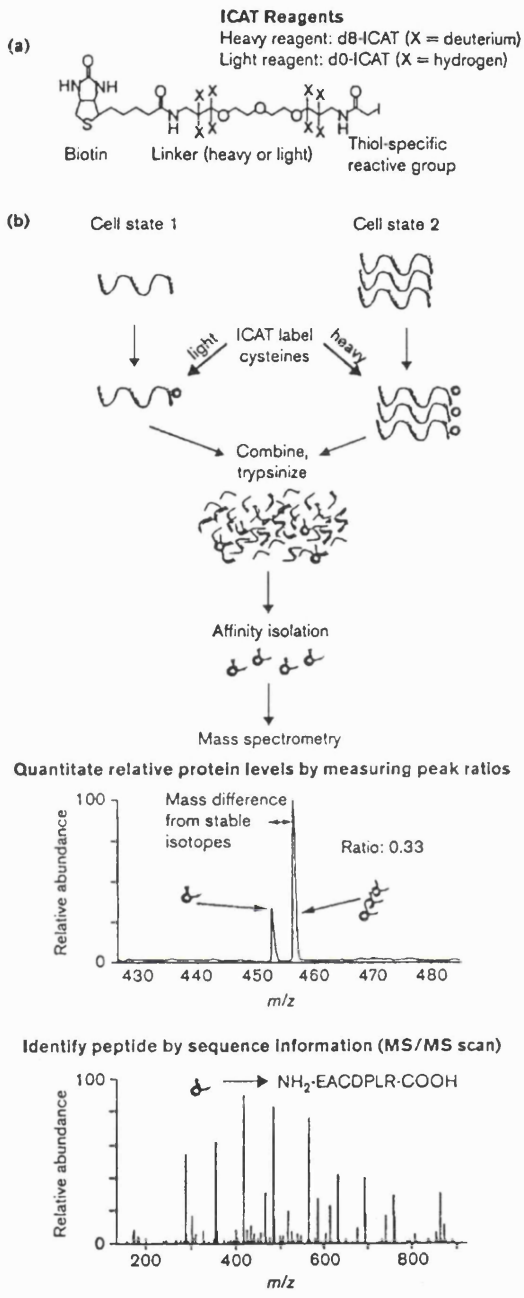


Figure 1.20 The ICAT strategy for quantifying differential protein expression. **(a)** Structure of the original ICAT reagent; **(b)** schematic of the ICAT strategy. Taken from Gygi *et al* [38].

Metabolic isotope labelling was, until recently, limited to the study of single-celled organisms or cells grown in culture. The "stable isotope labelling with amino acids in cell culture", or SILAC, approach [86], involves growing mammalian cell lines in media lacking a standard essential amino acid but supplemented with an isotopically labelled form of that amino acid, for example lysine deuterated at its side-chain methylenes. Potentially, all peptides can be labelled [86]. In other studies [87-90], cells were grown in media enriched with stable isotopes such as ^{15}N. Krijgsveld *et al* [91] have taken this approach one step further by metabolically labelling the nematode *Caenorhabditis elegans* and the common fruit fly *Drosophilia melanogaster* by feeding them on ^{15}N-labelled *Escherichia coli* (*E. coli*) and yeast, respectively. After two generations ^{15}N incorporation was at 98% for the nemotodes and 94-95% for the fruit flies.

A simple approach for absolute quantitation has recently been described by Gerber *et al* [92]. This involves the use of synthetic peptides, specific to the protein of interest, with incorporated stable isotopes. Samples are separated by 1DE and the bands of interest subjected to in-gel tryptic digestion in the presence of known amounts of the labelled peptide(s). Absolute quantitation is then achieved using LC-MS/MS analysis. There is also the potential for whole proteins to be expressed using only ^{15}N or ^{13}C-containing amino acids; these proteins could be introduced to samples prior to SDS-PAGE separation, thereby eliminating any differences between sample and standard due to incomplete trypsinisation.

Stable-isotope dilution methods in combination with LC-ES-MS/MS are increasingly being used to detect changes in protein profiles, for example between diseased and normal tissue states, and to infer biological function from the observed patterns.

1.2.6 Protein interactions

The majority of proteins exert their function via protein-protein interactions. To determine a protein's binding partners by MS, the protein itself can be used as an affinity reagent. There are three crucial components to these MS-based protein interaction experiments: bait presentation, affinity purification of the complex, and

analysis of the bound proteins [26]. If antibodies to the protein of interest are available, these can be used to specifically isolate the protein with its bound partners. Alternatively, the protein of interest can be "tagged" with a sequence recognised by an antibody specific for the tag. The tagged protein can be expressed in stable cell lines and allowed to bind to any other proteins with which it interacts. The protein and its binding partners can be separated from the mixture using the antibody, with subsequent denaturing of the entire complex for analysis by MS. The tandem affinity purification, or TAP, approach combines the use of two different tags on the same protein, separated by an enzyme-cleavable linker sequence [93,94]. Proteins are tagged with a calmodulin binding peptide linked to a tobacco etch virus (TEV) protease site linked to a protein A domain (Figure 1.21). The protein A domain binds tightly to a solid support modified with immunoglobulins. After washing, treatment with TEV protease allows the elution of the bound material, which is then incubated with calmodulin beads in the presence of calcium. This second affinity step allows the removal of any remaining contaminants as well as the TEV protease. The bound material is released with ethylene glycol-bis(2-aminoethylether)-N,N,N',N'-tetraacetic acid (EGTA). The use of TAP tags reduces background noise considerably compared with tags supporting single step purification, but may result in the loss of some of the weaker and more transient binding partners in the second affinity step [26].

Introducing a tagged protein into a cell system can cause problems because tagged protein expression levels are usually different from the expression levels of the untagged endogenous counterparts: artefacts can arise due to non-physiological levels of the tagged protein [26]. Also, many protein-protein interactions are transient, of low affinity and dependent on a specific cellular environment, therefore MS-based experiments of the type described above will only detect a subset of the interactions that actually occur. This is illustrated by a comparison of the results obtained in two different studies. Gavin *et al* [95] introduced 1,739 TAP-tagged genes into the yeast genome; 232 stable complexes were isolated and the proteins identified by MALDI PMF after 1DE separation. In a similar study, Ho *et al* [96] expressed Flag® epitope-tagged proteins; single-step immunopurification was used to isolate the complexes, followed by LC-MS/MS for identification of proteins separated by 1DE. Both studies reported a large number of interacting proteins, but, interestingly, for the group of

bait proteins that was common to both studies, surprisingly little overlap of data was observed, either between the two MS studies or between these and previous yeast two-hybrid studies [97]. Both MS studies reported results consistent with existing literature for known protein complexes. These results point towards the potentially huge scale of the "interactome", and indicate that it is largely under-sampled in current studies.

In the future, stable isotope methods are likely to revolutionise the study of protein-protein interactions. Stable isotope ratios can be used to distinguish between the protein composition of two or more complexes, and between true complex components and non-specifically associated proteins [98,99]. This ability to distinguish complex components from a background of non-specific binding will allow for fewer purification steps and less washing, thereby increasing the chances of detecting weak and transient interactions.

MS-based proteomics has also been used for the study of large protein complexes such as the yeast and human spliceosomes [100,101], the yeast nuclear pore complex [102] and the human nucleolus [103].

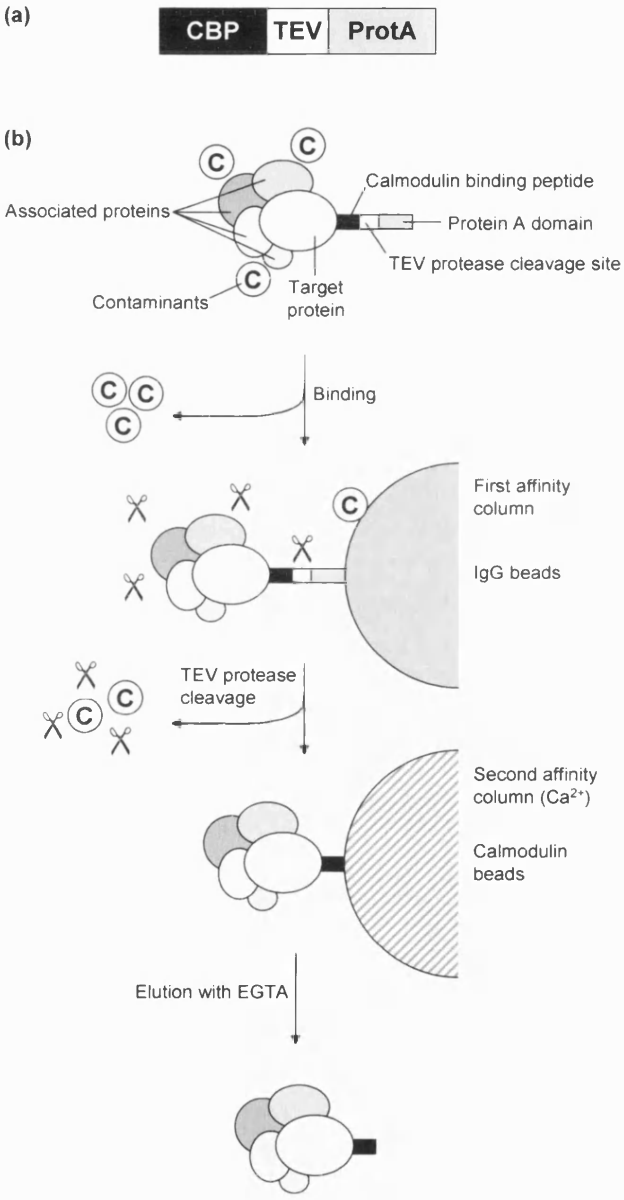


Figure 1.21 The TAP approach to the detection of protein interactions. **(a)** Schematic representation of the TAP tag; **(b)** overview of the TAP purification strategy. Adapted from Puig *et al* [94].

1.2.7 Protein modifications

Post-translational modifications (PTMs) are chemical processing events that alter the properties of a protein after its translation, either by proteolytic cleavage or by addition of a modifying group to one or more amino acids. PTMs include phosphorylation, glycosylation, acetylation, methylation, sulphation, disulphide bond formation, deamidation and ubiquitination [104]. Most eukaryotic proteins are post-translationally modified, and many of these PTMs are regulatory and reversible, most notably protein phosphorylation, which is a dynamic process with complex kinetics involving several amino acids in a single protein, and which controls biological function through many different mechanisms [26].

The application of mass spectrometry to the identification of PTMs ranges from the study of single, purified proteins through the search for one type of modification on all the proteins in a sample to scanning for all modifications on a proteome-wide scale. The complexities involved in identifying all the modifications even on a single protein mean that proteome-wide scanning is, at present, not comprehensive; nevertheless large amounts of biologically useful information can be generated. To determine the sites of modification, maximum protein sequence coverage is desirable; for this purpose, peptide mass mapping using two or more different enzymes can be employed, for example trypsin, Asp-N and Glu-C. Protein modifications are then identified from the measured masses and fragmentation spectra using manual or computer-assisted interpretation [26]. However, the introduction of a number of possible modifications into computer search programs such as Sequest and Mascot can result in a "combinatorial explosion" due to the need to consider all possible modifications for all peptides in the database [104]. To avoid this the experiment can be divided into a two stage process whereby a set of proteins is first identified on the basis of non-modified peptides, then only these proteins are searched for modifications. MacCoss *et al* [105] used an extension of the MudPIT approach to identify 73 sites of modification on 11 different crystalline proteins from human lens tissue.

Other approaches that attempt to address the low-stoichiometry and high-complexity problems associated with the analysis of PTMs, and protein phosphorylation in particular, involve the selective enrichment of modified proteins. These techniques are generally based on some form of affinity selection that is specific for the modification of interest, and that is used for the purification of modified proteins. Pandey *et al* [106-108] immunoprecipitated tyrosine phosphorylated proteins using antibodies specific for phosphotyrosine. Affinity purification has been combined with chemical modification, for example Oda *et al* [80] replaced phosphate moieties with affinity tags; Zhou *et al* [79] used a sequence of chemical reactions to isolate phosphopeptides, and Goshe *et al* [109,110] employed a phosphoprotein isotope-coded affinity tag. Probably the most extensive characterisation of the phosphoproteome was achieved by Ficarro *et al* [111], who esterified peptide mixtures to nullify negatively charged carboxyl groups, then captured the phosphopeptides by immobilised metal-affinity chromatography; more than 1000 phosphopeptides were detected from analysis of a whole-cell lysate from *Saccharomyces cerevisiae*, and a total of 216 peptide sequences defining 383 sites of phosphorylation were determined. Peng *et al* [112] used affinity purification to isolate the ubiquitinated proteins from yeast cells expressing 6xHis-tagged ubiquitin; 1075 proteins were identified, including 110 precise ubiquitination sites in 72 ubiquitin-protein conjugates. Two papers that identify N-glycosylated proteins have recently been published by Zhang *et al* [81] and Kaji *et al* [113]; both describe methods to immobilise the carbohydrate chain-containing proteins then, after tryptic digestion, release the N-glycosylated peptides from their carbohydrate chains (using peptide-N-glycosidase F, or PNGase F) for analysis.

Developing mass spectrometric technologies are expected to substantially accelerate the analysis of PTMs. FTMS instruments can be coupled with electron capture dissociation (ECD), which has been shown to have profound potential for the characterisation of PTMs [114]. ECD occurs after recombination of multiply protonated protein or peptide molecules with thermal electrons. After electron capture, the resultant $[M + nH]^{(n-1)+\bullet}$ ion dissociates via energetic $H^{\bullet}$ transfer to the backbone carbonyl to form c (-CHR-C(OH)=NH) and $z^{\bullet}$ ($^{\bullet}$CHR'-) ions [115,116]. Labile PTMs, which are easily lost under CAD conditions, remain intact during the ECD process, allowing the positions of modifications to be easily established; this is attributed to

the high rate of bond cleavage and moderate amount of excess energy during ECD compared with the vibrational excitation involved in CAD [114]. FTMS and ECD have also been used for "top-down" protein sequencing, in which the intact protein ion is fragmented inside the mass spectrometer, theoretically allowing the mapping of all PTMs [117,118].

1.2.8 Future challenges and expectations

"The specific objective of proteomics is to concurrently identify, quantify and analyse a large number of proteins in a functional context" [26]. The global focus of analyses on a proteome-wide scale threatens to result in information overload, and throws up many challenges in terms of data collection, analysis and interpretation, visualisation, storage, and data publication and sharing.

Proteomic studies generate huge amounts of data: in a typical LC-MS/MS experiment over 1000 CAD spectra of varying degrees of quality are generated in an hour. However, even if each one of these spectra led to the successful identification of a peptide, it would take a long time to analyse complete proteomes. High throughput collection of consistently high quality data remains a challenge, and it has been suggested [119] that "national proteome centres" be established to ensure availability of expertise and equipment.

The analysis and interpretation of the massive volumes of proteomic data is a major bottleneck in proteomics today. Expert manual analysis of data is incompatible with the thousands of spectra produced in a proteomics experiment; in addition, inconsistency exists between individuals. Therefore, the development of robust, transparent tools for the statistical analysis of proteomic data is crucial. Only once these tools have become standardised and widely accepted will the comparison of complementary proteomic data sets generated in different laboratories be possible. Such comparisons will also depend on well organised and accessible systems for data storage, communication and visualisation; the development of these tools for proteomics is still in its infancy [26].

The current system for publication of scientific work does not lend itself well to the reporting of large datasets generated by proteomic experiments. Data tables are frequently published as supplementary to the main paper, and validation and discussion limited to only a few conclusions drawn from the data. In order to make the publication of proteomics data more useful, new ways must be devised to review and validate large data sets and to make their content electronically searchable. Preliminary developments by a few publishers and journals are described by Mann [120]; however, this problem remains essentially unsolved.

The sheer scale of proteomics research necessitates a community effort. Public access to online raw data would facilitate research in all areas, and enable those researchers with no access to MS data to lend their expertise to the development of more advanced computational methods and software tools. In a recent *Nature* commentary [121], Aebersold proposes a community-wide strategy involving the synthesis of one unique heavy stable isotope-tagged peptide marker for each protein, protein isoform or specifically modified form of a protein, which could then be used as a definitive marker for that protein. These markers would be added in precise amounts to samples where the proteins of interest had been labelled with natural-isotopic tags, and used for protein quantitation. This would drastically reduce the number of peptides that need to be analysed, and would trivialise the data analysis itself since *de novo* identification would no longer be necessary. The method would be easily standardised between laboratories, and the absolute quantitation would make different data sets easily comparable. Any subset of proteins could be interrogated, and, providing an appropriate reference peptide could be synthesised, any protein isoform, polymorphism or modified form could be analysed. Finally, once the initial cost of synthesising the reference peptides had been met, the method would be relatively cheap since only tiny amounts of the peptide standards would be needed for each assay.

MS-based proteomics is, however, still an emerging technology where radical change is possible, and for this reason the call for the standardisation of research approaches may be premature. Current instrumentation is far from optimal, partly because in the rapid explosion of proteomics research manufacturers have not had the time to fully develop instruments that are perfectly tailored for protein analysis.

Mass spectrometers are still far from the physical limit of the few ions needed to register a peak, so a massive increase in performance can be expected in the next few years. There are several nascent techniques with the potential to revolutionise current MS-based proteomic practices. One of these, the analysis of intact proteins, was mentioned in Section 1.2.7. The emergence of whole-protein proteomics is being enabled by the development of mass spectrometers with large mass ranges, extremely high mass accuracy and resolution, and ionisation/fragmentation techniques compatible with large proteins; this has the potential to provide insights into modifications that are difficult to characterise by peptide analysis. Another promising technique is that of imaging mass spectrometry, in which thin tissue sections are mounted onto a target plate, spotted with matrix then analysed by MALDI [122]. From the intensity of a given *m/z* value, a density map or image is constructed; in this way, profiles of the proteins contained in the section can be obtained. The technique has already been applied to the assessment of protein patterns in multiple diseased tissues including human gliomas and nonsmall cell lung cancer [123]. A third concept is the use of peptide "accurate mass and time" (AMT) tags for high-throughput proteome characterisation [124-127]. This strategy involves the use of high mass accuracy and resolution FTMS to validate peptide AMT tags from "potential mass tags" identified from global enzymatic digestion of a specific organism, tissue or cell type using conventional MS/MS techniques. Once the peptides have been validated they can be identified in subsequent experiments by simply correlating the accurate mass and the retention time with the database of previously determined AMT tags, thus obviating the need to perform MS/MS analyses. The advantage of using this method over MS/MS-based techniques is the large number of protein identifications that can be obtained in a single LC run: the number of peptide identifications from an LC-MS/MS run is restricted because of the time needed to fragment the peptides individually.

1.3 Cytochromes P450

The cytochromes P450 (P450s) are a superfamily of haem-thiolate enzymes, members of which are present in almost all living organisms. As of January 19th 2004 there are 3043 named P450 sequences: 1277 in animals, 1098 in plants, 207 lower eukaryote sequences and 461 bacterial sequences; there are many more that are not yet named [128]. The majority of P450s are known only from DNA sequences extracted from genome databases; a small proportion have been purified or expressed in recombinant systems [129]. Only an extremely small number of P450s have been crystallised, including five mammalian P450s: rabbit CYP2C5 [130,131] and CYP2B4 [132], and human CYP2C8 [133], CYP2C9 [134] and CYP3A4 [135]. Mammalian P450s are membrane-bound and hydrophobic in character, and are therefore very difficult to crystallise. The structure of the most recent human P450 to be crystallised, CYP2C9, is shown in Figure 1.22.

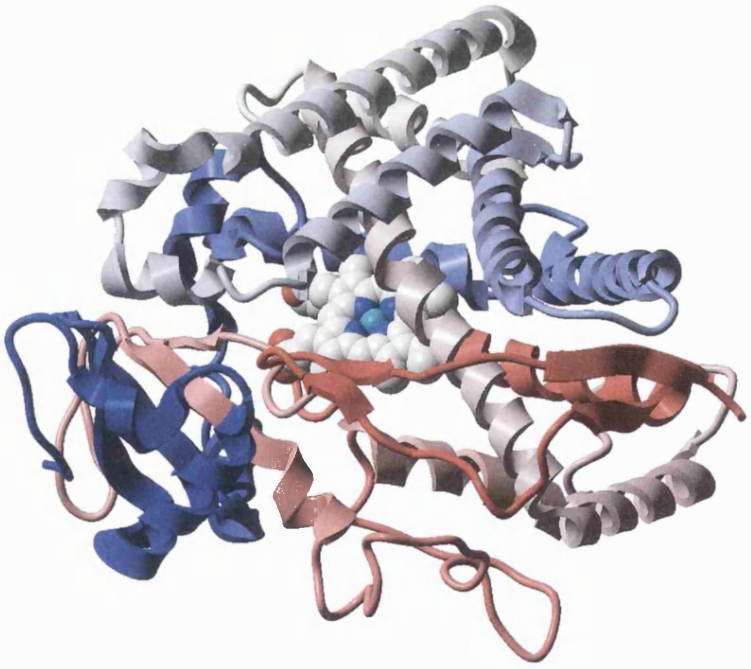


Figure 1.22 The crystal structure of human CYP2C9. Taken from Williams *et al* [134].

P450 enzymes are the most important drug-metabolising enzymes in mammals. Although they are concerned mainly with the oxidative detoxification of drugs and other xenobiotics, they are also involved in the oxidative metabolism of endogenous substrates such as fatty acids, steroids, bile acids, fat soluble vitamins and

eicosanoids. They are found in virtually all mammalian tissues in two cellular locations, the endoplasmic reticulum (ER) and the mitochondria; both types are membrane bound. The highest concentrations of P450s are found in the liver, but they are also present in the kidney, lung, gonads, adrenal, brain, breast, nasal epithelium, placenta, pancreas, spleen, gastrointestinal tract and skin [136]. The completion of the draft of the human genome revealed the presence of about 106 different P450 genes, of which 57 are functional (50 ER P450s and 7 mitochondrial P450s) whereas the remainder are pseudogenes [137,138].

1.3.1 Origins and evolution

It is generally believed that the entire P450 family of enzymes evolved over at least 1400 million years from an ancestral P450 gene, most probably arising in a primeval prokaryotic organism in response to the increasing availability of atmospheric oxygen for the metabolism of carbon sources [136]. The original biological role of cytochrome P450 was the detoxification of tissue dioxygen (O_2), eventually to form water. As primitive organisms evolved, a differentiation into plant and animal species occurred, leading to an era of co-evolution, or "plant/animal warfare," that lasted hundreds of millions of years. This resulted in the plant species synthesising protective toxins (phytoalexins) to prevent their consumption by animal species, and the animal species in turn developing defensive measures to detoxicate the phytoalexins, thereby ensuring survival. Cytochrome P450 played a major role in both these evolutionary processes, developing into a huge family of closely related, yet distinctly different, enzymes [136].

1.3.2 P450 nomenclature

The name P450 comes from P for pigment and 450 for the absorption maximum of the complex formed when CO binds to the reduced (Fe^{2+}) state of the protein to give the Fe^{II}-CO complex. The absorption maximum at ~450 nm results from the presence of the thiolate ligand; the corresponding absorption maximum in haemoproteins with the more common histidine ligand is at ~420 nm [139].

P450s are characterised into families and subfamilies by their sequence similarities: sequences that have greater than 40% amino acid homology belong to the same family; sequences that have greater than 55% amino acid homology belong to the same subfamily [137]. Individual proteins within subfamilies may have up to 97% homology [129]. P450 enzymes are named using a number to designate the family, followed by a letter for the subfamily and finally another number for the individual protein. A polymorphism within a gene is defined as such if it occurs at an allele frequency of at least 1% [129]. There are over 360 different families of P450s [128]. Humans have 18 families of P450 genes and 43 subfamilies [137]; these are shown in Table 1.1.

Table 1.1 Human P450 families and their functions. Adapted from Nelson [137] and Ingelman-Sundberg [138].

Family	No. of Sub-families	No. of genes	No. of pseudo-genes	Function
CYP1	3	3	1	Drug and pre-carcinogen metabolism
CYP2	13	16	20	Drug, pre-carcinogen, solvent and steroid metabolism
CYP3	1	4	5	Drug and pre-carcinogen metabolism
CYP4	5	12	18	Fatty acid and drug ω-side chain hydroxylation
CYP5	1	1	0	Thromboxane A2 synthesis
CYP7A	1	1	0	Bile acid biosynthesis, 7-α hydroxylation of steroids
CYP7B	1	1	0	Brain-specific 7-α hydroxylation of steroids
CYP8A	1	1	0	Prostacyclin synthesis
CYP8B	1	1	0	Bile acid biosynthesis
CYP11[a]	2	3	0	Steroid biosynthesis
CYP17	1	1	0	Steroid biosynthesis
CYP19	1	1	0	Steroid biosynthesis
CYP20	1	1	0	Unknown
CYP21	1	1	1	Steroid biosynthesis
CYP24[a]	1	1	0	Vitamin D degradation
CYP26A	1	1	1	Retinoic acid hydroxylation
CYP26B	1	1	0	Probable retinoic acid hydroxylation
CYP26C	1	1	0	Probable retinoic acid hydroxylation
CYP27A[a]	1	1	0	Bile acid biosynthesis
CYP27B[a]	1	1	0	Vitamin D3 1-α hydroxylation
CYP27C[a]	1	1	0	Unknown
CYP39	1	1	0	7-α hydroxylation of 24-hydroxy cholesterol
CYP46	1	1	0	Cholesterol 24-hydroxylation
CYP51	1	1	3	Cholesterol biosynthesis, lanosterol 14-α demethylation

[a] Mitochondrial P450s

1.3.3 P450 catalytic cycle and oxygenation mechanism

P450s are haemoproteins containing a single haem prosthetic group (iron protoporphyrin IX). The haem Fe atom is ligated to a cysteine residue, which is situated close to the C-terminus of a single polypeptide chain of apoprotein, composed of approximately 400 to 500 amino acids [136]. In the normal resting state of the enzyme, the haem moiety comprises low spin Fe^{3+}, primarily with the bound cysteine in the deprotonated anionic form, and with water occupying the haem pocket and probably also bound as the distal 6th ligand [136] (Figure 1.23).

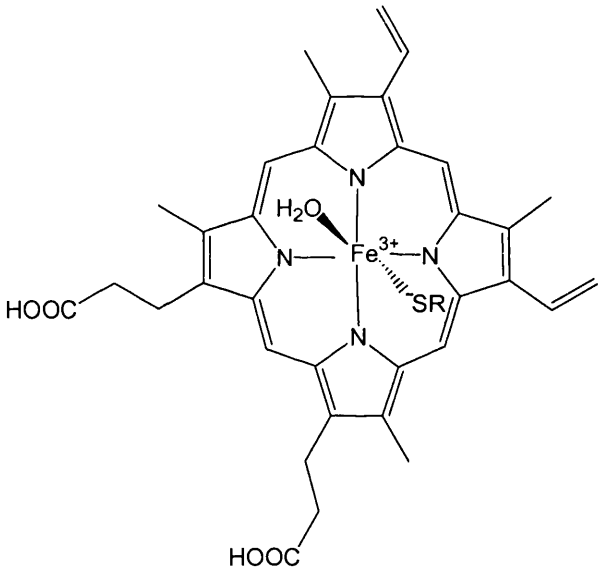


Figure 1.23 Haem group with S from bound cysteine in the deprotonated form and water as the 6th distal ligand. R = P450 apoprotein.

The stoichiometric equation for P450-catalysed monooxygenation of a typical substrate (RH) to form a hydroxyl metabolite (ROH) can be written as follows:

$$RH + O_2 + 2e^- + 2H^+ \longrightarrow ROH + H_2O$$

The sequential input of two electrons required to permit the reaction are supplied via the mediation of a redox partner; in microsomal systems two electrons are accepted sequentially from NADPH via cytochrome P450 reductase (a flavin adenine dinucleotide (FAD)- and flavin mononucleotide (FMN)-containing flavoprotein), although mammalian microsomal P450s also accept electrons from NADPH via

cytochrome b_5 for the second electron transfer step [140]. In mitochondrial and most bacterial systems, the redox partner utilised is an iron-sulphur redoxin, which employs an FAD-dependent reductase to transfer electrons from NADPH (mitochondrial systems) or NADH (bacterial systems) [141]. The crystal structures of the P450 enzymes solved to date do not reveal the presence of a distinct acidic group within close proximity of the haem to act as a proton donor for the reaction; however, it has been proposed that water molecules involved in a hydrogen bonded network between active site residues could serve this purpose [129].

The P450 catalytic cycle (Figure 1.24) involves binding of the substrate to the active site of the enzyme, bringing about the displacement of the bound water molecule and shifting the haem Fe spin state equilibrium in favour of high spin Fe^{3+}. The accompanying conformational change in the P450 protein brings about favourable binding with its redox partner and lowers the P450 redox potential, thus facilitating electron transfer from NADPH via the reductase protein. Oxygen binds rapidly to the reduced (Fe^{2+}) protein in the sixth ligand site, and the Fe^{2+} returns to its low-spin state. The P450-dioxygen complex is likely to be a mixture of the two resonance forms $Fe^{2+}O_2$ and $Fe^{3+}O_2^-$ [141]. A second electron transfer process then results in the formation of $Fe^{2+}O_2^-$, which is likely to be in equilibrium with $Fe^{3+}O_2^{2-}$. The ferric peroxide state ($Fe^{3+}O_2^{2-}$) is unstable and can dissociate to form the resting Fe^{3+} enzyme and hydrogen peroxide; in the absence of bound substrate this readily occurs, but the dissociation is diminished by substrate occupancy of the haem site. The final steps in the P450 catalytic cycle, involving the insertion of an oxygen atom into the substrate, are the least well understood and are the subject of ongoing mechanistic debate; hard experimental evidence has been difficult to obtain because of the rapid rates of the reactions occurring [141]. Several structures have been put forward as candidates for the active oxygenating species; three of these are shown in Figure 1.25. The particular intermediate involved may depend on the type of reaction being catalysed.

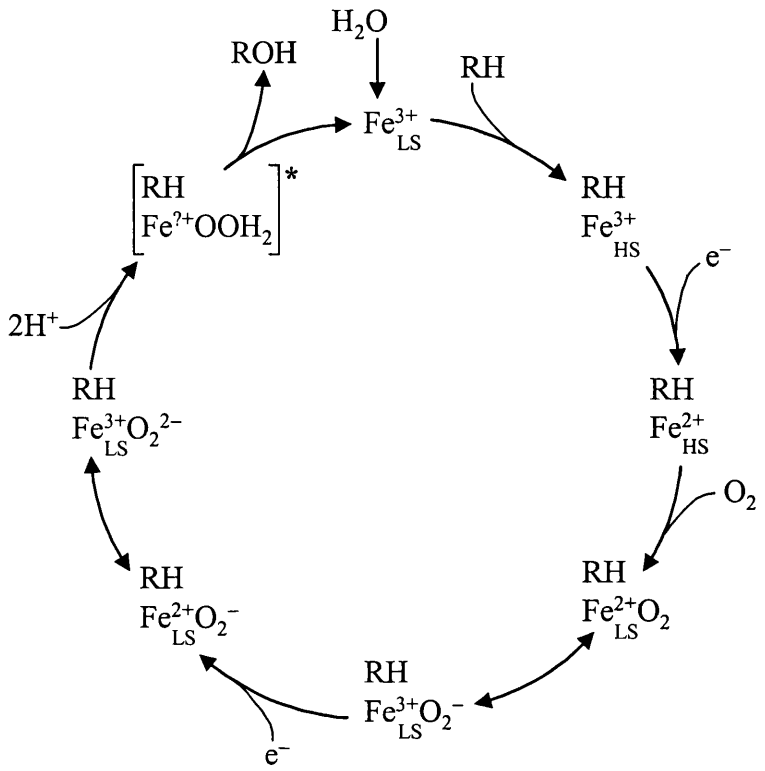


Figure 1.24 The P450 catalytic cycle, showing stages in the overall process for oxygenation of a typical substrate. HS = high spin; LS = low spin; RH = substrate; ROH = product. * For possible structures for the active oxygenating species, see Figure 1.24. Adapted from Lewis [141].

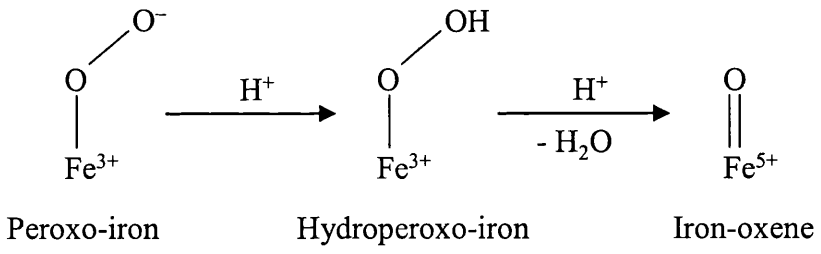


Figure 1.25 Likely active species in the P450 oxygenation reaction. Adapted from Lewis [141].

1.3.4 P450s in drug metabolism and pharmacogenetics

Evolution of the P450 system has led to a super-family of enzymes capable of metabolising practically any lipophilic molecule [129]. The human P450s can be divided into three major groups [138]:

1. Those in CYP families 5 to 51, important for the metabolism of endogenous compounds, usually with high affinity for their substrates and relatively well conserved during evolution.
2. Those in CYP families 1 to 3, usually with less affinity for their substrates, less well conserved during evolution and which exhibit important genetic polymorphisms; primarily involved in xenobiotic metabolism but also in some endogenous metabolism.
3. Those in CYP family 4, with roles in both fatty acid metabolism and metabolism of some xenobiotics.

The P450s in families 1 to 3 are responsible for 70-80% of all phase I dependent metabolism of clinically used drugs and participate in the metabolism of an enormous number of xenobiotic chemicals [138,142] (Figure 1.26).

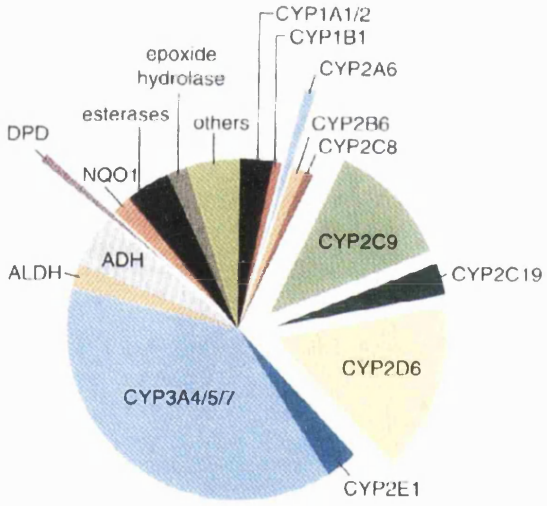


Figure 1.26 Enzymes contributing to the phase I metabolism of drugs. The percentage of phase I metabolism of drugs that each enzyme contributes can be estimated by the relative size of each section of the chart. Enzymes with polymorphisms that have been associated with changes in drug effects are separated from the pie chart. ADH = alcohol dehydrogenase; ALDH = aldehyde dehydrogenase; DPD = dihydropyrimidine dehydrogenase; NQO1 = NADPH:quinine oxidoreductase or DT diaphorase. Taken from Evans and Relling [142].

CYP3A4, the most abundant P450 protein, shows the greatest substrate diversity; this has been ascribed to a large lipophilic binding site capable of binding substrates in multiple orientations [129]. In addition, many substrates can be metabolised by more than one P450 enzyme or at more than one site by the same enzyme. P450-mediated reactions are primarily detoxification processes; however, certain substrates are

metabolically activated, resulting in the generation of reactive intermediates with increased toxicity or mutagenicity, such as pre-carcinogens. There are also the effects of induction and inhibition: enzyme inducers are often metabolised by the enzymes they induce and so accelerate their own metabolism. However, inducers will also affect the metabolism of other substrates of the same enzyme. Enzyme inhibition is usually the result of competition between two substances for metabolism by the same P450 enzyme [143]. Usually inhibition is reversible, but irreversible inhibition does occur, causing destruction of the enzyme. The subsequent short-term impairment in metabolism may be compensated by increased production of the affected enzyme, but longer-term problems can ensue [143]. These factors can all lead to adverse drug reactions.

Comparison of the number of pseudogenes for the different P450 families (Table 1.1) shows a difference between the characteristics of P450 genes important for xenobiotics metabolism (CYP1-3) and those encoding proteins that metabolise endogenous compounds: the CYP2 family has the largest number of pseudogenes (20), followed by the CYP4 family (18) and the CYP3 family (5); only four pseudogenes have been found among genes encoding P450s metabolising endogenous substrates [138]. This indicates a high level of gene inactivation in genes encoding xenobiotics metabolism, presumably as a result of adaptation to the environment. In addition, all the P450 genes in families 1-3 are functionally polymorphic, with the exception of CYP1A1, 2E1 and 3A4, which are relatively well conserved; in contrast, only 6 of the P450s of importance for the metabolism of endogenous compounds have been found to exhibit genetic polymorphism. This is because genes whose products are active in the metabolism of environmental agents tend to exhibit higher genetic variation [138].

Interindividual variation in P450 genes and P450 expression has implications for drug and xenobiotic response, tolerance and clearance, and can have toxic consequences.

1.3.4.1 Interindividual variation in levels of P450 enzymes

Shimada *et al* [144] and Snawder and Lipscomb [145] both studied interindividual variations in human liver P450 enzymes. Shimada *et al* [144] characterised the level and catalytic activity of P450 enzymes CYP1A2, 2A6, 2B6, 2C, 2D6, 2E1 and 3A in liver microsomes from 60 humans, 30 Japanese and 30 Caucasians, using immunoblotting and activity assays. They found large interindividual variations in relative levels of different P450 enzymes, for example the relative content of CYP1A2 as a percentage of the total P450 content in liver ranged from around 3% to almost 30%. Correlations were observed between the levels of individual P450 forms and the activities of several P450-dependent drug monooxidations and pre-carcinogen activations. They also found that the total P450 content in liver was higher in Caucasian (mean 0.43 ± 0.2 nmol/mg protein) than in Japanese populations (mean 0.26 ± 0.1 nmol/mg protein) and that clear race-related differences were observed in the relative levels of CYP2B6: Japanese samples contained only about a quarter of the level of CYP2B6 of that in the Caucasian samples. Oxidations of several drugs and activation of several pre-carcinogens showed significantly higher activities in Caucasian than Japanese samples.

Snawder and Lipscomb [145] used enzyme-linked immunosorbent assays (ELISAs) to quantify P450 forms 1A, 2B6, 2C, 2E1 and 3A in microsomal samples from 40 human donors; they also found correlation between the activities of P450-dependent drug monooxidations and the levels of individual P450 enzymes.

1.3.4.2 Polymorphism in P450 genes

Almost all genes encoding P450 enzymes in families 1 to 3 are polymorphic. The major polymorphic forms of the P450 enzymes that are important in drug metabolism are shown in Table 1.2.

Table 1.2 Polymorphisms in P450 enzymes important in xenobiotic metabolism and their functional effects. Adapted from Ingelman-Sundberg [138,146,147].

Enzyme	Substrates	Major allelic variants	Frequency of variant alleles	Clinical effects of polymorphism
CYP1A2	Drugs, carcinogens	*CYP1A2*1K*	Relatively high	Less enzyme expression and inducibility
CYP2A6	Nicotine, drugs, carcinogens	*CYP2A6*4, CYP2A6*9*	High in Orientals; less frequent in Caucasians	Altered nicotine metabolism
CYP2B6	Drugs	-	Relatively low	Reduced drug metabolism
CYP2C8	Drugs	*CYP2C8*3*	High	Altered taxol metabolism
CYP2C9	Drugs	*CYP2C9*2, CYP2C9*3*	Relatively low	Drug dosage [a]
CYP2C19	Drugs	*CYP2C19*2, CYP2C19*3*	High	Drug dosage [a], drug efficacy
CYP2D6	Drugs	*CYP2D6*2xn, CYP2D6*4, CYP2D6*10, CYP2D6*17, CYP2D6*41*	High	Drug dosage [a]
CYP2E1	Carcinogens, solvents, drugs	-	High	No conclusive studies
CYP3A4	Drugs, carcinogens	Rare	Low	No conclusive studies
CYP3A5	Drugs	*CYP3A5*3*	High	No conclusive studies

[a] The dose of the drug should be adjusted depending on the genotype with respect to the individual enzyme.

Polymorphic enzymes mediate approximately 40% of P450-dependent drug metabolism, which can make drug dosing problematic. Clinically, the most important polymorphisms are seen with CYP2C9, 2C19 and 2D6 [147]. Polymorphic forms of P450 enzymes can result in four general phenotypes: poor metabolisers (PMs), who lack the functional enzymes; intermediary metabolisers (IMs), who are either heterozygous for one deficient allele or carry two alleles that cause reduced activity; extensive metabolisers (EMs), who have two normal alleles; and ultra-rapid metabolisers (UMs), who have multiple gene copies. For CYP2D6, the rate of metabolism for a certain drug can differ 1000-fold between PMs and UMs [147]. However, at present, dosing levels for drugs dependent on CYP2D6 are not adjusted for phenotype but are instead based on population averages; hence the 5.5% of the European population who are UMs will not exhibit a response to a drug metabolised by CYP2D6 because of too rapid metabolism of the drug, and the 7% of the European population who are PMs are likely to experience adverse drug reactions as a result of dosages that are too high. Twenty to 30% of clinically used drugs are metabolised by CYP2D6, therefore 35 to 50 million people in Europe are excluded from the therapeutic benefits of these drugs simply because they do not receive dosages adjusted for their CYP2D6 phenotype [147].

There is no doubt that the polymorphism of metabolising enzymes, particularly P450s, has the largest effect on interindividual variability of drug response [138,142,147]. These polymorphisms affect the response of individuals to drugs used in the treatment of many disorders, including depression, psychosis, cancer, cardiovascular disorders, ulcer and gastrointestinal disorders, pain and epilepsy [147].

1.3.5 P450s in cancer

Historically, liver P450s were thought to contribute predominantly to the metabolic conversion of anticancer drugs, either to produce inactive metabolites or necessary bioactivation products [148]. P450s are expressed in a wide range of tumours, including colon, breast, lung, liver, kidney, prostate, ovarian, bladder and gastrointestinal tract (Table 1.3), yet their role in drug metabolism has had little attention. Assuming that the P450 forms expressed in tumours are functionally active, then the relative activity

of liver and tumour drug metabolism will have a major impact in determining therapeutic outcome. In addition, the presence of active P450 enzymes in tumours raises the attractive possibility of the development of prodrugs designed to be selectively activated by tumour P450s, thus maximising the cancer-directed cell kill whilst leaving normal tissues relatively unharmed.

Table 1.3 P450 expression in tumours. Taken from Patterson and Murray [148].

Tumour type, no. of tumours	P450 expression: % of positive tumours [a]		
	CYP1A	CYP2C	CYP3A
Bladder (invasive transitional cell carcinoma), n=25	68%	28%	68%
Breast (invasive ductal carcinoma), n=54	40%	0%	30%
Colon (invasive adenocarcinoma), n=28	75%	n.d.	61%
Kidney (clear cell carcinoma), n=30	100%	0%	100%
Oesophagus, n=50	64%	26%	72%
Liver (hepatoma), n=31	n.d.	n.d.	42%
Ovary (including serous and mucinous carcinoma), n=170	n.d.	n.d.	60%
Prostate (adenocarcinoma), n=51	63%	25%	60%
Sarcomas, n=37	70%	n.d.	78%
Stomach (adenocarcinoma), n=39	60%	0%	30%

[a] Determined by immunohistochemistry; n.d. = not determined.

Several chemotherapeutic agents have retrospectively been identified as P450-activated prodrugs; some examples are shown in Table 1.4. It is likely that the contribution of tumour P450 activation to the efficacy of P450-activated prodrugs is largely underestimated [148].

Table 1.4 Prodrugs activated by P450s. Adapted from Rooseboom *et al* [149] and Patterson and Murray [148].

Prodrug	Drug	P450 enzyme	Mode of action
Cyclophosphamide	Phosphoramide mustard	CYP2B6, CYP2C9	DNA alkylation
Ifosphamide	Isophophamide mustard	CYP2B6, CYP3A4, CYP2C9	DNA alkylation
Tegafur	5-Fluorouracil	CYP2A6, CYP1A1, CYP2C9	Thymidylate synthase inhibitor/incorporated into DNA and RNA
Dacarbazine	HHMTIC [a]	CTP1A1, CYP1A2, CYP2E1	DNA alkylation
Procarbazine	Not known	CYP1A, CYP2B	DNA alkylation
Flutamide	2-Hydroxy flutamide	CYP1A2, CYP3A4	Non-steroidal anti-androgen
Tamoxifen	4-Hydroxy tamoxifen	CYP2C9, CYP2D6, CYP3A4	Oestrogen receptor agonist
MMDX [b]	Not known	CYP3A	DNA cross-linking
DF 203 [c]	Not known	CYP1A1	DNA adduction
AQ4N [d]	AQ4 [e]	CYP3A4	Topoisomerase II inhibitor

[a] HHMTIC = 5-(3-hydroxymethyl-3-methyltriazen-1-yl)imidazole-4-carboxamide; [b] MMDX = methoxymorpholinyldoxorubicin; [c] DF 203 = 2-(4-amino-3-methylphenyl)benzothiazole; [d] AQ4N = 1,4-bis{[2-(dimethylamino-*N*-oxide)ethyl]amino}-5,8-dihydroxyanthracene-9,10-dione, or anthraquinone di-N oxide; [e] AQ4 = 1,4-bis{[2-(dimethylamino)ethyl]amino}-5,8-dihydroxyanthracene-9,10-dione.

A cancer chemotherapeutic prodrug would ideally be completely inactive until metabolised by a tumour-specific enzyme or by an enzyme that is only able to metabolise the prodrug under physical conditions unique to the tumour. The development of 2-(4-amino-3-methylphenyl)benzothiazole (DF 203) (Figure 1.27) and related compounds [150-152] illustrates the opportunity for the development of tumour-selective agents. These compounds are able to induce CYP1A1 in a number of cancer cell lines, including some originating from breast, ovarian, renal, colon and melanoma. After pre-treatment of cells with DF 203 and concurrent induction of CYP1A1, DNA adduction occurs, causing acute cytotoxicity [148]. The active intermediate is not known. Several human tumours have been shown to constitutively express CYP1A1 [148], although this may not be a necessary requirement given the induction of CYP1A1 in tumour cells by these agents. The absence of CYP1A1 in liver demonstrates that hepatic activation would not thwart this opportunity for tumour-specific prodrug activation.

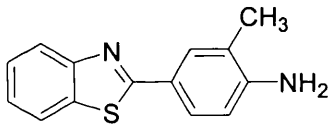


Figure 1.27 Structure of the selective anti-tumour agent DF 203.

Anthraquinone di-N oxide (AQ4N) was designed as a tumour-specific prodrug that makes use of the hypoxic conditions in many solid tumours. In the absence of oxygen, P450 enzymes can act as reductases [153]. AQ4N is inactive as an antitumour agent until it is 4-electron reduced to its cytotoxic metabolite AQ4 (Figure 1.28), a high affinity DNA binding agent and potent topoisomerase II inhibitor. Bioactivation of AQ4N to AQ4 is catalysed in humans by CYP3A4 [154]. The reduction of AQ4N is inhibited in the presence of oxygen since the reduced haem preferentially binds O_2 (see the P450 catalytic cycle, Figure 1.24). AQ4N is currently in Phase I clinical trials in oesophageal cancer, which is one of a large number of tumours known to express CYP3A4.

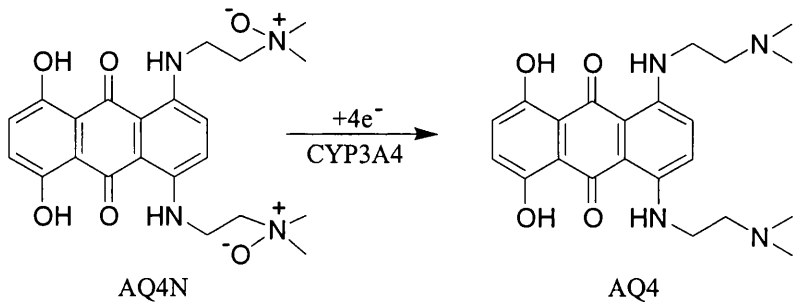


Figure 1.28 The anticancer prodrug AQ4N and its reduction to active metabolite AQ4.

It is important to determine the contribution of tumour P450s to the activation of currently used clinical agents such as cyclophosphamide, dacarbazine and tegafur (Table 1.4). The recognition of P450 expression in tumours should facilitate the search for tumour-selective prodrugs. However, not all tumours are positive for P450 expression (Table 1.3), therefore tumour P450 expression should be determined for individual patients prior to P450-based prodrug therapy. Furthermore, the possibility that polymorphic expression of P450s in tumours could result in some tumours having increased rates of metabolism should be explored, since this is apparent for P450 expression in normal tissues (Section 1.3.4.2).

1.4 Proteomics of membrane proteins and cytochromes P450

Mammalian P450s are integral membrane proteins. They have a membrane anchor at their N-terminus, a hydrophobic region 20-40 amino acid residues in length, which spans the phospholipid bilayer [136]. The other parts of the protein protrude from the membrane into the cytoplasm. Consequently, parts of the molecule are hydrophilic and parts are hydrophobic.

It has been estimated that 20-30% of the human genome encodes membrane proteins [155], yet their amphipathic nature makes them notoriously difficult to study, keeping them consistently underrepresented in proteomic studies [156]. The key challenge for the analysis of membrane proteins is solubilisation. This is true for both gel-based and chromatography-based proteomics techniques. The traditional proteomics approach of 2DE followed by MALDI-TOF PMF has major limitations when it come to the study of membrane proteins because many hydrophobic proteins are not

solubilised in the non-detergent isoelectric focussing sample buffer, and in addition, solubilised proteins are prone to precipitation at their isoelectric point [156]. The limited dynamic range of 2DE also discriminates against membrane proteins, since they are typically lower in abundance than soluble proteins [37,157]. Refinements to the 2DE procedure to allow the study of membrane proteins include the use of nonionic or zwitterionic detergents to improve the solubilisation of membrane fractions [158-160] and the use of subcellular fractionation and directed biochemical enrichments to overcome abundance issues [157]. Alternatively, the isoelectric focussing step can be eliminated altogether and proteins can be separated by SDS-PAGE with MS detection [161-163]. This, however, is limited by the increased protein complexity in each one-dimensional gel band; an additional chromatography step is desirable [164].

For multi-dimensional chromatography-based proteomics methods, membrane solubilisation strategies that have recently been employed for the analysis of enriched membrane fractions include the use of detergents [65], organic solvents [165-167] and organic acids [49] that are compatible with the subsequent proteolytic digestion/ chemical cleavage step and analysis by LC-MS/MS. An alternative strategy has been described [168] that takes a multi-step approach to the analysis of membrane proteins, allowing the determination of protein topology in membranes.

P450s have one membrane spanning domain. They therefore do not present the same intractability for analysis as proteins with large numbers of transmembrane domains [167]. Crystallisation of mammalian P450s has been achieved after removal of the membrane anchors [130-134].

1.5 Aims of this project

Traditional methods for the detection of P450s rely on indirect techniques such as immunoblotting, activity assays, and the detection of P450 mRNA. These techniques have significant limitations. Immunoblotting, whilst being very sensitive, relies on the availability of enzyme-specific antibodies; it is necessary to pre-select which P450s are to be investigated, and to identify each enzyme in turn. Activity assays that are designed to interrogate the activity of a P450 enzyme invariably require multiple

analysis techniques, since different assays must be developed for different target substrates; even then, they may not be totally enzyme-specific. Measurement at the expression level is fraught with uncertainty since the presence and abundance of a particular type of mRNA does not necessarily imply the presence and abundance of the corresponding protein [169-172]. Mass spectrometry offers major advantages over these techniques, with the ability to directly detect low levels of multiple proteins simultaneously and without pre-selecting the proteins anticipated.

To date, there have been few reports of the analysis of multiple P450s by MS. There are no comprehensive studies profiling the expression of P450 proteins in normal or diseased tissue. The use of MS for P450 identification should provide an unbiased profile of the proteins present, which could then be used for drug development and to improve patient treatment. A sensitive method to detect functional P450s on an individual patient basis prior to therapy could ultimately lead to patient-specific medicines for cancer treatment. In this way only patients shown to have the correct P450 expression profile would be offered treatment which was dependent on a specific P450 enzyme.

The aims of this project are:

1. To develop and optimise a method for the identification of P450 proteins using mass spectrometry.
2. To apply this method to the analysis of multiple P450 enzymes in biological samples.
3. To investigate the P450 profile of normal and cancerous tissue and the consequences thereof.
4. To investigate possibilities for the determination and quantification of functional P450 protein.

Chapter 2

Materials and methods

2.1 Materials

Purified recombinant human P450 enzymes CYP1A2, 2E1 and 3A4 were purchased from PanVera (Invitrogen, Paisley, UK). Purified recombinant human cytochrome P450 reductase, human CYP3A4/3A5 antibody (the primary antibody used for Western blotting) and horseradish peroxidase (HRP)-conjugated anti-rabbit (the secondary antibody for CYP3A4 Western blotting) were obtained from Gentest (BD Biosciences, Oxford, UK). Mouse monoclonal antibody to human CYP1B1 was a kind gift from Dr. Graham Murray (University of Aberdeen, UK); HRP-conjugated sheep anti-mouse (secondary antibody for CYP1B1 Western blotting) was from Amersham Bioscience (Little Chalfont, Buckinghamshire, UK). CYP1B1 supersomes were purchased from Gentest. Sequencing grade modified trypsin (porcine) was purchased from Promega (Southampton, UK). ProtoGel gel electrophoresis reagent (containing 30% w/v acrylamide, 0.8% w/v N,N'-methylene bisacrylamide, ratio 37.5:1) was obtained from National Diagnostics (Hessle Hull, UK). Prestained protein markers (6-175 kDa) for SDS-PAGE were purchased as a mixture from New England Biolabs (Hitchin, Hertfordshire, UK), and contained *E. coli* maltose-binding protein (MBP)-β-galactosidase (175 kDa), *E. coli* MBP-paramyosin (83 kDa), bovine liver glutamic dehydrogenase (62 kDa), rabbit muscle aldolase (47.5 kDa), rabbit muscle triosephosphate isomerase (32.5 kDa), bovine milk β-lactoglobulin A (25 kDa), chicken egg white lysozyme (16.5 kDa) and bovine lung aprotinin (6.5 kDa). Ammonium bicarbonate was purchased from Fisher (Loughborough, Leicestershire, UK). Benzylisothiocyanate (BITC) was purchased from Aldrich (Poole, Dorset, UK). Citric acid monohydrate was from Acros Organics, Geel, Belgium. 3,5-Dimethoxy-4-hydroxycinnamic acid (sinapinic acid) and bovine serum albumin (BSA) MALDI standard were obtained from Applied Biosystems (Warrington, UK). ECL Western blot detection reagents were from Amersham Bioscience (Little Chalfont, Buckinghamshire, UK). Heptafluorobutyric acid (99%) (HFBA) was obtained from Lancaster Chemicals (Morecambe, Lancashire, UK). Trifluoroacetic acid (TFA), bio-grade, was from Avocado Research Chemicals (Heysham, Lancashire, UK). C_4 and C_{18} ZipTips were purchased from Millipore (Bedford, MA, USA). Acetic and formic acids, AnalaR grade, bromophenol blue, ethylenediaminetetra-acetic acid (EDTA), glycine, phosphoric

acid, potassium ferricyanide, silver nitrate, sodium carbonate, sodium chloride, sodium dodecyl sulphate (SDS), sucrose and tris(hydroxymethyl)methylamine (Tris) were purchased from BDH (Lutterworth, Leicestershire, UK). Ammonium persulphate (APS), aprotinin, benzamidine, BSA for Bradford assay and Western blotting, catalase (from bovine liver), chymostatin, Coomassie Brilliant Blue R-250 and G-250, cytochrome c (from horse heart), 1,2-didecanoyl-sn-glycero-3-phosphocholine (DLPC), dithiothreitol (DTT), formaldehyde (37% w/v solution), iodoacetamide, leupeptin, myoglobin (from horse skeletal muscle), β-nicotinamide adenine dinucleotide phosphate, reduced form, tetrasodium salt (β-NADPH), nonidet P-40 (NP-40), pepstatin, phenylmethanesulphonyl fluoride (PMSF), potassium chloride (KCl), SigmaFast DAB tablets for Western blot detection, sodium thiosulphate, soya-bean trypsin inhibitor (SBTI) and N,N,N',N'-tetramethylethylenediamine (TEMED) were all purchased from Sigma Chemical Company (Poole, Dorset, UK). Glycerol was from BDH or Sigma. HPLC, AnalaR and GPR grade solvents were obtained from BDH, Fisher or Merck. Recombinant CYP1B1 was expressed in-house at the School of Pharmacy.

2.2 S9 fractions of CYP3A4 transfected and control RIF-1 murine tumours

S9 fractions from human CYP3A4-transfected and control RIF-1 murine tumours were from the Radiation Biology Group, School of Biomedical Sciences, University of Ulster at Jordanstown, Northern Ireland. They had been prepared in the following manner: 2 x 10^5 RIF-1 tumour cells were implanted on the rear dorsum of 8-12 week old mice. Established tumours were injected with 25 μg of plasmid DNA bearing the CYP3A4 gene, using a liposome-based gene delivery kit. Tumours were excised 24 hours after transfection and snap-frozen in liquid N_2. They were thawed on ice and homogenised in phosphate buffer solution using a hand-held homogeniser. The homogenate was spun at 9000 *g* to get the S9 fractions.

2.2.1 Preparation of microsomes from S9 fractions of CYP3A4 transfected and control RIF-1 murine tumours

The method used was based on that of McFadyen *et al* [173]. 250 µl of each of the CYP3A4 transfected and control S9 fractions were made up to 1 ml with 10 mM Tris-HCl, pH 7.4, containing 1.15% KCl. Samples were centrifuged at 180,000 *g* for 1 hour at 4 °C using a MLA 80 rotor in a Beckman Optima MAX-E Ultracentrifuge (Beckman Coulter, High Wycombe, Buckinghamshire, UK). Supernatants (cytosolic fraction) were decanted, and each microsomal pellet was suspended in 500 µl of 0.1 M Tris-HCl, pH 7.4, containing 15% glycerol and 1 mM EDTA. Suspended microsomal pellets, cytosolic fractions and remaining S9 fractions were stored at -80 °C.

2.3 Human tissue samples

Resected human liver and liver colorectal metastatic samples were from The Royal Free and University College Medical School.

2.3.1 Subjects

The protocol for the study was approved by the Ethics Committee of the Royal Free Hospital and University College School of Medicine. Informed patient consent for use of tissues was obtained in all cases. Tissues for the study were accessed from the resected and discarded masses of tumour and surrounding liver, which were removed as part of the surgical treatment for hepatic metastases arising from colon cancers.

2.3.2 Tissue sampling

Samples of normal liver and of tumour (approximately 0.5 g) were taken in duplicate, and immediately snap frozen in liquid nitrogen for further processing. The normal liver tissues were taken from the region of the resected mass most distant

from the tumour. In each case the tumour tissues were easily identifiable as hard, white deposits of diameters of 2 cm or greater. The metastatic nature of the deposits was confirmed by histopathological examination.

2.3.3 Preparation of microsomes

Samples had not been screened against hepatitis and human immunodeficiency virus (HIV), therefore protocols for the safe conduct of the work were agreed and adhered to strictly. Samples were processed in a Class 2 biological safety cabinet in a HSE Containment Level 2 laboratory.

Before processing, all samples were weighed. Frozen samples were ground either in a percussion mortar and pestle, or in a Mikro-Dismembrator U (B. Braun Biotech International, Melsungen, Germany), which powders the frozen tissue using a ball bearing. Ground tissue was then placed into an ice-cold glass homogeniser, along with approximately 1 ml of homogenisation buffer for every 0.1 g of tissue, and homogenised using a PTFE-head pestle. The homogenisation buffer contained 0.25 M sucrose, 50 mM Tris-HCl (pH 7.2 at 4 °C, pH 6.9 at room temperature), 1 mM EDTA, 100 mM sodium chloride, 0.1 mM DTT, 0.1 mM benzamidine, 0.1 μg/mL SBTI and 0.1 mM PMSF. Microsomes were prepared using differential centrifugation as follows: an initial centrifugation at 2400 *g* for 20 minutes in a Sigma 6K10 centrifuge (Sigma Laboratory Centrifuges, Osterode am Harz, Germany) was used to sediment the cell debris, nuclei, and unbroken cells. The supernatant was centrifuged at 12,000 *g* for 20 minutes at 4 °C using a 70.1-TI rotor in a Beckman L8-60M centrifuge (Beckman Coulter, High Wycombe, Buckinghamshire, UK). Supernatants from this step were centrifuged at 180,000 *g* for 60 minutes at 4 °C in the L8-60M centrifuge. The resultant microsomal pellets were suspended in 0.1 M Tris-HCl, pH 7.4, containing 15% glycerol, 1 mM EDTA, 0.1 mM PMSF, 0.1 mM benzamidine, 0.1 μg/mL SBTI, 7 μM pepstatin, 5 μg/mL chymostatin, 10 μM leupeptin and 5 μg/ml aprotinin, and then recentrifuged at 180,000 *g* for 1 h. The final pellet was resuspended in 0.1 M Tris-HCl, pH 7.4, containing 15% glycerol, 1 mM EDTA, 0.1 mM PMSF, 0.1 mM benzamidine, 0.1

μg/mL SBTI, 7 μM pepstatin, 5 μg/mL chymostatin, 10 μM leupeptin and 5 μg/ml aprotinin, and stored at -80 °C.

2.4 Bradford assay for protein determination

Microsomal protein content was determined using the Bradford assay [174]. The absorbance maximum of Coomassie Brilliant Blue G-250 in acidic solution shifts from 465 to 595 nm after the addition of protein due to stabilisation of the anionic form of the dye by hydrophobic and ionic interactions. The dye reacts primarily with arginine residues and to a lesser extent with histidine, lysine, tyrosine, tryptophan and phenylalanine residues [175].

The Coomassie reagent contained 0.03 mg/ml Coomassie Brilliant Blue G-250, 10% absolute ethanol, 5% concentrated phosphoric acid in ultra-pure water. The reagent was stored at 4 °C in the dark.

A calibration curve was constructed using BSA protein standards, and the protein content of microsomal samples was determined relative to this. BSA standards at 10, 20, 30, 40, 50, and 60 μg/ml in water were used, as well as a water blank. 1 ml of Coomassie reagent (warmed to room temperature) was added to100 μl of each BSA standard (or blank). The absorbance at 595 nm was measured using a Cecil CE 1011 1000 Series spectrophotometer (Cecil Instruments, Cambridge, UK), which had been set to zero using the blank sample. The standard curve was plotted using these measurements. Microsomal protein samples were diluted appropriately, and 1 ml Coomassie reagent was added to 100 μl of each diluted sample. The absorbance at 595 nm was measured, and the amount of protein determined from the calibration curve. All readings (standards and samples) were made in duplicate.

2.5 Covalent adduction of human CYP2E1 by benzylisothiocyanate

The method for the incubation of CYP2E1 with BITC and NADPH-dependent P450 reductase was based on that used by Moreno *et al* [176]. CYP2E1 (1 μM as determined

spectrophotometrically) was incubated with 1 μM cytochrome P450 reductase, 15 μg DLPC, 22 units of Catalase, 50 μM BITC, in the presence or absence of 1 mM NADPH, in a total volume of 200 μl for 15 minutes at 30 °C. The final reaction mixture contained 5.6 mM KPO_4, 26 μM EDTA, 120 μM DTT and 2.9% glycerol from the CYP2E1 and P450 reductase buffers, and 1% methanol, which was used to dissolve the BITC. The reaction was halted by placing the samples on ice. Samples were stored at -80 °C.

2.5.1 Microcon purification of incubations of CYP2E1 with BITC and P450 reductase

Incubations of CYP2E1 with BITC and P450 reductase, with and without NADPH supplement, were purified using Microcon YM-3 centrifugal filter devices with a molecular weight cut off of 3,000 Da (Millipore). Approximately 80 μl of assay was loaded into a Microcon sample reservoir, which was placed in a filtrate vial and the assembled device centrifuged at 10,000 rpm in a Biofuge Pico benchtop centrifuge (Heraeus Instruments, Germany) for 30 minutes. The filtrate vial was changed and 500 μl of 25 mM NH_4HCO_3 was added to the sample reservoir. The assembled device was then centrifuged at 10,000 rpm for 1 hour, and the contents of the filtrate vial removed. The retentate was washed in this way with two further aliquots of 500 μl of 25 mM NH_4HCO_3, with removal of the filtrate after each wash. The sample reservoir was then placed upside down in a new filtrate vial and centrifuged at 2,000 rpm for 3 minutes to transfer the retentate to the filtrate vial. The sample reservoir was washed with 30 μl of 25 mM NH_4HCO_3, with gentle agitiation, and this wash was transferred to the filtrate vial containing the retentate by centrifuging at 2,000 rpm for 3 minutes. The final volumes of retentate after washing were approximately 37 μl (NADPH-supplemented incubation) and 46 μl (NADPH-absent incubation), from starting volumes of approximately 80 μl. All filtrates were kept. Retentates and filtrates were stored at -80 °C.

2.6 In-solution tryptic digestion

A number of published protocols for tryptic digestion of proteins in solution were obtained. The procedure used for the in-solution tryptic digestion of proteins was adapted from B. Zhang *et al* [177] and L. K. Lightning *et al* [178]. Myoglobin (0.1 mg/ml in 25 mM ammonium bicarbonate), which is resistant to proteolysis [179], was thermally denatured before digestion by incubating at 90 °C for one hour.

0.1 mg/ml protein was incubated with 0.002 mg/ml trypsin, or 0.2 mg/ml protein with 0.004 mg/ml trypsin (such that the w/w ratio of substrate to enzyme was 50:1) in 25 mM ammonium bicarbonate at 37 °C for 24 hours. Control samples, containing enzyme but no substrate, were also incubated.

2.6.1 In-solution tryptic digestion of incubations of CYP2E1 with BITC and P450 reductase

The total protein content of both the NADPH-supplemented and NADPH-absent incubations of CYP2E1 with BITC was 0.22 mg/ml (CYP2E1, P450 reductase and catalase). In-solution digestion was performed such that the ratio of substrate to trypsin was 50:1 (by mass): trypsin in 25 mM NH_4HCO_3 was added to 40 µl of each incubation such that the final concentration of trypsin was 0.004 mg/ml (and the total concentration of CYP2E1, P450 reductase and catalase in the assay was reduced to 0.20 mg/ml). Digests were incubated for 24 hours at 37 °C, then stored at -80 °C.

2.7 SDS-PAGE

SDS-PAGE was performed by the Laemmli method [180] using the Hoefer Mighty Small gel system (Amersham Bioscience). Samples were diluted with denaturing sample buffer to contain 0.1 M DTT, 3 mg/ml SDS, 0.02 mg/ml bromophenol blue, 5% glycerol and 8.3 mM Tris-HCl pH 6.8, and heated at 95 °C for 10 minutes. Prestained protein marker (6-175 kDa) (New England Biolabs) was also heated at 95

°C for 10 minutes, and resolved on the same gels as the samples to provide an indication of molecular weight. P450 samples were resolved on 10% w/v acrylamide gels, and cytochrome c samples were resolved on 18% w/v acrylamide gels, each containing 0.4 M Tris-HCl pH 8.8, 0.08% w/v SDS, 0.08% TEMED and 0.08% APS. A less dense stacking gel was loaded on top of the running gel to preconcentrate the samples; this contained 5% acrylamide, 0.13 M Tris-HCl pH 6.8, 0.1% w/v SDS, 0.1% TEMED and 0.1% w/v APS. A Tris-glycine buffer system was used, containing 25 mM Tris, 0.25 M glycine and 0.1% SDS. Gels were run at 70 V, 20 mA until the samples had crossed over into the resolving gel, at which time the voltage was increased to 110 V and the current to 30 mA. Gels were allowed to run until the dye front had reached the base of the resolving gel. Gels were stained either with Coomassie Blue or silver nitrate.

2.7.1 Coomassie Blue staining

Gels were stained with 2 mg/ml Coomassie Brilliant Blue R-250 in 30% methanol, 10% acetic acid in water, and destained with the same solvent, excluding the dye. Gels were stained for approximately 1 hour. Gels that were to be used for in-gel tryptic digestion were destained for up to three hours (with changing) then left overnight in water or in 2% glycerol, 35% ethanol in water. Reference gels were left in destain solution overnight.

2.7.2 Silver staining

The silver staining protocol was chosen to omit gluteraldehyde, a cross-linking and sensitising agent [181]. Gels were fixed in 5% acetic acid, 50% methanol for 30 minutes. They were washed in 50% methanol for 10 minutes, in 5% methanol for 10 minutes, then incubated in 40 μM DTT for 30 minutes, and rinsed in water. Gels were incubated in 0.2% w/v silver nitrate for 30 minutes, rinsed well in water, and developed in 0.0185% w/v formaldehyde in 0.28 M sodium carbonate. The reaction was halted by the addition of citric acid monohydrate. Gels were left in 2% glycerol, 35% ethanol overnight.

2.8 Western blotting

SDS-PAGE resolving gels, run to completion but not stained, were separated from their stacking gels and soaked in transfer buffer containing 25 mM Tris, 0.19 M glycine and 20% methanol for approximately 15 minutes. They were then transferred onto a piece of nitrocellulose membrane, also soaked in transfer buffer. A 'sandwich' was assembled in a support cassette, consisting of a piece of pre-soaked sponge, overlaid with three pieces of pre-soaked 3 MM paper followed by the SDS-PAGE gel, the nitrocellulose membrane, three more pieces of pre-soaked 3 MM paper and another piece of sponge. This assembly was placed into an electroblotting tank containing transfer buffer, such that the nitrocellulose faced towards the anode, and the gel faced towards the cathode, and the direction of protein transfer was therefore from the gel to the nitrocellulose. Gels were electroblotted for 1.5 hours at 100 V, 150 mA, with cooling. The nitrocellulose membrane was removed and submerged in blocking solution (10 mg/ml BSA in 10 mM Tris-HCl pH 7.0, 150 mM sodium chloride) for 1 hour (CYP3A4 Western) or overnight (CYP1B1 Westerns). The blocking solution was then replaced with primary antibody diluted 1 in 2000 (anti-CYP3A4) or 1 in 100 (anti-CYP1B1) with blocking solution containing 0.05% NP-40, and incubated on a rocking platform overnight at 4 °C (CYP3A4) or for two hours at room temperature (CYP1B1). The membrane was washed in Tris-saline buffer (150 mM sodium chloride in 10 mM Tris-HCl pH 7.0) for 10 minutes, followed by two changes of 0.05% NP-40 in Tris-saline buffer (each for 10 minutes), and finally in Tris-saline buffer again (10 minutes). The membrane was placed in secondary antibody diluted 1 in 7500 (CYP3A4) or 1 in 1000 (CYP1B1) with blocking solution containing 0.05% NP-40, and incubated on a rocking platform for 1.5 hours. It was then washed as before.

The CYP3A4 Western blot protein bands were detected using ECL western blocking detection reagents (Amersham Bioscience) according to manufacturer's instructions, and a GeneGnome chemiluminescent imaging system (Syngene Bio-Imaging, Cambridge, UK). The CYP1B1 Western blot protein bands were developed using SigmaFast DAB tablets.

The blotted gel was stained with Coomassie Blue dye (Section 2.7.1) to check for any untransferred proteins.

2.9 In-gel tryptic digestion

The method used for in-gel tryptic digestion was continually modified as part of the optimisation process; aspects of this optimisation process are described Chapter 3. The original method was obtained from ThermoElectron, Hemel Hempstead, UK. The final method is detailed below.

Blank gel samples were routinely analysed along with protein-containing gel bands.

2.9.1 Preparation of gel slices for tryptic digestion

2.9.1.1 Coomassie blue-stained gels

Gel bands were excised using a scalpel or razor blade, washed with water until their pH was neutral (as estimated by spotting samples on pH paper), and completely destained using sequential washings of 50 mM ammonium bicarbonate (pH 7.8) in 40% ethanol, with mixing on a rotary mixer. Bands were cut into fine pieces to increase surface area, then dried with three to four sequential 15-minute incubations with acetonitrile at room temperature, followed by 30 minutes in a CentriVap (Labconco, Kansas City, Missouri, USA).

2.9.1.2 Silver-stained gels

Gel bands were excised using a scalpel or razor blade and destained in a freshly prepared solution containing 30 mM potassium ferricyanide mixed in a 1:1 ratio with 100 mM sodium thiosulphate. Bands were washed twice with water to remove the yellow colour, then washed two to three times with *either* 50 mM ammonium bicarbonate, pH 7.8, in 40% ethanol *or* 0.1 M ammonium bicarbonate, with mixing on a rotary mixer (to help remove SDS and salts). Bands were cut into fine pieces to

increase surface area, then dried with three to four sequential 15-minute incubations with acetonitrile at room temperature, followed by 30 minutes in a CentriVap.

2.9.2 Preparation of trypsin stock solution

Sequencing grade trypsin was made up at a concentration of 200 ng/µl, divided into 5 µl (i.e. 1 µg) aliquots and immediately snap-frozen in liquid nitrogen. Aliquots were stored at -80 °C.

2.9.3 Digestion of gel slices with trypsin

Dried gel pieces were separated using a needle. The volume of trypsin solution needed to rehydrate the gel pieces was estimated, and the appropriate dilution of stock trypsin solution was calculated such that 400 ng trypsin was added to each sample. Rehydration volumes ranged from 5 to 10 µl. Samples were left on ice for 30 minutes after addition of trypsin solution, then covered with 25 mM ammonium bicarbonate and incubated at 37 °C overnight.

2.9.4 Peptide extraction

25-30 µl of 5% TFA in 50% acetonitrile was added to each digest mixture. Samples were sonicated for 6 minutes to extract the peptides, after which the supernatant was transferred to a separate tube. This was repeated twice, then the extracts were combined and dried in the CentriVap to complete dryness. The resulting peptide mixtures were stored at -80 °C.

Samples that were not desalted were reconstituted in appropriate solvent (usually 0.1% TFA or 0.5% formic acid in 70% methanol) and sonicated for 2 x 5 minutes immediately prior to analysis.

2.10 Desalting for nano-electrospray

Samples that are to be analysed by ES-MS should not contain involatile solvents, salts or buffers because they inhibit ionisation and can cause instrument blockages. A ZipTip (Millipore, Bedford, MA, USA) is a 10 µl pipette tip with a bed of chromatography media immobilised in the tip; essentially, it is a scaled down version of a solid-phase extraction column. ZipTips are used for purifying and/or concentrating small volumes (less than 20 µl) of peptides, proteins and oligonucleotides; they were employed here for the removal of salts and involatile components from samples that were to be analysed by ES-MS.

ES-MS analysis was performed in two different ways: sample was either analysed as a mixture by nano-electrospray (nano-ES) with no prior chromatographic separation, or separated by on-line nano-scale reversed phase liquid chromatography prior to nano-electrospray analysis (nano-LC-ES). Nano-LC-ES employed on-line sample clean-up via the use of a guard column, which was placed before the main analytical column (Section 2.11), whereas there was no on-line sample clean-up for nano-ES.

Initially, all samples were desalted whether they were to be subjected to nano-ES or nano-LC-ES analysis. Eventually, however, desalting was only carried out on samples to be analysed by nano-ES (Section 3.8.3).

2.10.1 Peptide desalting using C_{18} ZipTips

Peptides that had been extracted from polyacrylamide gels were rehydrated in 10 µl of 2% HFBA and sonicated for 2 x 5 minutes. Peptide samples in solution were acidified to contain 0.1% TFA. C_{18} ZipTips were prepared by wetting firstly with 50% acetonitrile, then with 100% methanol, and finally equilibrating with 0.05% TFA. Samples were loaded onto the tip 10 times, then washed 7 times with 0.05% TFA. Peptides were eluted from the tip with 10 - 30 µl of a solution of 0.5% formic acid in 70% methanol. Samples that were not to be analysed immediately were dried in a CentriVap and stored at -80 °C. Samples were reconstituted in an appropriate

solvent (0.1% TFA or 0.5% formic acid in 70% methanol) and sonicated for 2 x 5 minutes immediately prior to analysis.

2.10.2 Desalting of recombinant P450 proteins using C_4 ZipTips

Purified human recombinant CYP1A2, 2E1 and 3A4 were provided in a matrix containing 20 mM potassium phosphate buffer, 0.2 mM EDTA, 1 mM DTT and 20% glycerol. This matrix contains involatile components and so is not well-suited for electrospraying directly into the mass spectrometer without prior sample clean-up. C_4 ZipTips, containing C_4 reversed-phase media, are recommended by the manufacturer for proteins of molecular weight 25 to 100 kDa, therefore they were chosen for sample clean-up of the intact P450 proteins.

Maximum binding to the ZipTip pipette tip is achieved in the presence of 0.1% TFA or other ion-pairing agents. Sample solutions should have a pH of less than 4 (manufacturer's recommendation). To this end, 2 µl of each P450 solution was diluted to 20 µl using 0.1% TFA. C_4 ZipTips were prepared by wetting firstly with 50% acetonitrile, then with 100% methanol, and finally equilibrating with 0.05% TFA. Samples were loaded onto the tip 10 times, then washed 7 times with 0.05% TFA. Proteins were eluted from the tip with 10 µl of a solution of either 0.5% formic acid in 70% methanol or 0.5% formic acid in 70% acetonitrile.

2.11 Nano-LC

Nano-scale reversed phase liquid chromatography (nano-LC) was performed using an LC Packings Ultimate Capillary HPLC system with FAMOS autosampler (Dionex, Camberley, Surrey, UK). A separate Ultimate Micropump was employed as a loading pump: samples were loaded onto a guard column at a flow rate of 20 µl/min and washed before being transferred onto the analytical column by means of a ten-port valve fitted to the side of the FAMOS® autosampler. The instrument was used in several different configurations; those most commonly employed will be described here.

Sample (1 µl) was loaded via either a 1 µL or 10 µL sample loop (using 0.1 % TFA in water as loading buffer) onto a 1 mm x 300 µm PepMap™ C18 guard column (5 µm, 100 Å) (LC Packings), using full loop or µl pick-up injections. Sample was washed with 0.1% TFA on the guard column before being transferred onto a 15 cm x 75 µm PepMap™ C18 column (3 µm, 100 Å) (LC Packings), equilibrated with mobile phase at a flow rate of approximately 200 nL/min.

The most commonly used gradient program is shown in Table 2.1 and Figure 2.1.

Table 2.1 Gradient program used for the separation of tryptic digests using nano-LC.

Time (hh:mm:ss)	Flow rate (nL/min)	Valve %A [a]	Valve %B [b]
00:00:00	200	95	5
00:05:00	200	95	5
00:30:00	200	50	50
00:30:01	200	5	95
00:40:00	200	5	95
00:40:01	200	95	5
00:60:00	200	95	5

[a] Mobile phase A: 0.05% TFA (UV detection) or 0.1% formic acid (MS detection), 5% acetonitrile in water; [b] mobile phase B: 0.04% TFA (UV detection) or 0.1% formic acid (MS detection), 80% acetonitrile in water.

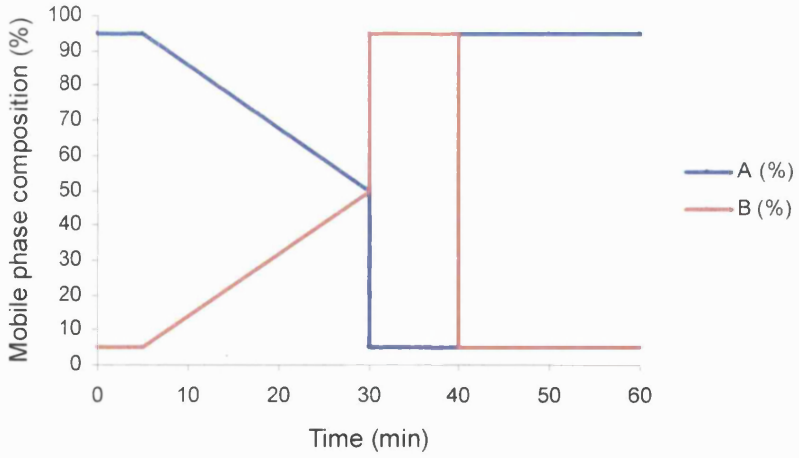


Figure 2.1 Gradient program used for the separation of tryptic digests using nano-LC. Mobile phase A: 0.05% TFA (UV detection) or 0.1% formic acid (MS detection), 5% acetonitrile in water; mobile phase B: 0.04% TFA (UV detection) or 0.1% formic acid (MS detection), 80% acetonitrile in water. Flow rate 200 nl/min.

Detection was initially performed using the built-in ultra-violet (UV) detector, set to 214 nm. The nano-LC system was later coupled to the LCQ mass spectrometer (Section 2.12), and plumbed to bypass the UV detector.

2.12 Electrospray mass spectrometry

All ES mass spectra were obtained using an LCQ^{duo} mass spectrometer fitted with a nanospray source (ThermoElectron, Hemel Hempstead, UK). Sample was either analysed as a mixture, with no prior chromatographic separation (nano-ES), or separated by on-line nano-LC prior to mass spectrometric analysis (nano-LC-ES). Many different instrumental conditions were used during the course of this work. Those described below represent the final conditions decided upon after much experimentation, optimisation and external recommendation.

2.12.1 Nano-ES

Approximately 5 to 10 µl sample was loaded into a capillary glass PicoTip™ emitter (1.2 mm outer diameter (OD), 0.69 µm inner diameter (ID), 1 ± 0.5 µm tip ID, end-coated with conductive material) (New Objective, Inc., Woburn, MA, USA) using a Gel-Saver pipette tip (Fisher). The PicoTip™ emitter was loaded into the nanospray probe, and spectra were acquired. The instrument was operated in positive ion mode using the following settings: capillary temperature 180 °C; spray voltage 1.00 to 1.10 kV; no sheath or auxiliary gas used. The acquisition mass range in the full scan mode was *m/z* 400-2000; centroid data was collected. Other conditions were optimised automatically using the autotune function on samples chemically and physically similar to the sample to be analysed (for example, a CYP2E1 in-gel digest, a cytochrome c in-solution digest, or an intact P450 protein).

Data was collected either manually or automatically. For automated data collection whilst running protein digests, the instrument was operated in "Big 3" mode: over a specified time period (30 to 60 minutes), the following four scan events were repeated:

1. Full scan of *m/z* 400-2000, consisting of eight averaged "microscans" with a maximum injection time of 200 ms.
2. Data dependent MS/MS on the most intense ion from 1.
3. Data dependent MS/MS on the second most intense ion from 1.
4. Data dependent MS/MS on the third most intense ion from 1.

MS/MS scans consisted of three averaged "microscans", each with a maximum injection time of 200 ms. The isolation width for the parent ion isotopic cluster in MS/MS scans was *m/z* 3.0 and the normalised collision energy was set to 35%. The minimum signal required to trigger MS/MS data collection was 100,000. The following contaminant peaks were added to the reject mass list: *m/z* 430, 446, 463 and 476; no MS/MS scans were performed on these ions at any time during the data acquisition. If an ion appeared as one of the top three most intense ions for more than 5 full scans in a time period of thirty minutes, it was put on an exclusion list for 30 minutes; during this time, no MS/MS scans would be performed on that ion, whatever its intensity. The exclusion mass width was *m/z* 2.0.

2.12.2 Nano-LC-ES

The column eluate from the nano-LC (Section 2.11) was directed into the ion source of the mass spectrometer via a fused silica PicoTip™ emitter (360 μm OD, 50 or 75 μm ID, 15 μm tip ID, either end- or distal-coated) (New Objective, Inc.). The instrument was operated in positive ion mode using the following settings: capillary temperature 180 °C; spray voltage 2.00 kV; no sheath or auxiliary gas used. The acquisition mass range in the full scan mode was *m/z* 400-2000; centroid data was collected. Other conditions were optimised automatically using nano-ES performed on samples chemically and physically similar to the sample to be analysed.

The instrument was operated in "Big 3" mode: over the 60 minute time period of the nano-LC gradient, the following four scan events were repeated:

1. Full scan of *m/z* 400-2000.

2. Data dependent MS/MS on the most intense ion from 1.
3. Data dependent MS/MS on the second most intense ion from 1.
4. Data dependent MS/MS on the third most intense ion from 1.

Full and MS/MS scans consisted of three averaged "microscans", each with a maximum injection time of 200 ms. The isolation width for the parent ion in MS/MS scans was *m/z* 3.0 and the normalised collision energy was set to 35%. The minimum signal required to trigger MS/MS data collection was 100,000. The following contaminant peaks were added to the reject mass list: *m/z* 415.95, 429.95, 445.92, 462.87, 503.92, 519.87, 536.95 and 610.99. If an ion appeared as one of the top three most intense ions for more than 3 full scans in a time period of 30 seconds, it was put on an exclusion list for 1.25 minutes. The exclusion mass width was *m/z* 2.0.

2.13 Deconvolution of ES mass spectra of intact proteins

ES mass spectra of intact proteins were deconvoluted using Biomass software (ThermoElectron). This uses an algorithm to transform an ES mass spectral plot of relative abundance versus *m/z* to a plot of relative abundance versus mass.

2.14 Database searching and peptide identification

MS/MS spectra were searched using Sequest Browser software [52,182] (ThermoElectron).

2.14.1 FASTA databases

Sequest requires FASTA formatted databases to search against. FASTA formatted databases are in ASCII text format. There is a single header or description line per sequence entry, followed by lines of sequence data. The header line starts with a "greater than" symbol (>). The end of the header is denoted by a carriage return-line feed. Lines of text in the sequence data should contain no more than 80 characters.

The sequence ends if there is another "greater than" symbol (>) at the beginning of a line and another sequence begins. Sequences are expected to be represented in the standard IUB/IUPAC amino acid and nucleic acid codes, with these exceptions: lower-case letters are accepted and are mapped into upper-case; a single hyphen or dash can be used to represent a gap of indeterminate length; and in amino acid sequences, U and * are acceptable letters (see below).

The accepted amino acid codes are:

A → alanine	P → proline
B → aspartate or asparagine	Q → glutamine
C → cysteine	R → arginine
D → aspartate	S → serine
E → glutamate	T → threonine
F → phenylalanine	U → selenocysteine
G → glycine	V → valine
H → histidine	W → tryptophane
I → isoleucine	Y → tyrosine
K → lysine	Z → glutamate or glutamine
L → leucine	X → any
M → methionine	* → translation stop
N → asparagine	- → gap of indeterminate length

Examples of protein sequence entries in FASTA format are shown in Figure 2.2.

```
>gi|9313018|gb|AAC50052.2| cytochrome P450 4F2 [Homo sapiens] [MASS=59858]
MSQLSLSWLGLCRVAASPWLLLLLVGASWLLAHVLAWTYAFYDNCRRLRCFPQPPRRNWFWGHQGMVNPTEEGMRVLTQL
VATYPQGFKVWMGPISPLLSLCHPDIIRSVINASAAIAPKDKFFYSFLEPWLGDGLLLSAGDKWSRHRRMLTPAFHFNIL
KPYMKIFNESVNIMHAKWQLLASEGSACLDMFEHISLMTLDSLQKCVFSFDSHCQEKPSEYIAAILELSALVSKRHHEIL
LHIDFLYYLTPDGQRFRRACRLVHDFTDAVIQERRRTLPSQGVDDFLQAKAKSKTLDFIDVLLLSKDEDGKKLSDEDIRA
EADTFMFEGHDTTASVLSWVLYHLAKHPEYQERCRQEVQELLKDREPKEIEWDDLAHLPFLTMCMKESLRVHPPVPVISR
HVTQDIVLPDGRVIPKGIICLISVFGTHHNPAVWPDPEVYDPFRFDPENIKERSPLAFIPFSAGPRNCIGQTFAMAEMKV
VLALTLLRFRVLPDHTEPRRKPELVLRAEGGLWLRVEPLS
>gi|6470135|gb|AAF13598.1|AF182273_1 cytochrome P450-3A4 [Homo sapiens] [MASS=57253]
MALIPDLAMETRLLLAVSLVLLYLYGTHSHGLFKKLGIPGPTPLPFLGNILSYHKGFCMFDMECHKKYGKVWGFYDGQQP
VLAITDPDMIKTVLVKECYSVFTNRRPFGPVGFMKSAISIAEDEEWKRLRSLLSPTFTSGKLKEMVPIIAQYGDVLVRNL
RREAETGKPVTLKDVFGAYSMDVITSTSFGVNIDSLNNPQDPFVENTKKLLRFDFLDPFFLSITVFPFLIPILEVLNICV
FPREVTNFLRKSVKRMKESRLEDTQKHRVDFLQLMIDSQNSKETESHKALSDLELVAQSIIFIFAGCETTSSVLSFIMYE
LATHPDVQQKLQEEIDAVLPNKAPPTYDTVLQMEYLDMVVNETLRLFPIAMRLERVCKKDVEINGMFIPKGVVVMIPSYA
LHRDPKYWTEPEKFLPERFSKKNKDNIDPYIYTPFGSGPRNCIGMRFALMNMKLALIRVLQNFSFKPCKETQIPLKLSLG
GLLQPEKPVVLKVESRDGTVSGA
```

Figure 2.2 Two protein sequence database entries in FASTA format.

The human database initially used contained 56,106 entries (Finnigan Xcalibur, Revision 1.0 P/N XCALI-64012, July 2000). In July 2003, a non-redundant protein database was downloaded from the National Center for Biotechnology Information (NCBI) (Bethesda, MD, USA) and modified for use with searching for P450 proteins (Section 3.5). Cytochrome c, rat, mouse and *E. coli* samples were searched against equine, rat, mouse and *E. coli* databases (Finnigan Xcalibur, Revision 1.0 P/N XCALI-64012, July 2000).

2.14.2 Sequest

Sequest correlates uninterpreted MS/MS peptide fragmentation spectra with spectra generated from a theoretical digest of each protein in the database. This is a complicated process involving several stages; these are detailed below.

Step 1: MS/MS data reduction

The first element of the program involves pre-search analysis and reduction of the MS/MS data. ".Dta" files are created from the raw MS/MS data using the following user-specified parameters (the values used for this work are indicated in brackets):

- Start scan/end scan (*first to last scan, i.e. the entire file*): scans from the raw data file which are to be processed.
- Bottom MW/Top MW (*400 Da/4000 Da*): precursor mass range to search.
- Minimum # ions (*15*): the minimum number of ions which must appear in the MS/MS spectrum.
- Minimum TIC (*500,000, unless otherwise indicated*): the minimum intensity of the MS/MS spectrum.
- Mass, or precursor mass tolerance (*m/z 1.4*): mass measurement error to allow grouping of scans.
- Intermediate scans (*25*): the difference between the first and last scan numbers that can be grouped together to create a single .dta file.

- Grouped scans (*l*): the minimum number of MS/MS scans of the same precursor *m/z* value that are needed to create a .dta file.

Scans are grouped together in order to decrease the search time and increase the quality of the MS/MS data. There is, however, the possibility that two different peptides with similar precursor *m/z* values, which elute very close together in time, could be grouped together, mixing their MS/MS spectra.

.Dta files are plain text files; an example is shown in Figure 2.3.

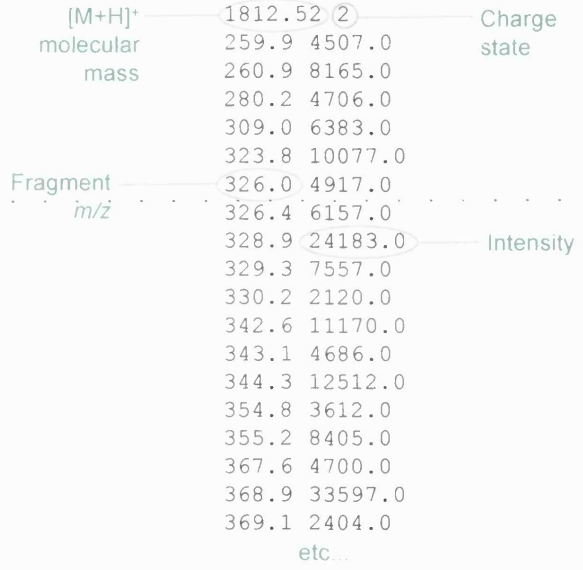


Figure 2.3 Part of a Sequest .dta file. .Dta files are created from raw MS/MS data and used by Sequest to search against protein databases.

Step 2: .Dta file searching and initial scoring

.Dta files are searched in a multi-step procedure. Initially, candidate peptide sequences are found in the database on the basis of intact peptide *m/z* values. Masses for the amino acid sequences are summed until the mass of the peptide falls within a user-specified mass tolerance. The *m/z* values for the fragment (b and y) ions predicted for each amino acid sequence are calculated as follows:

$$b_n = \sum a_n + 1 \qquad (1)$$

$$y_n = MW - \sum a_n \quad (2)$$

where a_n is the mass of the amino acid residue, b_n is a type b ion, y_n is a type y ion and MW is the molecular weight of the peptide. These calculated *m/z* values are used in a second step to compare sequences to the *m/z* list in the .dta file. Once an amino acid sequence fits the defined mass tolerance, it is given a primary score (S_p) based on:

- the number (n_i) of predicted fragment ions that match ions observed in the spectrum within ±1 Da and their abundances (i_m).
- the continuity of the ion series: this is considered by incrementing a component of the score (β) for each consecutive fragment ion matched. The value of 0.075 is used for β.
- the presence/absence of an immonium ion for the amino acids His, Tyr, Trp, Met or Phe, along with the associated amino acid. If the immonium ion is present, then an additional component, ρ, of the score is added. If the amino acid is not present in the sequence then ρ is decreased. The value of 0.15 is used for ρ.
- the total number of predicted sequence ions, n_t.

The following formula is used to calculate S_p:

$$S_p = (\sum i_m) n_i (1+\beta)(1+\rho)/n_t \quad (3)$$

S_p provides a reasonably accurate method for the assessment of sequence matches; however, the search is less successful for small peptides, especially when using a large database, and false positives occasionally result [52]. To address this, the top 500 identified amino acid sequences are subjected to a cross correlation analysis, which takes individual fragment ion peak intensities into account.

The following parameters were used when searching .dta files using Sequest: fragment ion tolerance 0.00 Da (fragment ion mass values are rounded to the nearest Da, therefore an automatic fragment ion tolerance of ± 0.5 Da is included in this

parameter); peptide mass tolerance 1.00 Da; average mass; maximum number of internal cleavage sites 2.

Step 3: Cross-correlation analysis

Cross correlation analysis is used to compare the top 500 amino acid sequences identified in the database search. A theoretical spectrum is reconstructed from the character-based amino acid sequences of the candidate peptide. This reconstructed spectrum not only contains predicted *m/z* values for fragment ions, but also assigns them each a magnitude component. This magnitude component does not try to predict relative abundances of b and y ions, but rather accounts for ions that may be formed from sequential losses from these ions. All *m/z* values that can be assigned to b and y fragment ions are assigned a magnitude of 50.0. Mass to charge values within ± 1 of the b or y ion values are assigned a magnitude of 25.0. The neutral losses of ammonia, water and carbon monoxide (a ions), which are generally less facile processes at the collision energies used to generate MS/MS spectra of peptides, and the *m/z* ratios ± 1 of the neutral loss values, are assigned a magnitude of 10.0.

In order to compare the original MS/MS spectrum with the reconstructed spectrum, the original spectrum is processed in the following manner: the *m/z* value corresponding to the precursor ion is removed, and the spectrum is divided into 10 equal regions. The ions in each region are then normalized to a value of 50.0. The removal of the precursor ion allows the correlation between the reconstructed spectrum and the original spectrum to be calculated on the basis of fragment ions alone.

The cross-correlation between two continuous signals *x(t)* and *y(t)* can be calculated using the following formula:

$$C_{x,y} = \int_{-\infty}^{+\infty} x(t)y(t+\tau)dt \qquad (4)$$

where τ is a displacement value between the two signals. The correlation function measures the coherence of two signals by, effectively, translating one signal across

the other. τ, the amount by which the signal is offset during the translation, is varied over a range of different values. If the two signals are identical, the correlation function will maximize at τ = 0, where there is no offset between the two signals. However, equation (4) applies to continuous functions only. The reconstructed spectrum, x_i, and the normalized experimental spectrum, y_i, are discrete functions, therefore the following form of the cross-correlation is used:

$$R_\tau = \sum_{i=0}^{n-1} x[i]y[i+\tau] \tag{5}$$

where τ is again a displacement value.

The discrete correlation, R_τ, of two real functions x and y, is one member of the discreet Fourier transform pair:

$$R_\tau \leftrightarrow X_\tau Y_\tau^* \tag{6}$$

Where X and Y are discrete Fourier transforms of x and y, and Y^* is the complex conjugate of Y. The cross correlations are computed by fast Fourier transformation of the two data sets x_i and y_i to get X and Y, followed by multiplication of one with the complex conjugate of the other, and inverse Fourier transformation of the resulting product. The graphical output of the cross correlation of two spectra x_i and y_i resembles a Gaussian curve. The final score, Xcorr, assigned to each candidate peptide sequence is the value of the function when τ = 0 minus the mean of the cross-correlation function over the range $-75 < \tau < 75$. This process is illustrated in Figure 2.4.

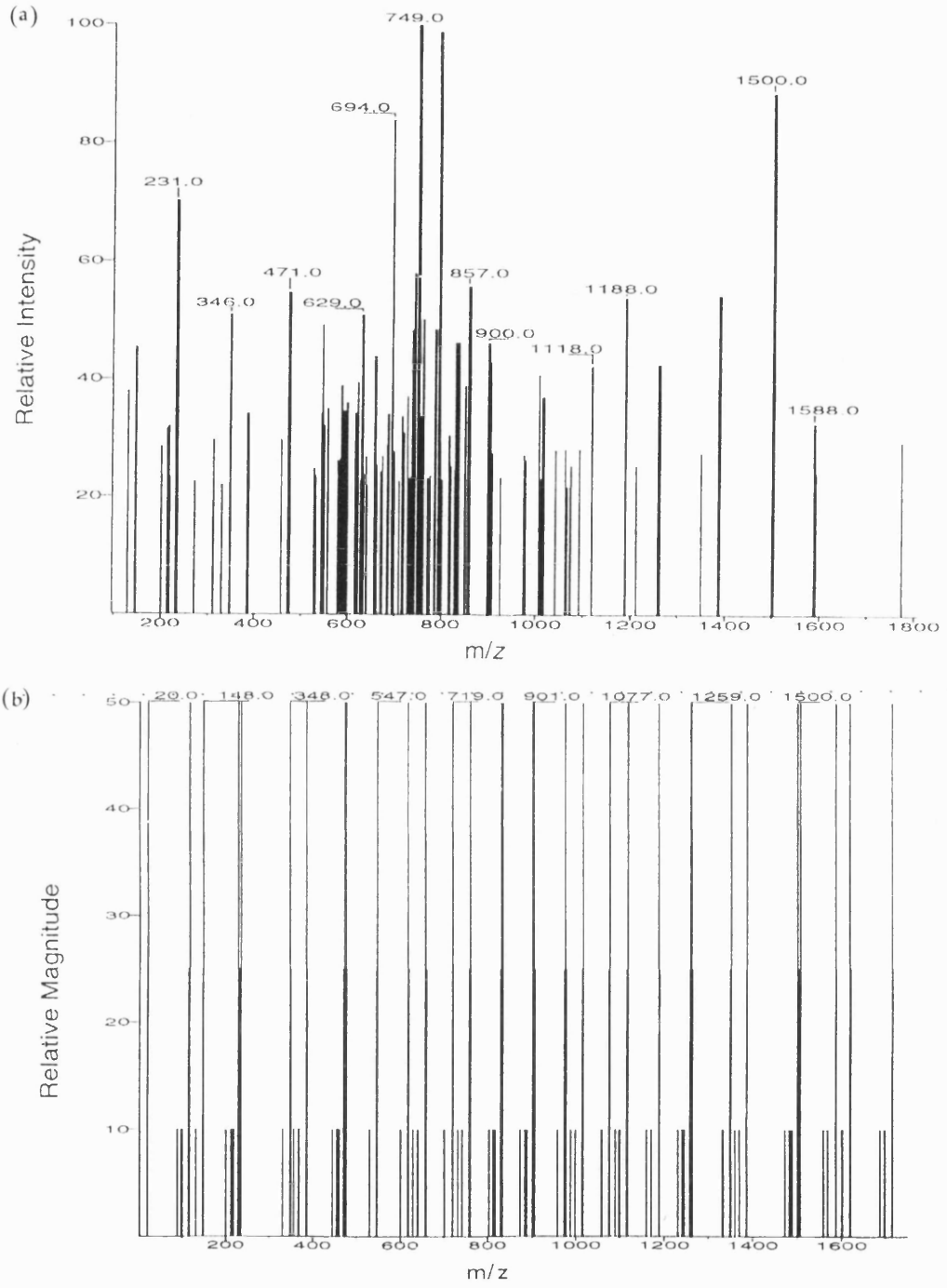


Figure 2.4 Cross correlation analysis used in Sequest searching. Taken from Eng *et al* [52]. **(a)** Processed MS/MS data for the doubly charged $[M + 2H]^{2+}$ ion of the peptide DLRSWTAADTAAQISQ. The 200 most abundant ions are shown in the graphical display. A 10 *m/z* window around the precursor ion at *m/z* 868 has been removed. The abundances of fragment ions within 1 *m/z* of each other are equalised to the higher value. **(b)** A graphical display of the reconstructed data used for the correlation analysis for the amino acid sequence DLRSWTAADTAAQISQ. A magnitude of 50.0 is assigned to the predicted *m/z* values for the type *b* and *y* ions; ions with *m/z* ratios ±1 *m/z* from the *b* and *y* ions are assigned a value of 25.0. The neutral losses of water and ammonia and the *a* type ions are assigned values of 10.0.

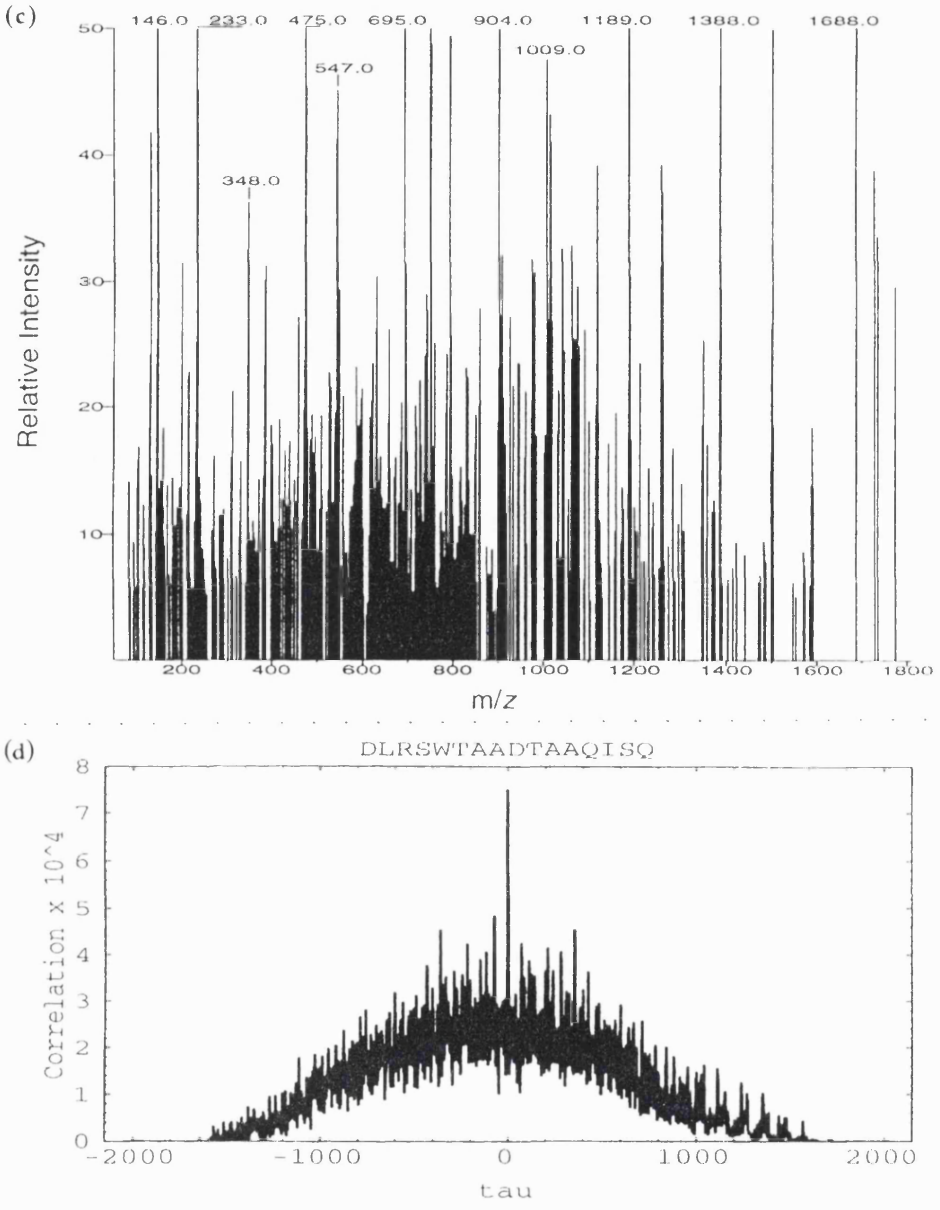


Figure 2.4 continued. (c) A graphical display of the processed MS/MS spectrum used in the correlation analysis. The spectrum is divided into 10 equal sections and the ion abundances in each section are normalised to 50.0. **(d)** A graphical display of the result of the cross-correlation function for the spectrum displayed in (b) and (c). The final score attributed to the analysis is the value at $\tau = 0$ minus the mean of the cross-correlation function over the range $-75 < \tau < 75$.

2.14.2.1 Acceptance criteria for protein identification

The output of a Sequest search is displayed on a *Sequest Summary* page; an example is shown in Figure 2.5. Manual inspection of this page is necessary to determine whether correct peptide and protein identifications have been made. Generally, proteins that were matched by two or more peptides with Xcorr values ≥ 2.5 were considered conclusively identified, provided that the peptides were unique to that protein in the database (or that the protein was the only one common to all the peptides identified), and that the difference in the Xcorr scores for the top two ranked peptides (ΔC_n) was ≥ 0.1. For proteins that were identified but did not match this criteria, the MS/MS data was assessed manually.

Identifications were confirmed by manual examination of the MS/MS spectra.

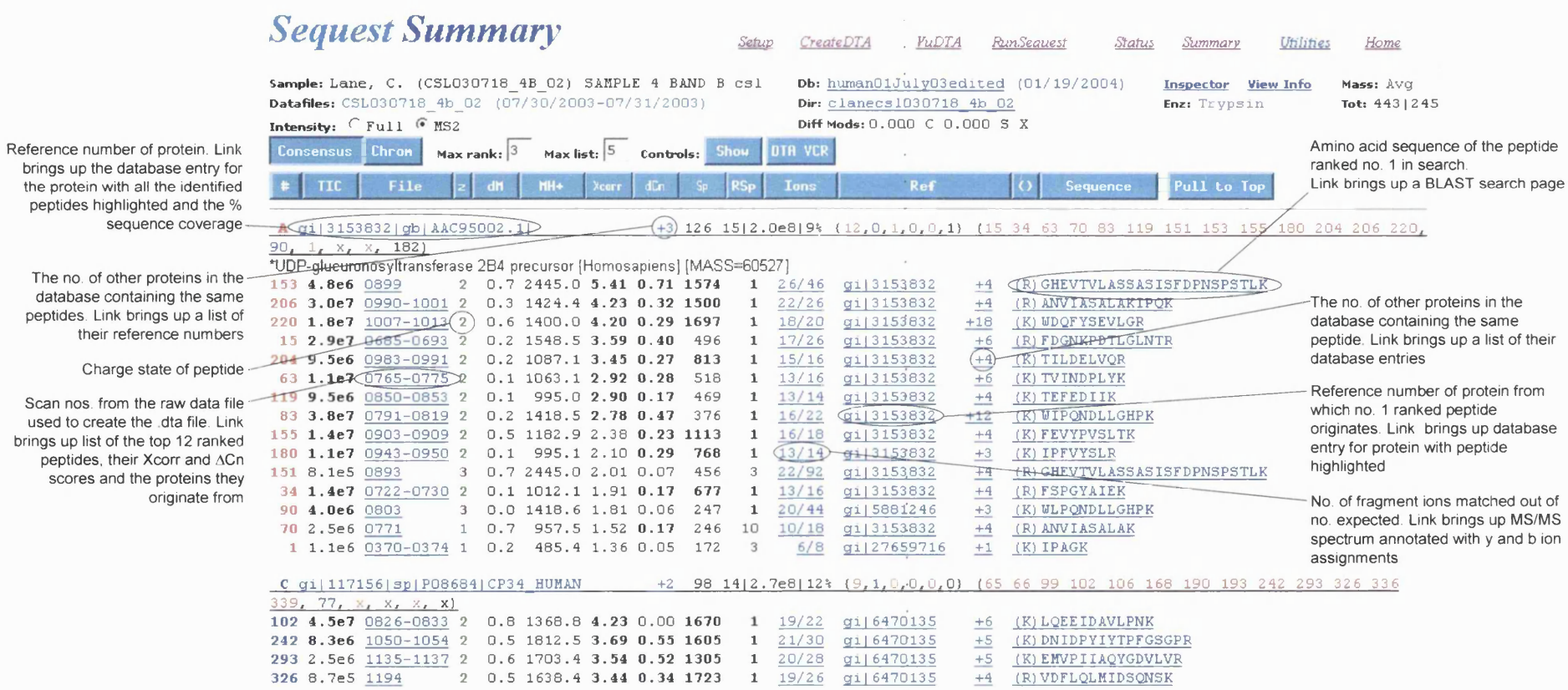


Figure 2.5 Part of a *Sequest Summary* page: the final output of a Sequest search.

2.14.3 GPMAW

GPMAW Version 3.04 (Lighthouse Data, Odense, Denmark) is a software program that predicts the peptides that are expected to result from the digestion of a given protein by a chosen protease, and displays their masses. GPMAW was used to calculate the molecular masses of truncated and modified proteins, to aid in the selection of peaks for MS/MS and zoom scan when data was being collected manually, and to assign peptide peaks in cytochrome c and myoglobin tryptic digest mass spectra before Sequest was available.

2.15 MALDI-TOF mass spectrometry

MALDI-TOF mass spectrometry was performed on a Voyager DE-PRO instrument from Applied Biosystems (Warrington, UK). For the analysis of intact proteins, the instrument was operated in positive linear mode with the following parameters: accelerating voltage 25,000 V, grid voltage 94%, guide wire 0.3%, extraction delay time 750 ns. The matrix solution contained 10 mg/ml sinapinic acid in 0.3% TFA, 30% acetonitrile. Internal, or close external calibration with BSA was used. 1 µl recombinant CYP1A2, 2E1 or 3A4 was mixed with 1 µl of 22 pmol/µl BSA standard in water and 4 µl matrix solution, and 1 µl of this was spotted onto a MALDI plate and allowed to dry. For the mechanism-based inactivation study, incubations were mixed with 1.2 pmol/µl BSA and matrix in the ratio 1:1:2 assay:BSA:matrix, and 1 µl was spotted onto a MALDI plate and allowed to dry.

Stainless steel flat plates were used for analysis of recombinant P450s 1A2, 2E1 and 3A4. For analysis of the mechanism-based inactivation study Teflon-coated plates were used, which have a hydrophilic mask to produce hydrophilic wells: spotted sample will form a bead above a well; as the sample dries the analyte becomes concentrated into a small diameter spot, affording an increase in sensitivity.

2.15.1 On-plate washing with 0.1% TFA

Sample was spotted onto the MALDI plate and allowed to dry. Spots were then washed with 0.1% TFA in water to remove small, water-soluble impurities. Each spot was overlaid with 2 µl of cold (4 °C) 0.1% TFA and left for approximately 10 seconds. The 0.1% TFA was removed with a Gilson pipette.

Chapter 3

Optimisation and assessment of methods for the analysis of cytochrome P450 proteins by mass spectrometry

3.1 Introduction

The cytochromes P450 are a super-family of haemoproteins that are responsible for the oxidative metabolism of a large number of drugs, xenobiotics and endogenous compounds. Mammalian P450s are membrane bound and hydrophobic in character. Although classed as one super-family of enzymes, different P450 proteins often share little amino acid sequence homology (Section 1.3.2). Cytochrome c and myoglobin, two water-soluble haemoproteins, were employed as P450 "surrogates" in order to optimise the various techniques used, and to develop and implement working protocols that could then be applied to P450s. Eventually work focussed on cytochrome c because it is more amenable to proteolysis than myoglobin [179]. In later analyses cytochrome c in-solution tryptic digest was used as the 'test standard' to ensure that the nano-LC-ES system was operating correctly. After sufficient progress had been made with cytochrome c, purified recombinant human P450 enzymes 1A2, 2E1 and 3A4 were used. Further development and optimisation was performed on these in preparation for work on biological tissues.

3.2 Intact protein analysis

3.2.1 Analysis of cytochrome c and myoglobin by nano-ES

Intact cytochrome c and myoglobin were used to test the use of the LCQ mass spectrometer for intact protein analysis. A nano-ES mass spectrum of intact cytochrome c is shown in Figure 3.1(a). The spectrum is typical of electrospray analyses of proteins, showing a series of ions of increasing charge with decreasing *m/z*. The deconvoluted spectrum is shown in Figure 3.1(b). The calculated mass of the protein agrees very well with the theoretical mass of 12360 Da.

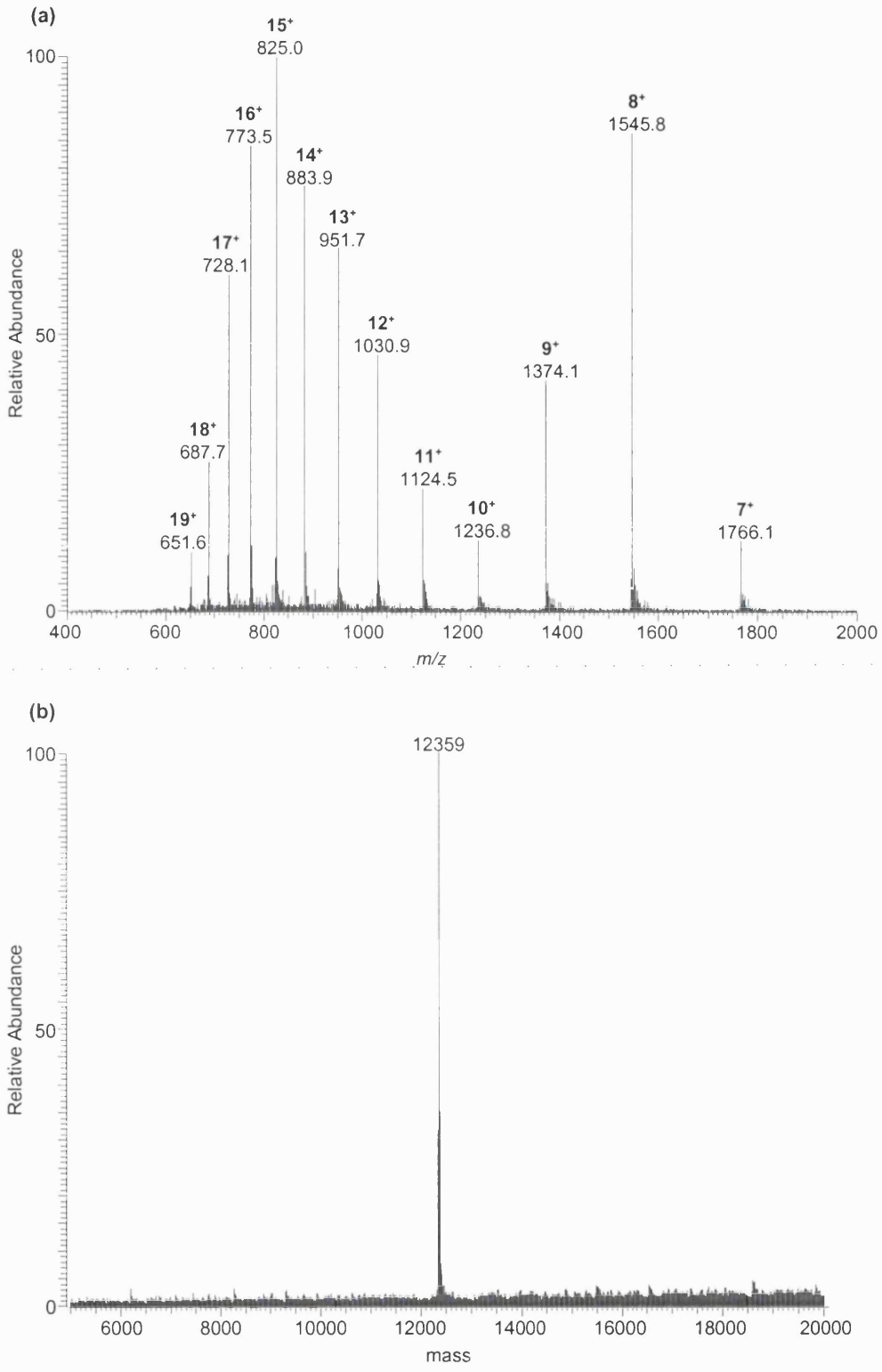


Figure 3.1 (a) A nano-ES mass spectrum of cytochrome c (8 pmol/µl in 0.5% formic acid in 70% methanol). The degree of protonation is shown in bold above each protein peak. **(b)** Deconvoluted mass spectrum, showing the mass of the protein. An average of 10 scans is shown. Instrumental conditions were optimised for cytochrome c.

3.2.2 Analysis of recombinant CYP1A2, 2E1 and 3A4 by nano-ES

Recombinant CYP1A2, 2E1 and 3A4 were analysed by nano-ES in order to determine their molecular masses. Initially, sample clean-up was attempted using C_4 ZipTips (Section 2.10.2) because the recombinant P450s were provided in a matrix containing 20 mM potassium phosphate buffer, 0.2 mM EDTA, 1 mM DTT and 20% glycerol, which is unsuitable for analysis by electrospray. However, later analyses were performed after dilution of the P450s but without prior sample clean-up.

3.2.2.1 Analysis of CYP1A2, 2E1 and 3A4 by nano-ES after sample clean-up using ZipTips

Recombinant CYP1A2, 2E1 and 3A4 were desalted using C_4 ZipTips (Section 2.10.2), eluted in either 0.5% formic acid in 70% methanol or 0.5% formic acid in 70% acetonitrile, and analysed by nano-ES (Figures 3.2 to 3.4). The haem group, which does not remain attached to the apoprotein during the electrospray process, can be seen at *m/z* 616.3 in each spectrum.

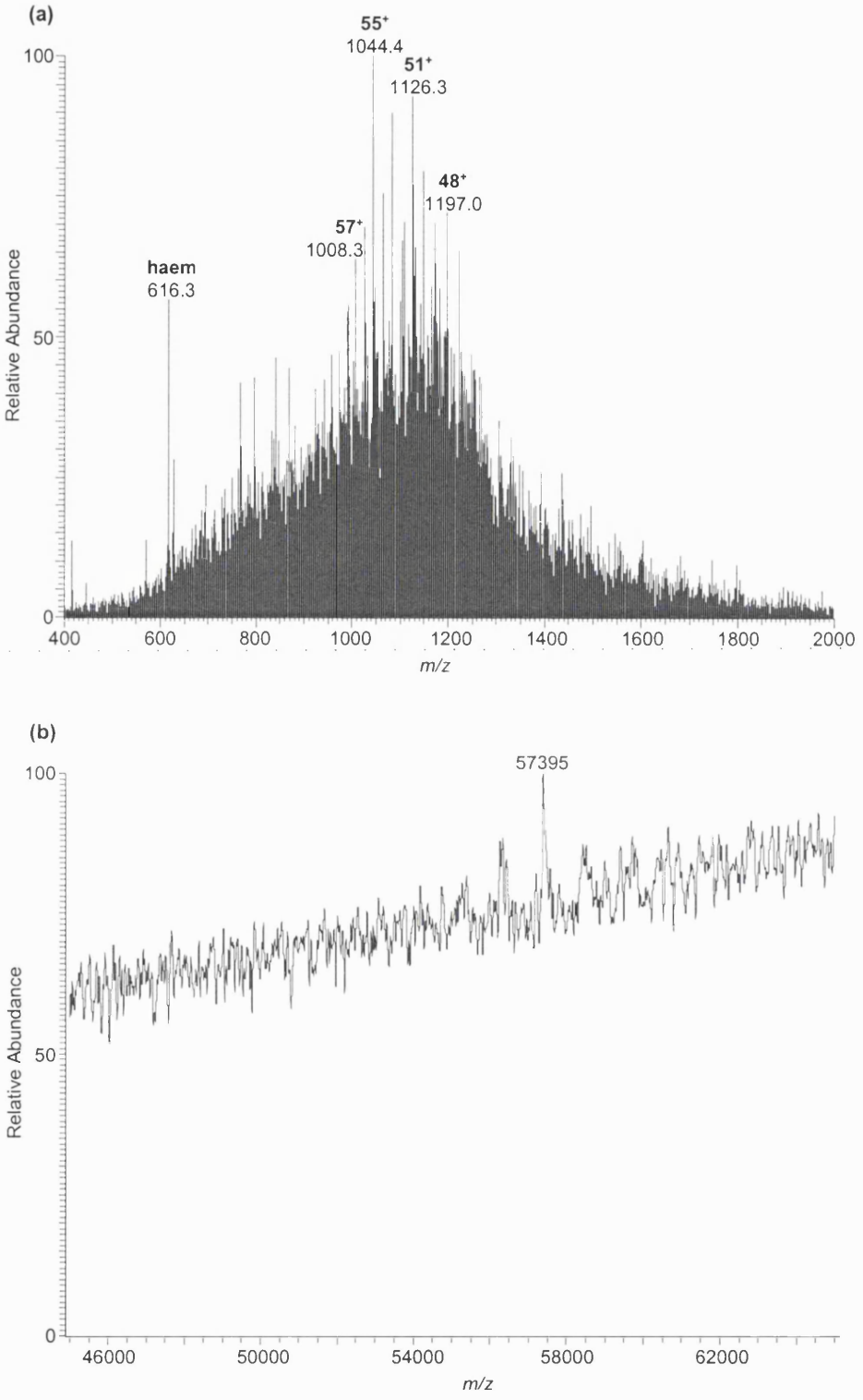


Figure 3.2 (a) A nano-ES mass spectrum of intact CYP1A2, acquired after sample clean-up using ZipTips. Some of the charge states on the protonated molecules are indicated in bold above the peaks. **(b)** Deconvoluted spectrum showing the measured mass of the protein. An average of 62 scans is shown. Sample was desalted using C4 ZipTips (Section 2.10.2) and eluted in 0.5% formic acid in 70% methanol. The concentration of desalted CYP1A2 was 4.4 pmol/μl, assuming all the protein was eluted from the Ziptip. Instrumental conditions were optimised using a desalted CYP2E1 sample.

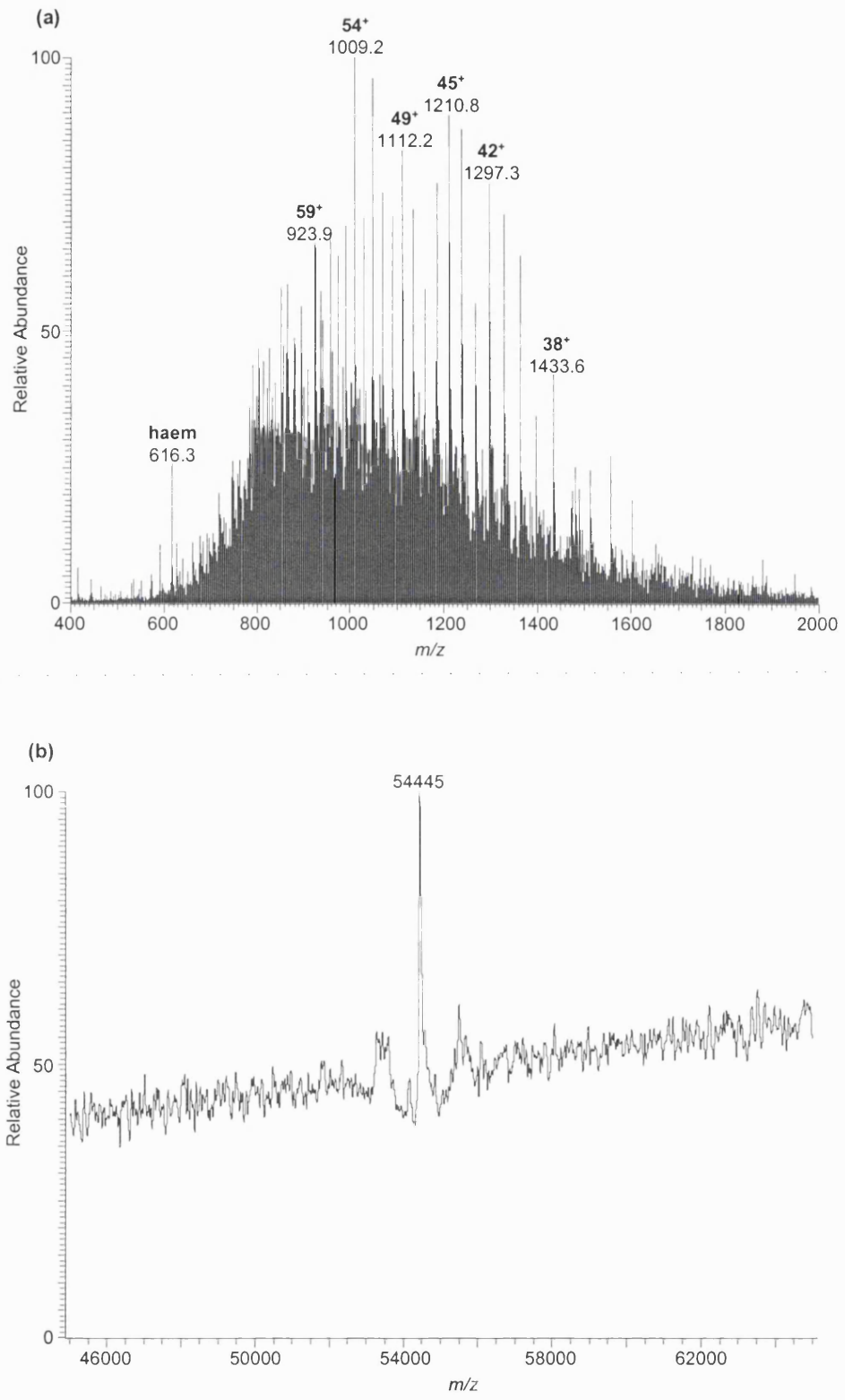


Figure 3.3 (a) A nano-ES mass spectrum of intact CYP2E1, acquired after sample clean-up using ZipTips. Some of the charge states on the protonated molecules are indicated in bold above the peaks. **(b)** Deconvoluted spectrum showing the measured mass of the protein. An average of 71 scans is shown. Sample was desalted using C_4 ZipTips (Section 2.10.2) and eluted in 0.5% formic acid in 70% methanol. The concentration of desalted CYP2E1 was 4.8 pmol/µl, assuming all the protein was eluted from the Ziptip. Instrumental conditions were optimised for CYP2E1.

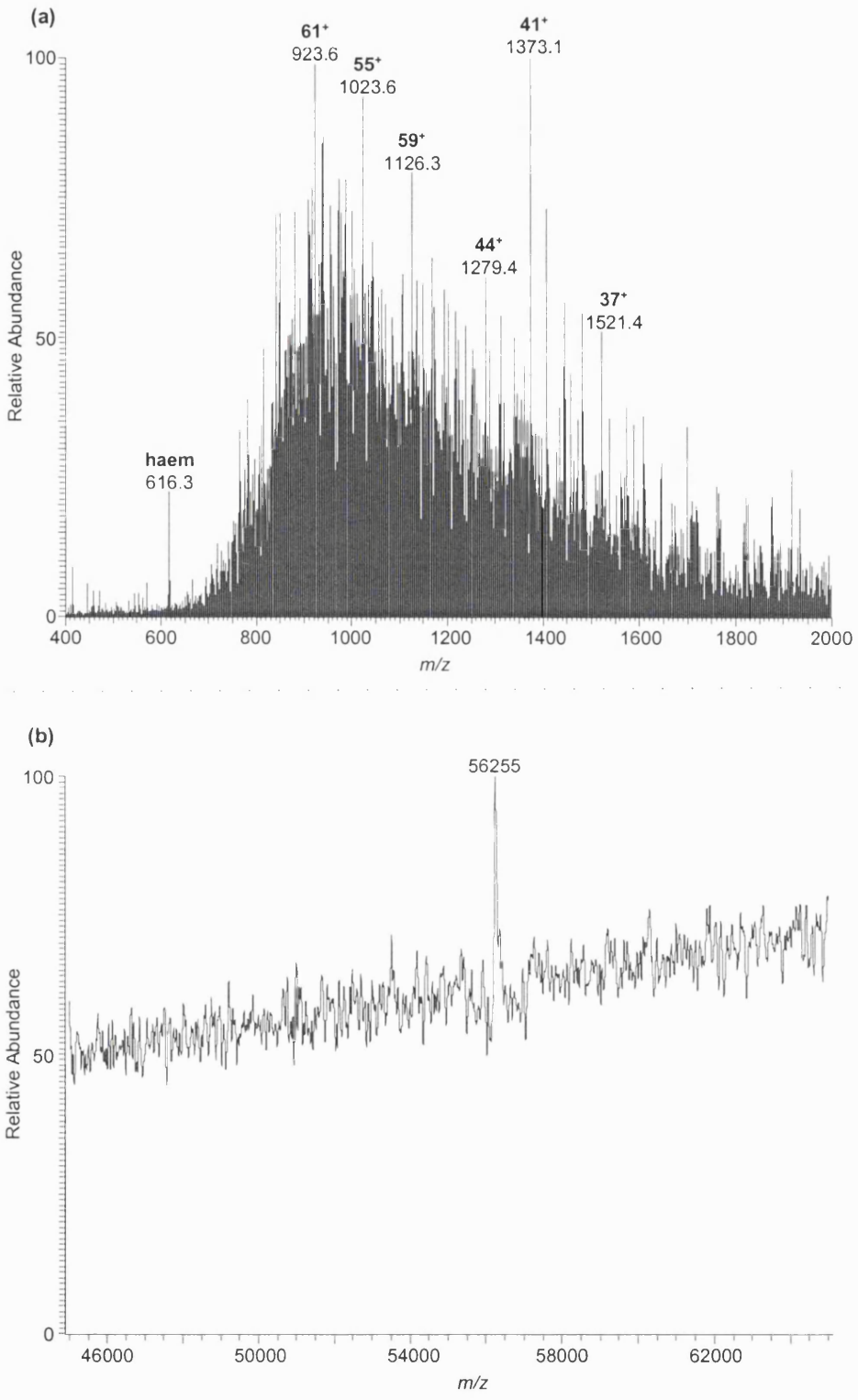


Figure 3.4 (a) A nano-ES mass spectrum of intact CYP3A4, acquired after sample clean-up using ZipTips. Some of the charge states on the protonated molecules are indicated in bold above the peaks. **(b)** Deconvoluted spectrum showing the measured mass of the protein. An average of 42 scans is shown. Sample was desalted using C_4 ZipTips (Section 2.10.2) and eluted in 0.5% formic acid in 70% acetonitrile. The concentration of desalted CYP3A4 was 5.9 pmol/μl, assuming all the protein was eluted from the Ziptip. Instrumental conditions were optimised using a desalted CYP2E1 sample.

The mean intact mass of CYP2E1 was 54446 Da (range 54443 – 54452), calculated from 10 different deconvolutions of protein spectra of sample eluted in both 0.5% formic acid in 70% methanol and 0.5% formic acid in 70% acetonitrile. An average mass of 57410 (range 57395 – 57422) was calculated from deconvolutions of 8 spectra of CYP1A2 (eluted in 0.5% formic acid in 70% methanol), and an average mass of 56252 (range 56250 – 56255; 3 deconvoluted spectra) was calculated from the spectra recorded for CYP3A4 eluted in 0.5% formic acid in 70% acetonitrile. These values are compared with values from literature in Section 3.2.2.3.

The use of either methanol or acetonitrile in the elution solvent did not affect the MS signal obtained for intact CYP2E1 analysis; however, the use of methanol produced better results for CYP1A2 and the use of acetonitrile produced better results for CYP3A4. The MS data for CYP1A2 was not of sufficient quality to calculate an intact mass after elution with 0.5% formic acid in 70% acetonitrile, and the MS data for CYP3A4 was not of sufficient quality to calculate an intact mass after elution with 0.5% formic acid in 70% methanol. However, assuming that all the protein was eluted from the ZipTip resin, the concentrations of CYP1A2, 2E1 and 3A4 after desalting would have been 4.4, 4.8 and 5.9 pmol/µl, respectively. These levels should be detectable by static nanospray without difficulty.

The effect of ZipTip desalting on the intact P450 proteins was investigated by SDS-PAGE. CYP1A2, 2E1 and 3A4 were desalted as described in Section 2.10.2 and eluted in 0.5% formic acid in 70% methanol. All the solutions that could potentially contain the P450 proteins were analysed for evidence of protein by SDS-PAGE, along with control samples prepared from CYP1A2, 2E1 and 3A4 without desalting. The gel lanes containing the control samples showed very strong bands. Very faint protein bands were visible in the lanes containing the elution solvent and in the lanes containing the original sample matrix. No protein was evident in the lanes containing the washes. It was concluded that the majority of the protein remains bound to the ZipTip resin, and hence the C_4 chromatography medium used in the ZipTips is not suitable for the clean-up of P450 proteins, probably due to their hydrophobic nature.

3.2.2.2 Analysis of recombinant CYP1A2, 2E1 and 3A4 by nano-ES without prior sample clean-up

Analysis of recombinant CYP1A2, 2E1 and 3A4 by nano-ES after desalting with ZipTips was only partially successful because most of the protein was retained on the ZipTip resin and could not be eluted (Section 3.2.2.1). Attempts were therefore made to analyse the three recombinant proteins without prior sample clean-up. Recombinant P450 enzymes were diluted 1 in 10 with 10% formic acid (the high concentration of acid was used to solubilise the proteins) and analysed by nano-ES.

Nano-ES mass spectra of CYP2E1 and CYP3A4, along with the corresponding deconvoluted spectra, are shown in Figures 3.5 and 3.6. No data could be obtained for CYP1A2.

Mean intact masses of 54476 (range 54471 – 54479, 4 deconvolutions) and 56253 (range 56251 – 56255, 3 deconvolutions) were measured for 2E1 and 3A4, respectively.

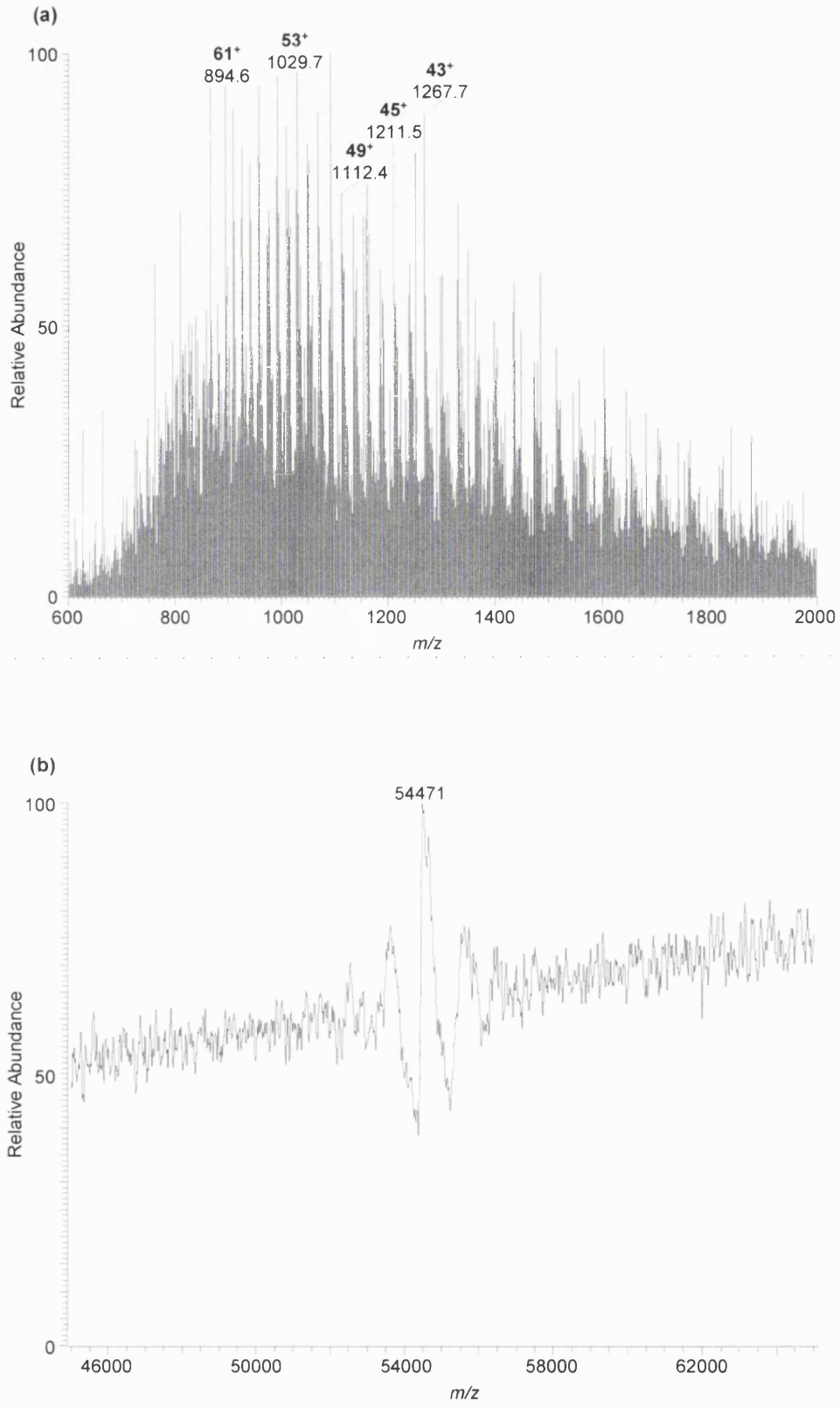


Figure 3.5 (a) A nano-ESI mass spectrum of intact CYP2E1 (1.6 pmol/µl), acquired with no sample clean-up prior to analysis. Some of the charges on the protein molecules are shown in bold above the peaks. **(b)** The corresponding deconvoluted spectrum, showing the mass calculated for the protein. An average of 66 scans is shown. Sample was diluted 1 in 10 with 10% formic acid. Instrumental conditions were optimised using CYP2E1.

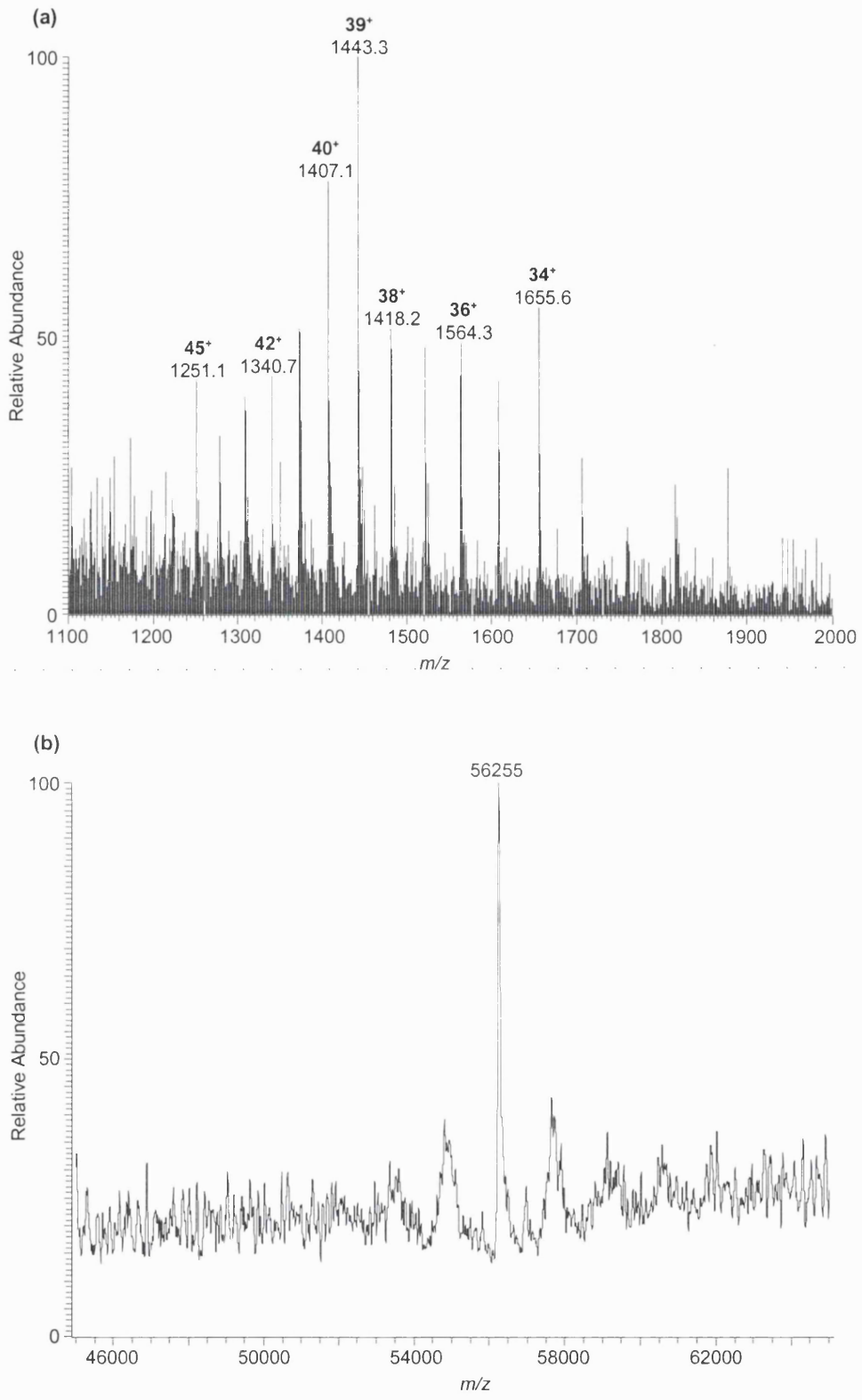


Figure 3.6 (a) A nano-ESI mass spectrum of intact CYP3A4 (3.0 pmol/µl), acquired with no sample clean-up prior to analysis. Some of the charges on the protein molecules are shown in bold above the peaks. **(b)** The corresponding deconvoluted spectrum, showing the mass calculated for the protein. An average of 25 scans is shown. Sample was diluted 1 in 10 with 10% formic acid. Instrumental conditions were optimised using CYP2E1.

Sample clean-up is usually considered essential for analysis by electrospray because the presence of salts such as K_3PO_4 and involatile buffers and solvents such as glycerol inhibit ionisation. Salts can form ion pairs and dilute the sample signal by the formation of adduct ions. Non-volatile solvents and buffers inhibit ion desolvation during electrospray and can lead to clogging and blockages in the instrument. However, very little analyte is used during nano-ES, therefore instrument blockages are unlikely to occur. Blockage of the nano-ES needle is more likely, but these are disposable. The proteins were diluted 1 in 10 prior to analysis, thereby reducing the concentrations of phosphate buffer, EDTA, DTT and glycerol to 2 mM, 0.02 mM, 0.1 mM and 2%, respectively (Section 2.10.2). This was successful for the analysis of CYP3A4, for which excellent signal to noise ratio was attained (Figure 3.6), and CYP2E1 (Figure 3.5) but not for CYP1A2, for which no protein signal was observed under these conditions.

3.2.2.3 A comparison of the observed masses of recombinant CYP1A2, 2E1 and 3A4 with the theoretical masses

The measured intact masses of CYP1A2, 2E1 and 3A4 are compared with the theoretical masses in Table 3.1.

Table 3.1 Theoretical and experimental molecular masses of recombinant CYP1A2, 2E1 and 3A4.

P450	Molecular mass (Da)		
	Calculated sequence mass [a]	Nano-ES with desalting	Nano-ES with no sample clean-up
1A2	58287	57410	not determined
2E1 (batch 1) [b]	54569 [c]	54446	not determined
2E1 (batch 2) [b]	54569 [c]	not determined	54476
3A4	57205 [d]	56252	56253

[a] Calculated using GPMAW (Section 2.14.3) from the Swissprot amino acid sequences (http://us.expasy.org/sprot/). Values are for the sequence mass, i.e. without the mass of the haem. [b] Two different batches of CYP2E1were analysed. [c] N-terminal modifications for CYP2E1 were as described in Gillam *et al* [183]; the mass value was adjusted accordingly. [d] Analysis of tryptic digests of CYP3A4 showed Trp391 → Val; this is a Swissprot-registered conflict. The calculated intact mass of the protein was adjusted to account for this.

The nano-ES measured mass of CYP2E1 batch 1 is approximately 30 Da lower than that of CYP2E1 batch 2. This could be due to formylation of CYP2E1 from batch 2 (which was diluted with 10% formic acid), resulting in the addition of a mass of 28 Da to the N-terminal of the protein, or to potassium ion adduction, since batch 2 was

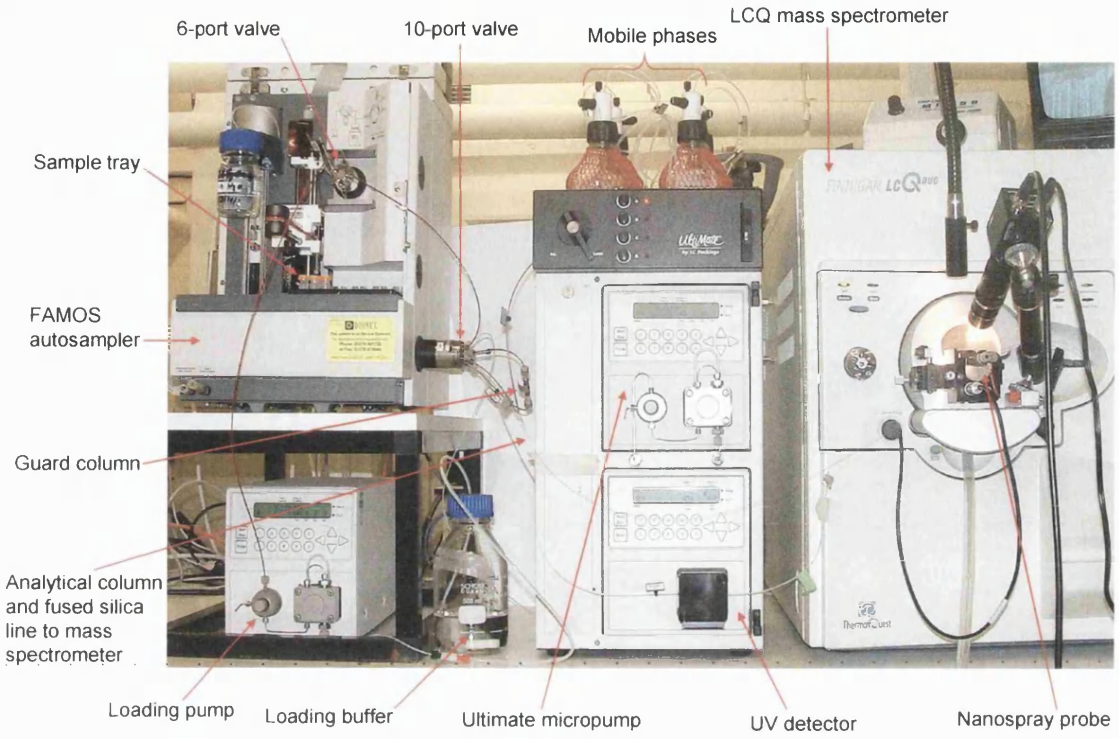
analysed without prior desalting. Alternatively, the presence of adduct peaks in the electrospray raw data for CYP2E1 batch 2 could have resulted in loss of accuracy during the deconvolution step. It is also possible that the two different batches of CYP2E1 had different masses.

The measured masses of the P450 proteins do not agree well with the calculated masses. The recombinant proteins are therefore likely to be truncated forms of the wild-type proteins. An average mass of 57473 can be calculated for CYP1A2 by removing the first 8 N-terminal amino acids; alternatively, an average mass of 57442 can be calculated by removing the first 7 C-terminal amino acids. An average mass of 54438 can be calculated for CYP2E1 by removing the first N-terminal amino acid, or 54482 by removing the first C-terminal amino acid; an average mass of 56251 Da can be calculated for CYP3A4 by removing the first 9 N-terminal amino acids, or 56245 Da by removing the first 10 C-terminal amino acids. Neither the N- or C-terminal peptides were observed during analysis of recombinant CYP1A2, 2E1 or 3A4. There is no indication from the commercial provider of the recombinant P450s that CYP1A2 and CYP3A4 are truncated or that CYP2E1 was modified in any way other than that described by Gillam *et al* [183] but, given their recombinant nature, this is plausible.

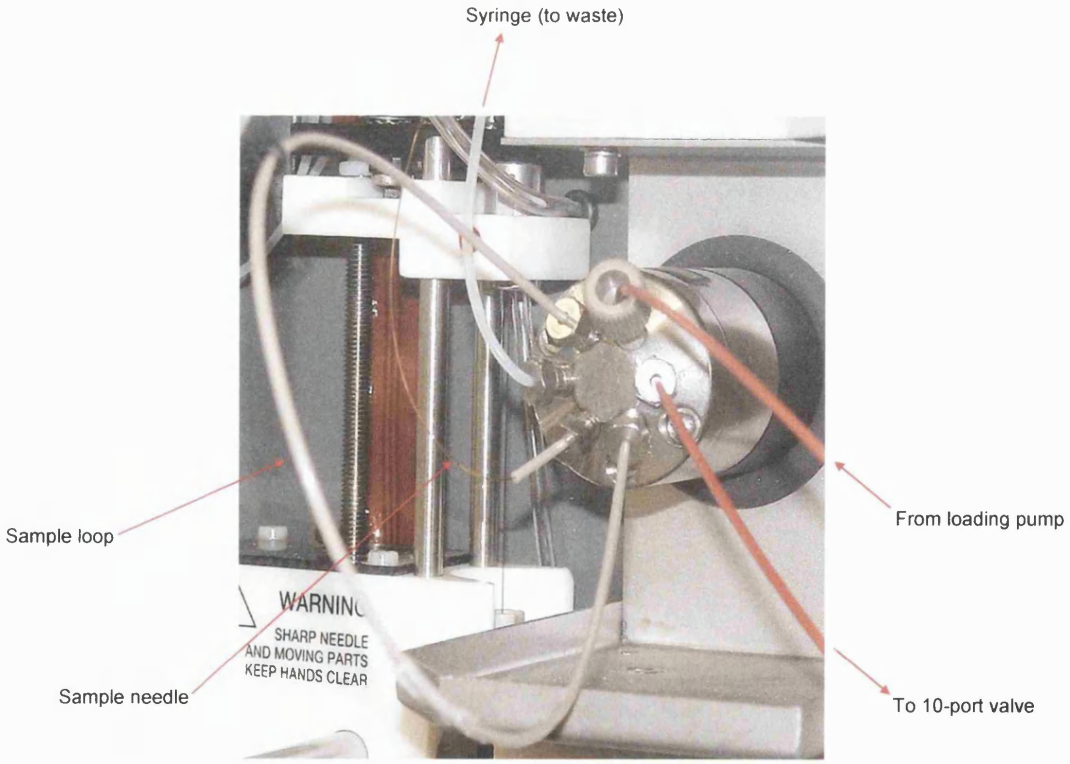
3.3 Configuration of the nano-LC for online connection to the mass spectrometer

The nano-LC was used for separation of tryptic peptides prior to mass spectrometric analysis, although initially it was configured as a stand-alone instrument with UV detection. Figure 3.7(a) shows the nano-LC connected to the LCQ mass spectrometer. Close-up views of the 6- and 10-port valves are shown in Figure 3.7(b) and (c); schematic diagrams of the 6- and 10-port valves are shown in Figure 3.8. During the configuration of the nano-LC for connection to the mass spectrometer, three areas were focussed on for optimisation: the volume of sample taken for each injection, the dead volume in the connecting tubing, and the time at which the 10-port valve was switched to allow sample to be transferred from the guard column onto the analytical column.

(a)


6-port valve
10-port valve
Mobile phases
LCQ mass spectrometer
Sample tray
FAMOS autosampler
Guard column
Analytical column and fused silica line to mass spectrometer
Loading pump
Loading buffer
Ultimate micropump
UV detector
Nanospray probe

(b)


Syringe (to waste)
Sample loop
Sample needle
From loading pump
To 10-port valve
SHARP NEEDLE
AND MOVING PARTS
KEEP HANDS CLEAR

(c)

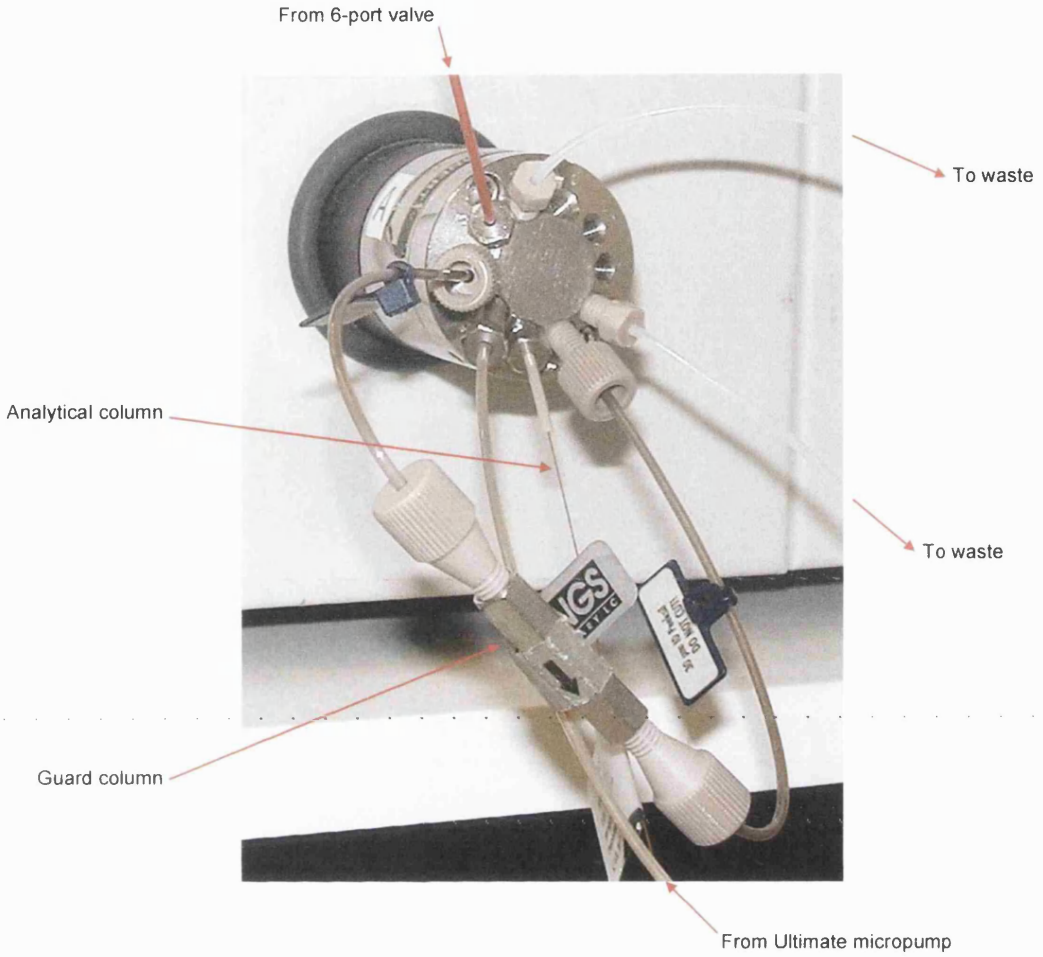


Figure 3.7 The instrumental setup used for the analysis of tryptic peptides. **(a)** Overall view of the FAMOS autosampler and Ultimate nano-LC connected to the LCQ mass spectrometer. **(b)** The FAMOS 6-port valve. **(c)** The FAMOS 10-port valve.

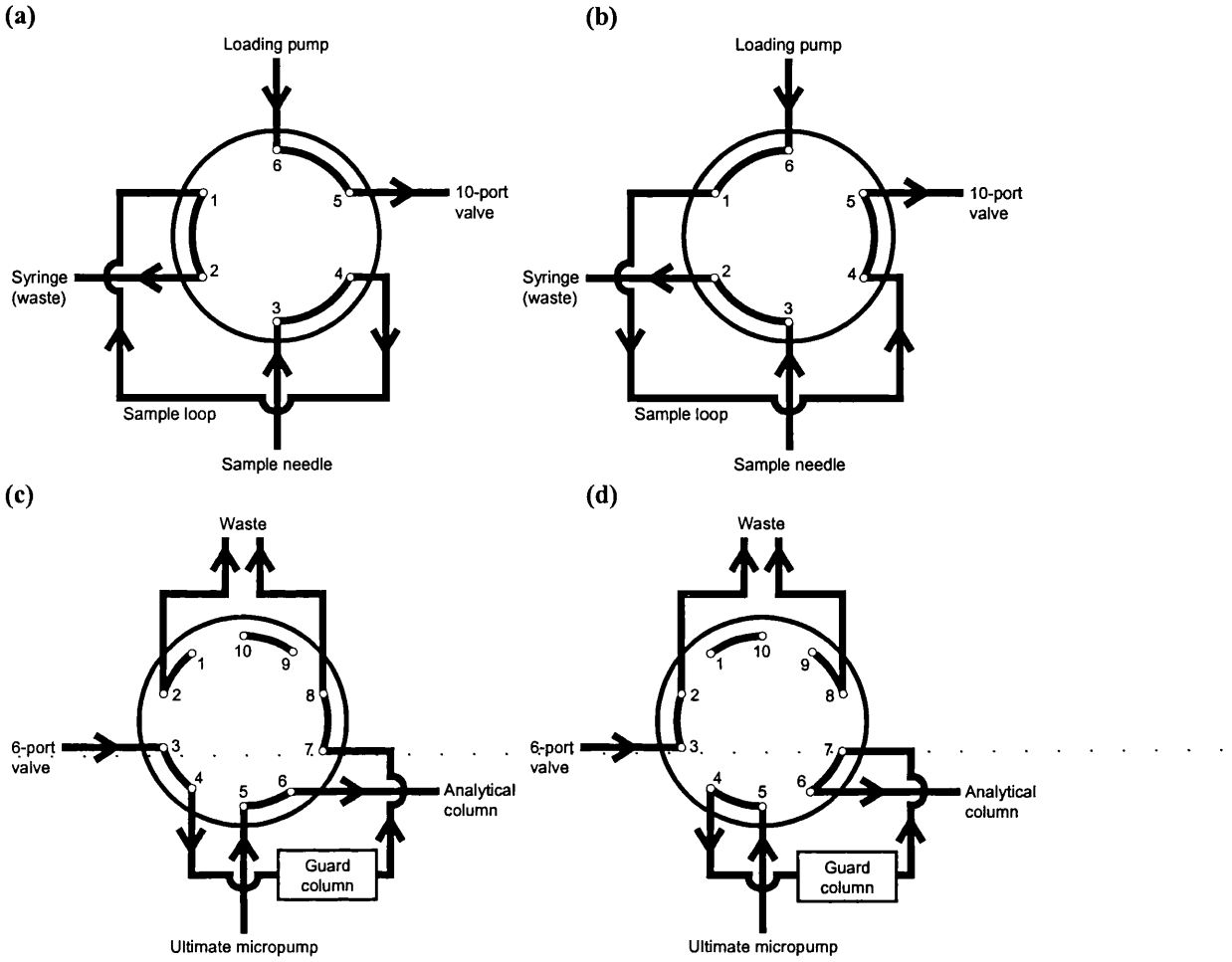


Figure 3.8 Schematic diagrams of the 6- and 10-port valves on the FAMOS autosampler. **(a)** 6-port valve: sample is loaded onto the sample loop; **(b)** 6-port valve: the 6-port valve is switched and sample is loaded onto the 10-port valve; **(c)** 10-port valve: sample is loaded from the 6-port valve onto the guard column and washed with loading buffer; the loading buffer is directed to waste; **(d)** after a user-specified period of time the 10-port valve is switched and sample is directed onto the analytical column by the mobile phases coming from the Ultimate micropump.

3.3.1 The effect of minimising sample volume

The FAMOS autosampler was originally configured with a 1 µl sample loop, and full loop injections were performed. Full loop injection gives maximum reproducibility, but considerable sample volume is wasted because sample is used to flush the lines and fill the sample loop. An alternative to the full loop injection is the "µl pick-up" injection, for which only the volume of sample to be injected is taken, preceded and followed by transport liquid. The maximum injection volume for injection by µl pick-up is calculated using the following formula:

$$\text{Maximum injection volume} = \frac{\text{loop volume} - (3 \times \text{needle tubing volume})}{2}$$

The needle tubing volume is 2.4 µl, therefore a 10 µl sample loop (Figure 3.7(b)) was installed to enable sample injections of up to 1.4 µl.

3.3.2 The effect of minimising dead volume

Dead volume in the LC system causes chromatographic band broadening and time delays, especially at flow rates as low as 200 nl/min. In order to minimise delay in the gradient the autosampler was raised so that the length of the tubing from the 10-port valve to the Ultimate micropump could be minimised (Figure 3.9). In order to avoid band broadening after the sample had left the column, the UV detector was bypassed so that the distance between the end of the analytical column and the LCQ source could be minimised (Figure 3.9). The lengths of the two pieces of tubing connecting the guard column to the 10-port valve were also minimised (Figure 3.9).

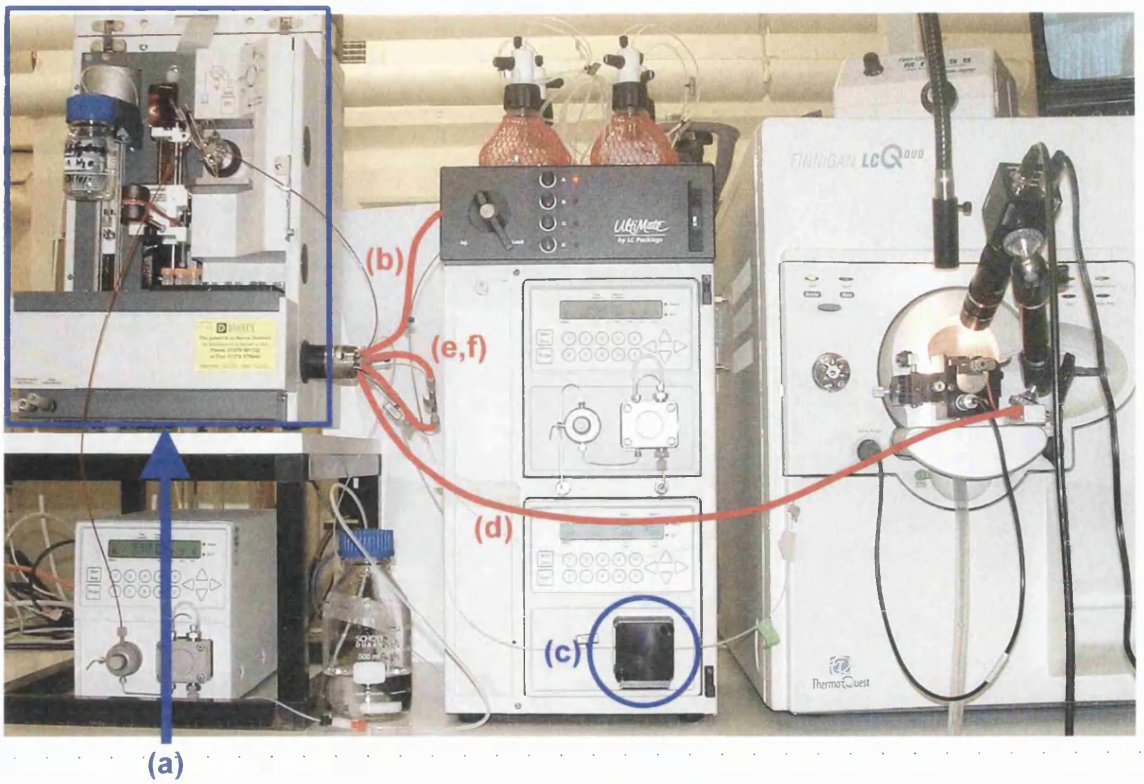


Figure 3.9 The FAMOS autosampler, Ultimate nano-LC and LCQ mass spectrometer, showing the areas where adjustments were made in order to minimise dead volume in the connecting tubing. The FAMOS autosampler was raised (a) so that the length of tubing connecting the Ultimate micropump to the 10-port valve could be minimised (b). The UV detector was bypassed (c) so that the distance between the end of the analytical column and the LCQ source could be minimised (d). The lengths of the two pieces of tubing connecting the guard column to the 10-port valve were minimised (e, f).

3.3.3 Optimisation of the 10-port valve switching time

Before being transferred to the LC column, samples were loaded onto a guard column and washed with loading buffer for a user-specified period of time; the loading buffer was directed to a waste container. The 10-port valve was then switched so that sample was brought in-line with the Ultimate micropump and directed by the mobile phases onto the analytical column (Figure 3.8). Replacement of the 1 µl sample loop by the 10 µl sample loop (Section 3.3.1) and reduction of the volume of the tubing connecting the 10-port valve to the guard column (Section 3.3.2) affects the time taken for the sample to reach the guard column, and therefore the valve switching time needed to be altered accordingly.

The valve switching time should take into account the time taken for the sample to travel to the guard column, and should be long enough to allow the sample to be

washed before it is transferred onto the analytical column in time for the start of the LC gradient. The volume of loading buffer used to wash the sample on the guard column can be calculated from the volume of tubing in the system (sample loop plus tubing going from the 6-port valve to the 10-port valve plus tubing connecting the 10-port valve to the guard column, Figure 3.10), the flow rate on the loading pump and the switching time. Substituting the 1 µl sample loop for the 10 µl sample loop (Section 3.3.1) introduces 9 µl extra volume in the system. Some of this extra volume was negated by shortening the length of the tubing running from the 6-port to the 10-port valve (Figure 3.10); however, an increase in the 10-port valve switching time was necessary in order to account for the extra volume in the system. The parameters before and after optimisation are shown in Table 3.2.

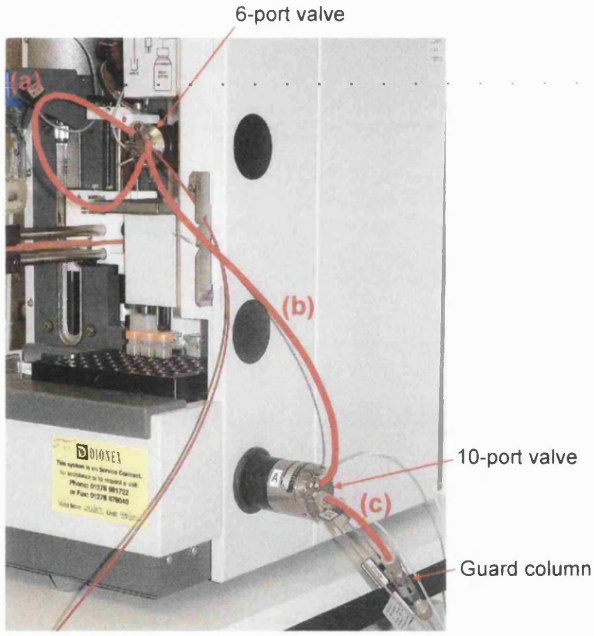


Figure 3.10 The 6- and 10-port valves on the FAMOS autosampler. The volume of the tubing shown in red was taken into account for the calculations for the optimisation of the 10-port valve switching time. (a) 10µl sample loop, (b) tubing connecting the 6- and 10-port valves, and (c) tubing connecting the 10-port valve to the guard column.

Table 3.2 Volume of tubing in the nano-LC system, 10-port valve switching times, loading pump flow rates and corresponding sample wash volumes before and after optimisation.

	Before optimisation	After optimisation
Sample loop volume (µl)	1	10
Total tubing volume (µl) [a]	9.8	14.8
Valve switching time (min)	1.5	3.5
Loading pump flow rate (µl/min)	30	20
Sample wash volume (µl)	45	70
Sample wash volume/total tubing volume [b]	4.6	4.7

[a] Total tubing volume = volume of sample loop + volume of tubing from 6-port to 10-port valve + volume of tubing connecting 10-port valve to guard column (Figure 3.10). [b] The ratio of sample wash volume to total tubing volume remained approximately the same before and after optimisation.

The new valve switching time was calculated so that the ratio of sample wash volume to total tubing volume remained approximately the same before and after optimisation. A series of analyses of cytochrome c in-solution tryptic digest was performed using different switching times in order to investigate the effect of switching time on peptide chromatography: 0.8 pmol cytochrome c tryptic digest in 0.1% TFA was analysed three times at six different switching times (2.5 min, 3 min, 3.5 min, 4 min, 4.5 min and 5 min). Eight chromatographic peptide peaks were integrated, and their heights and areas analysed to establish whether there was a difference in the amount of peptide eluted from the guard column at different switching times. No major difference in peptide amounts was noted, therefore the valve switching time was kept at 3.5 min.

3.4 Optimisation of the LCQ instrument parameters for the collection of protein digest data

There are several options available when deciding on an instrument method for the collection of protein digest data. Data can be collected manually or the mass spectrometer can be configured for automatic data collection. When using ion trap mass spectrometers such as the LCQ, which do not have high mass accuracy, MS/MS analysis must be carried out: ions of the *m/z* ratio of interest are isolated in the ion trap and excited so that they collide with the helium buffer gas present in the trap. The collisions of the parent ions cause them to fragment to produce product ions, which are then scanned out of the ion trap and detected. However, there is also the option of performing a second type of scan, known as a "zoom scan", to

determine the charge states of the ions: a scan of 10 *m/z* width is conducted, which shows the $^{12}C/^{13}C$ isotopic separation of a specified ion. An ion with charge state +1 will have isotopic separation of 1 *m/z*; an ion with charge state +2 will have isotopic separation of 0.5 *m/z*; an ion with charge state +3 will have isotopic separation of 0.33 *m/z*, and so on. MS/MS scans, zoom scans and full MS scans can be combined in various different ways for the collection of data. Samples can be analysed as mixtures without prior chromatographic separation, or on-line LC can be employed to simplify the samples prior to mass spectrometric analysis. The following sections address some of these options.

3.4.1 Manual versus automatic data collection

Initially, data was collected manually. This involved electrospraying a sample into the mass spectrometer using nano-ES, selecting peaks for MS/MS and zoom scan analysis by eye, then manually performing these scans. GPMAW (Section 2.14.3) printouts, predicting the *m/z* values of the peptide ions expected from tryptic digestion of the proteins being analysed, were used as an aid in peak selection. However, the collection of data in this way was found to be extremely labour-intensive and also inefficient, since it was often difficult to distinguish peptide ion peaks from background noise and contaminants. Programming the LCQ mass spectrometer to automatically select ions and perform the necessary analysis was found to be a far more effective way of obtaining useful data in much greater amounts than could be collected manually.

3.4.2 "Tripleplay" versus "Big 3" data collection

Several approaches to automated data collection can be designed using different combinations of full MS, MS/MS and zoom scans. The data dependent "Triple Play" experiment is a commonly used alternative to the "Big 3" experiment described in Section 2.12. The "Triple Play" experiment involves repeating the following three scan events for the duration of the analysis time:

1. Full scan MS
2. Data dependent zoom scan on the most intense ion from 1.
3. Data dependent MS/MS on the most intense ion from 1.

However, Sequest does not need zoom scan data to identify a peptide sequence: if no interpretable zoom scan data is available, it will assign either a +1 charge state, creating one .dta input file, or a +2 and +3 charge state, creating two .dta input files. A +1 charge state is assigned when there are no fragment ions with *m/z* values above the *m/z* value of the precursor ion; for spectra where fragment ions are present with *m/z* values higher than the *m/z* value of the precursor ion, the program calculates the molecular weight assuming both a +2 and +3 charge state and creates a .dta file for each. The most useful information during data acquisition is collected during MS/MS scans. Therefore, in order to maximise the number of MS/MS scans collected, the "Triple Play" experimental setup was rejected in favour of the "Big 3" setup.

3.4.3 A comparison of nano-ES with nano-LC-ES

The peptide identifications made after nano-ES analysis of a protein digest mixture were compared with those made after nano-LC-ES analysis of the same sample. The use of on-line nano-LC separation prior to electrospray ionisation and mass spectrometric detection simplifies the sample: preferential ionisation effects, which can be observed in complex mixtures, should be minimised for nano-LC-ES compared with nano-ES. However, the use of nano-ES rather than nano-LC-ES removes the time constraint of an LC peak-width from the analysis by decoupling the ES from the LC: 5 µl of sample can last all day with nano-ES, affording plenty of time to analyse all the components in a mixture.

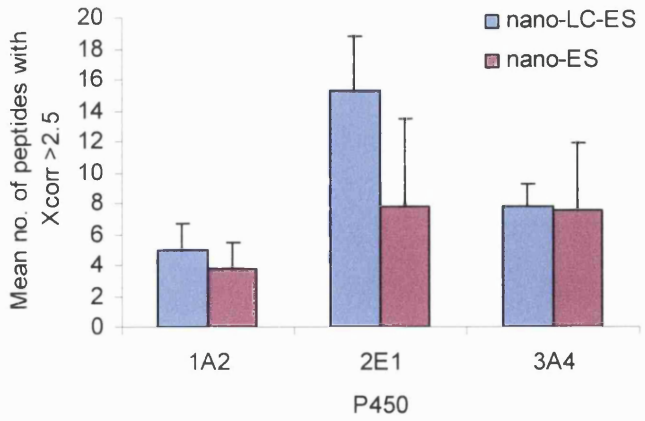
The method comparison was performed using a mixture of CYP1A2, 2E1 and 3A4 digested in solution. P450 tryptic digests were mixed together such that each digest was present at a concentration of approximately 290 fmol/µl. Samples were analysed by nano-ES and nano-LC-ES using the methods described in Sections 2.12.1 and 2.12.2, respectively, and the results searched using Sequest (Section 2.14.2). Four analyses were made using nano-ES and four using nano-LC-ES. A table of results is

shown in Appendix 1, Table A1.1, along with mean and standard deviation values (Table A1.2).

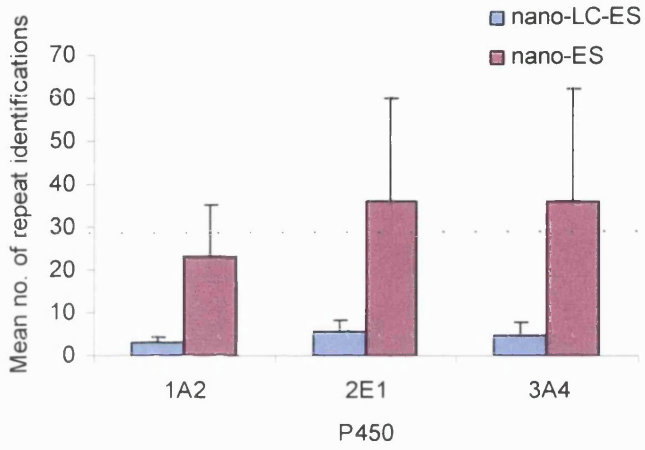
Results generated by Sequest searching were evaluated using Student's two-sampled t-tests for the difference in two means. For the work described here, proteins that were matched by two or more peptides with Xcorr scores ≥ 2.5 were considered to be conclusively identified, provided the peptides were unique to that protein in the database and the ΔC_n scores were ≥ 0.1 (Section 2.14.2.1). The statistical analysis indicated that there is no significant difference ($P < 0.05$) in the mean number of peptides identified (Xcorr ≥ 2.5) after nano-LC-ES compared with nano-ES analysis (Figure 3.11(a)), but that nano-LC-ES analysis generates a smaller number of repeat and incorrectly listed peptide identifications than nano-ES analysis ($P < 0.05$) (Figure 3.11(b) and (c)). Incorrectly listed peptide identifications are defined as peptide sequences that are listed on the *Sequest Summary* page under the heading of a P450 protein from which they do not originate. This occurs because Sequest lists 1st choice peptide sequence identifications under the heading of the protein from which the corresponding 2nd or 3rd choice peptide sequences originate (See Section 3.6.3).

For the analysis of the three-protein digest mixture both nano-ES and nano-LC-ES techniques have equal value for peptide identification (Xcorr ≥ 2.5); however, the results generated by nano-LC-ES are easier to interpret than those generated by nano-ES due to the significantly greater number of repeat and incorrect peptide sequence identifications generated by nano-ES compared with nano-LC-ES (Figure 3.11).

(a)



(b)



(c)

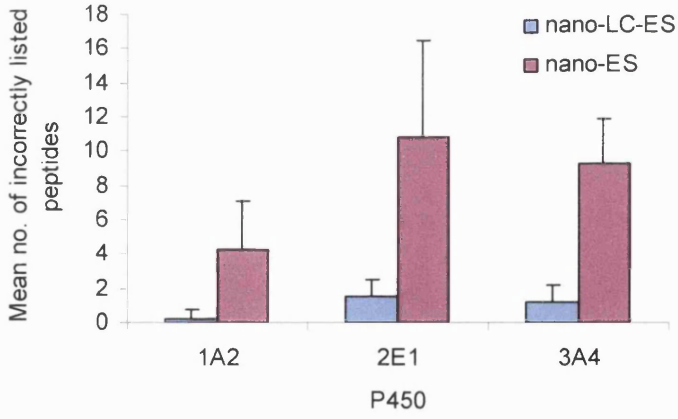


Figure 3.11 Comparison of nano-LC-ES with nano-ES. The mean results from four replicate nano-LC-ES analyses and four replicate nano-ES analyses are displayed. Error bars show +1 standard deviation of the mean. **(a)** Mean number of peptides with Xcorr $\geq$ 2.5. There was no significant difference ($P < 0.05$) in the nano-ES and nano-LC-ES means for CYP1A2, 2E1 or 3A4. **(b)** Mean number of repeat peptide identifications (all Xcorr values). The differences in the nano-ES and nano-LC-ES means were significant for CYP1A2 and 2E1 ($P < 0.05$), but not for CYP3A4 ($P < 0.05$). **(c)** Mean number of incorrectly listed peptide identifications on the *Sequest summary* page under the P450 heading (all Xcorr values). The differences in the nano-ES and nano-LC-ES means were significant ($P < 0.05$) for all three P450s. For analysis by nano-ES the digest mixture was acidified to contain 0.1% TFA and desalted using ZipTips (Section 2.10.1). For analysis by nano-LC-ES sample was not desalted; 1 µl of sample was injected via a 10 µl sample loop. MS/MS data were searched with Sequest (Section 2.14.2) against the Finnigan Xcalibur human database (Section 2.14.1).

3.5 Interrogation and modification of a NCBI FASTA database for use with Sequest searching for P450 proteins

MS/MS spectra were searched using Sequest software, which requires FASTA formatted databases (Section 2.14.1). The databases initially used were purchased with the Sequest software and were compiled in July 2000 (Section 2.14.1). However, there was a need to update the entries in the protein database because new P450 sequences are continually being reported. The NCBI builds its databases from a variety of sources, including submissions by individual authors. Sequence submissions are assigned accession numbers and entered into the database; NCBI databases are continually updated with new sequence information [184]. In July 2003, a non-redundant protein database in FASTA format was downloaded from the NCBI and modified for use with searching for P450 proteins.

Using ThermoElectron Xcalibur software, a subset database was constructed from the NCBI database, which contained 196,330 protein entries, all with the words "human" or "homo sapiens" in their descriptor. A subset database was then created from this human database, which contained all entries with the terms "P450", "CYP" or "P-450" in their descriptor. This subset database contained 251 proteins. Not all were P450 entries; other proteins such as cytochrome P450 reductase and calcyphosine were included. However, most P450 enzymes did have more than one entry in the database; for example, there were three entries for CYP3A4, four entries for CYP4A11, five entries for CYP2A6 and four entries for CYP4B1. These replicate entries could not be distinguished from one another by their descriptors. A sequence comparison was performed to check whether these replicate entries were redundant or fragments of longer entries: in most cases it was found that the replicate entries were different from each other, and usually by more than one amino acid. Only eight entries were found to be fragments of longer entries; no redundant entries were found.

The subset human P450 database was then edited to exclude the following:

- Generic P450 entries, i.e. those proteins which were described simply as "Cytochrome P450", with no indication of family or subfamily.
- Protein entries that were described as "similar to" a P450 enzyme already described in the database.
- Entries that were fragments of P450 sequences already described in the database.

The human subset database (containing 196, 330 entries) was then modified to exclude all entries containing the terms "P450", "CYP" or "P-450", and the edited human P450 database was appended to it. The final human database contained 196,278 protein entries.

After the NCBI database had been modified as described above, it was interrogated in order to ascertain which and how many P450 forms it contained. The following P450 protein sequences were included (note that this list does not include archaic P450 names such as P-450LBTV or CYPXIA, or generic subfamily names such as CYP26): 1A1, 1A2, 1B1, 2A3, 2A4, 2A6, 2A7, 2A13, 2B6, 2C8, 2C9, 2C10, "similar to" 2C13, 2C17, 2C18, 2C19, 2D6, 2E1, 2F1, 2J2, 2R1, 2S1, 3A3, 3A4, 3A5, 3A7, 3A43, 4A11, 4B1, 4F2, 4F3, 4F8, 4F11, 4F12, 4Z1, 5A, 7A1, 7B1, 8B1, 11A1, 11B1, 11B2, 17A1, 19A1, 20A1, 21A2, 24A1, 26A1, 26A2, "similar to" 26B1, 27A1, 27B1, 39A1, 46A1, 51A1, i.e. 55 P450 sequences in total. Humans have 57 full length functional P450 genes [137]; however, CYP2A3, 2A4, 2C10, 2C13, 2C17 and 3A3 are redundant forms that are not included in the list of 57 functional P450 proteins [137].

The database downloaded from NCBI is missing sequence entries for the following P450s: CYP2U1, 2W1, 4A22, 4F22, 4V2, 4X1, 8A1, 26C1, 27C1 and perhaps CYP26B1, for which there is a sequence described as "similar to CYP26B1".

3.6 The pitfalls of interpreting Sequest results when searching for P450 identifications using the modified NCBI database

Sequest is a relatively straightforward program to use. However, the results of a Sequest search can be complex and always require manual confirmation. Some examples of this are discussed below.

3.6.1 How to recognise whether or not a P450 sequence is present in the results of a Sequest search

The *Sequest Summary* page (Figure 2.5) can be configured to display the results of the .dta searches in several different ways. For the work carried out in this project, the most useful configuration sorts the results by protein, so that each .dta file is listed under the header of the protein entry from the database from which its identified peptide sequence originates. For some entries, the protein header is listed along with the first few words of the protein descriptor. Within the protein groups of peptide sequences, the results are listed in order of Xcorr scores, with the peptide with the highest Xcorr score listed first. This listing by protein, however, is complicated by the fact that most of the proteins in the FASTA database downloaded from NCBI (Section 2.14.1) do not have descriptive headers. For example, database entries for CYP2C9, 3A4 and 4A11 may appear on *Sequest Summary* pages as gi|13699818|ref|NP_000762.2|, gi|13435386|ref|NP_059488.2|, and gi|2117372|pir||I65981, respectively, with no further description. A *Sequest Summary* page will usually contain many of these non-descriptive headers. It is not possible to tell which proteins these headers correspond to without referring to a linked page entitled "Flicka", which shows the entire database entry for the protein, including the header, full description line, protein sequence and peptide sequences that have been identified. An example of a *Flicka* page is shown in Figure 3.12. Usually it is possible to identify the protein from the description line on the *Flicka* page. However, where the protein is not the only one in the database common to the group of peptides identified (indicated by a number to the right of the protein header on the *Sequest Summary* page, see Figure 2.5), only one of these proteins will be shown on

the *Flicka* page. Some entries in the database are listed as "unknown" or "hypothetical" proteins. If this appears on the *Flicka* page instead of a more descriptive database entry then there is no way the protein can be identified except by following the link on the number to the right of the protein header on the *Sequest Summary* page; this brings up a further page with a list of (often non-descriptive) headers on it. Following one of these links will change the entry on the *Sequest Summary* page. It is then possible to view the full database entry for the new header protein by following the modified link on the *Sequest Summary* page. This is a very protracted process. In addition, some P450 proteins have archaic names that do not comply with the usual P450 nomenclature comprising family number, sub-family letter and individual protein number. For example, CYP2C19 is also known as S-mephenytoin 4-hydroxylase, CYP4A11 as fatty acid omega-hydroxylase and CYP4F2 as leukotriene-β4 omega-hydroxylase. "P450-LTBV" is alternately CYP4F2 or CYP4F3. If one is not familiar with this alternative nomenclature then it is possible to overlook protein identifications.

Figure 3.12 A *Flicka* page showing the database entry for the protein CYP4A11, the peptides identified in the Sequest search and the sequence coverage.

3.6.2 How to determine if a P450 protein has been identified during a Sequest search

If a peptide sequence is present in more than one protein in the database, this is indicated on the *Sequest Summary* page by a number to the left of the peptide's sequence (Figure 2.5). Following this link brings up a list of all the proteins in the database that contain the peptide sequence. P450 enzymes from the same family have greater than 40% sequence homology, and P450 enzymes from the same subfamily have greater than 55% sequence homology. It is therefore not surprising that many of the peptides identified are common to more than one P450 enzyme. In order to ascertain whether or not a protein is conclusively identified, it is necessary to follow these links and check that at least two of the peptide sequences identified are unique to the protein, *or* that the protein is the only one common to all the peptides. CYP3A4, for example, has 94% sequence homology with CYP3A3 and 89% sequence homology with CYP3A7. In some cases there may be no CYP3A4-specific peptides, but CYP3A4 is the only protein common to all the peptides identified, therefore CYP3A4 is considered to be identified. If CYP3A3 or CYP3A7 were present, but the only peptides identified were also common to CYP3A4 then it would not be possible to identify them. Problems with high sequence homology between different P450 enzymes are compounded by the fact that most of the P450 enzymes have more than one entry in the database and that these replicate entries are not distinguished from each other in any meaningful way (Section 3.5), for example as polymorphisms of the same enzyme. In many cases, peptides that are listed on the *Sequest Summary* page as being present in more than one protein in the database will actually be from one P450 enzyme that has several entries in the database.

Ideally, a new database should be designed with descriptive headers and entries that can be easily distinguished, including polymorphisms. This would provide a more simple and efficient basis to Sequest searching.

3.6.3 How to identify incorrectly listed peptides on a *Sequest Summary* page

A further problem with deciphering the output of Sequest searches was mentioned in Section 3.4.4. The *Sequest Summary* page can be configured so that the peptide sequences identified during a Sequest search are listed underneath the header for the protein from which they originate. However, sometimes a peptide sequence is listed on the *Sequest Summary* page under the heading of a protein from which it does not originate. When this occurs, the 2nd or 3rd choice peptide sequence will be found to be from the header protein. In theory, these peptide sequences should be easily distinguished from the peptide sequences that do originate from the header protein because they will have a different protein "reference number" (Figure 2.5). However, where the header protein itself has more than one entry in the database and therefore more than one reference number, it is not possible to distinguish which peptide sequences are from the same protein (but with a different reference number), which peptide sequences are from a related protein, and which are from a different protein altogether. For P450 proteins any of these three scenarios is possible. This is illustrated in Figure 3.13.

Following the link on the protein header on the *Sequest Summary* page brings up the *Flicka* page (Figure 3.12), which shows the protein entry from the database and lists the peptide sequences identified, also giving a percentage sequence coverage for the protein. On the *Flicka* page, the peptide sequences listed on the *Sequest Summary* page that do not originate from the header protein are absent; instead the 2nd or 3rd choice peptide sequence identifications that do originate from the header protein are listed. Therefore an apparent discrepancy exists between the peptide sequences listed on the *Sequest Summary* page and those listed on the *Flicka* page. Often the ΔC_n values separating the 1st and 2nd (or 3rd) choice peptide sequences are < 0.1, indicating that their Xcorr values are very similar and hence it is difficult to establish a clear 1st choice peptide sequence for the MS/MS spectrum, but this is not always the case: ΔC_n values of > 0.2 have been observed. Ideally there should be some indication on the *Sequest Summary* and *Flicka* pages, other than the protein reference number, that these peptide sequence identifications are not 1st choice for the header protein.

In conclusion, it should be stated that considerable care is needed in the interpretation of Sequest results: manual verification is always required.

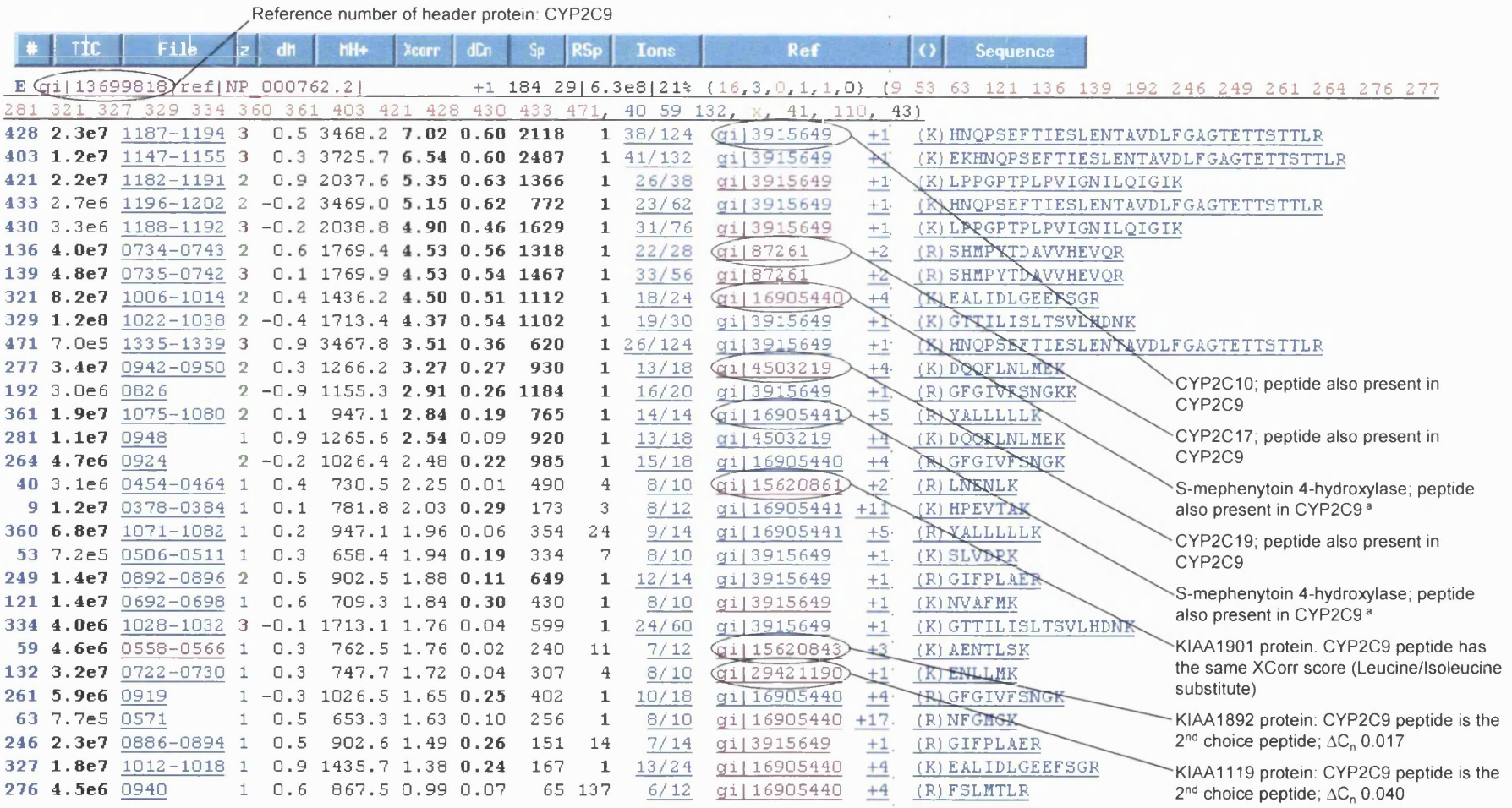


Figure 3.13 An entry for CYP2C9 on a *Sequest Summary* page, showing the different protein reference numbers listed for peptides under the CYP2C9 heading. None of the peptides listed have the same protein reference number as the header protein. [a] S-mephenytoin 4-hydroxylase is an alternative name for CYP2C19; both the entries shown here were found to be fragments of the CYP2C19 entries and subsequently removed from the database (Section 3.5).

3.7 The effect of altering the peptide mass tolerance and isotopic mass on the outcome of searching with Sequest

Altering the parameters used for a Sequest search can have a considerable effect on the outcome of the search: peptide sequence identifications made using a particular set of parameters may be missed if one or more of these parameters is altered. The Sequest parameters used for searches of MS/MS data in this project are listed in Section 2.14.2. The effects of altering two of these parameters, the peptide mass tolerance and the isotopic mass, on the identification of the CYP2E1 active site peptide ^{283}LYTMDGITVTVADLFFAGTETTSTTLR^{309} was investigated. This peptide was of particular interest in the study of the modification of CYP2E1 by benzyl isothiocyanate (Section 6.2).

Trypsin-digested incubations of CYP2E1 with BITC and P450 reductase (Section 2.5) were analysed four times by nano-LC-ES (Section 2.12.2) (the NADPH-supplemented and NADPH-absent digests were each analysed twice). Each data file was searched with Sequest four times, using a peptide mass tolerance of either 1 or 0.5 Da, and either average or mono-isotopic mass. Other search parameters were as described in Section 2.14.2.

3.7.1 The identification of a CYP2E1 active site peptide from four nano-LC-ES sample analyses using a peptide mass tolerance of 1 Da and average isotopic mass

A peptide mass tolerance of 1 and average isotopic mass were used for Sequest searches throughout this project. However, after searching the four data files using these parameters the target peptide, LYTMDGITVTVADLFFAGTETTSTTLR, was identified from only two of the four sample analyses: the $[M+2H]^{2+}$ ion (Xcorr score of 3.35) and the $[M+3H]^{3+}$ ion (Xcorr score of 5.37) were each identified once. Both peptide ions had retention times of approximately 40 minutes. The Xcorr scores for the identified peptides are high, and therefore it seems surprising that, out of 8 chances to identify the peptide (i.e. two possible ions in each of four sample

analyses), the peptide was identified only twice. The raw data files for each of the four analyses were examined between the times of 38 and 43 minutes to ascertain whether ions with *m/z* values of the $[M + 2H]^{2+}$ and $[M + 3H]^{3+}$ ions had been selected for MS/MS analysis (Table 3.3). For three of the four sample analyses, ions of *m/z* 1463, corresponding to the *m/z* value of the $[M + 2H]^{2+}$ ion, were fragmented in the time period 38-43 mins, yet only one of the resulting MS/MS spectra was identified to originate from the $[M + 2H]^{2+}$ ion. The MS/MS spectra demonstrate a high degree of similarity (Figure 3.14), indicating that they all derive from the same peptide; the series of y ions that dominates the spectrum strongly suggests that this peptide is the $[M+2H]^{2+}$ ion of LYTMDGITVTVADLFFAGTETTSTTLR. Yet two of the spectra have been assigned to different peptide sequences with poor Xcorr scores (Table 3.3).

MS/MS spectra from triply charged peptide ions are complex and assessment by eye is difficult. All four of the sample analyses resulted in MS/MS analysis of ions of the correct *m/z* value (within approximately ± 0.5 *m/z*) for the $[M + 3H]^{3+}$ ion of LYTMDGITVTVADLFFAGTETTSTTLR, yet the peptide sequence was identified only once (Table 3.3). Manual examination of the MS/MS spectra led to the conclusion that misidentifications were made for the MS/MS data from two of the sample analyses, whilst the MS/MS data and ion retention times from the fourth sample injection did not indicate that the LYTMDGITVTVADLFFAGTETTSTTLR $[M + 3H]^{3+}$ ion had been fragmented.

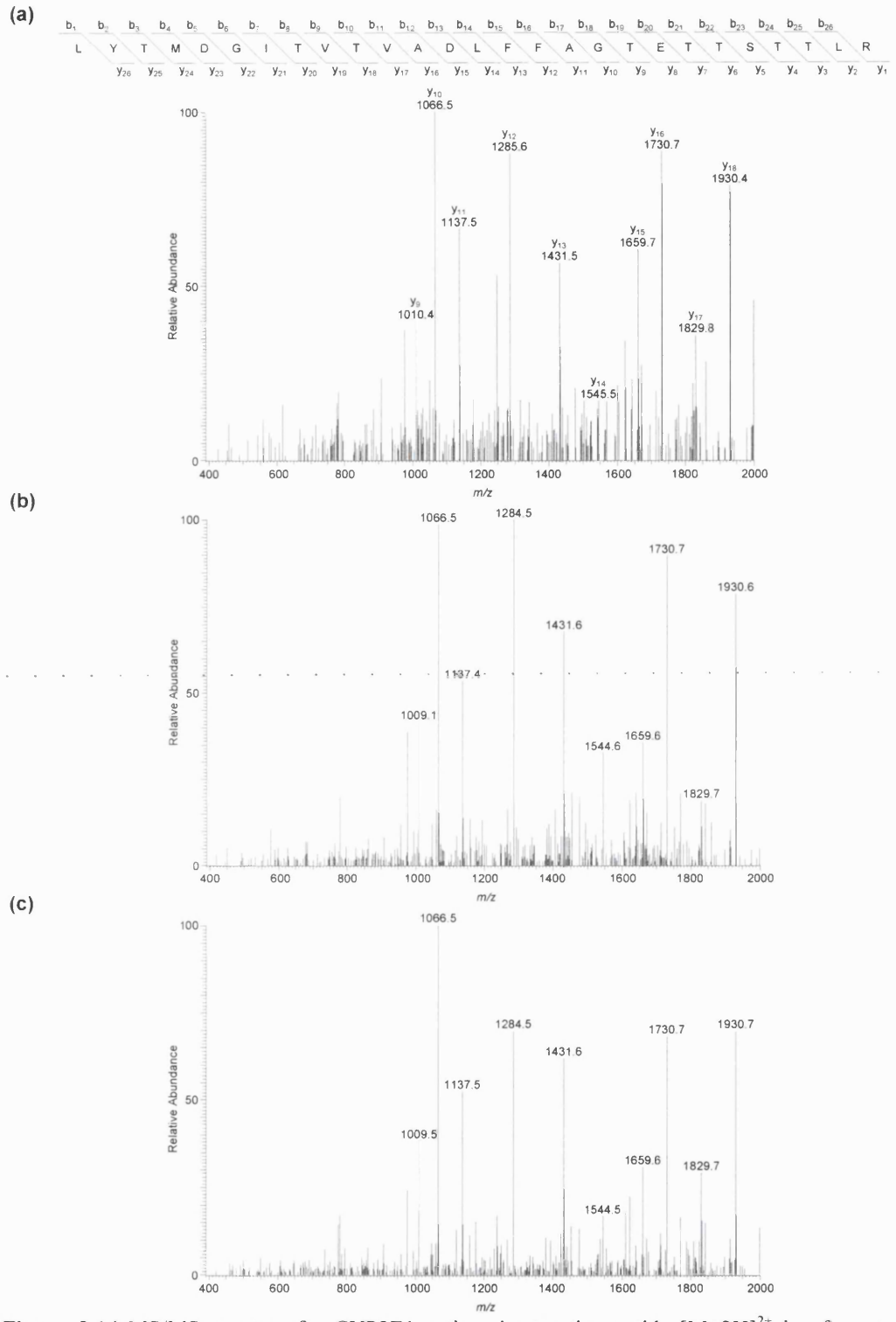


Figure 3.14 MS/MS spectra of a CYP2E1 active site tryptic peptide $[M+2H]^{2+}$ ion from trypsin-digested incubations of CYP2E1 with BITC and P450 reductase (Section 2.5). **(a)** Spectrum from the second analysis of the NADPH-supplemented digest (Table 3.3); parent ion *m/z* value 1463.5. The amino acid sequence of the peptide is shown above the spectrum; some of the y ions are indicated. The spectrum was correctly identified by Sequest. **(b)** Spectrum from the first analysis of the NADPH-absent digest (Table 3.3). The spectrum was incorrectly identified by Sequest. **(c)** Spectrum from the second analysis of the NADPH-absent digest (Table 3.3). The spectrum was incorrectly identified by Sequest. Digests were diluted in 0.1% TFA such that CYP2E1 was at a concentration of approximately 500 fmol/µl and analysed by nano-LC-ES (Section 2.12.2): 1 µl of sample was injected via a 10 µl sample loop. Data was searched using Sequest (Section 2.14.2) against the Finnigan Xcalibur human database (Section 2.14.1). Sample was not desalted before analysis.

Table 3.3 Ions with *m/z* values matching those for the $[M+2H]^{2+}$ and $[M+3H]^{3+}$ ions of the CYP2E1 peptide LYTMDGITVTVADLFFAGTETTSTTLR eluting in the time period 38-43 minutes from four sample analyses. **(a)** Ions with *m/z* values matching those for the $[M+2H]^{2+}$ ion. **(b)** Ions with *m/z* values matching those for the $[M+3H]^{3+}$ ion.

(a)

Sample [a]	Nano-LC-ES analysis no.	*m/z* values of fragmented ion(s) [b]	retention time(s) (min)	Peptide identified	Xcorr	Correctly identified? (Y/N) [c]
NADPH +ve	01	-	-	-	-	-
NADPH +ve	02	1463.5	39.76	LYTMDGITVTVADLFFAGTETTSTTLR	3.35	Y
NADPH -ve	01	1462.8, 1463.5, 1462.9	40.11, 40.22 40.31	FGFSLLAAGRSVWTLEMDAGAGVLTGRLIR	1.34	N
NADPH -ve	02	1462.8, 1462.9	39.97, 40.03	ILQGGAKGPGLFFILPCTDSFIKVDMR	0.80	N

(b)

Sample [a]	Nano-LC-ES analysis no.	*m/z* values of fragmented ion(s) [b]	retention time(s) (min)	Peptide identified [d]	Xcorr	Correctly identified? (Y/N) [c]
NADPH +ve	01	977.1	41.44	None	-	-
NADPH +ve	01	975.8	43.18	None	-	-
NADPH +ve	02	976.6, 975.7, 975.6, 975.8	39.50, 39.72, 39.84, 39.96	NDTVKNATNTNNSSWER/	1.68/	N
				ETTIRDTGEYTLELKNVTGTTLETIK	2.02	N
NADPH -ve	01	975.5, 976.3, 975.7	39.95, 40.08, 40.19	PCVKLTPLCVTLSCSDIR/	1.40/	N
				IKQIINLWQEVGEAMYAPPIRGQIR	1.51	N
NADPH -ve	02	975.9, 976.3, 976.8	39.90, 40.06, 40.15	LYTMDGITVTVADLFFAGTETTSTTLR	5.37	Y

[a] NADPH +ve = NADPH-supplemented digest; NADPH -ve = NADPH-absent digest. [b] Ions in the same row were combined by Sequest to make one .dta file. [c] Identifications were made after Sequest searching using the following parameters: peptide mass tolerance 1 Da; average isotopic mass; fragment ion tolerance 0 Da; maximum number of internal cleavage sites 2. [d] Where two peptides are listed, these are the identifications made from the two .dta files constructed by Sequest, one assuming a charge state of +2 and the other assuming a charge state of +3.

Table 3.4 The effect of altering peptide mass tolerance and isotopic mass on the identification of a CYP2E1active site peptide after Sequest searching of four data files. $[M + H]^+$ ion theoretical average isotopic mass = 2926.31; theoretical mono-isotopic mass = 2924.46.

Sample [a]	Nano-LC-ES analysis no.	.dta file	Charge from .dta file	$[M + H]^+$ mass from .dta file (Da)	Parent ions used for .dta file		Sequest search parameters [b]: identification? (Y/N)			
					m/z value	Equivalent $[M + H]^+$ mass (Da)	Mass = average Tolerance = 1 Da	Mass = average Tolerance = 0.5 Da	Mass = mono Tolerance = 1 Da	Mass = mono Tolerance = 0.5 Da
NADPH +ve	02	1345.1345.2	2	2925.98	1463.49	2925.98	Y	Y	N	N
NADPH +ve	02	1337.1352.3	3	2927.77	976.59	2927.77	N	N	N	N
					975.57	2924.71				
					975.74	2925.22				
					975.83	2925.49				
NADPH –ve	01	1370.1377.2	2	2924.63	1462.81	2924.63	N	N	Y	Y
					1463.53	2926.06				
					1462.87	2924.74				
NADPH –ve	01	1365.1373.3	3	2924.55	975.52	2924.55	N	N	Y	Y
					976.31	2926.93				
					975.67	2925.01				
NADPH –ve	02	1378.1380.2	2	2924.67	1462.84	2924.67	N	N	Y	Y
					1462.93	2924.86				
NADPH –ve	02	1376.1384.3	3	2925.78	975.93	2925.78	Y	N	N	N
					976.30	2926.90				
					976.81	2928.43				

[a] NADPH +ve = NADPH-supplemented digest; NADPH -ve = NADPH-absent digest. [b] mass = isotopic mass; tolerance = peptide mass tolerance. Other parameters: fragment ion tolerance 0 Da; maximum number of internal cleavage sites 2.

3.7.2 The identification of a CYP2E1 active site peptide from four nano-LC-ES sample analyses after varying the peptide mass tolerance and isotopic mass

The four data files were searched again using a peptide mass tolerance of 0.5 Da and mono-isotopic masses. The results are shown in Table 3.4. The identification of a particular ion after searching with Sequest depends on several factors. When Sequest creates a .dta file it calculates a value for the mass of the intact peptide $[M + H]^+$ ion. This mass is calculated from the *m/z* value of the parent ion from the *first* scan used for the .dta file (Table 3.4). More than one scan can be grouped together to create a .dta file (Section 2.14.2) but the *m/z* value of the parent ion from the first scan is the only one taken into account. After creating a .dta file, Sequest employs a two-stage searching procedure (Section 2.14.2): candidate peptide sequences are initially found from the database on the basis of how similar their intact peptide $[M + H]^+$ ion masses are to the $[M + H]^+$ mass calculated for the .dta file, and given an S_p score (based on several factors; see Section 2.14.2). If the $[M + H]^+$ ion mass for the candidate peptide sequence does not fall within a specified user mass tolerance of the $[M + H]^+$ ion mass calculated for the .dta file it will not be given an S_p score; it therefore cannot progress onto cross correlation analysis (the next stage in the Sequest scoring procedure) and cannot be identified.

The identification of the CYP2E1 active site peptide LYTMDGITVTVADLFFAGTETTSTTLR $[M + 2H]^{2+}$ and $[M + 3H]^{3+}$ ions from data from the three nano-LC-ES sample analyses for which the appropriate MS/MS fragmentation data was collected can be explained by considering the $[M + H]^+$ masses calculated for the appropriate .dta file. Only when the calculated $[M + H]^+$ mass is within the user-specified peptide mass tolerance of the specified isotopic mass will the ion be identified (Table 3.4). For example, the .dta file 1370.1377.2 from sample injection NADPH-absent 01 had an assigned $[M + H]^+$ ion mass of 2924.63 (Table 3.4). This mass is within 0.5 Da of the theoretical mono-isotopic mass of 2924.46 Da but more than 1 Da different from the theoretical average isotopic mass of 2926.31 Da. Therefore the peptide was identified for both of the Sequest searches where mono-isotopic mass was employed. The .dta file 1376.1384.3 from sample injection NADPH-absent 02 had an assigned $[M + H]^+$ ion

mass of 2925.78 Da (Table 3.4). This value is greater than 0.5 Da from the theoretical average isotopic mass therefore was not identified when searched with Sequest using average isotopic mass and a peptide mass tolerance of 0.5 Da. But it is within 1 Da of the average isotopic mass, therefore it was identified when the peptide mass tolerance was increased to 1 Da. The .dta file 1337.1352.3 from sample injection NADPH-supplemented 02 had an assigned $[M + H]^+$ ion mass of 2927.77 Da (Table 3.4). This is more than 1 Da from both of the theoretical isotopic masses therefore it was not identified. However, if the $[M + H]^+$ mass of the parent ion from the 2nd, 3rd or 4th MS/MS scan used to create the .dta file had been used for the overall .dta $[M + H]^+$ ion mass, the correct peptide would have been identified in at least one of the four Sequest searches performed.

It may seem that the way to get the most identifications is to use wide error tolerances at all times. However, only the candidate peptide sequences with the top 500 S_p scores are subjected to cross-correlation analysis. If the correct amino acid sequence is not in the top 500 sequences ranked by S_p score, it will not be subjected to cross-correlation analysis and therefore cannot be identified. Increasing the error tolerances increases the chance that the correct peptide sequence will not be among the top 500 S_p scored sequences, thereby increasing the chance of false positives. Therefore a balance must be found between allowing for the relatively poor mass accuracy of the LCQ mass spectrometer and minimising the number of spurious sequences assigned S_p scores so that the correct amino acid amino acid sequences have the greatest chance of progressing to cross correlation analysis.

An alternative to increasing the error tolerances for the Sequest search would be to reduce the precursor mass tolerance used in the creation of .dta files (Section 2.14.2). The precursor mass tolerance is the difference in the parent ion *m/z* value allowed for MS/MS scans to be grouped into a single .dta file. A value of 1.4 *m/z* was used throughout the work described here. Average and mono-isotopic masses of tryptic peptides can differ by 1.4 Da, therefore this value is reasonable for the analysis of singly charged ions. However, this allows for MS/MS spectra of doubly charged parent ions with differences in mass values of up to 2.8 Da to be grouped together, or MS/MS spectra of triply charged parent ions with differences in mass values of up to 4.2 Da to be grouped together. This explains how the MS/MS scan for a parent ion

with calculated $[M + H]^+$ mass 2927.77 Da came to be grouped in the same .dta file as the MS/MS scan for a parent ion with calculated $[M + H]^+$ mass 2924.71 Da (Table 3.4). The difference between the average and mono-isotopic masses for the majority of the peptides observed after tryptic digestion will not be this large.

The window of error in measured mass values on an LCQ mass spectrometer necessitates the use of statistical analysis in the interpretation of data. This is in contrast with a high mass accuracy instrument, for which theoretical masses can be expected to match experimental values to several decimal places.

When searching with Sequest, one must always be aware that there are several possible reasons for a particular peptide not being identified: the peptide may not be present in the sample, or the peptide may be present in the sample but was not selected for MS/MS analysis. Alternatively, the peptide may be present in the sample, selected for MS/MS analysis, but either not identified or incorrectly identified by Sequest.

3.7.3 The effect of varying the peptide mass tolerance and isotopic mass on the CYP2E1 protein coverage obtained after searching with Sequest

A comparison of the number of peptides identified and equivalent percentage sequence coverages obtained for CYP2E1 in the four sample analyses after varying the isotopic mass and peptide mass tolerance is shown in Table 3.5. The data suggests that maximum sequence coverage results from the use of average isotopic mass and a peptide mass tolerance of 1 Da over using mono-isotopic mass and a peptide mass tolerance of 0.5 Da.

Table 3.5 The effect of varying peptide mass tolerance and isotopic mass on the CYP2E1 protein coverage obtained after searching with Sequest.

Sample [a]	Nano-LC-ES analysis no.	Peptide mass tolerance (Da)	Mono/ average isotopic mass	No. of CYP2E1 peptides with Xcorr≥2.5	Total no. of CYP2E1 peptides	% CYP2E1 sequence coverage (peptides with Xcorr≥2.5)	% CYP2E1 sequence coverage (all peptides)
NADPH +ve	01	1	average	15	23	38.3	49.5
NADPH +ve	02	1	average	14	25	38.1	53.3
NADPH -ve	01	1	average	17	24	33.7	42.6
NADPH -ve	02	1	average	18	29	45.0	57.0
NADPH +ve	01	0.5	average	11	21	31.4	45.2
NADPH +ve	02	0.5	average	10	20	26.6	38.3
NADPH -ve	01	0.5	average	15	24	33.7	43.8
NADPH -ve	02	0.5	average	13	23	30.6	42.6
NADPH +ve	01	1	mono	13	21	25.2	36.1
NADPH +ve	02	1	mono	11	18	29.0	38.3
NADPH -ve	01	1	mono	17	24	40.2	47.7
NADPH -ve	02	1	mono	14	24	34.9	46.9
NADPH +ve	01	0.5	mono	5	12	14.2	23.7
NADPH +ve	02	0.5	mono	3	9	9.9	17.0
NADPH -ve	01	0.5	mono	4	12	12.6	24.7
NADPH -ve	02	0.5	mono	6	15	17.8	29.0

[a] NADPH +ve = NADPH-supplemented digest; NADPH -ve = NADPH-absent digest.

3.8 In-gel tryptic digestion

After it was established that in-gel tryptic digestion of cytochrome c and recombinant CYP1A2, 2E1 and 3A4 produced satisfactory identification of these proteins, experiments were designed to assess various aspects of the in-gel digestion procedure. The percentage recoveries were calculated for several peptides after in-gel digestion; the limit of detection was determined for in-gel digestion of recombinant P450s; the effect of the use of ZipTips on small pmol quantities of proteins digested in-gel was investigated; and the addition of a S-carbamidomethylation step to prevent the formation of disulphide bonds was considered.

3.8.1 Evaluation of the efficiency of in-gel tryptic digestion in terms of peptide recovery

The efficiency of the in-gel digestion process was evaluated by calculating the percentage recoveries for several cytochrome c peptides after in-gel digestion. An experiment was designed to quantify the cytochrome c peptides recovered after in-gel tryptic digestion. Nano-LC with UV rather than MS detection was used for quantification so that chromatographic peaks would not be affected by variations in

the sample spray. In order to determine a suitable concentration range for the calibrants for peptide quantification, the LC column loading capacity and the linearity of the response from the UV detector were investigated.

3.8.1.1 Investigation of the linearity of response from the UV detector and the LC column loading capacity

Cytochrome c in-solution tryptic digest was made up freshly to avoid any errors in peptide quantification due to degradation during storage. Samples containing between 0.4 and 4 pmol of cytochrome c in-solution digest were analysed by nano-LC with UV detection. Peak heights and areas for two cytochrome c peptides (c and d, see Figure 3.15) were subjected to regression analysis to assess the linearity of the response from the detector. Plots of peak height and area against amount of cytochrome c digest for the two peptides are shown in Figure 3.16.

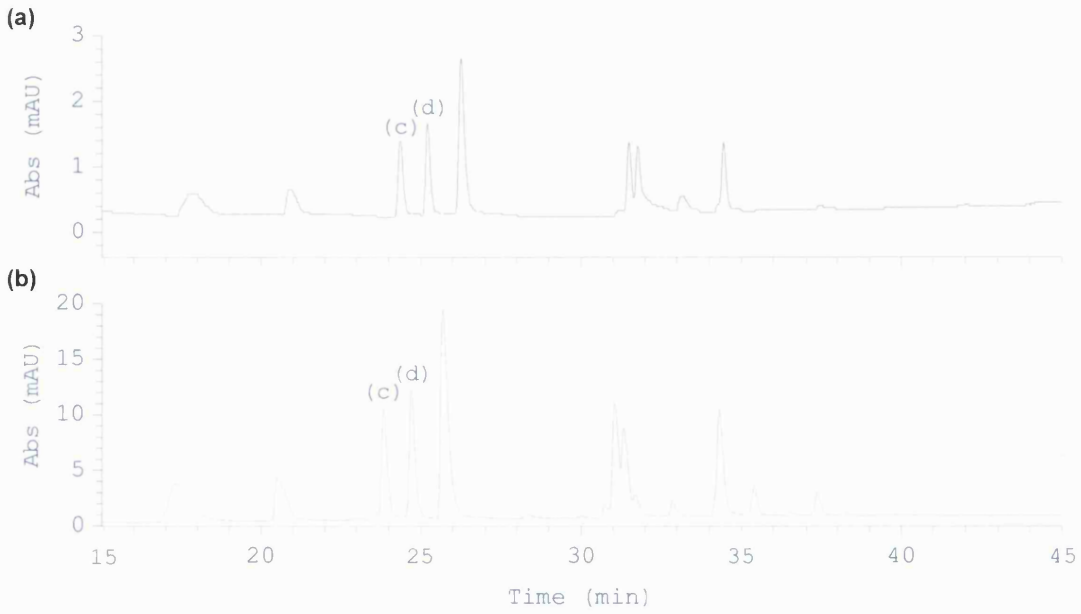


Figure 3.15 (a) Nano-LC chromatogram of 0.4 pmol cytochrome c in-solution digest. **(b)** Nano-LC chromatogram of 4 pmol cytochrome c in-solution digest. Cytochrome c in-solution tryptic digest was diluted in 0.1% TFA to concentrations of 0.4, 0.8, 1.6 and 4 pmol/µl, and analysed using full-loop injection (with a 1 µL sample loop) by nano-LC with UV detection. Each sample was analysed three times. The heights and areas of peaks c and d (indicated) were used to assess the linearity of the response from the detector.

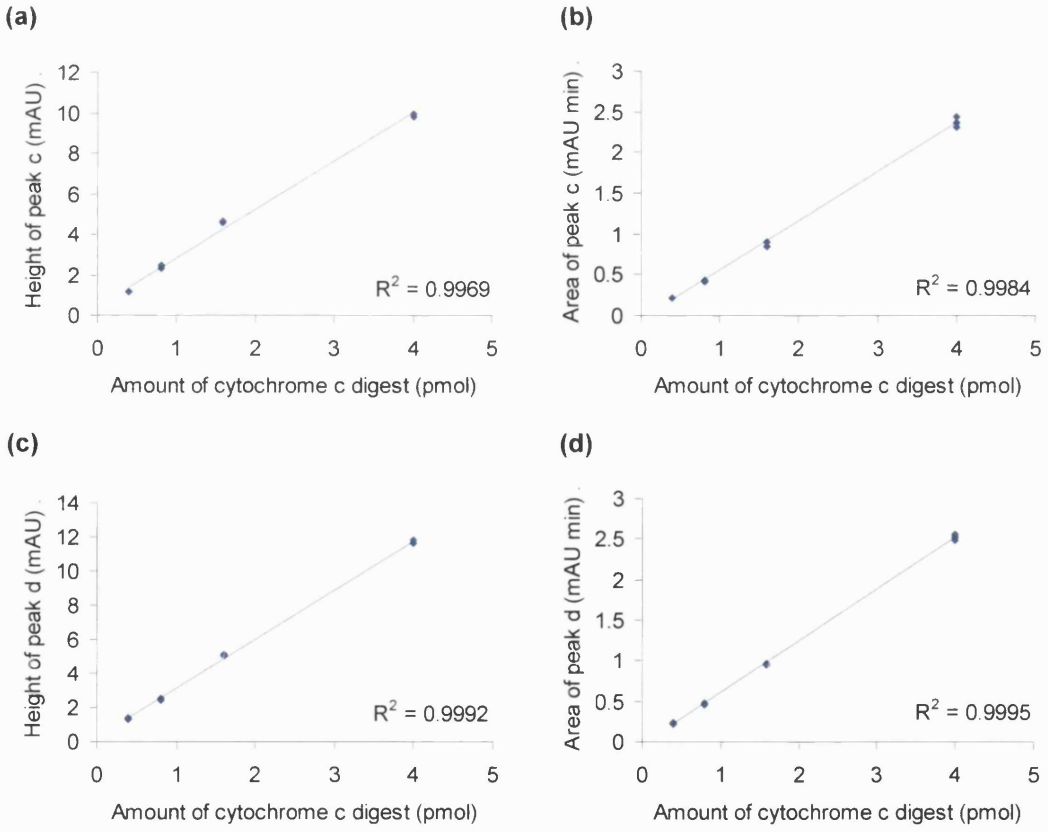


Figure 3.16 Plots of peak height and area against amount of cytochrome c in-solution digest for two cytochrome c tryptic peptides, c and d (Figure 3.15). **(a)** Height of peak c; **(b)** area of peak c; **(c)** height of peak d; **(d)** area of peak d. R^2 values are inset.

The R^2 values for the plots shown in Figure 3.16 indicate a high degree of linearity in the response of the detector between 0.4 and 4 pmol of cytochrome c in-solution digest loaded onto the column.

Even loading 4 pmol of cytochrome c in-solution digest onto the column did not result in the peaks being overloaded (Figure 3.15(b)); only a very small amount of carry-over was observed for the more hydrophobic peptides in a blank sample of 0.1% TFA analysed directly afterwards.

3.8.1.2 Evaluation of the percentage recoveries of cytochrome c peptides after in-gel tryptic digestion

A gel loaded with eight lanes of 240 pmol cytochrome c per lane was run by SDS-PAGE. The protein bands were subjected to in-gel digestion and analysed by nano-LC with UV detection, along with calibrants prepared from cytochrome c in-solution tryptic digest. Five peptides were chosen for quantification (Figure 3.17). Calibration lines based on peak heights and areas were constructed, the amounts of the different peptides recovered from the gel were determined, and the percentage recoveries calculated. Two of the calibration lines are shown in Figure 3.18. All calibration lines had R^2 values of ≥ 0.99 with the exception of the plot for the height of peak a (Figure 3.17), which had a R^2 value of 0.93. The percentage recoveries for the five peptides ranged from 13% to 43% (Figure 3.19).

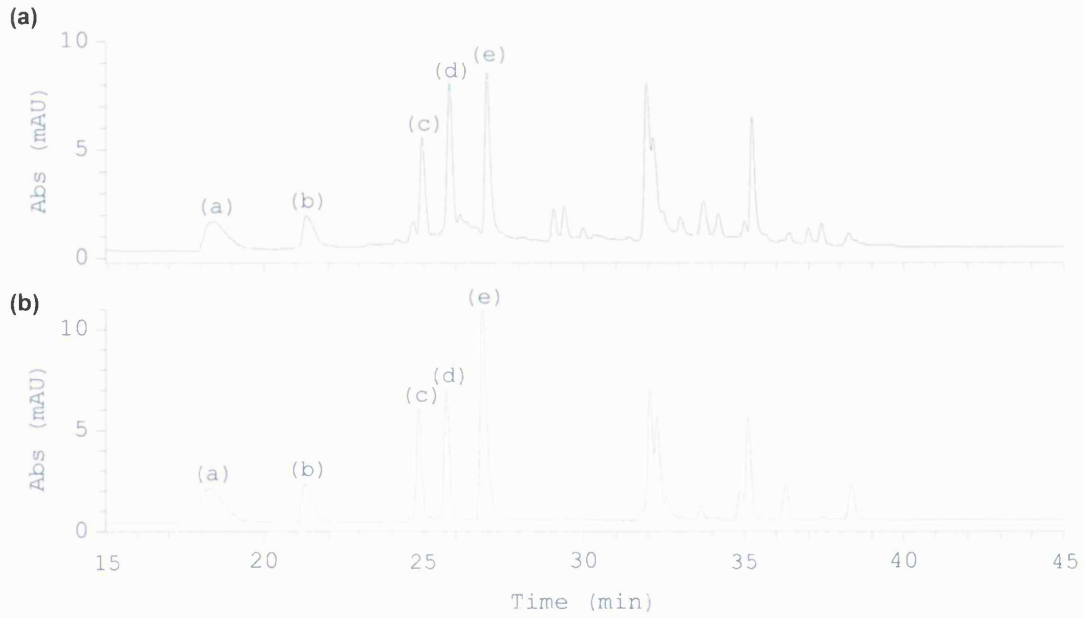


Figure 3.17 (a) Nano-LC chromatogram of cytochrome c in-gel digest; **(b)** Nano-LC chromatogram of 2.0 pmol of cytochrome c in-solution digest. A gel loaded with eight lanes of 240 pmol cytochrome c was run by SDS-PAGE. The gel was left in water overnight, and the protein bands subjected to in-gel digestion (Section 2.9), with desalting (Section 2.10.1). The resultant peptides were dried and stored at -20 °C. Samples were reconstituted in 40 µl of 5 mM ammonium bicarbonate in 0.1% TFA for analysis. Cytochrome c in-solution digest was used to calibrate the quantitation. Calibration standards were made up at concentrations of 0.4, 0.8, 1.2, 1.6, 2.0, 2.5, 3.0, 3.5 and 4.0 pmol/µl; each calibrant contained 0.1% TFA. Samples and calibrants were analysed using full-loop injection (with a 1 µL sample loop) by nano-LC with UV detection. Each sample and calibrant was analysed at least twice. The peptide peaks chosen for quantification are labelled a to e.

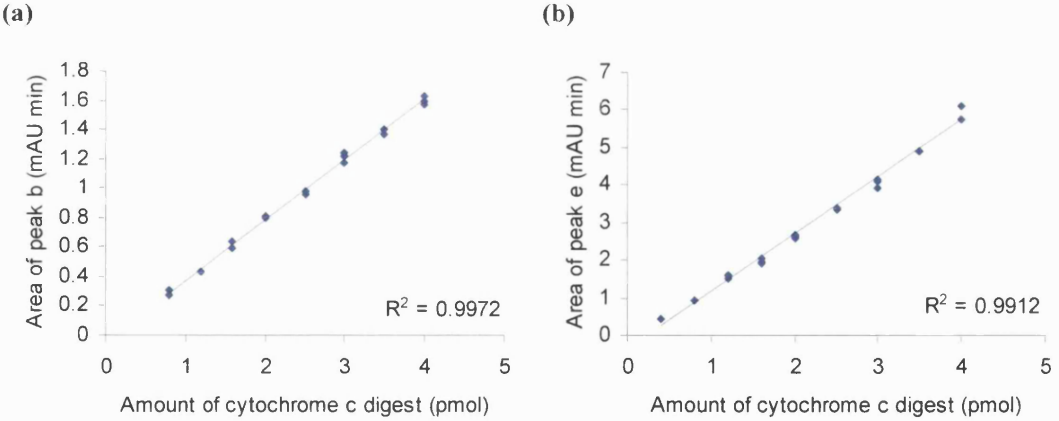


Figure 3.18 Two of the calibration lines used for the quantification of cytochrome c peptides recovered after in-gel digestion. **(a)** Area of peak b (Figure 3.17) against amount of cytochrome c in-solution digest; **(b)** area of peak e (Figure 3.17) against amount of cytochrome c in-solution digest. R^2 values are inset. Calibration lines were prepared from peak heights and peak areas.

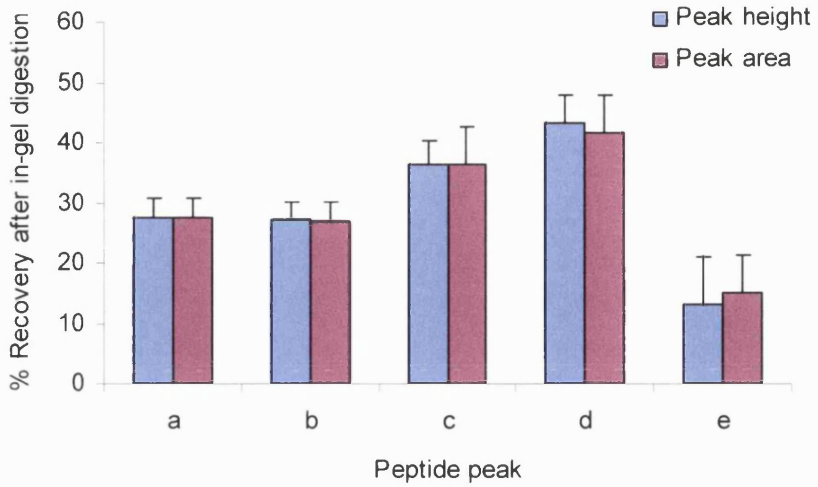


Figure 3.19 Mean percentage recoveries for 5 cytochrome c tryptic peptides, a to e (Figure 3.17), after in-gel digestion. Error bars show +1 standard deviation of the mean. Eight lanes of cytochrome c were run by SDS-PAGE and subjected to in-gel digestion. Each sample was analysed twice. Peptide recoveries from duplicate analyses of the same sample were averaged; the overall mean percentage recoveries for the eight cytochrome c lanes were calculated from the averaged percentage recoveries.

Contributions to the incomplete recovery of peptides after in-gel digestion could come from a combination of sources, including incomplete tryptic digestion and incomplete extraction of the peptides from the gel pieces after digestion.

3.8.2 Determination of the limit of detection for recombinant P450 proteins after in-gel digestion

The limit of detection for recombinant P450s after in-gel digestion was determined to investigate whether detection of these haemoproteins was possible at concentrations likely to be found in biological samples.

SDS-PAGE gels were loaded with low pmol amounts of recombinant CYP1A2 and CYP3A4. The gels were stained with silver (Section 2.7.2) and the protein bands subjected to in-gel digestion. Resultant peptides were analysed by nano-LC-ES. CYP3A4 was identified at 0.6 pmol on-gel (one peptide with Xcorr $\geq$ 2.5 only, Figure 3.20) and 1.5 pmol on-gel (up to 3 peptides with Xcorr scores $\geq$ 2.5). CYP1A2 was identified at 1 pmol on-gel (one peptide with Xcorr $\geq$ 2.5) and 2 pmol on-gel (up to 6 peptides with Xcorr scores $\geq$ 2.5).

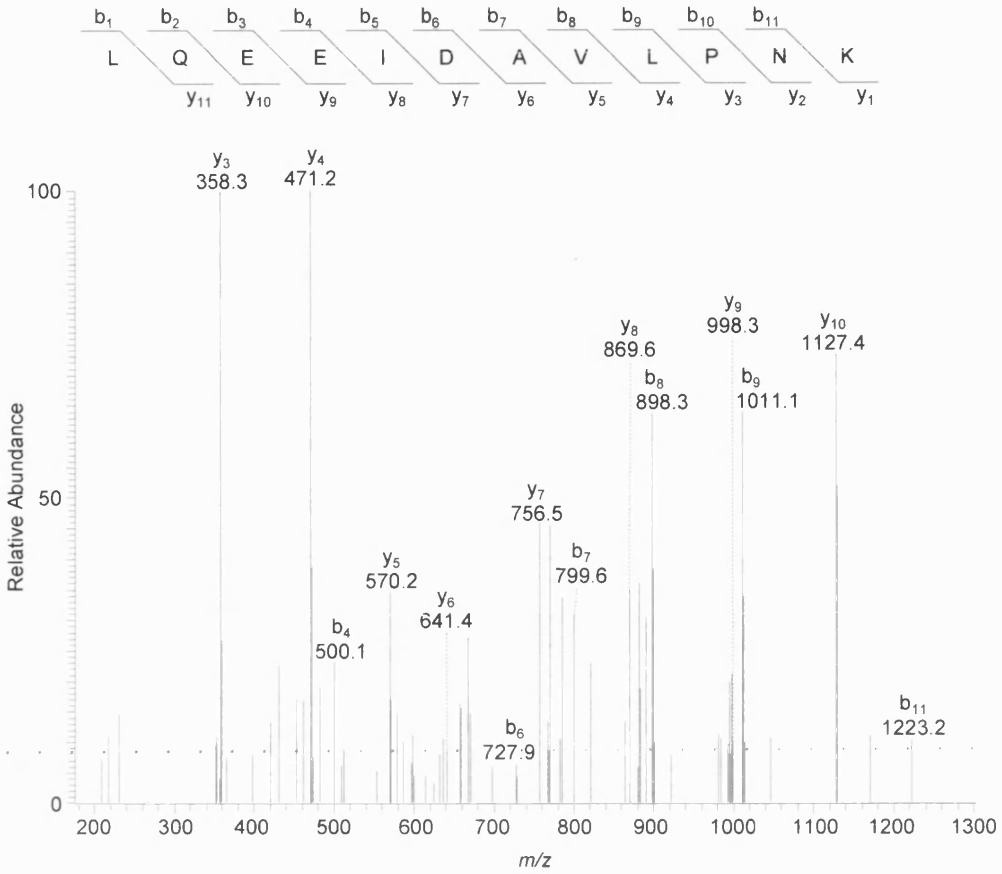


Figure 3.20 MS/MS spectrum of the CYP3A4 tryptic peptide LQEEIDAVLPNK $[M+2H]^{2+}$ ion of *m/z* 685.2, identified from 0.6 pmol CYP3A4 on-gel. The amino acid sequence of the peptide is shown above the spectrum. y and b ions are formed by peptide bond cleavage with charge retention on the C-terminus and N-terminus, respectively. Samples containing between 0.2 and 6 pmol CYP1A2 and 3A4 were run by SDS-PAGE, with silver staining (Section 2.7.2). Protein bands were subjected to in-gel digestion (Section 2.9) with desalting (Section 2.10.1) and the resultant peptides were dried and stored at -80 °C. Samples were reconstituted in 20 µl of 0.1% TFA and analysed by nano-LC-ES: 1 µl of sample was injected via a 10 µl sample loop. Data files were searched using Sequest software (Section 2.14.2) against the Finnigan Xcalibur human database (Section 2.14.1).

The limit of detection for recombinant P450s after in-gel digestion was determined to be approximately 1 pmol on-gel. There is considerable variability in the concentrations of mammalian P450s according to specific P450 enzyme and tissue location. Liver, the most P450-rich organ, has concentrations of between 1 and 100 pmol/mg microsomal protein for selected P450 enzymes [144,145,185]. Approximately 25 µg of microsomal protein was loaded onto SDS-PAGE gels for the tissue samples described in Chapter 5, which equates to between 0.025 and 2.5 pmol P450. CYP2B6 and CYP2D6, which have concentrations of between 1 and 10 pmol pmol/mg microsomal protein [144,185], would therefore have been present on SDS-PAGE gels at 0.025 to 0.25 pmol after loading 25 µg microsomal protein. These two proteins were

identified in the study described in Chapter 5, which suggests that the on-gel limit of detection for P450 proteins after in-gel tryptic digestion is lower than 1 pmol.

There are several aspects of this study that may have elevated the limit of detection relative to the tissue samples analysed in Chapter 5. The SDS-PAGE gels used to determine the limits of detection of CYP1A2 and CYP3A4 were stained with silver (Section 2.7.2). Previous attempts to stain the gels with Coomassie Blue R-250 were not successful because the Coomassie Blue stain was not sufficiently sensitive to visualise protein bands containing less than approximately 3 pmol P450; in this regard silver stain is more than 10 times more sensitive than Coomassie Blue stain [186]. However, concern has been raised regarding the use of silver staining when followed by protein analysis because silver staining includes protein treatment with the strong oxidising agent Ag^+, which is generally thought to cause oxidative attack on the protein, leading to chemical modification or destruction [187]. The silver staining protocol used for this study omitted gluteraldehyde (Section 2.7.2), a cross-linking and sensitising agent [187], yet the limit of detection may have been elevated due to the use of Ag^+. ZipTips were used in the limit of detection studies to desalt peptide mixtures, whilst in later work samples for analysis by nano-LC-ES were not desalted (Section 3.8.3). Some sample could therefore have been lost on the ZipTips (Section 3.8.3). Finally, the samples used for the determination of the limit of detection were reconstituted in a volume of 20 µl. The tissue samples analysed in Chapter 5 were reconstituted in a volume of 12 µl, therefore slightly lower concentrations of P450s would have been detectable.

The limit of detection for P450s from biological matrices will vary greatly depending on factors such as the nature and complexity of the matrix and the effect of competitive ionisation between sample components. The limit of detection for some P450s will be greater than for others, since some peptides are more easily ionised by the electrospray process than others. Ideally, the detection limit for each P450 in the tissue of interest should be investigated. However, the present study provides an approximate value for the P450 detection limit after in-gel digestion.

3.8.3 The effect of the use of ZipTips on the number of CYP1A2 and CYP3A4 peptides identified after in-gel digestion

The effect of the use of ZipTips on the number of peptides identified after in-gel digestion of low pmol levels of P450 protein was investigated. Previously, in-gel digest samples were desalted prior to analysis by nano-ES or nano-LC-ES (Section 2.10.1). This is necessary for the analysis of in-gel digest samples by nano-ES because they may contain SDS and other ionic components which are unsuitable for electrospray analysis. However, the guard column employed for nano-LC-ES analysis (Section 2.11) provides on-line sample clean-up as an alternative to the use of ZipTips, although ZipTips can still be employed to prolong the life of the guard column. The effect of the ZipTip desalting step on the analysis of P450s after in-gel digestion was investigated.

CYP1A2 (2.2 pmol) and CYP3A4 (3.0 pmol) were subjected to SDS-PAGE analysis; six replicate lanes of each P450 were analysed. The protein bands were subjected to in-gel tryptic digestion (Section 2.9); half of the samples for each P450 were desalted whilst the other half were not. Each sample was analysed at least twice by nano-LC-ES and the results were searched using Sequest. A table of results is shown in Appendix 1, Table A1.3.

Student's two-sampled t-tests were performed to ascertain whether there were any significant differences in the mean numbers of peptide identifications with Xcorr $\geq$ 2.5 made by Sequest, and in the mean total numbers of peptide identifications (all Xcorr scores), with and without desalting for both CYP1A2 and CYP3A4. The mean values are plotted in Figure 3.21.

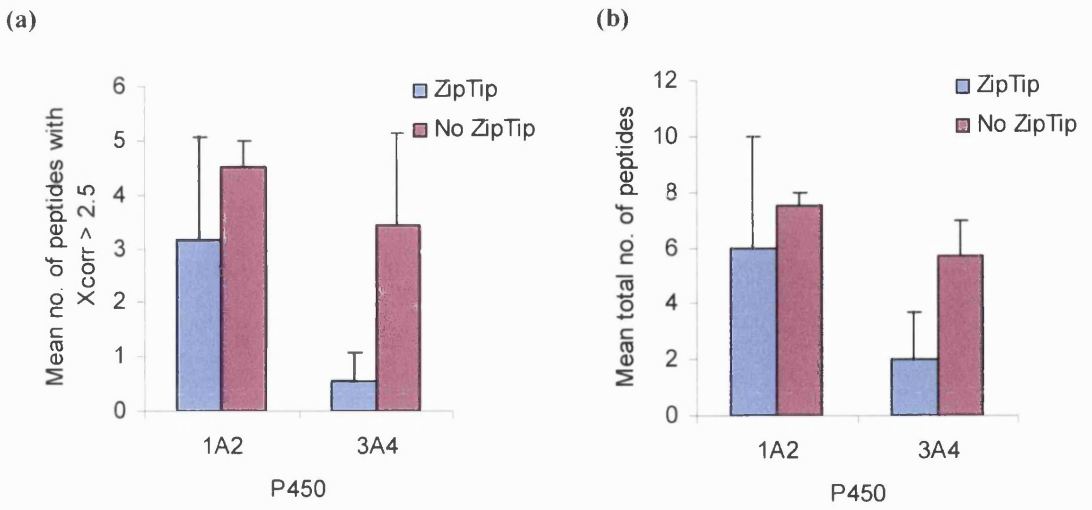


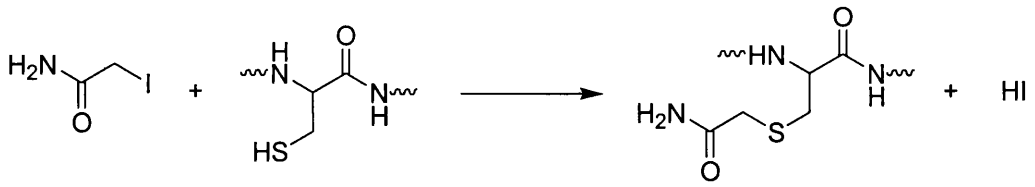
Figure 3.21 The effect of the use of ZipTips on the number of CYP1A2 and CYP3A4 peptides identified after in-gel digestion. **(a)** The mean numbers of peptides with Xcorr score ≥ 2.5. The difference between the means was not significant for CYP1A2, but was significant for CYP3A4 (Student's t-test, $P < 0.05$). **(b)** The mean total number of peptides identified. The difference between the means was not significant for CYP1A2, but was significant for CYP3A4 (Student's t-test, $P < 0.05$). Six lanes each of CYP1A2 (2.2 pmol) and CYP3A4 (3.0 pmol) were analysed by SDS-PAGE and the protein bands subjected to in-gel digestion. Half the samples were desalted whilst the other half were not. Samples were reconstituted in 20 µl of 0.1% TFA and analysed by nano-LC-ES (Section 2.12.2): 1 µl of sample was injected via a 10 µl sample loop. Data was searched against the Finnigan Xcalibur human database (Section 2.14.1) using Sequest (parameters as described in Section 2.14.2, except the minimum TIC for the preparation of .dta files was set to 1×10^5 rather then 5×10^5). Results for replicate nano-LC-ES sample analyses (Appendix 1, Table A1.3) were averaged, then the overall mean values were calculated from the averaged values. Error bars show +1 standard deviation of the mean.

Omission of the desalting step results in the identification of more CYP3A4 peptides ($P < 0.05$), but with CYP1A2 it makes no difference to the number of peptides identified (Figure 3.21). This provides evidence that, at least for some P450s, sample is lost in the desalting step. The desalting step after in-gel digestion was consequently omitted from future samples analysed by nano-LC-ES.

3.8.4 Modification of cysteine residues of proteins prior to in-gel digestion

Modification of the cysteine groups of proteins before in-gel digestion has been described in order to prevent disulphide bonds from reforming after the proteins have been denatured and reduced for gel electrophoresis [188]. The reformation of disulphide bonds can result in incomplete tryptic digestion and cross-linked peptides that will not be recognised by database search programs. The S-carbamidomethylation of

protein cysteine residues (Figure 3.22) before in-gel digestion was investigated as a possible improvement of the in-gel digestion protocol.



Iodoacetamide Cysteine S-Carbamidomethyl cysteine

Figure 3.22 The modification of cysteine with iodoacetamide to produce S-carbamidomethyl cysteine.

The S-carbamidomethylation of the cysteine residues of CYP1A2 and 3A4 was attempted in order to increase the identification of cysteine-containing peptides. Protein samples were prepared in the usual manner for analysis by SDS-PAGE (Section 2.7). Samples to be loaded onto the gel were treated with iodoacetamide (200 mg/ml in water) to give a concentration of approximately 20 mg/ml iodoacetamide and incubated for 30 minutes (in the dark) to allow the S-carbamidomethylation reaction to take place. The samples were analysed by SDS-PAGE and subjected to in-gel digestion with trypsin (Section 2.9), with analysis by nano-LC-ES. When searching with Sequest, the modification of cysteine residues (by a mass of 57.05 Da) was included to account for possible S-carbamidomethylation.

CYP1A2 and CYP3A4 each have 7 cysteine groups. However, no cysteine-containing peptides were identified in any of the samples. Therefore, it was not possible to draw any conclusions regarding the value of the S-carbamidomethylation step. Regardless, the modification of cysteine residues before in-gel digestion is a procedure that is widely used and, in principle, has value in increasing the identification of cysteine-containing peptides.

Chapter 4

Analysis for recombinant CYP1B1 and transfected CYP3A4

4.1 Introduction

Previous work established optimal conditions for MS analysis of purified recombinant CYP1A2, 2E1 and 3A4. This chapter describes the application of the nano-LC-ES and nano-ES methods to the analysis of P450s in complex biological samples.

4.2 The identification of recombinantly expressed CYP1B1

CYP1B1, the only member of the CYP1B subfamily, was identified in 1994 [189]. There has been particular interest in CYP1B1 in the field of cancer research because studies have shown that it is capable of metabolising several potential human carcinogens in normal tissue and shows increased expression in a range of human tumours, including breast, colon, lung, oesophagus, skin, lymph node, brain and testis [190]. This makes it an attractive potential target for anticancer drugs and as a biomarker for tumours.

Highly purified, functionally active expressed human CYP1B1 would be invaluable in the investigation of its use as a drug target and biomarker. With that aim, attempts were made the School of Pharmacy to express human CYP1B1 in *E. coli*. Mass spectrometry was used for the identification of recombinantly expressed proteins.

4.2.1 The identification of *E. coli* outer membrane protein C precursor and *E. coli* tryptophanase

On two occasions during attempts to express CYP1B1, SDS-PAGE analysis showed very dark Coomassie-blue stained bands, indicating over-expressed proteins. The proteins were subjected to in-gel tryptic digestion (Section 2.9), with desalting (Section 2.10.1). Peptides were analysed by nano-ES with manual collection of MS/MS and zoom scan data. Data was searched against the Finnigan Xcalibur human protein database (Section 2.14.1) using Sequest (Section 2.14.2). No

identifications were made, so data was searched against the *E. coli* database (Section 2.14.1).

The proteins were identified as *E. coli* outer membrane protein c precursor (ompC) and *E. coli* tryptophanase (tnaA). Both are membrane-bound proteins, which is consistent with their identification from *E. coli* membranes. TnaA and ompC were the first proteins to be identified without pre-selection of the proteins anticipated, and provided early indication that the methods developed here could be applied to complex samples.

4.2.2 The identification of CYP1B1

CYP1B1 was identified from an *E. coli* solubilised membrane fraction. Solubilised membranes were run by SDS-PAGE and stained with Coomassie blue dye (Section 2.7.1). Two bands were selected as candidates for CYP1B1, a large diffuse band and a fainter, tighter band at slightly lower molecular weight.

Identification was initially made after analysis by nano-ES. Both bands were subjected to in-gel tryptic digestion (Section 2.9), with desalting (Section 2.10.1). The two samples were analysed by nano-ES (Section 2.12.1) and searched using Sequest (Section 2.14.2) against the Finnigan Xcalibur human protein database (Section 2.14.1). The higher molecular weight diffuse protein band was identified as CYP1B1. Three peptides were identified with Xcorr scores ≥ 2.5, with 9 peptides identified in total (Figure 4.2). All the peptides were unique to CYP1B1.

The identification of CYP1B1 from *E. coli* solubilised membranes was repeated with analysis by nano-LC-ES. The gel lane from which the CYP1B1 band was removed is shown in Figure 4.1.

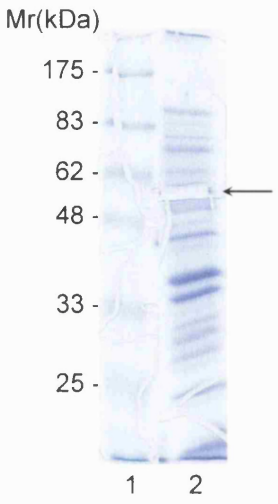


Figure 4.1 SDS-PAGE gel of *E. coli* solubilised membranes from which CYP1B1 was identified. *Lane 1* Protein molecular weight markers; *Lane 2* CYP1B1 solubilised membrane fraction. The position from which the CYP1B1 band was removed is indicated by an arrow.

Searching with Sequest after in-gel digestion and analysis by nano-LC-ES led to the identification of 11 peptides with Xcorr values ≥ 2.5, and 14 peptides in total (Figure 4.2). This corresponds to percentage sequence coverages (by amino acid count) of 23% (Xcorr ≥ 2.5) and 34% (All Xcorr values). All the peptides were unique to CYP1B1. The MS/MS spectrum of the doubly charged $[M+2H]^{2+}$ ion of the peptide VQAELDQVVGR is shown in Figure 4.3.

---MLSPNDP WPLNPLSIQQ TTLLLLLSVL ATVHVGQRLL RQRRRQLRSA PPGPFAWPLI
GNAAAVGQAA HLSFARLARR YGDVFQIRLG SCPIVVLNGE RAIHQALVQQ GSAFADRPAF
ASFRVVSGGR SMAFGHYSEH WKVQRRAAHS MMRNFFTRQP RSRQVLEGHV LSEARELVAL
LVRGSADGAF LDPRPLTVVA VANVMSAVCF GCRYSHDDPE FRELLSHNEE FGRTVGAGSL
VDVMPWLQYF PNPVRTVFRE FEQLNRNFSN FILDKFLRHC ESLRPGAAPR DMMDAFILSA
EKKAAGDSHG GGARLDLENV PATITDIFGA SQDTLSTALQ WLLLLFTRYP DVQTRVQAEL
DQVVGRDRLP CMGDQPNLPY VLAFLYEAMR FSSFVPVTIP HATTANTSVL GYHIPKDTVV
FVNQWSVNHD PVKWPNPENF DPARFLDKDG LINKDLTSRV MIFSVGKRRC IGEELSKMQL
FLFISILAHQ CDFRANPNEP AKMNFSYGLT IKPKSFKVNV TLRESMELLD SAVQNLQAKE
TCQHHHHHH

XXXX Xcorr ≥ 2.5
XXXX Xcorr < 2.5

Figure 4.2 The amino acid sequence of recombinant CYP1B1, showing the peptides identified after in-gel digestion of *E. coli* solubilised membranes used for CYP1B1 expression with analysis by nano-ES and nano-LC-ES. The first 4 amino acids were omitted from the N-terminal of the recombinantly expressed protein; a tag consisting of 6 His residues was added to the C-terminal for purification purposes.

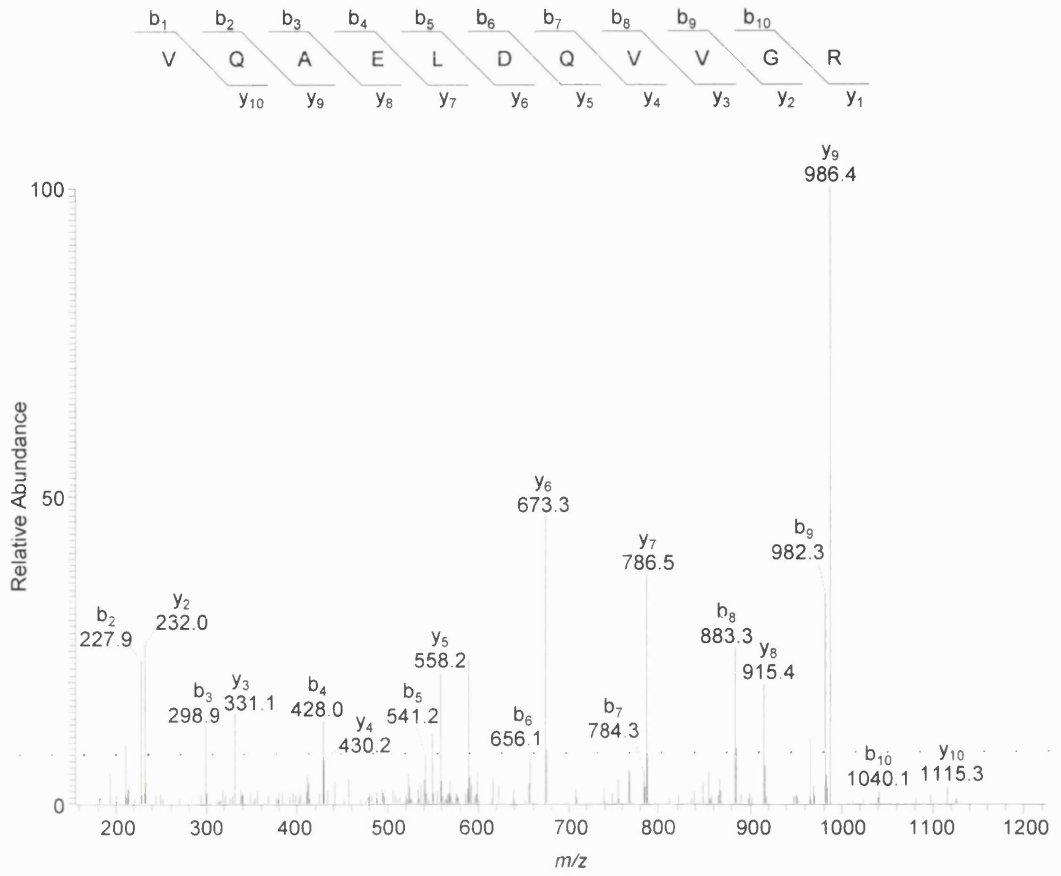


Figure 4.3 MS/MS spectrum of the tryptic peptide VQAELDQVVGR $[M+2H]^{2+}$ ion of *m/z* 607.7, identified to originate from CYP1B1. The amino acid sequence of the peptide is shown above the spectrum. y and b ions are formed by peptide bond cleavage with charge retention on the C-terminus and N-terminus, respectively. *E. coli* solubilised membranes were analysed by SDS-PAGE. The CYP1B1 protein band was subjected to in-gel tryptic digestion (Section 2.9) with desalting (Section 2.10.1) and the resultant peptides were dried and stored at -80 °C. Samples were reconstituted in 20 µl of 0.1% TFA and analysed by nano-LC-ES: 1 µl of sample was injected via a 10 µl sample loop. Data files were searched using Sequest software (Section 2.14.2) against the Finnigan Xcalibur human database (Section 2.14.1).

4.3 The analysis of CYP3A4-transfected and control RIF-1 mouse tumours for P450 proteins

Mice with CYP3A4-transfected RIF-1 tumours were used for an investigation into the role of CYP3A4 in the metabolism of AQ4N, a bioreductive prodrug that is metabolically activated under hypoxic conditions to a cytotoxic product [191]. Previously, it was found that antitumour efficacy in mice transfected with CYP3A4 and exposed to AQ4N and radiation was significantly increased over the control group of mice, whose tumours were not transfected with CYP3A4. Three out of nine

of the tumours from CYP3A4-transfected mice shrank completely to the point where they were not detectable by gross examination [191].

S9 fractions from human CYP3A4-transfected and control RIF-1 murine tumours (Section 2.2) were analysed with the aim of identifying the product of the transfected CYP3A4 gene. Microsomes were prepared from the S9 fractions in order to simplify the samples (Section 2.2.1). Cytosolic fractions and microsomes prepared from the S9 fractions were analysed by SDS-PAGE, with Coomassie Blue staining (Figure 4.4). The gel region between 48 and 62 kDa was divided into five bands and subjected to in-gel tryptic digestion (Section 2.9). Samples were analysed by nano-LC-ES (Section 2.12.2), and the data was searched with Sequest (Section 2.14.2).

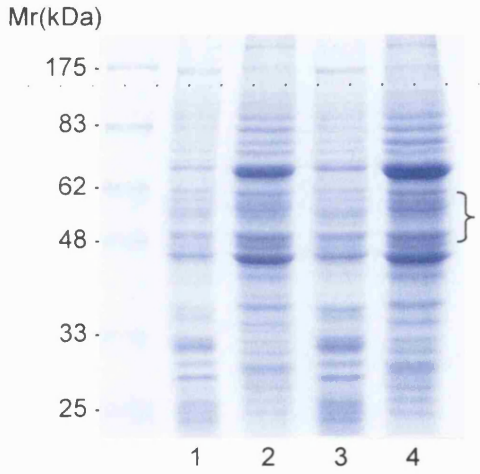


Figure 4.4 SDS-PAGE analysis of cytosolic and microsomal fractions prepared from CYP3A4 transfected and control RIF-1 murine tumours. *Lane 1* CYP3A4 transfected tumour, microsomal fraction; *Lane 2* CYP3A4 transfected tumour, cytosolic fraction; *Lane 3* control tumour, microsomal fraction; *Lane 4* control tumour, cytosolic fraction. The molecular weight area removed for analysis is indicated by a bracket on the right. This area was divided into 5 bands for each sample and the proteins were subjected to in-gel tryptic digestion, with desalting. The resultant peptides were dried and stored at -80 °C. Samples were reconstituted in 20 µl of 0.1% TFA for analysis by nano-LC-ES: 1 µl of sample was injected onto the nano LC via a 10 µl sample loop. Data was searched against the Finnigan Xcalibur human and mouse databases (Section 2.14.1) using Sequest (Section 2.14.2).

No P450s were identified in any of the samples. RIF-1 was chosen as a carrier because it is naturally low in P450 expression, therefore the absence of CYP3A4 was not unexpected in the control samples. However, no CYP3A4 was detected in the microsomes prepared from the CYP3A4-transfected sample. This is consistent with the observation that, after destaining of the gel, there was no discernible difference

between the microsomes from the CYP3A4 transfected tumour and those from the control tumour in the CYP region of the gel (Figure 4.4).

Several possible explanations were investigated for the failure to identify CYP3A4. It was feasible that the area of the gel containing CYP3A4 had not been selected for analysis. The microsomal fraction of the CYP3A4-transfected tumour was spiked with 11 pmol of recombinant CYP3A4 and run by SDS-PAGE. After Coomassie Blue staining, the spiked CYP3A4 band was clearly visible in the region of the gel that had been excised and analysed. Another possibility was that any CYP3A4 originally present in the tumour had been degraded by proteases; alternatively, CYP3A4 may be present, but at levels below the limit of detection of the nano-LC-ES technique. To test these theories, Western blotting was carried out (Section 4.3.1).

4.3.1 Western blotting of S9, microsomal and cytosolic fractions prepared from control and CYP3A4-transfected RIF-1 mouse tumours

Western blots for CYP3A4 were prepared for the S9, cytosolic and microsomal fractions (Section 2.8). Gel lanes containing 11 pmol and 5.5 pmol of recombinant CYP3A4 were blotted as positive controls. Two identical gels were run; one was stained with Coomassie blue dye and the other was analysed using a human CYP3A4/3A5 antibody (Figure 4.5). The gel used for Western blotting was later stained to check for any protein left behind by the blotting process: very little protein remained on the gel.

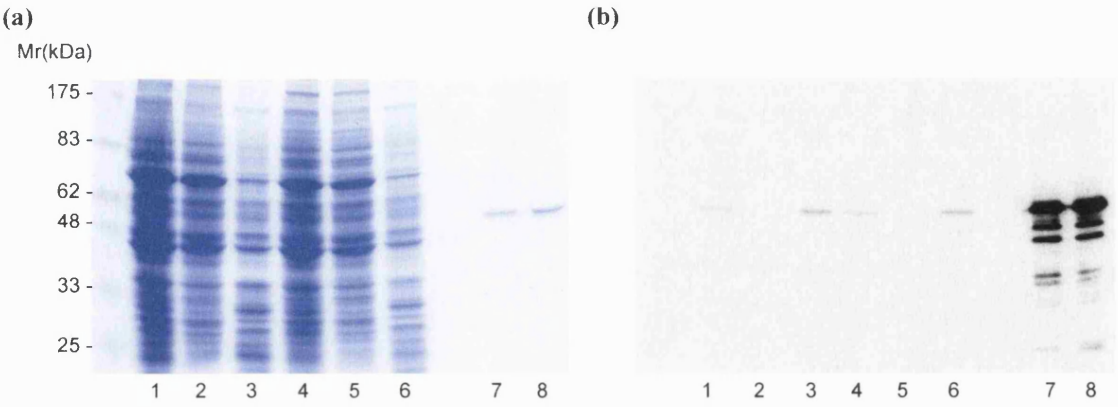


Figure 4.5 SDS-PAGE and Western blotting for CYP3A4 of S9, cytosolic and microsomal fractions prepared from CYP3A4 transfected and control murine RIF-1 tumours. **(a)** SDS-PAGE gel; **(b)** Western blot of an SDS-PAGE gel. **(a)** and **(b)** *Lane 1* control tumour S9 fraction; *Lane 2* control tumour cytosolic fraction; *Lane 3* control tumour microsomal fraction; *Lane 4* CYP3A4-transfected tumour S9 fraction; *Lane 5* CYP3A4-transfected tumour cytosolic fraction; *Lane 6* CYP3A4-transfected tumour microsomal fraction; *Lane 7* recombinant CYP3A4 (5.5 pmol); *Lane 8* recombinant CYP3A4 (11 pmol).

The Western blot shows that, although CYP3A4 is present in the S9 and microsomal fractions, the amount of the protein is quantitatively much less than 5.5 pmol as judged from the authentic CYP3A4 samples analysed. Furthermore, there is no obvious difference between the amount of CYP3A4 in the control tumour and the amount in the CYP3A4-transfected tumour. This indicates either that transfection of CYP3A4 has not been successful, or that the CYP3A4 has been degraded, perhaps by proteases. Protease inhibitors were not used in the preparation of the S9, microsomal or cytosolic fractions. The limit of detection for in-gel tryptic digestion of CYP3A4 with mass spectrometric analysis is approximately 0.6 pmol on-gel (Section 3.8.2), therefore it is likely that the levels of CYP3A4 in the tumour fractions are lower than those that can be identified by nano-LC-ES. Preconcentration of the microsomal fractions before analysis would increase the likelihood of detecting CYP3A4. In addition, the nano-LC-ES instrument program could be modified so that ions with *m/z* values corresponding to peptides known to be from CYP3A4 are selected for MS/MS fragmentation over the other ions present in the sample, thereby increasing the chances of detecting CYP3A4 (for an example of this, see Section 6.2.2.2). CYP3A4 was successfully identified in other biological tissues, for example the human colorectal metastatic and surrounding liver samples described in Chapter 5.

Chapter 5

The identification of P450s in human colorectal metastases and surrounding liver

5.1 Introduction

P450s in liver are known to be of major importance to the fate of anticancer agents; however, their expression and role in tumours has had little attention. P450 mediated metabolism is generally viewed as a route to drug detoxification and increased elimination, although P450 activation of certain anticancer drugs, e.g. cyclophosphamide, dacarbazine and thiotepa, has long been known and the importance of this process as a way of targeting novel anticancer therapy is being explored [148]. P450s are also responsible for reductive metabolism and can contribute to the activation of tumour hypoxia targeted cytotoxins [192]. The fate of cancer therapeutic outcome in part relies on the expression profile of an array of different P450 enzymes, especially in the liver, contributing to the metabolic fate of administered drugs. The heterogeneity of the human population means that a P450 profile in any one individual is unique due to a combination of genetic and environmental factors including the plethora of drugs used in treating disease. Determination of the P450 protein expression profile on an individual patient basis prior to cancer chemotherapy could provide important information regarding the fate of the selected drugs and hence outcome of therapy. In addition, the relative activity of liver and tumour drug metabolism will have a major impact in determining therapeutic outcome and in the development of tumour-specific prodrugs.

The aim of the present study was to apply the methodology developed in Chapter 3 to the proteomic analysis of P450s in human tumour and liver samples, and to reveal its potential in the assessment of which P450 enzymes could contribute to the outcome of chemotherapy. A collaboration was established with surgeons at the Royal Free and University College Medical School: six sets of colorectal liver metastases and corresponding surrounding liver samples from patients with metastatic colorectal cancer of the liver were provided. Stomach and colorectal cancers are among the most common cancers worldwide, with the only curative approach being surgical resection. For patients with metastatic disease, the 5-year survival rate is less than 5% [193]. There is clearly a need for novel anticancer drugs that would change the survival rate in this disease.

5.2 Sample processing

Sample details are shown in Table 5.1 Frozen samples were ground, homogenised and then subjected to differential ultracentrifugation for the preparation of microsomes (Section 2.3.3). The microsomal pellets prepared from the tumour samples were much paler in colour than those prepared from the liver samples, and also much smaller. For sample 5T the pellet consisted only of a very small (approximately 2 x 5 mm) smear on the side of the centrifuge tube. Resuspension volumes for the pellets were chosen depending on the size of the pellet obtained: all the livers except for sample 4L were suspended in 1 ml of suspension buffer; sample 4L was suspended in 800 µl of buffer; tumour samples 1T, 3T, 4T and 6T were suspended in 300 µl of buffer; sample 2T was suspended in 200 µl of buffer and sample 5T was suspended in 100 µl of buffer. Microsomal protein concentrations were determined using the Bradford assay (Section 2.4); the Bradford assay calibration line is shown in Figure 5.1. Tumour and liver patient details and microsomal protein concentrations are shown in Table 5.1.

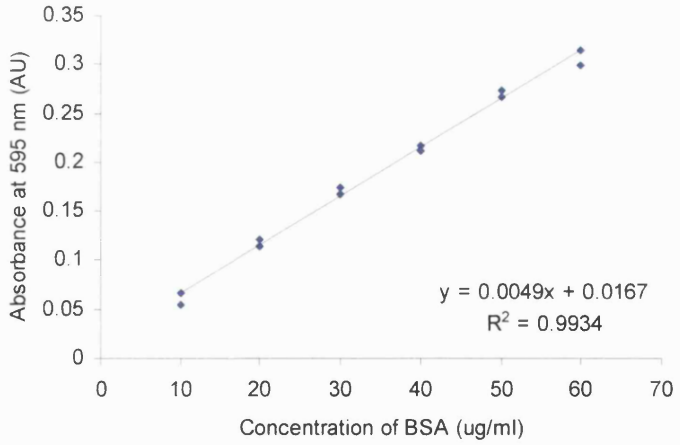


Figure 5.1 The Bradford assay calibration line used to calculate protein concentrations of microsomes prepared from human colorectal metastases and liver samples.

Microsomal protein (approximately 25 µg for all samples except for sample 5T, where 2.6 µg was used; see Table 5.1) was resolved on 10% acrylamide gels. Figure 5.2 shows the separation of the microsomal samples by SDS-PAGE; the 48-62 kDa region can be seen to be very rich in protein, especially for the liver microsomes. The concentration of protein in microsomes prepared from tumour sample 5T was

insufficient for analysis. For each of the other samples, six bands of approximately equal size covering the molecular weight range of the P450s (48-62 kDa) were removed and subjected to in-gel digestion with trypsin (Section 2.9), followed by analysis by nano-LC-ES.

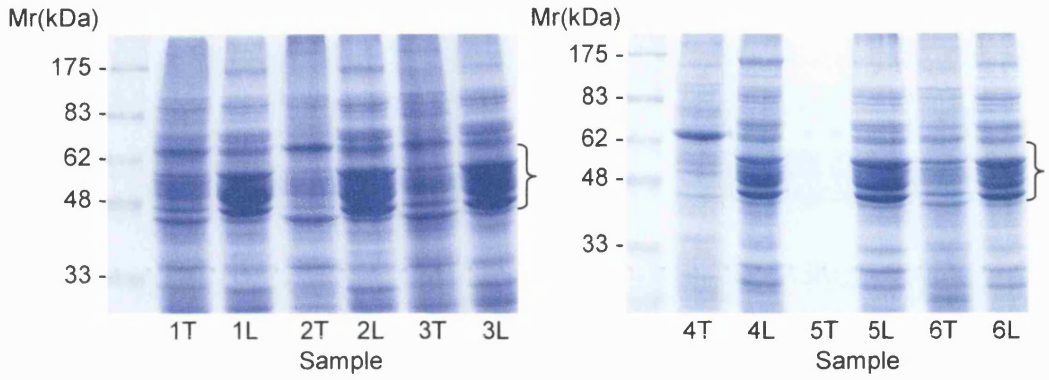


Figure 5.2 SDS-PAGE analysis of microsomal protein from six pairs of liver and tumour samples. Molecular weight markers are indicated on the left (in kDa). Approximately 25 µg protein was loaded onto the gels for all samples except 5T (2.6 µg). The area selected for analysis is indicated by brackets. This area (all samples except 5T) was divided into six bands of approximately equal size and the proteins present subjected to in-gel digestion with trypsin (Section 2.9), without desalting. Peptide mixtures were stored at -80 °C and reconstituted in 12 µl of 0.1% TFA immediately prior to analysis by nano-LC-ES (Section 2.12.2): 1 µl of sample was injected onto the nano-LC via a 10 µl sample loop. Each sample was analysed at least twice. Data was searched using Sequest (Section 2.14.2) against the human database obtained from NCBI and customised (Section 3.5). For sample details, refer to Table 5.1.

Table 5.1 Patient details and microsomal protein concentrations for samples taken from six patients with metastatic colorectal cancer of the liver.

Sample[a]	Age of patient (years)	Gender (M/F)	Smoker/non-smoker (S/NS)	Alcohol intake	Medical history	Drug history	Mass of tissue used (g)	Protein concentration in microsomes (μg μl^{-1})[b]	Calculated microsomal protein concentration in tissue (mg g^{-1})
1T	66	M	NS	None	Myocardial infarction (MI), atrial fibrillation	Warfarin, amiodarone, lisinopril, fluvastatin, aspirin	1.6[c]	9.0	1.7
1L							1.9[c]	14.3	7.6
2T	42	F	NS	None	None	None	0.37[c]	1.3	0.67
2L							1.5[c]	14.0	9.7
3T	60	F	NS	None	None	None	1.7[c]	2.9	0.53
3L							1.8[c]	12.8	7.0
4T	79	F	NS	Up to 20 units/week	MI, osteoarthritis, hypertension, portal vein embolisation (PVE) 132 days prior to surgery	Amlodipine, thyroxine, isosorbide mononitrate, ranitidine, aspirin	1.1[d]	1.1	0.29
4L							1.0[d]	3.9	3.1
5T	62	M	NS	Up to 20 units/week	None	None	0.28[d]	0.14	0.05
5L							1.3[d]	5.4	4.3
6T	69	F	NS	Up to 20 units/week	Hysterectomy, PVE 81 days prior to surgery	Warfarin, fluoxetine, cyclizine	1.1[d]	4.1	1.1
6L							1.2[d]	4.7	3.9

[a] T = tumour tissue; L = liver tissue; [b] determined using the Bradford assay; [c] ground using percussion mortar and pestle; [d] ground using a tissue dismembrator (Mikro-Dismembrator U, B. Braun Biotech International).

Table 5.2 P450 enzymes identified in the microsomal fractions of liver and tumour samples from six patients with metastatic colorectal cancer of the liver.

Sample [a]	P450s identified: number of matched peptides (% sequence coverage by amino acid) [b]													
	1A2	2A6	2B6	2C8	2C9	2C19	2D6	2E1	3A4	4A11	4F2	4F11	8B1	27A1
1T	6 (16)	1[c] (2.6)	-	6 (14)	10 (29)	-	1 (2.8)	2 (3.0)	6 (16)	3 (8.1)	3 (7.5)	-	-	1 (4.7)
1L	7 (22)	11 (30)	1 (2.4)	10 (33)	12 (38)	-	1 (3.6)	6 (16)	7 (21)	6 (21)	6 (16)	-	-	-
2T	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2L	7 (18)	16 (42)	-	8 (28)	8 (24)	3 (9.0)	-	12 (31)	9 (25)	6 (17)	7 (16)	2[d] (6.1)	3 (8.4)	-
3T	1 (3.1)	5 (13)	-	4 (16)	4 (17)	2 (6.3)	-	6 (18)	5 (13)	2 (16)	1[e] (2.3)	1[e] (2.3)	-	-
3L	5 (14)	18 (47)	-	9 (30)	9 (29)	6 (18)	1 (3.6)	9 (24)	13 (36)	4 (10)	4 (10)	2 (5.0)	3 (11)	-
4T	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4L	7 (17)	9 (24)	1 (2.4)	9 (30)	6 (22)	2 (5.7)	3 (15)	10 (25)	10 (31)	6 (14)	9 (20)	2 (5.9)	-	-
5T[f]	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5L	11 (32)	19 (51)	1 (3.5)	12 (36)	9 (29)	3 (9.0)	-	10 (26)	15 (46)	7 (15)	5 (13)	2 (5.9)	1 (1.9)	-
6T	1 (3.1)	3 (7.9)	-	2 (5.9)	7 (21)	-	-	9 (24)	3 (7.8)	2 (4.9)	-	1[d] (3.6)	-	-
6L	1 (3.1)	12 (35)	-	7 (24)	10 (31)	4 (15)	1 (3.6)	4 (13)	8 (22)	4 (10)	11 (28)	1[d] (3.6)	-	-

[a] T = tumour tissue; L = liver tissue; [b] the number of matched peptides and % sequence coverage refer to identifications with Xcorr scores of ≥2.5. [c] Peptide common to CYP2A6 and 2A13; [d] peptides could be from CYP4F8 and/or 4F11; [e] peptide common to CYP4F2, 4F3, 4F11 and 4F12; [f] the microsomal protein concentration of sample 5T was too low for analysis. - indicates that the P450 protein was not identified.

5.3 P450 enzymes identified

In total, 14 distinct P450 enzymes were identified, 13 from the liver samples and 12 from three of the tumours (Table 5.2). The liver samples yielded identifications of CYP1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1, 3A4, 4A11, 4F2, 4F11 and 8B1. Eight of these proteins, CYP1A2, 2A6, 2C8, 2C9, 2E1, 3A4, 4A11 and 4F2, were detected in each of the 6 livers; CYP2C19 and 4F11 were identified in all of the liver samples except for that of patient 1; CYP2D6 was identified in four of the livers and CYP2B6 and 8B1 were each identified in half of the liver samples. An example MS/MS spectrum, used for the identification of CYP2E1, is shown in Figure 5.3. In two cases it was not possible to distinguish between P450 enzymes: the two CYP4F11 peptides detected in liver sample 2L are also common to CYP4F8 (Table 5.2). No peptides were identified in the study that are unique to CYP4F8, whilst several were identified that are unique to CYP4F11; nevertheless, it is not possible to distinguish between CYP4F8 and CYP4F11 in sample 2L. This is also the case for the single CYP4F8/4F11 peptide identified in sample 6L.

Twelve P450 enzymes were identified in three of the five tumour samples analysed (1T, 3T and 5T). The three tumour samples in which P450s were detected yielded identifications of CYP1A2, 2A6, 2C8, 2C9, 2C19, 2D6, 2E1, 3A4, 4A11, 4F2, 4F11 and 27A1 (Table 5.2). Seven of these P450 enzymes (CYP1A2, 2A6, 2C8, 2C9, 2E1, 3A4 and 4A11) were found in all three of the tumours; CYP4F2 and 4F11 were found in two of the tumours; CYP2C19, 2D6 and 27A1 were each found in one tumour sample only. CYP2B6 and 8B1, which were identified from some of the liver samples, were not detected in any of the tumours. As was the case for two of the liver P450 identifications, it was not always possible to distinguish between P450 enzymes (Table 5.2): the single CYP2A6 peptide identified in tumour sample 1T is also common to CYP2A13; identification of CYP4F2/4F11 in sample 3T was made on the basis of one peptide that is also common to CYP4F3 and 4F12; the single CYP4F11 peptide identified from sample 6T is also common to CYP4F8. Since there were no peptides unique to CYP2A13, 4F3, 4F8 or 4F12 identified in the study, these proteins were not included as separate columns in Table 5.2. Two of the tumour samples (2T and 4T) did not express any P450 enzymes in quantities sufficient for

detection. In the case of sample 2T, this can perhaps be attributed to the low mass of tissue available for analysis. The tissue microsomal protein concentration (Table 5.1) was comparable to that of sample 3T, therefore, if more tissue had been available then P450s may have been detected. Sample 4T was low in microsomal protein (Table 5.1) and, in addition, was morphologically necrotic.

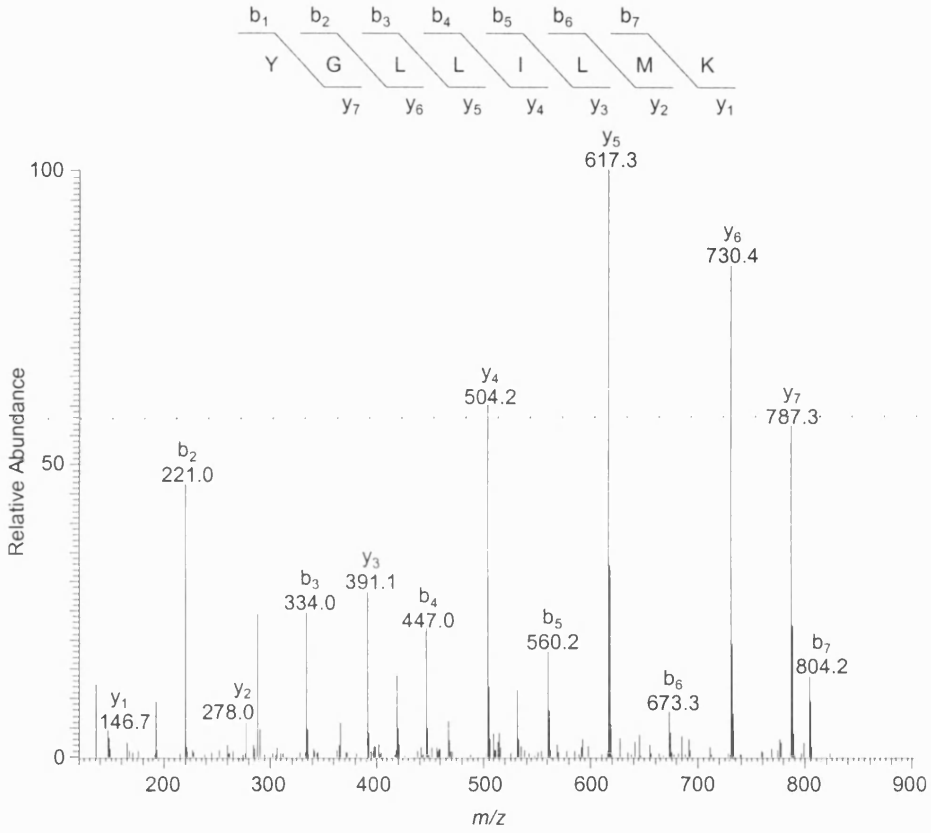


Figure 5.3 MS/MS spectrum of the tryptic peptide YGLLILMK $[M+2H]^{2+}$ ion of *m/z* 476.0, identified to originate from CYP2E1. The amino acid sequence of the peptide is shown above the spectrum. y and b ions are formed by peptide bond cleavage with charge retention on the C-terminus and N-terminus, respectively.

The numbers of peptides identified for each P450 protein in the different samples ranges from a maximum of 19 (51% sequence coverage by amino acid count) to a minimum of 1 (1.9% to 3.6% sequence coverage) (Table 5.2). This variation may be due to a number of factors, including protein abundance, the accessibility of the protein to complete trypsin digestion, and mass spectrometric sensitivity. Whilst the non-detection of a particular P450 does not mean that it is not present in a given sample, this does indicate that, if present, it is at a relatively low abundance. Hence, where a P450 enzyme is only ever identified on the basis of one or two peptides (for

example, CYP2B6, 4F11 or 27A1), the failure to identify the same enzyme in other samples does not necessarily imply that the enzyme is absent in these samples. Similarly, at least to a first approximation, amino acid sequence coverage relates to relative protein abundance (Section 5.5).

One of the challenges when investigating P450 expression is distinguishing between P450 enzymes that share a high level of sequence homology. For example, CYP2C9 and CYP2C19 share 91% sequence homology (Figure 5.4); nevertheless, it was possible to differentiate them. The overall sequence coverages for CYP2C9 and 2C19 were 41% and 26%, respectively; the corresponding peptides are shown in bold in Figure 5.4. MS/MS spectra for the two closely related CYP2C9 and CYP2C19 peptides GTTILISLTSVLHDNK and GTTILTSLTSVLHDNK are shown in Figure 5.5.

```
2C9  MDSLVVLVLCLSCLLLLSLWRQSSGRGKLPPGPTPLPVIGNILQIGIKDISKSLTNLSKV  60
2C19 MDPFVVLVLCLSCLLLLSIWRQSSGRGKLPPGPTPLPVIGNILQIDIKDVSKSLTNLSKI  60
       **              *                          *   *         *

2C9  YGPVFTLYFGLKPIVVLHGYEAVKEALIDLGEEFSGRGIFPLAERANRGFGIVFSNGKKW 120
2C19 YGPVFTLYFGLERMVVLHGYEVVKEALIDLGEEFSGRGHFPLAERANRGFGIVFSNGKRW 120
                ***       *                *                   *

2C9  KEIRRFSLMTLRNFGMGKRSIEDRVQEEARCLVEELRKTKASPCDPTFILGCAPCNVICS 180
2C19 KEIRRFSLMTLRNFGMGKRSIEDRVQEEARCLVEELRKTKASPCDPTFILGCAPCNVICS 180

2C9  IIFHKRFDYKDQQFLNLMEKLNENIKILSSPWIQICNNFSPIIDYFPGTHNKLLKNVAFM 240
2C19 IIFQKRFDYKDQQFLNLMEKLNENIRIVSTPWIQICNNFPTIIDYFPGTHNKLLKNLAFM 240
        *                      * * *         **               *

2C9  KSYILEKVKEHQESMDMNNPQDFIDCFLMKMEKEKHNQPSEFTIESLENTAVDLFGAGTE 300
2C19 ESDILEKVKEHQESMDINNPRDFIDCFLIKMEKEKQNQQSEFTIENLVITAADLLGAGTE 300
     * *             *   *       *      *  *      * **  *  *

2C9  TTSTTLRYALLLLLKHPEVTAKVQEEIERVIGRNRSPCMQDRSHMPYTDAVVHEVQRYID 360
2C19 TTSTTLRYALLLLLKHPEVTAKVQEEIERVIGRNRSPCMQDRGHMPYTDAVVHEVQRYID 360
                                           *

2C9  LLPTSLPHAVTCDIKFRNYLIPKGTTILISLTSVLHDNKEFPNPEMFDPHHFLDEGGNFK 420
2C19 LIPTSLPHAVTCDVKFRNYLIPKGTTILTSLTSVLHDNKEFPNPEMFDPRHFLDEGGNFK 420
      *           *               *                    *

2C9  KSKYFMPFSAGKRICVGEALAGMELFLFLTSILQNFNLKSLVDPKNLDTTPVVNGFASVP 480
2C19 KSNYFMPFSAGKRICVGEGLARMELFLFLTFILQNFNLKSLIDPKDLDTTPVVNGFASVP 480
       *               *  *        *          *   *

2C9  PFYQLCFIPV 490
2C19 PFYQLCFIPV 490
```

Figure 5.4 Sequence alignment for human CYP2C9 and CYP2C19. Peptides identified during the study are shown in bold. Sequence differences are marked with asterisks.

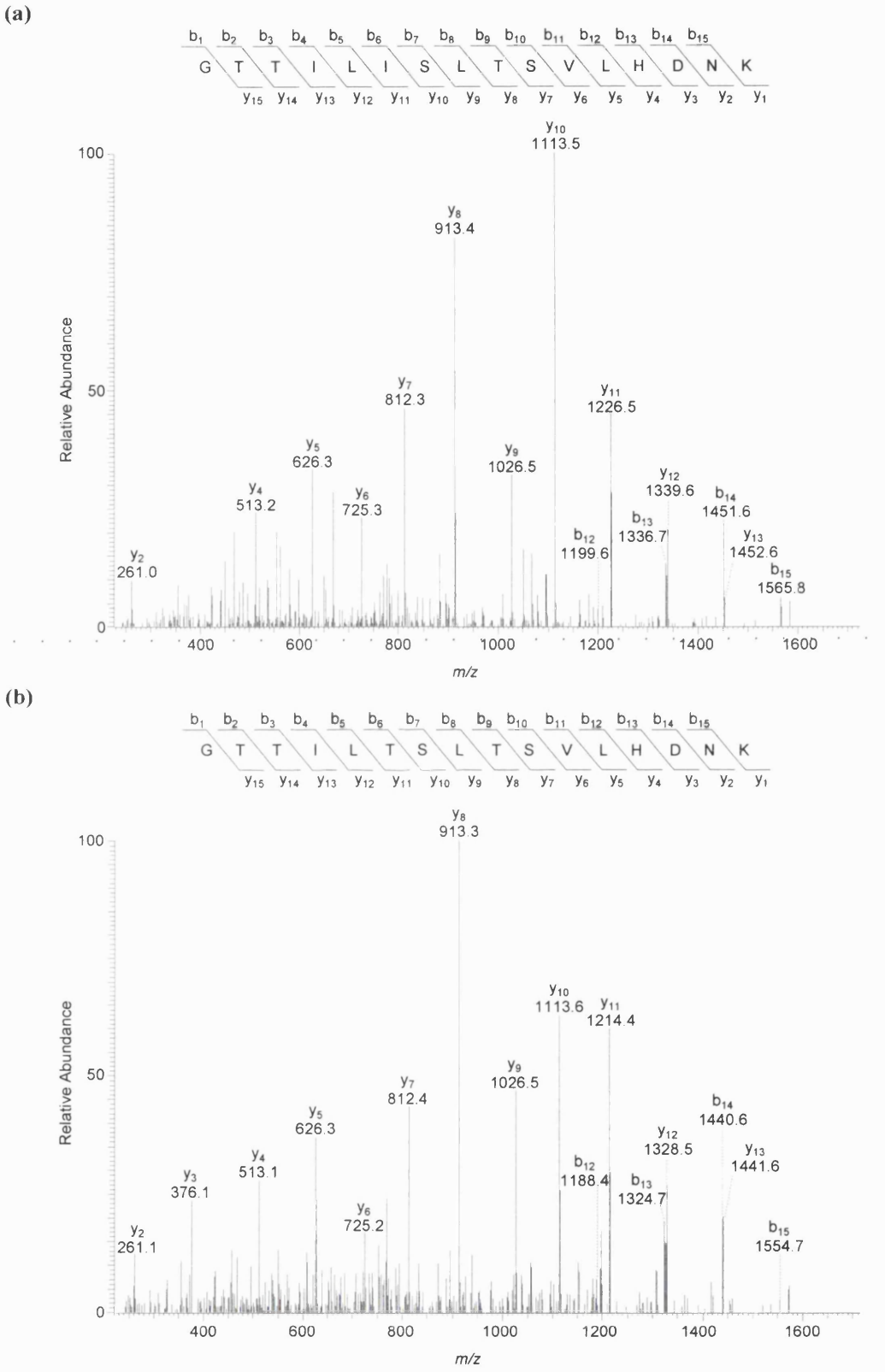


Figure 5.5 MS/MS spectra of two closely related CYP2C9 and CYP2C19 peptides. **(a)** CYP2C9 peptide GTTILISLTSVLHDNK $[M+2H]^{2+}$ ion of *m/z* 856.8. The spectrum is an average of 3 scans. **(b)** CYP2C19 peptide GTTILTSLTSVLHDNK $[M+2H]^{2+}$ ion of *m/z* 851.0. The spectrum is an average of 3 scans. The amino acid sequences of the peptides are shown above the spectra. y and b ions are formed by peptide bond cleavage with charge retention on the C-terminus and N-terminus, respectively.

The P450s identified include all of the major P450s involved in drug metabolism (CYP1A2, 2A6, 2B6, 2C8 2C9, 2C19, 2D6, 2E1, 3A4) and some of those involved in bile acid synthesis and fatty acid metabolism (CYP4A11, 4F2, 4F11, 8B1 and 27A1). Figure 5.6 shows the mean number of peptides identified in the liver and three of the tumour samples for each P450 enzyme detected. The three tumour samples that yielded P450 identifications were found to express the same pattern of P450s as the livers, although generally fewer peptides were found in the tumours. It is intriguing that essentially the same P450 profile was identified in both the liver and three of the tumour samples, given that they originated in the colon. The metastases were shown to be free from liver tissue by histopathological examination, therefore this similarity cannot be caused by hepatocyte contamination. Massaad *et al* [194] assayed for CYP1A1, 1A2, 2B, 2C, 2E1 and 3A by Western blotting in seven colorectal primary tumours and corresponding peritumoural tissues, but could only detect CYP3A4. The possible influence of the liver environment on the P450 expression profile of deposited tumours invites further investigation.

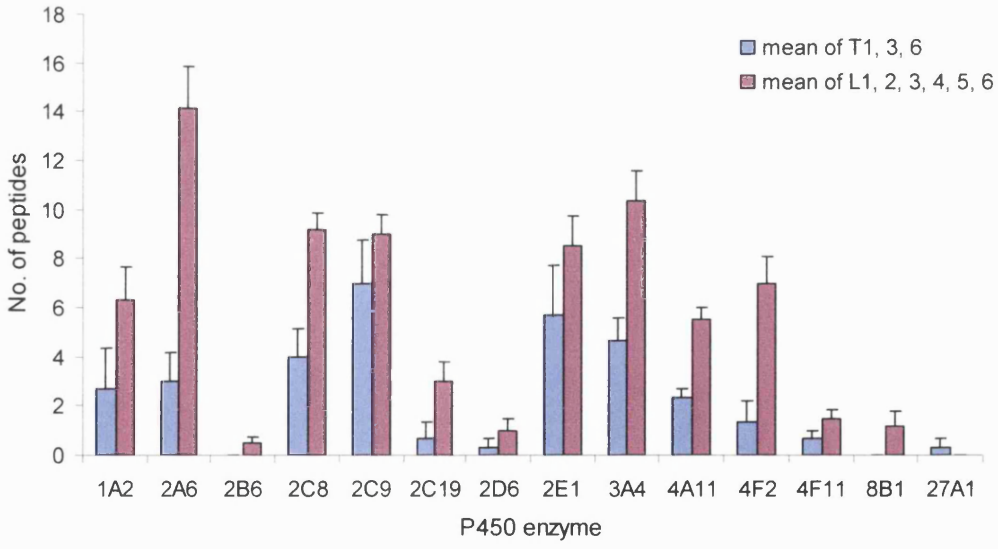


Figure 5.6 The mean P450 content of six liver samples and three tumour samples from patients with metastatic colorectal cancer of the liver. Error bars show +1 standard error of the mean. No P450 enzymes were identified in samples T2, T4 or T5.

The subject of this study was the detection of P450s; detailed analysis was not performed on other identified proteins. Nevertheless, examples of other proteins that were identified with high peptide numbers/sequence coverage include adenosine triphosphate (ATP) synthases and protein disulphide isomerases, which were present

in all samples analysed except for the necrotic tumour sample 4T, and flavin-containing monooxygenases and uridine diphosphate (UDP) glycosyltransferases, which were identified in all samples analysed except 2T and 4T (in which no P450 enzymes were identified) (Table 5.3).

Table 5.3 Examples of other proteins identified with high peptide number/sequence coverage in the microsomal fractions of liver and tumour samples from six patients with metastatic colorectal cancer of the liver.

Sample [a]	Protein: identified? (Y/N)			
	ATP synthase	Protein disulphide isomerase	Flavin-containing mono-oxygenase 3 or 5	UDP glycosyl transferase
1T	Y	Y	Y	Y
1L	Y	Y	Y	Y
2T [b]	Y	Y	N	N
2L	Y	Y	Y	Y
3T	Y	Y	Y	Y
3L	Y	Y	Y	Y
4T [c]	N	N	N	N
4L	Y	Y	Y	Y
5T [d]	-	-	-	-
5L	Y	Y	Y	Y
6T	Y	Y	Y	Y
6L	Y	Y	Y	Y

[a] T = tumour tissue, L = liver tissue; [b] no P450s were identified from sample 2T; [c] necrotic tumour, no P450s identified; [d] the microsomal protein concentration of sample 5T was too low for analysis.

5.4 A comparison of the methods used to grind the samples

An interesting pattern was observed by comparing the two methods used to grind the liver samples (Table 5.1). Samples 1, 2 and 3 were ground using a percussion mortar and pestle, whilst samples 4, 5 and 6 were ground using a tissue dismembrator (Mikro-Dismembrator U), which pulverises the tissue using a ball-mill. The calculated tissue microsomal protein concentrations for the livers ground using the tissue dismembrator are considerably less than those of the livers ground using a percussion mortar and pestle. The difference in the means of the protein concentrations of the two sets of livers is significant (Student's t-test, $P < 0.005$). This suggests that tissue disruption using the tissue dismembrator is not as effective as tissue disruption using the percussion mortar and pestle, assuming all the livers have approximately the same microsomal protein concentrations.

5.5 The contribution of each P450 enzyme to the total P450 content in liver: a comparison of the results from nano-LC-ES analysis with the immunochemical results from published studies

The contribution of each P450 enzyme to the total P450 content in human liver has been investigated previously by immunochemistry [144,185]. The data of the current study are of a qualitative nature; nevertheless, the relative expression of P450s based on the average number of peptides detected in the six liver samples in this mass spectrometric approach was compared with previous accounts. The immunochemically derived results of two reports are shown in Table 5.4, along with values calculated from the current study based on the average number of peptides detected in the six liver and three tumour samples. Figure 5.7 compares the three sets of liver results for CYP1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1 and 3A4. Values based on the average number of peptides detected compare very favourably with the literature values, based on immunoblotting. This indicates that, to a first approximation, the number of tryptic peptides identified by Sequest to originate from a given protein can be correlated to the expression level of that protein, at least where the proteins compared are from the same super-family. The observation that data from Sequest search results can be used in a semi-quantitative manner is supported by previous work by Ducret *et al* [195].

Table 5.4 The contribution to the total P450 content by different enzymes in liver (and tumour) microsomes.

P450	% of total P450 content of liver microsomes [a] (pmol P450/mg microsomal protein)		% of total P450 content by peptide count	
	Rodrigues 1999 [b]	Shimada *et al* 1994 [c]	Liver	Tumour
1A2	8.0 (45)	13 (42)	8.2	8.9
2A6	13 (68)	4 (14)	18	9.2
2B6	7.0 (39)	0.2 (1.0)	0.6	n.d.
2C8	12 (64)		12	12
2C9	18 (96)	18[d] (60)	12	21
2C19	4.0 (19)		3.9	2.0
2D6	2.0 (10)	1.5 (5.0)	1.3	1.0
2E1	9.0 (49)	6.6 (22)	11	17
3A4	20 (108)	29 (96)	13	14
4A11	-	-	7.1	7.1
4F2	-	-	9.1	4.1
4F11	-	-	1.9	2.0
8B1	-	-	1.5	n.d.
27A1	-	-	n.d.	1.0
Total	93 (498)	72 (240)	100	100

[a] Determined by immunoblotting; [b] Gentest figures, using a pool of liver microsomes from 12 different organ donor subjects [185]; [c] mean data from 60 different livers [144]; [d] 2C combined total. - indicates that the P450 was not analysed; n.d. = not detected.

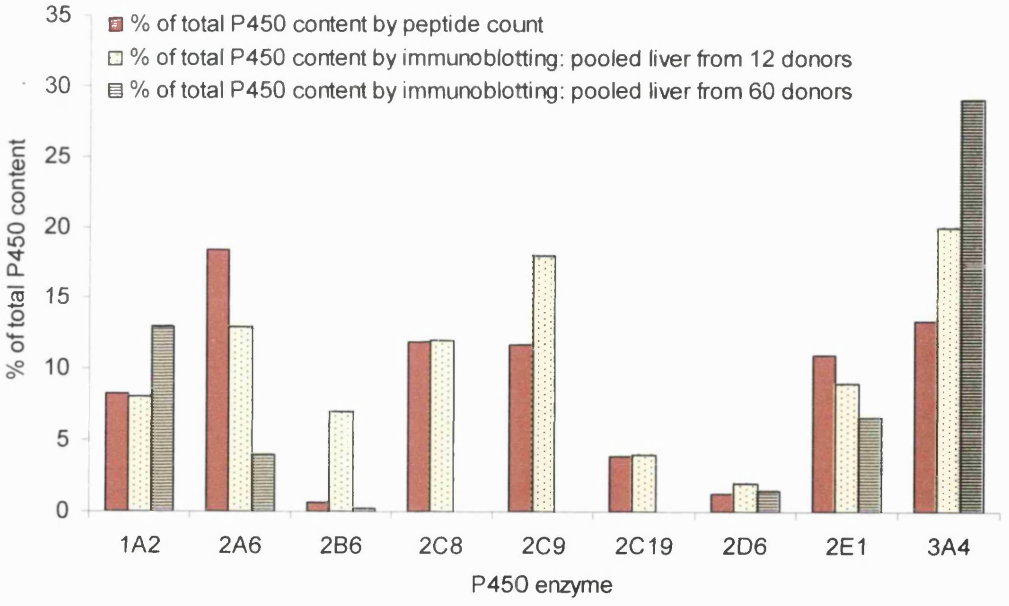


Figure 5.7 The contribution of different P450 enzymes to the total P450 content in liver: comparison with published studies of 12 donors [185] and 60 donors [144]. The percentage of total P450 content by peptide count was calculated using the average number of peptides for each P450 enzyme found in the liver samples (shown in Figure 5.6), taking the total number of P450 peptides identified as 100%.

The P450 enzymes with the lowest percentage contribution from studies by Rodrigues [185] and Shimada *et al* [144] are P450s 2B6, 2C19 and 2D6. These P450 proteins were also those for which the lowest numbers of peptides were identified (excluding P450s from families 4, 8 and 11, which are not accounted for in the

published studies). Rodrigues [185] and Shimada *et al* [144] measured the levels of each P450 in terms of pmol/mg microsomal protein. In the present study CYP2B6 and CYP2D6 were identified in some samples but not in others. Only one 2B6 peptide was identified for each sample in which it was detected, and up to three CYP2D6 peptides. Therefore 2B6 and 2D6 were likely to have been present in concentrations at the very limit of detection for the method. Approximately 25 μg of microsomal protein was loaded onto SDS-PAGE gels for analysis. This should contain 975 fmol (according to Rodrigues [185]) or 25 fmol (according to Shimada *et al* [144]) CYP2B6 and 250 fmol (Rodrigues [185]) or 125 fmol (Shimada *et al* [144]) CYP2D6. Therefore the limit of detection for these two proteins is well under 1 pmol on-gel, i.e. substantially lower than that estimated previously for recombinant CYP1A2 and CYP3A4 (Section 3.8.2).

5.6 Relating the P450s identified to patients' histories

The P450 profile in any one individual will be unique due to a combination of genetic and environmental factors. An effort was made to relate the patients' gender, drug history and lifestyle (Table 5.1) to the P450 proteins identified. Patients 2, 3 and 5 had no recorded drug histories.

Patient 1 was taking amiodarone, fluvastatin, lisoprinil, aspirin and warfarin (Table 5.1). Fluvastatin and warfarin are mainly metabolised by CYP2C9. Amiodarone, a class III antiarrhythmic drug, is mainly metabolised by CYP3A4 and is also a potent inhibitor of CYP1A2, 2C9, 2D6 and 3A4 [196]; the concurrent use of amiodarone and warfarin inhibits metabolism of S-warfarin by CYP2C9, thereby increasing the anti-coagulant effect of warfarin [197]. More 2C9 peptides were identified in liver and tumour from patient 1 than in any other liver or tumour samples. Aspirin has been shown to induce the in-vivo activity of CYP2C19 [198]. CYP2C19 was identified in every patient except patient 1.

Patient 4 was taking amlodipine, thyroxine, isosorbide mononitrate, aspirin and ranitidine (Table 5.1). Amlodipine, a calcium channel blocker, has been shown to

inhibit CYP2B6 activity [199]. It is metabolised by CYP3A4 [200]. CYP2B6 was identified in only three of the patients, including patient 4, by one peptide only.

Patient 6 was taking cyclizine, fluoxetine and warfarin (Table 5.1). Warfarin is metabolised mainly by CYP2C9; fluoxetine, a selective serotonin reuptake inhibitor, has been shown to inhibit CYP2C9, 2C19, 2D6 and 3A4 [201,202]. Patient 6 has the second highest number of detected 2C9 peptides in liver and tumour after patient 1.

Patients 1, 2 and 3 did not drink alcohol, whilst patients 4, 5 and 6 consumed up to 20 units of alcohol a week (Table 5.1). CYP2E1 is induced by chronic alcohol consumption [203], however up to 20 units a week does not constitute "chronic" consumption; indeed, no difference in the number of CYP2E1 peptides is observed in patients 1, 2 and 3 compared with 4, 5 and 6.

Two-fold higher CYP3A4 levels in female compared with male liver samples have been reported [204]. Patients 1 and 5 are male, and 2, 3, 4 and 6 are female (Table 5.1). Yet the highest number of CYP3A4 peptides was detected in male patient 5.

In conclusion, regarding the drug treatment, in some patients, of agents that can act as inducers and inhibitors of P450 subfamily members, there was no evidence that this influenced the P450 expression profiles in the liver or tumours investigated. This could be due to the small sample size coupled with the high number of different drugs taken by those patients for whom histories were recorded. In addition, when relating patients' gender, drug history and lifestyle to the P450 proteins identified one must be wary of over-interpretation of the data because, where a P450 enzyme has been identified in a sample on the basis of one or two peptides only, it is not possible to say whether the apparent absence of this P450 in other samples is real or whether it is due to the limit of detection of the technique employed.

5.7 Western blotting for CYP1B1

Studies have indicated that CYP1B1 is selectively over-expressed in several tumours [148], yet no CYP1B1 was detected in any of the colorectal metastases examined.

Western blotting, which can be more sensitive for the detection of single proteins than nano-LC-ES, was carried out to determine whether the apparent absence of CYP1B1 in the samples was due to the limit of detection of the nano-LC-ES technique.

Western blots (Section 2.8) were prepared for all of the tumour and liver microsomal samples; gels were run in duplicate, with one of each pair being stained with Coomassie blue dye (Section 2.7.1) and the other analysed using a monoclonal antibody to human CYP1B1. Supersomes containing 1 pmol of CYP1B1 were run as positive controls. The Western blots and duplicate gels are shown in Figure 5.8.

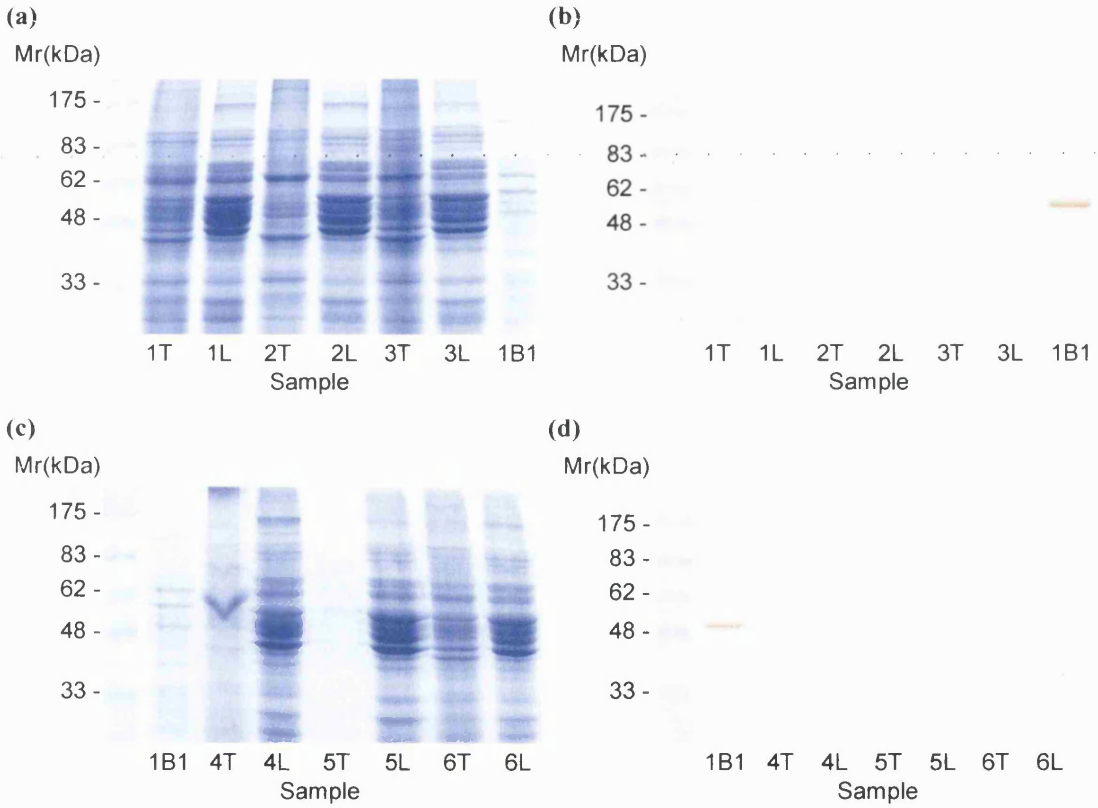


Figure 5.8 Western blotting for CYP1B1. **(a)** SDS-PAGE analysis of tumour and liver samples from patients 1 to 3 with metastatic colorectal cancer of the liver, and **(b)** the corresponding Western blot for CYP1B1. The CYP1B1 positive control is in the furthest right-hand lane. **(c)** SDS-PAGE analysis of tumour and liver samples from patients 4 to 6 with metastatic colorectal cancer of the liver, and **(d)** the corresponding Western blot for CYP1B1. The CYP1B1 positive control is in the lane to the right of the molecular weight markers. Approximately 25 μg of protein was loaded onto the gels for each microsomal sample except sample 5T (for which 2.6 μg was loaded).

There is no discernible CYP1B1 in any of the liver or tumour samples, whilst 1 pmol of CYP1B1 is clearly visible in the positive control lanes. This suggests that

CYP1B1 was not present in the samples, or that, if present, CYP1B1 is at levels well below 1 pmol/25 μg microsomal protein, i.e. 40 pmol/mg microsomal protein.

5.8 Conclusions

The importance of tumour P450 expression in influencing the outcome of chemotherapy through resistance or prodrug activation is poorly understood [148]. Hence there is a need to profile P450 expression in a way that does not rely on anticipating the P450s present. This is the first study that investigates P450 enzymes in human tissue without pre-selection of the proteins to be interrogated. In total, 14 P450 proteins were identified in colorectal metastases of human liver and matched liver samples. The use of the methods developed here for the study of P450s in metastatic tumour and liver samples provides an attractive alternative to traditional methods, offering uniquely the ability to directly detect multiple P450 enzymes simultaneously and without pre-selection. The P450 profile of the tumour samples demonstrates that metastatic cancers in liver potentially have extensive drug-metabolising capabilities, which are likely to be important in determining the metabolic fate of chemotherapeutic agents and hence the outcome of treatment.

Chapter 6:

The use of mechanism-based inactivators for the identification of functional cytochrome P450 proteins

6.1 Introduction

The methods described in this thesis can be used to identify multiple P450s in complex tissues (Chapter 5), but they do not identify the presence of functional P450s, nor are they quantitative. The aim of this study was to use nano-LC-ES as a means to ultimately quantify functional protein through the identification of covalently modified active site peptides.

Mechanism-based inactivators (MBIs) of P450s are substrates that are catalytically transformed in the active site of the enzyme to reactive intermediates that inactivate the enzyme without leaving the active site, leaving the enzyme permanently modified [205]. Mechanism-based inactivation is thought to be relatively unusual in most enzymatic reactions; however, it is observed somewhat more frequently in reactions catalysed by P450s, perhaps due to the reactivity of oxygenated intermediates [206]. Three general classes of P450 MBIs have been identified [207]:

1. Compounds that bind covalently to the apoprotein.
2. Compounds that alkylate or arylate the haem group.
3. Compounds that cause destruction of the haem group, often producing haem-derived products that covalently modify the apoprotein.

Although MBIs have been extensively used to investigate enzyme structure and mechanism [208], their potential for identification and quantification of active P450 isoforms has not been explored. As a first step in achieving this goal, the modification of human CYP2E1 as a consequence of the metabolism of benzyl isothiocyanate (BITC) was investigated.

6.2 Covalent modification of human CYP2E1 by benzyl isothiocyanate

In order to explore the use of MBIs for the identification and quantitation of functional P450 protein, the modification of human CYP2E1 by BITC was

investigated. Previous studies have shown that BITC forms covalent adducts with rabbit CYP2E1 [176,209].

CYP2E1 was incubated with BITC and NADPH-dependent P450 reductase as described in Section 2.5. Control experiments were carried out by incubating CYP2E1 with BITC and P450 reductase in the absence of NADPH.

6.2.1 MS analysis of intact human CYP2E1 after incubation with BITC and P450 reductase

Intact protein from incubations of human CYP2E1 with BITC and P450 reductase was analysed by MALDI-TOF MS and by nano-ES MS with the aim of measuring the mass difference between unmodified and BITC-modified CYP2E1.

MALDI-TOF MS analysis (Section 2.15) was performed on the unprocessed incubations, on the incubations after on-plate washing with 0.1% TFA (Section 2.15.1) and on the incubations after Microcon purification (Section 2.5.1). Mass spectra of NADPH-supplemented and NADPH-absent incubations of CYP2E1 with BITC and P450 reductase, obtained after on-plate washing with 0.1% TFA, are shown in Figure 6.1.

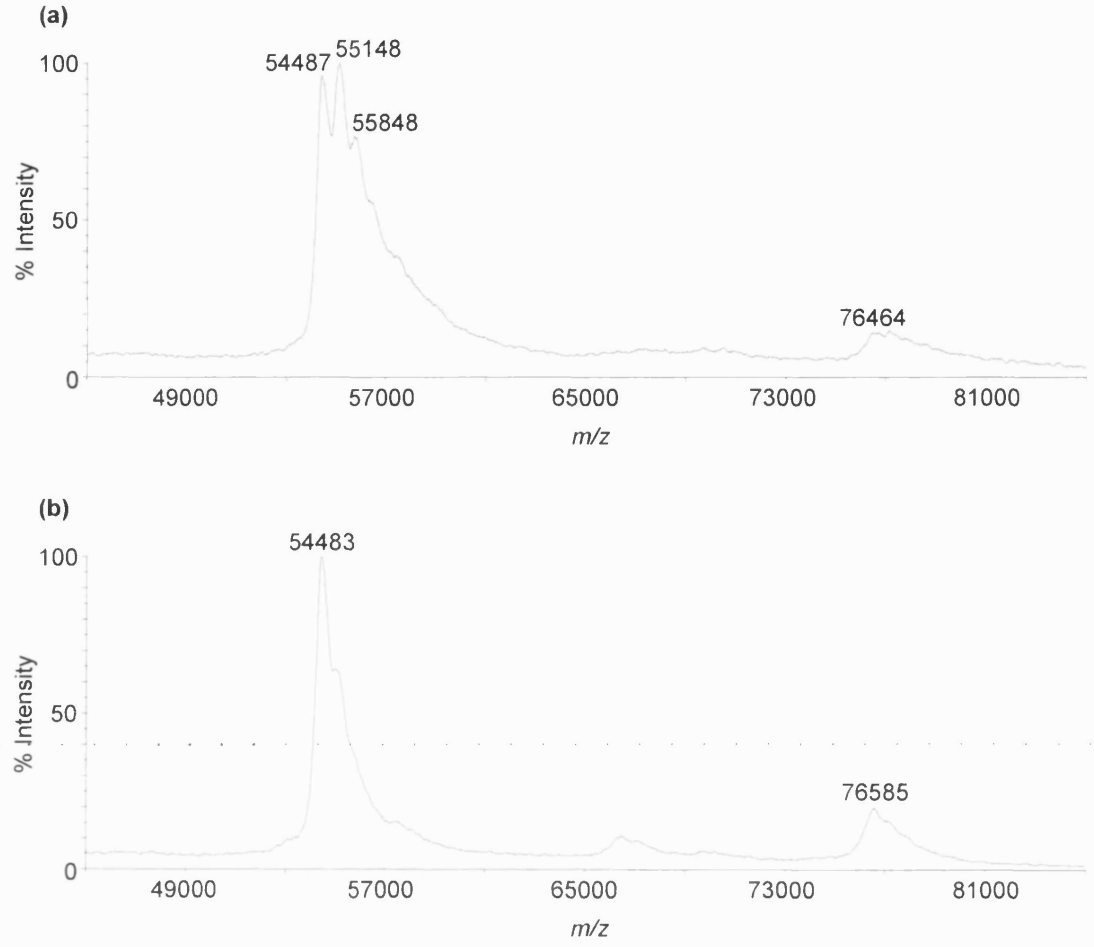


Figure 6.1 MALDI-TOF mass spectra of incubations of CYP2E1 with BITC and P450 reductase, obtained after on-plate washing with 0.1% TFA (Section 2.15.1). **(a)** NADPH-supplemented incubation; **(b)** NADPH-absent incubation. BSA was used for close-external calibration: 12 pmol/μl BSA was mixed with matrix in the ratio 1:1 and spotted onto the plate in positions adjacent to the incubation spots; calibration files were constructed and used for the acquisition of the incubation spectra. Spectra were processed using a baseline correction followed by a 25-point Gaussian smooth.

The CYP2E1 peak in the MS data for the NADPH-supplemented assay (Figure 6.1(a)) is split into 3 peaks at *m/z* 54487, 55148 and 55848. The leftmost peak of the CYP2E1 triplet occurs at approximately the same *m/z* value as the CYP2E1 peak in the data for the NADPH-absent incubation. These mass values correlate well with the value of 54476 Da obtained for the same batch of intact CYP2E1 after analysis by nano-ES (Section 3.2.2.3). Peaks at *m/z* 76464 in the data from the NADPH-supplemented incubation and *m/z* 76585 in the data from the NADPH-absent incubation are from P450 reductase.

Spectral peaks (Figure 6.1) are too broad to resolve a mass difference of 149 Da (M_r of BITC), which would be expected after adduction of one molecule of BITC to the

protein. The mass differences between the CYP2E1 triplet peaks in the MS data for the NADPH-supplemented incubation (Figure 6.1(a)) are consistent with the non-specific association of NADPH (M_r 745 Da).

Nano-ES analysis of intact protein after Microcon-purification of incubations of CYP2E1 with BITC proved difficult. This, and the poor quality of the MALDI-TOF MS data from Microcon-purified incubations, indicates that the majority of the protein was retained on the Microcon sample reservoir membrane.

Since the analysis of intact modified proteins yielded little information of value in adduct identification, alternative methods for the detection of covalent modifications were pursued.

6.2.2 Analysis of incubations of CYP2E1 with BITC and P450 reductase after in-solution tryptic digestion

Incubations of CYP2E1 with BITC and P450 reductase were subjected to in-solution tryptic digestion (Section 2.9) and analysed by nano-LC-ES (Section 2.12.2), with Sequest searching for peptide identification (Section 2.14.2). Up to 57% of CYP2E1 (by amino acid count) was identified (all Xcorr scores).

In order for the Sequest search program to identify covalently modified peptides, specific masses for the modification(s) must be entered during the search process. Up to three amino acid modifications can be entered; Sequest then allows for these modifications whilst searching. One must therefore have a good idea of the mass of the expected modification. MBIs are catalytically transformed in the active site of the enzyme, leaving it permanently modified. It is therefore reasonable to hypothesize that modification will occur at an amino acid residue within the active site. The active site of human CYP2E1 has not been fully characterised; however, the threonine residue at position 303 of CYP2E1 corresponds to the conserved threonine identified in several other P450 enzymes [176,210], suggesting that it could be important for P450 catalysis. The human CYP2E1 tryptic peptide containing this residue, ^{283}LYTMDGITVTVADLFFAGTETTSTTLR^{309}, has previously been identified after

nano-ES and nano-LC-ES analysis of tryptic digests of CYP2E1. Both the $[M + 2H]^{2+}$ (Figure 6.2) and $[M + 3H]^{3+}$ ions of the peptide ^{283}LYTMDGITVTVADLFFAGTETTSTTLR^{309} were observed in initial nano-LC-ES analyses of tryptic digests of incubations of CYP2E1 with BITC and P450 reductase, eluting at approximately 40 minutes.

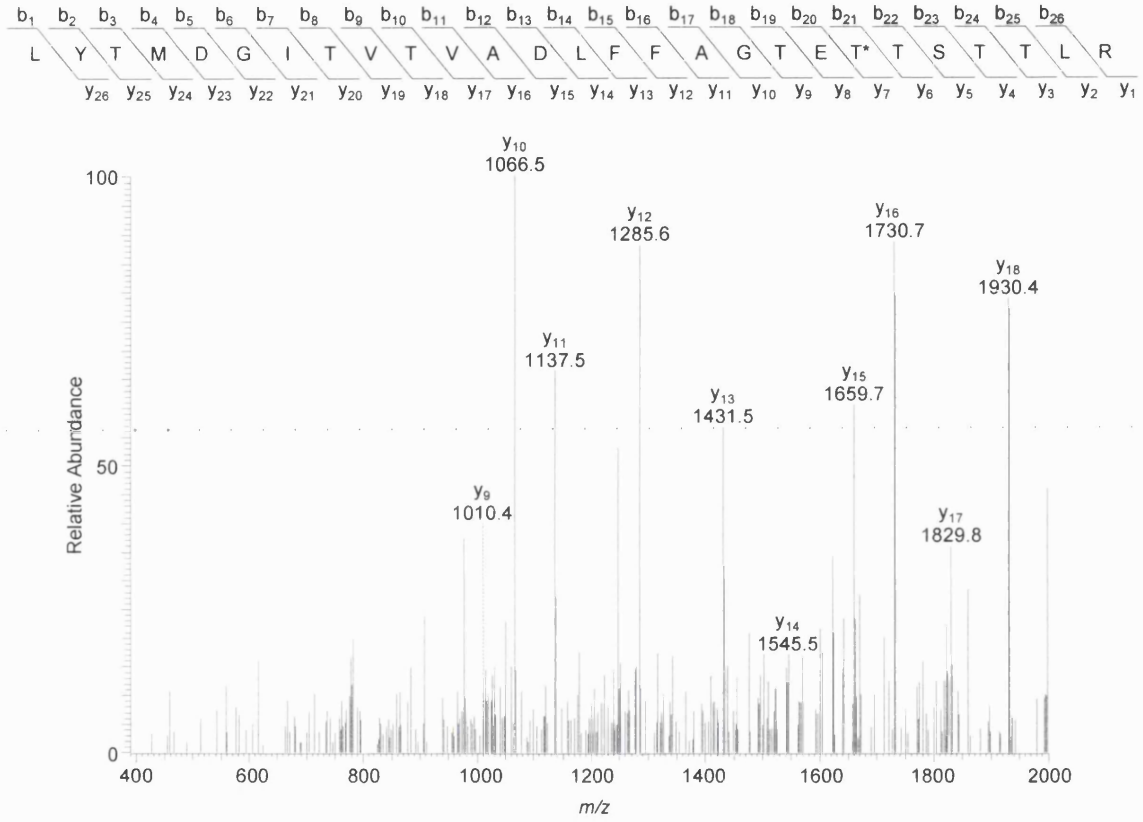


Figure 6.2 MS/MS spectrum of the CYP2E1 tryptic peptide LYTMDGITVTVADLFFAGTETTSTTLR $[M + 2H]^{2+}$ ion of *m/z* 1463.5, identified after in-solution tryptic digestion of NADPH-supplemented incubation of CYP2E1 with BITC and P450 reductase. Digest was diluted 1 in 2 with 0.1% TFA (such that CYP2E1 digest was at a concentration of 500 fmol/µl) and analysed by nano-LC-ES (Section 2.12.2): 1 µl was injected onto the nano LC via a 10 µl sample loop. The retention time of the ion was 39.8 minutes. Data was searched using Sequest (Section 2.14.2) against the Finnigan Xcalibur human database (Section 2.14.1).

6.2.2.1 Reaction scheme for the modification of human CYP2E1 by BITC

Modification of Thr303 of CYP2E1 by BITC could occur in a number of different ways. Possible reactive intermediates include benzyl isocyanate (BIC) and α-hydroxy BIC (Figure 6.3).

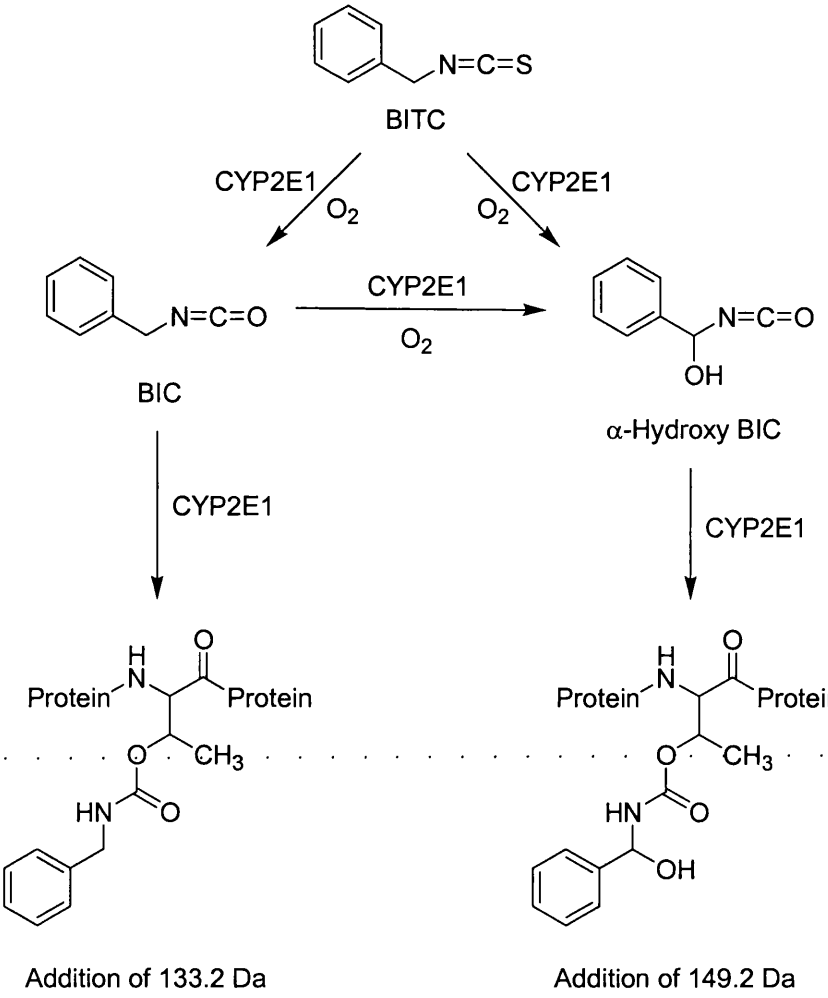


Figure 6.3 Possible mechanisms for the modification of CYP2E1 by BITC, showing adduction to a CYP2E1 threonine residue.

The involvement of a BIC reactive intermediate in the modification of CYP2E1 would result in an increase of 133.2 Da to the average mass of the active site peptide (Figure 6.3). The parent $[M+H]^+$ ion of the modified peptide would have a mass of 3059.5 Da (compared to the unmodified $[M+H]^+$ mass of 2926.3 Da); whilst the $[M+2H]^{2+}$ and $[M+3H]^{3+}$ ions would have *m/z* values of 1530.2 and 1020.5, respectively. Modification via an α-hydroxy BIC reactive intermediate would result in an increase of 149.2 Da to the average mass of the active site peptide (Figure 6.3). The parent $[M+H]^+$ ion of the modified peptide would have a mass of 3075.5 Da; the $[M+2H]^{2+}$ and $[M+3H]^{3+}$ ions would have *m/z* values of 1538.3 and 1025.8, respectively.

Modifications of 133.15 Da and 149.22 Da, corresponding to adduction of CYP2E1 by BIC and α-hydroxy BIC, were entered into Sequest and the nano-LC-ES data

from tryptic digests of the incubations of CYP2E1 with BITC and P450 reductase were searched, but no modifications were identified.

6.2.2.2 Identification of a BITC-modified CYP2E1 tryptic peptide after nano-LC-ES analysis

To facilitate the detection of BITC-modified peptides a new nano-LC-ES instrument method was implemented, with alterations to that previously described in Section 2.12.2: for the LCQ data dependent MS/MS scans, the parent *m/z* values of 1463.5, 1530.2 and 1538.3 (corresponding to the $[M+2H]^{2+}$ ions for the unmodified, BIC-modified and α-hydroxy BIC-modified peptides, respectively) were entered. The first MS/MS scan after a full scan was set to fragment the most intense ion from the parent *m/z* list; the second MS/MS scan after the full scan was set to fragment the 2nd most intense ion from the parent *m/z* list, and the third MS/MS scan was set to fragment the 3rd most intense ion from the parent *m/z* list. If no parent ions matching those on the list were detected, then the instrument was set to simply fragment the 1st, 2nd and 3rd most intense ions present (as usual). Ions with *m/z* values on the parent ion list would therefore be selected for MS/MS analysis even if they were not among the most intense ions in the spectrum. In addition, an extended nano-LC gradient (Figure 6.4), which ran to a higher percentage organic phase, was employed so that peptides with increased hydrophobicity as a consequence of covalent adduction with small lipophilic molecules were likely to be eluted from the column before the wash phase of the gradient program.

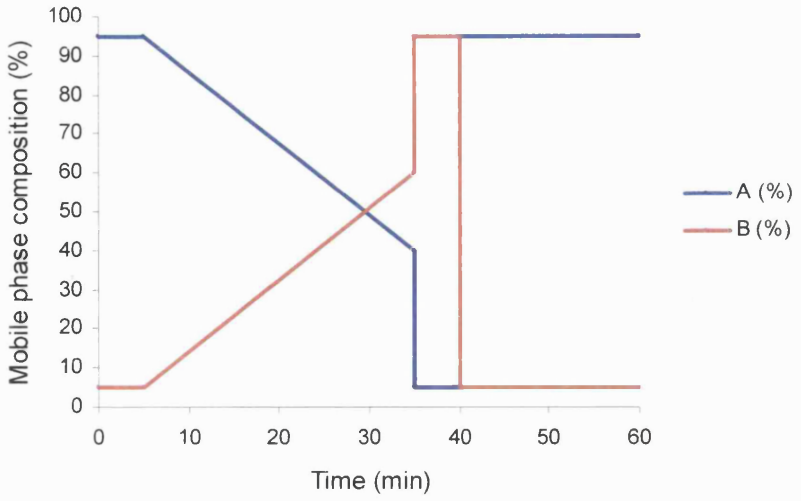


Figure 6.4 The nano-LC gradient program used for the analysis of tryptic digests of incubations of CYP2E1 with BITC and P450 reductase by nano-LC-ES. Mobile phase A: 0.1% formic acid, 5% acetonitrile in water; mobile phase B: 0.1% formic acid, 80% acetonitrile in water. Flow rate 200 nl/min.

In-solution digests of NAPDH-supplemented and NADPH-absent incubations of CYP2E1 with BITC and P450 reductase were analysed using the nano-LC-ES program described above. When searching data with Sequest two modifications were entered, 133.15 Da and 149.22 Da, corresponding to adduction of CYP2E1 by BIC and α-hydroxy BIC (Figure 6.3).

The peptide ^{283}LYTMDGITVTVADLFFAGTETTSTTLR^{309}, modified by a mass of 149.2 Da, was identified from the MS/MS data from the NADPH-supplemented incubation digest (Figure 6.5(a)) with Sequest Xcorr scores of 2.35 and 1.56. Possible sites of modification were Thr301, Thr303, Thr304, Ser305, Thr306 or Thr307. The unmodified peptide was identified with Xcorr scores ≥ 2.5 from the MS/MS data of both the NADPH-supplemented and the NADPH-absent digests of CYP2E1 with BITC and P450 reductase (Figure 6.5(b)). No BITC-modified CYP2E1 peptides were identified from the NADPH-absent incubation digest; no other modified peptides were identified by Sequest.

A modification of 149.2 Da (parent $[M + H]^+$ mass 3075.5 Da) corresponds to the mass of the α-hydroxy BIC modified peptide. The Sequest identification of the modified peptide and possible sites of modification were confirmed from the MS/MS data manually. Data was also checked for evidence of a BIC-modified peptide

(Parent $[M + H]^+$ mass 3059.5 Da); there was no MS/MS data to support this scenario.

An increase in retention time was observed for the BITC-modified CYP2E1 peptide (retention time 42 minutes) in comparison with the unmodified CYP2E1 peptide (retention time 37 minutes). This can be explained by the increased hydrophobicity of the modified peptide due to covalent adduction by α-hydroxy BIC.

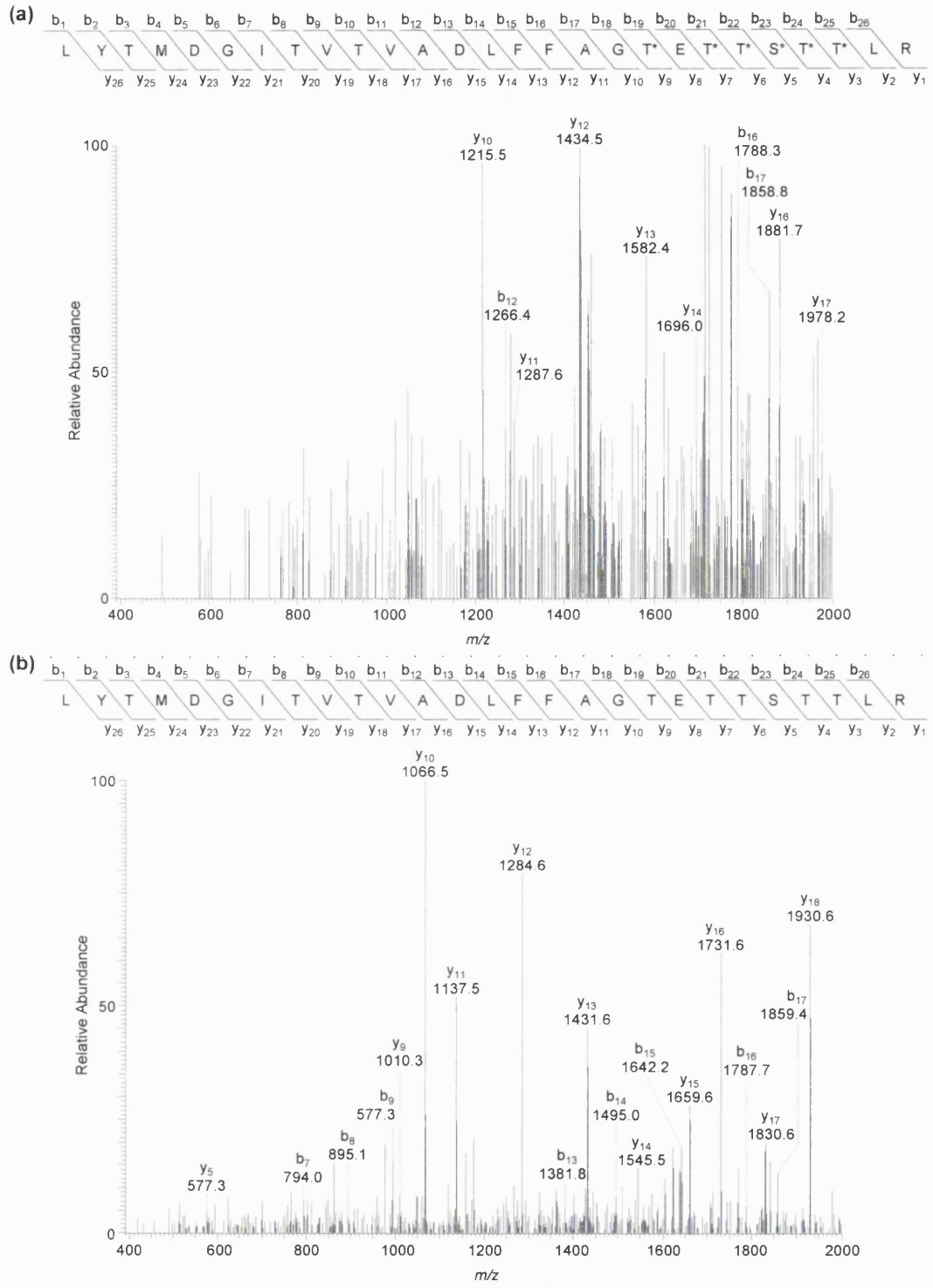


Figure 6.5 MS/MS spectra from analysis of a digest of an incubation of CYP2E1 with BITC and P450 reductase, NADPH-supplemented. **(a)** MS/MS spectrum of the BITC-modified CYP2E1 peptide LYTMDGITVTVADLFFAGTETTSTTLR $[M+2H]^{2+}$ ion of *m/z* 1538.3, retention time 42.2 minutes. Possible sites of BITC modification are marked with asterisks. The spectrum is an average of 5 scans. **(b)** MS/MS spectrum of the unmodified CYP2E1 peptide LYTMDGITVTVADLFFAGTETTSTTLR $[M+2H]^{2+}$ ion of *m/z* 1463.5, retention time 36.8 minutes. The spectrum is an average of 4 scans. Samples were acidified to contain 0.1% TFA such that CYP2E1 digest was at a concentration of 900 fmol/μl, and analysed by nano-LC-ES: 1 μl was injected via a 10 μl sample loop. Data files were searched using Sequest against the Finnigan Xcalibur human database (Section 2.14.1). Sequest searches allowed for modifications of 149.22 Da (α-hydroxy BIC) and 133.15 Da (BIC).

6.2.3 Discussion

The BITC modification of functional CYP2E1 has been demonstrated by nano-LC-ES analysis of the tryptic digest. This is the first reported MS/MS characterisation of a modified P450 active site peptide.

The results presented are consistent with those of Moreno *et al* [176,209], who studied the adduction of BITC with rabbit CYP2E1. Rabbit CYP2E1 shares 80% sequence homology with human CYP2E1 (Figure 6.6).

```
h MSALGVTVALLVWAAFLLLVSMWRQVHSSWNLPPGPFPLPIIGNLFQLELKNIPKSFTRL 60
r MAVLGITVALLGWMVILLFISVWKQIHSSWNLPPGPFPLPIIGNLLQLDLKDIPKSFGRL 60
   **  *     * ***  ** * * *                    *  *  *     *

h AQRFGPVFTLYVGSQRMVVMHGYKAVKEALLDYKDEFSGRGDLPAFHAHRDRGIIFNNGP 120
r AERFGPVFTVYLGSRRVVVLHGYKAVREMLLNHKNEFSGRGEIPAFREFKDKGIIFNNGP 120
   *       * *  * *  *      * *  ** *      **   **** *

h TWKDIRRFSLTTLRNYGMGKQGNESRIQREAHFLLEALRKTQGQPFDPTFLIGCAPCNVI 180
r TWKDTRRFSLTTLRDYGMGKQGNEDRIQKEAHFLLEELRKTQGQPFDPTFVIGCTPFNVI 180
     *         *         *   *       *             *   * *

h ADILFRKHFDYNDEKFLRLMYLFNENFHLLSTPWLQLYNNFPSFLHYLPGSHRKVIKNVA 240
r AKILFNDRFDYKDKQALRLMSLFNENFYLLSTPWLQVYNNFSNYLQYMPGSHRKVIKNVS 240
   *   ***   * ***    *      *        *    *** * *           *

h EVKEYVSERVKEHHQSLDPNCPRDLTDCLLVEMEKEKHSAERLYTMDGITVTVADLFFAG 300
r EIKEYTLARVKEHHKSLDPSCPRDFIDSLLIEMEKDKHSTEPLYTLENIAVTVADMFFAG 300
   *   ***      *    *    ** *  *    *   *     *** *     *

h TETTSTTLRYGLLILMKYPEIEEKLHEEIDRVIGPSRIPAIKDRQEMPYMDAVVHEIQRF 360
r TETTSTTLRYGLLILLKHPEIEEKLHEEIDRVIGPSRMPSVRDRVQMPYMDAVVHEIQRF 360
                 * *                   * ***  **

h ITLVPSNLPHEATRDTIFRGYLIPKGTVVVPTLDSVLYDNQEFPDPEKFKPEHFLNENGK 420
r IDLVPSNLPHEATRDTTFQGYVIPKGTVVIPTLDSLLYDKQEFPDPEKFKPEHFLNEEGK 420
   *                 *  *       *     *   *                   *

h FKYSDYFKPFSTGKRVCAGEGLARMELFLLLCAILQHFNLKPLVDPKDIDLSPIHIGFGC 480
r FKYSDYFKPFSAGKRVCVGEGLARMELFLLLSAILQHFNLKPLVDPEDIDLRNITVGFGR 480
             *     *             *               *    ** **   *

h IPPRYKLCVIPRS
r VPPRYKLCVIPRS
  *

h CYP2E1 human
r CYP2E1 rabbit
```

Figure 6.6 Sequence alignment for human and rabbit CYP2E1. The sequence of the tryptic peptide modified after reaction with BITC is underlined. Sequence differences are marked with asterisks.

Moreno *et al* [176] obtained intact protein ES data that indicated that a mass consistent with one molecule of BIC and oxygen was adducted to the enzyme. Work with a

rabbit CYP2E1 T303A mutant, in which the conserved threonine residue at position 303 was replaced with an alanine residue, supported the theory that the site of modification is Thr303: unlike for the wild-type enzyme, BITC did not inactivate the T303A mutant, but rather inhibited it in a competitive manner [176]. It has been proposed [210] that the conserved threonine residue is involved in a proton delivery network to the active site of P450 enzymes, with an important role in oxygen activation. The detection of a modified active site peptide in the present study supports the involvement of this amino acid in substrate turnover. Future work will involve the determination of the exact site of modification.

6.2.3.1 Development of a method to relatively quantify functional P450s in tissues

The results of the present study represent the first step towards the quantification of functional P450 protein. The study demonstrates that functional P450s can be identified by selective substrate modification of active site peptides. Further work is necessary to show that selected functional P450s within tissues containing multiple P450s could be detected. In addition, the use of several P450-modifying substrates used in combination should allow detection of functionality in multiple P450s. The ability to relatively quantify the level of functional P450s could be developed by the incorporation of stable (heavy) isotopes into the MBI structure (Figure 6.7(a)). Microsomes from one source would be incubated with natural isotopic MBI, whilst microsomes from a second source would be incubated with heavy isotope-labelled MBI. The two samples would be combined, separated by SDS-PAGE, subjected to in-gel tryptic digestion and analysed by nano-LC-ES. The heavy and light MBI-modified peptides would be expected to co-elute, and hence a measure of their relative ion currents would give a measure of the relative abundance of the parent functional proteins (Figure 6.7(b)).

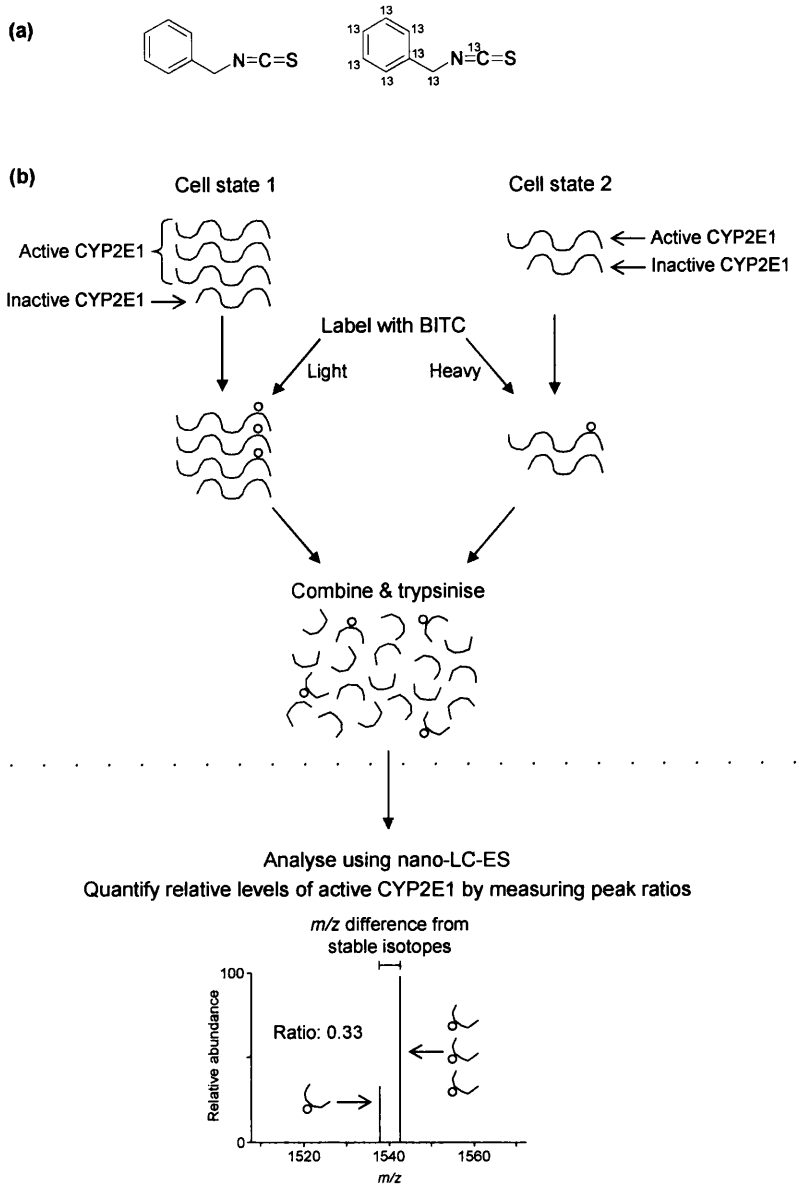
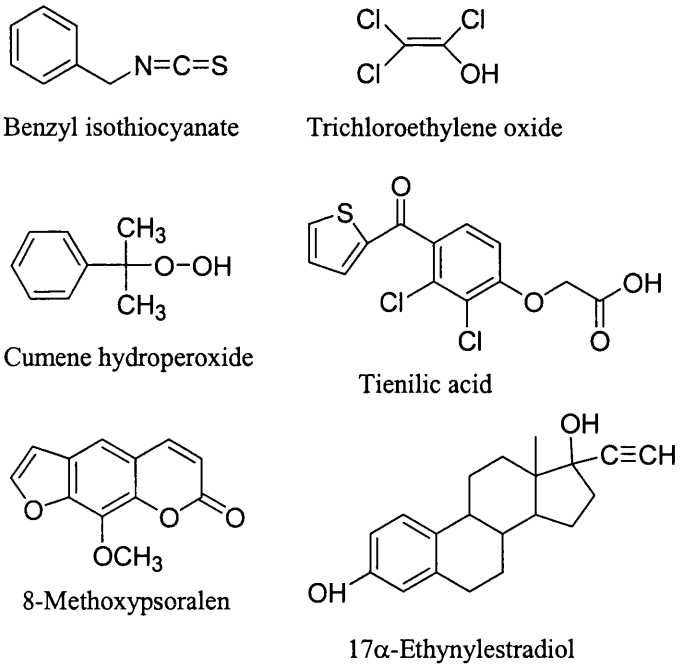


Figure 6.7 Proposed strategy for the quantification of functional CYP2E1 using BITC. **(a)** Structure of BITC and heavy isotope-labelled BITC; **(b)** schematic of the quantification strategy.

MBIs are generally specific to one or two P450s [208], therefore a range of MBIs will be required for the quantification of multiple functional P450 enzymes. It is possible that a different MBI will be needed for each P450 to be interrogated. A literature survey identified several candidate MBIs; the structures of some of these are shown in Figure 6.8.



Benzyl isothiocyanate

Trichloroethylene oxide

Cumene hydroperoxide

Tienilic acid

8-Methoxypsoralen

17α-Ethynylestradiol

Figure 6.8 Structures of candidate MBIs for use in the quantification of functional P450s.

As well as forming adducts with CYP2E1, BITC has been to shown to covalently modify human CYP2B6 and rat CYP2B1 [211,212]. Trichloroethylene oxide inactivates human CYP2E1 [213]; cumene hydroperoxide has been shown to form adducts with human CYP3A4 [214] and rat CYP2B1 [215]; tienilic acid modifies human CYP2C9 [216]; 8-methoxypsoralen inactivates human CYP2A6 and rat CYP2B1 [217-219]; and 17α-ethynylestradiol modifies human CYP2B6 and CYP3A4 and rat CYP2B1 [220,221]. Future work will involve the evaluation of these compounds for use in the identification and quantification of functional P450 enzymes.

Chapter 7

Summary and general discussion

Summary and general discussion

The last two decades have seen the emergence of the interdisciplinary field of proteomics, which has come to encompass the identification, characterisation and quantification of the complete set of proteins expressed by the entire genome in the lifetime of a given cell, tissue or organism. Mass spectrometry is the central pillar of proteomics, with an extensive range of existing and potential applications. In this study, mass spectrometry has been applied to the analysis of cytochrome P450 proteins, a superfamily of mixed function oxidases with a central role in the metabolism of a large number of drugs, xenobiotics and endogenous compounds.

Chapter 3 describes the optimisation and assessment of methods for the analysis of P450s. These methods were applied to the identification of P450s from various biological matrices (Chapter 4), culminating in a study in which the P450 expression profile was determined for six sets of colorectal liver metastases and corresponding liver samples from patients with metastatic colorectal cancer of the liver (Chapter 5). Fourteen distinct P450 enzymes were identified, including all the major drug-metabolising P450s and also some forms involved in bile acid synthesis and fatty acid metabolism. The P450 profiles of the tumours demonstrate that metastatic cancers in liver potentially have extensive drug-metabolising capabilities, which are likely to be important in determining the outcome of treatment. The nano-LC-ES approach adopted offers an attractive alternative to the techniques traditionally used for the detection of P450s because it has the ability to directly identify multiple proteins without pre-selection of the proteins anticipated. This is the first study in which P450 enzymes were investigated in human tissue without pre-selection. It is also the first demonstration of multiple P450 expression in colorectal metastases of human liver. Galeva *et al* [163] applied (1D) SDS PAGE to the separation of proteins from clofibrate-induced rat and phenobarbital-induced rabbit liver microsomes followed by in-gel digestion and peptide identification by MALDI-TOF PMF; up to eight P450s were positively identified. A recent report [222] described the identification of a novel rat microsomal vitamin D_3 hydroxylase, CYP2J3. Extensive purification was carried out to isolate vitamin D_3 hydroxylase from rat liver microsomes, followed by (1D) SDS-PAGE, in-gel tryptic digestion and MALDI-TOF PMF.

MALDI-TOF PMF is most reliable when samples containing only a few proteins are analysed; the presence of multiple proteins within a SDS-PAGE band may lead to spurious results. This is particularly the case when proteins are present in low abundance and their tryptic peptide sequence coverage is low [44,223,224]. It is therefore well suited for use with 2DE; however, the low performance of 2DE in the separation of membrane proteins, including P450s [225], precludes its use for P450 analysis. Very recently, Kislinger *et al* [72] applied the MudPIT technique to the analysis of liver microsomes from female mice and were able to detect 20 P450 enzymes. The MudPIT method [50] does not require pre-fractionation of the protein sample, but rather relies on multi-dimensional chromatographic separation of very complex mixtures of peptides generated by proteolysis of unseparated protein mixtures. The MudPIT approach may hold advantages for the unbiased analysis of protein mixtures; however, huge amounts of data are generated and one therefore runs the risk of information overload. For the work described in this thesis analysis was targeted to the identification of P450 proteins. In this regard, SDS-PAGE was used to separate microsomal proteins into discrete bands according to molecular weight, allowing specific selection of proteins in the P450 molecular weight range. The methods developed here have been applied to the identification of P450s in male and female rat liver [226]: twenty-four distinct P450 enzymes were identified, including gender-specific forms CYP2C11, 2C12 and 2C13.

By comparison with published studies it was found that, to a first approximation, the number of peptides identified in liver samples from a given P450 form correlated with protein abundance (Chapter 5). Since the primary goal of the study was the identification of P450 proteins, it was decided not to pursue a stable-isotope dilution approach, such as the ICAT method [73], for relative quantification. The ICAT approach requires dilution of one sample by another and the analysis of cysteine-containing peptides only; such a procedure would inherently limit differentiation between P450 enzymes with high sequence similarity and would also be likely to elevate the P450 protein detection limit, reducing the probability of detecting less abundant P450 forms. However, since the study described in Chapter 5 has identified extensive P450 expression in both the tumour and liver samples, future work will pursue fully quantitative methods of analysis. A particularly attractive approach has recently been described by Gerber *et al* [92], which involves the use of synthetic

peptides with incorporated stable isotopes: one or more tryptic peptides specific to the protein of interest are synthesised using isotopically enriched lysine. Samples are separated by 1D gel electrophoresis, and the bands of interest excised and subjected to in-gel tryptic digestion in the presence of known amounts of the labelled peptide(s). Absolute quantitation is then achieved using LC-MS/MS analysis. There is also the potential for whole proteins to be expressed using only ^{15}N or ^{13}C-containing amino acids; these proteins could be introduced to samples prior to SDS-PAGE separation and be carried through the entire procedure, thereby eliminating any differences between sample and standard due to incomplete trypsinisation.

A limitation of many proteomic approaches is the inability to distinguish between functional and inactive forms of a protein. This is particularly relevant in the field of cancer research where the role of tumour P450s in drug metabolism has had little attention. Determination of the functional activity of P450s in tumours would have wide reaching implications for therapeutic outcome and drug discovery, including the development of tumour-selective prodrugs. Chapter 6 details the preliminary stages of a study that aims to utilise MBIs for the identification and quantification of functional P450 proteins. Work with BITC and CYP2E1 led to the identification and characterisation of a BITC-modified CYP2E1 catalytic site peptide. Previously, intact modified CYP2B1 [217], 2C9 [227,228], 2E1 [176,229] and 3A4 [228] have been observed, along with their covalent adducting metabolites, by LC-ES-MS. MALDI-TOF MS analysis has been described of peptides generated by cyanogen bromide cleavage of covalently modified CYP3A4 [214,230] and trypsin proteolysis of modified CYP2E1 [213]. However, Chapter 6 describes the first characterisation of a modified P450 active site peptide by MS/MS. The site of modification was determined by MS/MS to be one of six C-terminal threonine and serine residues. Future work will involve the determination of the exact site of modification, the incorporation of stable isotopes into the BITC structure for relative quantification of active CYP2E1, and the evaluation of other MBIs for use in the identification and quantification of functional P450s.

In conclusion, the use of mass spectrometry for the analysis of cytochrome P450 proteins has proved to be a very powerful technique. However, we are only at the tip of the iceberg: MS-based proteomics is still very much a nascent technology, where

radical change is possible. Advances in instrumentation, experimental design and data handling could revolutionise current practices and enable the analysis of samples in ways that, at present, we can only dream of.

References

1. Wilkins M. R., Pasquali C., Appel R. D., Ou K., Golaz O., Sanchez J. C., Yan J. X., Gooley A. A., Hughes G., Humphery-Smith I., Williams K. L., Hochstrasser D. F. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology* **14**, 61-65 (1996).

2. Tyers M., Mann M. From genomics to proteomics. *Nature* **422**, 193-197 (2003).

3. Pearson H. Geneticists play the numbers game in vain. *Nature* **423**, 576 (2003).

4. Speicher D. W. Current challenges in proteomics: mining low abundance proteins and expanding protein profiling capacities. *16th International Mass Spectrometry Conference, Edinburgh, 31 August-05 September 2003.*

5. Thomson J. J. Rays of positive electricity and their applications to chemical analysis. Longmans Green, London (1913).

6. Aston F. W. Mass spectra and isotopes. Edward Arnold, London (1933).

7. Griffiths W. J., Jonsson A. P., Liu S., Rai D. K., Wang Y. Electrospray and tandem mass spectrometry in biochemistry. *Biochem. J.* **355**, 545-561 (2001).

8. Yamashita M., Fenn J. B. Electrospray ion source. Another variation on the free-jet theme. *J. Phys. Chem.* **88**, 4451-4459 (1984).

9. Yamashita M., Fenn J. B. Negative ion production with the electrospray ion source. *J. Phys. Chem.* **88**, 4671-4675 (1984).

10. Meng C. K., Mann M., Fenn J. B. Electrospray ionization of some polypeptides and small proteins. *Proceedings of the 36th American Society for Mass Spectrometry Conference on Mass Spectrometry and Allied Topics, San Francisco, CA, June 1988.*

11. Fenn J. B., Mann M., Meng C. K., Wong S. F., Whitehouse C. M. Electrospray ionization for mass spectrometry of large molecules. *Science* **246**, 64-71 (1989).

12. Karas M., Hillenkamp F. Ultraviolet laser desorption of ions above 10 kDa. *Abstracts of the 11th International Mass Spectrometry Conference, Bordeaux, abstract MOA OR, 29 August-2 September 1988.*

13. Karas M., Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10000 Da. *Anal. Chem.* **60**, 1299-2301 (1988).

14. Tanaka K., Waki H., Ido Y., Akita S., Yoshida Y., Yoshida T. Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**, 151-153 (1988).

15. Klose J. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **26**, 231-243 (1975).

16. O'Farrell P. H. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**, 4007-4021 (1975).

17. Dole M., Mach L. L., Hines R. L., Mobley R. C., Ferguson L. D., Alice M. B. Molecular beams of macroions. *J. Chem. Phys.* **49**, 2240-2247 (1968).

18. Gaskell S. J. Electrospray: principles and practice. *J. Mass Spectrom.* **32**, 677-688 (1997).

19. Willoughby R., Sheehan E., Mitrovich S. A global view of LC/MS: how to solve your most challenging analytical problems. Global View Publishing (1998).

20. Thompson B. A., Iribarne J. V. Field induced ion evaporation from liquid surfaces at atmospheric pressure. *J. Chem. Phys.* **71**, 4451-4463 (1979).

21. Iribarne J. V., Thompson B. A. On the evaporation of small ions from charged droplets. *J. Chem. Phys.* **64**, 2287-2294 (1976).

22. Gaskell S. J. Electrospray: principles and practise. *J. Mass Spectrom.* **32**, 677-688 (1997).

23. de Hoffmann E., Stroobant V. Mass Spectrometry: principles and applications. Wiley (2003).

24. Hager J. W. A new linear ion trap mass spectrometer. *Rapid Commun. Mass Spectrom.* **16**, 512-526 (2002).

25. Amster I. J. Fourier transform mass spectrometry. *J. Mass Spectrom.* **31**, 1325-1337 (1996).

26. Aebersold R., Mann M. Mass spectrometry-based proteomics. *Nature* **422**, 198-207 (2003).

27. Krutchinsky A. N., Kalkum M., Chait B. T. Automatic identification of proteins with a MALDI-quadrupole ion trap mass spectrometer. *Anal. Chem.* **73**, 5066-5077 (2001).

28. Medzihradszky K. F., Campbell J. M., Baldwin M. A., Falick A. M., Juhasz P., Vestal M. L., Burlingame A. L. The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal. Chem.* **72**, 552-558 (2000).

29. Loboda A. V., Krutchinsky A. N., Bromirski M., Ens W., Standing K. G. A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: design and performance. *Rapid Commun. Mass Spectrom.* **14** 1047-1057 (2000).

30. Wysocki V. H., Tsaprailis G, Smith L. L., Breci L. A. Mobile and localized protons: a framework for understanding peptide dissociation. *J. Mass Spectrom.* **35**, 1399-1406 (2000).

31. Dongre A. R., Jones J. L., Somogyi A., Wysocki V. H. Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: evidence for the mobile proton model. *J. Am. Chem. Soc.* **118**, 8365-8374 (1996).

32. Tang X. J., Boyd R. K. An investigation of fragmentation mechanisms of doubly protonated tryptic peptides. *Rapid Commun. Mass Spectrom.* **6**, 651-657 (1992).

33. Hunt D. F., Yates J. R. III, Shabanowitz J., Winston S., Hauer C. R. Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 6233-6237 (1986).

34. Loo J. A., Edmonds C. G., Smith R. D. Tandem mass spectrometry of very large molecules. 2. Dissociation of multiply charged proline-containing proteins from electrospray ionization. *Anal. Chem.* **65**, 425-438 (1993).

35. Vaisar T., Urban J. Probing the proline effect in CID of protonated peptides. *J. Mass Spectrom.* **31**, 1185-1187 (1996).

36. Breci L. A., Tabb D. L., Yates III J. R., Wysocki V. H. Cleavage N-terminal to proline: analysis of a database of peptide tandem mass spectra. *Anal. Chem.* **75**, 1963-1971 (2003).

37. Gygi S. P., Corthals G. L., Zhang Y., Rochon Y., Aebersold R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 9390-9395 (2000).

38. Gygi S. P., Rist B., Aebersold R. Measuring gene expression by quantitative proteome analysis. *Curr. Opin. Biotechnol.* **11**, 396-401 (2000).

39. Naaby-Hansen S., Waterfield M. D., Cramer R. Proteomics - post-genomic cartography to understand gene function. *Trends Pharmacol. Sci.* **22**, 376-384 (2001).

40. Fey S. J., Larsen P. M. 2D or nor 2D. *Curr. Opin. Chem. Biol.* **5**, 26-33 (2001).

41. Gauss C., Kalkum M., Lowe M., Lehrach H., Klose J. Analysis of the mouse proteome. (I) Brain proteins: separation by two-dimensional electrophoresis and identification by mass spectrometry and genetic variation. *Electrophoresis* **20**, 575-600 (1999).

42. Rabilloud T. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics* **2**, 3-10 (2002).

43. Unlu M., Morgan M. E., Minden J. S. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**, 2071-2077 (1997).

44. Clauser K. R., Baker P., Burlingame A. L. Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal. Chem.* **71**, 2871-2882 (1999).

45. Zhan Q., Gusev A., Hercules D. M. A novel interface for on-line coupling of liquid capillary chromatography with matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **13**, 2278-2283 (1999).

46. Gelpi E. Interfaces for coupled liquid-phase separation/mass spectrometry techniques. An update on recent developments. *J. Mass Spectrom.* **37**, 241-253 (2002).

47. Ericson C., Phung Q. T., Horn D. M., Peters E. C., Fitchett J. R., Ficarro S. B., Salomon A. R., Brill L. M., Brock A. An automated noncontact deposition surface for liquid chromatography matrix-assisted laser desorption/ionization mass spectrometry. *Anal. Chem.* **75**, 2309-2315 (2003).

48. Zhang B., McDonald C., Li L. Combining liquid chromatography with MALDI mass spectrometry using a heated droplet interface. *Anal. Chem.* **76**, 992-1001 (2004).

49. Washburn M. P., Wolters D., Yates III J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242-247 (2001).

50. Wolters D. A., Washburn M. P., Yates III J. R. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **73**, 5683-5690 (2001).

51. Nesvizhskii A. I., Aebersold R. Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem MS. *Drug Discov. Today* **9**, 173-181 (2004).

52. Eng J. K., McCormack A. L., Yates III J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976-989 (1994).

53. Perkins D. N., Pappin D. J., Creasy D. M., Cottrell J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).

54. Mann M., Wilm M. S. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **66**, 4390-4399 (1994).

55. Keller A., Purvine S., Nesvizhskii A. I., Stolyar S., Goodlett D. R., Kolker E. Experimental protein mixture for validating tandem mass spectral analysis. OMICS 6, 207-212 (2002).

56. Kapp E. A., Schutz F., Reid G. E., Eddes J. S., Moritz R. L., O'Hair R. A. J., Speed T. P., Simpson R. J. Mining a tandem mass spectrometry database to determine the trends and global factors influencing peptide fragmentation. *Anal. Chem.* **75**, 6251-6264 (2003).

57. Gentzel M., Kocher T., Ponnusamy S., Wilm M. Preprocessing of tandem mass spectrometric data to support automatic protein identification. *Proteomics* **3**, 1597-1610 (2003).

58. Beer I., Barnea E., Ziv T., Admon A. Improving large-scale proteomics by clustering of mass spectrometry data. *Proteomics* **4**, 950-960 (2004).

59. Moore R. E., Young M. K., Lee T. D. Method for screening peptide fragment ion mass spectra prior to database searching. *J. Am. Soc. Mass Spectrom.* **11**, 422-426 (2000).

60. Kolker E., Purvine S., Galperin M. Y., Stolyar S., Goodlett D. R., Nesvizhskii A. I., Keller A., Xie T., Eng J. K., Yi E., Hood L., Picone A. F., Cherny T., Tjaden B. C., Siegel A. F., Reilly T. J., Makarova K. S., Palsson B. O., Smith A. L. Initial proteome analysis of model microorganism Haemophilus influenzae strain Rd KW20. *J. Bacteriol.* **185**, 4593-4602 (2003).

61. Sadygov R. G., Eng J., Durr E., Saraf A., McDonald H., MacCoss M. J., Yates J. R. III. Code developments to improve the efficiency of automated MS/MS spectra interpretation. *J. Proteome Res.* **1**, 211-215 (2002).

62. Colinge J., Magnin J., Dessingy T., Giron M., Masselot A. Improved peptide charge state assignment. *Proteomics* **3**, 1434-1440 (2003).

63. Tabb D. L., Smith L. L., Breci L. A., Wysocki V. H., Lin D., Yates J. R. III. Statistical characterization of ion trap tandem mass spectra from doubly charged tryptic peptides. *Anal. Chem.* **75**, 1155-1163 (2003).

64. Tabb D. L., McDonald W. H., Yates III J. R. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **1**, 21-26 (2002).

65. Han D. K., Eng J., Zhou H., Aebersold R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* **19**, 946-951 (2001).

66. Eddes J. S., Kapp E. A., Frecklington D. F., Connolly L. M., Layton M. J., Moritz R. L., Simpson R. J. CHOMPER: a bioinformatic tool for rapid validation of tandem mass spectrometry search results associated with high-throughput proteomic strategies. *Proteomics* **2**, 1097-1103 (2002).

67. Nesvizhskii A. I., Keller A., Kolker E., Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646-4658 (2003).

68. Keller A., Nesvizhskii A. I., Kolker E., Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383-5392 (2002).

69. MacCoss M. J., Wu C. C., Yates III J. R. Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal. Chem.* **74**, 5593-5599 (2002).

70. Anderson D. C., Li W., Payan D. G., Noble W. S. A new algorithm for the evaluation of shotgun peptide sequencing in proteomics: support vector machine classification of peptide MS/MS spectra and SEQUEST scores. *J. Proteome Res.* **2**, 137-146 (2003).

71. Fenyo D., Beavis R. C. A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Anal. Chem.* **75**, 768-774 (2003).

72. Kislinger T., Rahman K., Radulovic D., Cox B., Rossant J., Emili A. PRISM, a Generic Large Scale Proteomic Investigation Strategy for Mammals. *Mol. Cell Proteomics* **2**, 96-106 (2003).

73. Gygi S. P., Rist B., Gerber S. A., Turecek F., Gelb M. H., Aebersold R. Quantitative analysis of protein mixtures using isotope coded affinity tags. *Nat. Biotechnol.* **17**, 994-999 (1999).

74. Oda Y., Owa T., Sato T., Boucher B., Daniels S., Yamanaka H., Shinohara Y., Yokoi A., Kuromitsu J., Nagasu T. Quantitative chemical proteomics for identifying candidate drug targets. *Anal. Chem.* **75**, 2159-2165 (2003).

75. Zhou H., Ranish J. A., Watts J. D., Aebersold R. Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. *Nat. Biotechnol.* **20**, 512-515 (2002).

76. Munchbach M., Quadroni M., Miotto G., James P. Quantitation and facilitated do novo sequencing of proteins by isotopic N-terminal labeling of peptides with a fragmentation-directing moiety. *Anal. Chem.* **72**, 4047-4057 (2000).

77. Liu Y., Patricelli M. P., Cravatt B. F. Activity-based protein profiling: the serine hydrolases. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 14694-14699 (1999).

78. Greenbaum D., Medzihradszky K. F., Burlingame A., Bogyo M. Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem. Biol.* **7**, 569-581 (2000).

79. Zhou H., Watts J. D., Aebersold R. A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.* **19**, 375-378 (2001).

80. Oda Y., Nagasu T., Chait B. T. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotechnol.* **19**, 379-382 (2001).

81. Zhang H., Li X., Martin D. B., Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **21**, 660-666 (2003).

82. Mirgorodskaya O. A., Kozmin Y. P., Titov M. I., Korner R., Sonksen C. P., Roepstorff P. Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using (18)O-labeled internal standards. *Rapid Commun. Mass Spectrom.* **14**, 1226-1232 (2000).

83. Yao X., Freas A., Ramirez J., Demirev P. A., Fenselau C. Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal. Chem.* **73**, 2836-2842 (2001).

84. Yao X., Afonso C., Fenselau C. Dissection of proteolytic 18O labeling: endoprotease-catalyzed 16O-to-18O exchange of truncated peptide substrates. *J. Proteome Res.* **2**, 147-152 (2003).

85. Heller M., Mattou H., Menzel C., Yao X. Trypsin catalyzed 16O-to-18O exchange for comparative proteomics: tandem mass spectrometry comparison using MALDI-TOF, ESI-QTOF, and ESI-ion trap mass spectrometers. *J. Am. Soc. Mass Spectrom.* **14**, 704-718 (2003).

86. Ong S. E., Blagoev B., Kratchmarova I., Kristensen D. B., Steen H., Pandey A., Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell Proteomics* **1**, 376-386 (2002).

87. Oda Y., Huang K., Cross F. R., Cowburn D., Chait B. T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 6591-6596 (1999).

88. Smith R. D., Pasa-Tolic L., Lipton M. S., Jensen P. K., Anderson G. A., Shen Y., Conrads T. P., Udseth H. R., Harkewicz R., Belov M. E., Masselon C., Veenstra T. D. Rapid quantitative measurements of proteomes by Fourier transform ion cyclotron resonance mass spectrometry. *Electrophoresis* **22**, 1652-1668 (2001).

89. Wang Y. K., Ma Z., Quinn D. F., Fu E. W. Inverse 15N-metabolic labeling/mass spectrometry for comparative proteomics and rapid identification of protein markers/targets. *Rapid Commun. Mass Spectrom.* **16**, 1389-1397 (2002).

90. Conrads T. P., Alving K., Veenstra T. D., Belov M. E., Anderson G. A., Anderson D. J., Lipton M. S., Pasa-Tolic L., Udseth H. R., Chrisler W. B., Thrall B. D., Smith R. D. Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and 15N-metabolic labeling. *Anal. Chem.* **73**, 2132-2139 (2001).

91. Krijgsveld J., Ketting R. E., Mahmoudi T., Johansen J., Artal-Sanz M., Verrijzer C. P., Plasterk R. H. A., Heck A. J. R. Metabolic labeling of C. elegans and D. melanogaster for quantitative proteomics. *Nat. Biotechnol.* **21**, 927-931 (2003).

92. Gerber S. A., Rush J., Stemman O., Kirschner M. W., Gygi S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 6940-6945 (2003).

93. Rigaut G., Shevchenko A., Rutz B., Wilm M., Mann M., Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**, 1030-1032 (1999).

94. Puig O., Caspary F., Rigaut G., Rutz B., Bouveret E., Bragado-Nilsson E., Wilm M., Seraphin B. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**, 218-229 (2001).

95. Gavin A. C., Bosche M., Krause R., Grandi P., Marzioch M., Bauer A., Schultz J., Rick J. M., Michon A. M., Cruciat C. M., Remor M., Hofert C., Schelder M., Brajenovic M., Ruffner H., Merino A., Klein K., Hudak M., Dickson D., Rudi T., Gnau V., Bauch A., Bastuck S., Huhse B., Leutwein C., Heurtier M. A., Copley R. R., Edelmann A., Querfurth E., Rybin V., Drewes G., Raida M., Bouwmeester T., Bork P., Seraphin B., Kuster B., Neubauer G., Superti-Furga G. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-147 (2002).

96. Ho Y., Gruhler A., Heilbut A., Bader G. D., Moore L., Adams S. L., Millar A., Taylor P., Bennett K., Boutilier K., Yang L., Wolting C., Donaldson I., Schandorff S., Shewnarane J., Vo M., Taggart J., Goudreault M., Muskat B., Alfarano C., Dewar D., Lin Z., Michalickova K., Willems A. R., Sassi H., Nielsen P. A., Rasmussen K. J., Andersen J. R., Johansen L. E., Hansen L. H., Jespersen H., Podtelejnikov A., Nielsen E., Crawford J., Poulsen V., Sorensen B. D., Matthiesen J., Hendrickson R. C., Gleeson F., Pawson T., Moran M. F., Durocher D., Mann M., Hogue C. W., Figeys D., Tyers M. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **415**, 180-183 (2002).

97. von Mering C., Krause R., Snel B., Cornell M., Oliver S. G., Fields S., Bork P. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**, 399-403 (2002).

98. Blagoev B., Kratchmarova I., Ong S. E., Nielsen M., Foster L. J., Mann M. A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nat. Biotechnol.* **21**, 315-318 (2003).

99. Ranish J. A., Yi E. C., Leslie D. M., Purvine S. O., Goodlett D. R., Eng J., Aebersold R. The study of macromolecular complexes by quantitative proteomics. *Nat. Genet.* **33**, 349-355 (2003).

100. Neubauer G., King A., Rappsilber J., Calvio C., Watson M., Ajuh P., Sleeman J., Lamond A., Mann M. Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat. Genet.* **20**, 46-50 (1998).

101. Neubauer G., Gottschalk A., Fabrizio P., Seraphin B., Luhrmann R., Mann M. Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 385-390 (1997).

102. Rout M. P., Aitchison J. D., Suprapto A., Hjertaas K., Zhao Y., Chait B. T. The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148**, 635-651 (2000).

103. Andersen J. S., Lyon C. E., Fox A. H., Leung A. K., Lam Y. W., Steen H., Mann M., Lamond A. I. Directed proteomic analysis of the human nucleolus. *Curr. Biol.* **12**, 1-11 (2002).

104. Mann M., Jensen O. N. Proteomic analysis of post-translational modifications. Nat. Biotechnol. **21**, 255-261 (2003).

105. MacCoss M. J., McDonald W. H., Saraf A., Sadygov R., Clark J. M., Tasto J. J., Gould K. L., Wolters D., Washburn M., Weiss A., Clark J. I., Yates III J. R. Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 7900-7905 (2002).

106. Pandey A., Podtelejnikov A. V., Blagoev B., Bustelo X. R., Mann M., Lodish H. F. Analysis of receptor signaling pathways by mass spectrometry: identification of Vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 179-184 (2000).

107. Steen H., Kuster B., Fernandez M., Pandey A., Mann M. Tyrosine phosphorylation mapping of the epidermal growth factor receptor signalling pathway. *J. Biol. Chem.* **277**, 1031-1039 (2002).

108. Pandey A., Fernandez M., Steen H., Blagoev B., Nielsen M. M., Roche S., Mann M., Lodish H. F. Identification of a novel immunoreceptor tyrosine-based activation motif-containing molecule, STAM2, by mass spectrometry and its involvement in growth factor and cytokine receptor signalling pathways. *J. Biol. Chem.* **275**, 38633-38639 (2000).

109. Goshe M. B., Veenstra T. D., Panisko E. A., Conrads T. P., Angell N. H., Smith R. D. Phosphoprotein isotope-coded affiniity tags: application to the enrichment and identification of low-abundance phosphoproteins. *Anal. Chem.* **74**, 607-616 (2002).

110. Goshe M. B., Conrads T. P., Panisko E. A., Angell N. H., Veenstra T. D., Smith R. D. Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analyses. *Anal. Chem.* **73**, 2578-2586 (2001).

111. Ficarro S. B., McCleland M. L., Stukenberg P. T., Burke D. J., Ross M. M., Shabanowitz J., Hunt D. F., White F. M. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat. Biotechnol.* **20**, 301-305 (2002).

112. Peng J., Schwartz D., Elias J. E., Thoreen C. C., Cheng D., Marsischky G., Roelofs J., Finley D., Gygi S. P. A proteomics approach to understanding protein ubiquitination. *Nat. Biotechnol.* **21**, 921-926 (2003).

113. Kaji H., Saito H., Yamauchi Y., Shinkawa T., Taoka M., Hirabayashi J., Kasai K., Takahashi N., Isobe T. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat. Biotechnol.* **21**, 667-672 (2003).

114. Kjeldsen F., Haselmann K. F., Budnik B. A., Sorensen E. S., Zubarev R. A. Complete characterization of posttranslational modification sites in bovine milk protein PP3 by tandem mass spectrometry with electron capture dissociation as the last stage. *Anal. Chem.* **75**, 2355-2361 (2003).

115. Mirgorodskaya E., Roepstorff P., Zubarev R. A. Localization of O-glycosylation sites in peptides by electron capture dissociation in a fourier transform mass spectrometer. *Anal. Chem.* **71**, 4431-4436 (1999).

116. Zubarev R. A., Kelleher N. L., McLafferty F. W. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **120**, 3265-3266 (1998).

117. Sze S. K., Ge Y., Oh H., McLafferty F. W. Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 1774-1779 (2002).

118. Shi S. D. H., Hemling M. E., Carr S. A., Horn D. M., Lindh I., McLafferty F. W. Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. *Anal. Chem.* **73**, 19-22 (2001).

119. Aebersold R., Watts J. D. The need for national centers for proteomics. *Nat. Biotechnol.* **20**, 651-651 (2002).

120. Mann M. A home for proteomics data? *Nature* **420**, 21-21 (2002).

121. Aebersold R. Constellations in a cellular universe. *Nature* **422**, 115-116 (2003).

122. Chaurand P., Schwartz S. A., Caprioli R. M. Imaging mass spectrometry: a new tool to investigate the spatial organization of peptides and proteins in mammalian tissue sections. *Curr. Opin. Chem. Biol.* **6**, 676-681 (2002).

123. Chaurand P., Schwartz S. A., Caprioli R. M. Assessing patterns in disease using imaging mass spectrometry. *J. Proteome Res.* **3**, 245-252 (2004).

124. Goodlett D. R., Bruce J. E., Anderson G. A., Rist B., Pasa-Tolic L., Fiehn O., Smith R. D., Aebersold R. Protein identification with a single accurate mass of a cysteine-containing peptide and constrained database searching. *Anal. Chem.* **72**, 1112-1118 (2000).

125. Smith R. D., Anderson G. A., Lipton M. S., Pasa-Tolic L., Shen Y., Conrads T. P., Veenstra T. D., Udseth H. R. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* **2**, 513-523 (2002).

126. Lipton M. S., Pasa-Tolic L., Anderson G. A., Anderson D. J., Auberry D. L., Battista J. R., Daly M. J., Fredrickson J., Hixson K. K., Kostandarithes H., Masselon C., Markillie L. M., Moore R. J., Romine M. F., Shen Y., Stritmatter E., Tolik N., Udseth H. R., Venkateswaran A., Wong K.-K., Zhao R., Smith R. D. Global analysis of the Deinococcus radiodurans proteome by using accurate mass tags. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11049-11054 (2002).

127. Stritmatter E. F., Ferguson P. L., Tang K., Smith R. D. Proteome analyses using accurate mass and elution time peptide tags with capillary LC time-of-flight mass spectrometry. *J. Am. Soc. Mass Spectrom.* **14**, 980-991 (2003).

128. Nelson D. Cytochrome P450 Homepage. http://drnelson.utmem.edu/CytochromeP450.html (2003).

129. Drug metabolising enzymes. Ed: Lee J. S., Obach R. S., Fisher M. B. Marcel Dekker (2003).

130. Williams P. A., Cosme J., Sridhar V., Johnson E. F., McRee D. E. Microsomal cytochrome P450 2C5: comparison to microbial P450s and unique features. *J. Inorg. Biochem.* **81**, 183-190 (2000).

131. Cosme J., Johnson E. F. Engineering microsomal cytochrome P450 2C5 to be a soluble, monomeric enzyme. Mutations that alter aggregation, phospholipid dependence of catalysis, and membrane binding. *J. Biol. Chem.* **275**, 2545-2553 (2000).

132. Scott E. E., He Y. A., Wester M. R., White M. A., Chin C. C., Halpert J. R., Johnson E. F., Stout C. D. An open conformation of mammalian cytochrome P450 2B4 at 1.6-A resolution. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13196-13201 (2003).

133. Schoch G. A., Yano J. K., Wester M. R., Griffin K. J., Stout C. D., Johnson E. F. Structure of human microsomal cytochrome P450 2C8. Evidence for a peripheral fatty acid binding site. *J. Biol. Chem.* **279**, 9497-9503 (2004).

134. Williams P. A., Cosme J., Ward A., Angove H. C., Matak Vinkovic D., Jhoti H. Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* **424**, 464-468 (2003).

135. Williams P. A., Cosme J., Vinkovic D. M., Ward A., Angove H. C., Day P. J., Vonrhein C., Tickle I. J., Jhoti H. Crystal Structures of Human Cytochrome P450 3A4 Bound to Metyrapone and Progesterone. *Science* **305**, 683-686 (2004).

136. Lewis D. F. V. Cytochromes P450: structure, function and mechanism. Taylor and Francis (1996).

137. Nelson D. Cytochrome P450s in humans. http://drnelson.utmem.edu/P450lect.html (2003).

138. Ingelman-Sundberg M. Human drug metabolising cytochrome P450 enzymes: properties and polymorphisms. *Naunyn Schmiedebergs Arch. Pharmacol.* **369**, 89-104 (2004).

139. Ortiz de Montellano P. R., De Voss J. J. Oxidizing species in the mechanism of cytochrome P450. *Nat. Prod. Rep.* **19**, 477-493 (2002).

140. Hlavica P., Lewis D. F. Allosteric phenomena in cytochrome P450-catalyzed monooxygenations. *Eur. J. Biochem.* **268**, 4817-4832 (2001).

141. Lewis D. F. P450 structures and oxidative metabolism of xenobiotics. *Pharmacogenomics.* **4**, 387-395 (2003).

142. Evans W. E., Relling M. V. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* **286**, 487-491 (1999).

143. Tredger J. M., Stoll S. Cytochromes P450 - their impact on drug treatment. Hospital Pharmacist **9**, 167-173 (2002).

144. Shimada T., Yamazaki H., Mimura M., Inui Y., Guengerich F. P. Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. *J. Pharmacol. Exp. Ther.* **270**, 414-423 (1994).

145. Snawder J. E., Lipscomb J. C. Interindividual variance of cytochrome P450 forms in human hepatic microsomes: correlation of individual forms with xenobiotic metabolism and implications in risk assessment. *Regul. Toxicol. Pharmacol.* **32**, 200-209 (2000).

146. Ingelman-Sundberg M. Polymorphism of cytochrome P450 and xenobiotic toxicity. *Toxicology* **181-182**, 447-452 (2002).

147. Ingelman-Sundberg M. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *Trends Pharmacol. Sci.* **25**, 193-200 (2004).

148. Patterson L. H., Murray G. I. Tumour cytochrome P450 and drug activation. *Curr. Pharm. Des.* **8**, 1335-1347 (2002).

149. Rooseboom M., Commandeur J. N., Vermeulen N. P. Enzyme-catalyzed activation of anticancer prodrugs. *Pharmacol. Rev.* **56**, 53-102 (2004).

150. Hutchinson I., Chua M. S., Browne H. L., Trapani V., Bradshaw T. D., Westwell A. D., Stevens M. F. Antitumor benzothiazoles. 14. Synthesis and in vitro biological properties of fluorinated 2-(4-aminophenyl)benzothiazoles. *J. Med. Chem.* **44**, 1446-1455 (2001).

151. Chua M. S., Kashiyama E., Bradshaw T. D., Stinson S. F., Brantley E., Sausville E. A., Stevens M. F. Role of Cyp1A1 in modulation of antitumor properties of the novel agent 2-(4-amino-3-methylphenyl)benzothiazole (DF 203, NSC 674495) in human breast cancer cells. *Cancer Res.* **60**, 5196-5203 (2000).

152. Bradshaw T. D., Stevens M. F., Westwell A. D. The discovery of the potent and selective antitumour agent 2-(4-amino-3-methylphenyl)benzothiazole (DF 203) and related compounds. *Curr. Med. Chem.* **8**, 203-210 (2001).

153. Patterson L. H., McKeown S. R., Robson T., Gallagher R., Raleigh S. M., Orr S. Antitumour prodrug development using cytochrome P450 (CYP) mediated activation. *Anti-Cancer Drug Des.* **14**, 473-486 (1999).

154. Raleigh S. M., Wanogho E., Burke M., McKeown S. R., Patterson L. H. Involvement of human cytochromes P450 (CYP) in the reductive metabolism of AQ4N, a hypoxia activated anthraquinone di-N-oxide prodrug. *Int. J. Radiat. Oncol. Biol. Phys.* **42**, 763-767 (1998).

155. Wallin E., von Heijne G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **7**, 1029-1038 (1998).

156. Santoni V., Molloy M., Rabilloud T. Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21**, 1054-1070 (2000).

157. Wu C. C., Yates III J. R. The application of mass specrtometry to membrane proteomics. *Nat. Biotechnol.* **21**, 262-267 (2003).

158. Henningsen R., Gale B. L., Straub K. M., DeNagel D. C. Application of zwitterionic detergents to the solubilization of integral membrane proteins for two-dimensional gel electrophoresis and mass spectrometry. *Proteomics* **2**, 1479-1488 (2002).

159. Carboni L., Piubelli C., Righetti P. G., Jansson B., Domenici E. Proteomic analysis of rat brain tissue: comparison of protocols for two-dimensional gel electrophoresis analysis based on different solubilizing agents. *Electrophoresis* **23**, 4132-4141 (2002).

160. Luche S., Santoni V., Rabilloud T. Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. *Proteomics* **3**, 249-253 (2003).

161. Ferro M., Salvi D., Riviere-Rolland H., Vermat T., Seigneurin-Berny D., Grunwald D., Garin J., Joyard J., Rolland N. Integral membrane proteins of the chloroplast envelope: identification and subcellular localization of new transporters. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11487-11492 (2002).

162. Ferro M., Seigneurin-Berny D., Rolland N., Chapel A., Salvi D., Garin J., Joyard J. Organic solvent extraction as a versatile procedure to identify hydrophobic chloroplast membrane proteins. *Electrophoresis* **21**, 3517-3526 (2000).

163. Galeva N., Yakovlev D., Koen Y., Duzhak T., Altermann M. Direct identification of cytochrome P450 isozymes by matrix-assisted laser desorption/ionization time of flight-based proteomic approach. *Drug Metab Dispos.* **31**, 351-355 (2003).

164. Wu C. C., Yates III J. R. The application of mass spectrometry to membrane proteomics. *Nat. Biotechnol.* **21**, 262-267 (2003).

165. Blonder J., Goshe M. B., Moore R. J., Pasa-Tolic L., Masselon C. D., Lipton M. S., Smith R. D. Enrichment of integral membrane proteins for proteomic analysis using liquid chromatography-tandem mass spectrometry. *J. Proteome Res.* **1**, 351-360 (2002).

166. Blonder J., Conrads T. P., Yu L. R., Terunuma A., Janini G. M., Issaq H. J., Vogel J. C., Veenstra T. D. A detergent- and cyanogen bromide-free method for integral membrane

proteomics: application to Halobacterium purple membranes and the human epidermal membrane proteome. *Proteomics* **4**, 31-45 (2004).

167. Goshe M. B., Blonder J., Smith R. D. Affinity labeling of highly hydrophobic integral membrane proteins for proteome-wide analysis. *J. Proteome Res.* **2**, 153-161 (2003).

168. Wu C. C., MacCoss M. J., Howell K. E., Yates III J. R. A method for the comprehensive proteomic analysis of membrane proteins. *Nat. Biotechnol.* **21**, 532-537 (2003).

169. Anderson L., Seilhamer J A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**, 533-537 (1997).

170. Chen G., Gharib T. G., Huang C.-C., Taylor J. M. G., Misek D. E., Kardia S. L. R., Giordano T. J., Iannettoni M. D., Orringer M. B., Hanash S. M., Beer G. D. Discordant protein and mRNA expression in lung adenocarcinomas. *Mol. Cell Proteomics* **1**, 304-314 (2002).

171. McFadyen M. C. E., Rooney P. H., Melvin W. T., Murray G. I. Quantitative analysis of the Ah receptor/cytochrome P450 CYP1B1/CYP1A1 signalling pathway. *Biochem. Pharmacol.* **65**, 1663-1674 (2003).

172. Gygi S. P., Rochon Y., Franza B. R., Aebersold R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol.* **19**, 1720-1730 (1999).

173. McFadyen M. C. E., Breeman S., Payne S., Stirk C., Miller I. D., Melvin W. T., Murray G. I. Immunohistochemical localization of cytochrome P450 CYP1B1 in breast cancer with monoclonal antibodies. *J. Histochem. Cytochem.* **47**, 1457-1464 (1999).

174. Bradford M. M. A rapid and sensitive method for the quantification of microgram quanities of protein utilizing the principle of dye-binding. *Anal. Biochem.* **72**, 248-254 (1976).

175. Compton S. J., Jones C. G. Mechanism of dye response and interference in the Bradford protein assay. *Anal. Biochem.* **151**, 369-374 (1985).

176. Moreno R. L., Goosen T., Kent U. M., Chung F. L., Hollenberg P. F. Differential effects of naturally occurring isothiocyanates on the activities of cytochrome P450 2E1 and the mutant P450 2E1 T303A. *Arch. Biochem. Biophys.* **391**, 99-110 (2001).

177. Zhang B., Liu H., Karger B. L., Foret F. Microfabricated devices for capillary electrophoresis-electrospray mass spectrometry. *Anal. Chem.* **71**, 3258-3264 (1999).

178. Lightning L. K., Jones J. P, Friedberg T., Pritchard M. P., Shou M., Rushmore T. H., Trager W. F. Mechanism-based inactivation of cytochrome P450 3A4 by L-754,394. *Biochemistry* **39**, 4276-4287 (2000).

179. Park Z. Y., Russell D. H. Thermal denaturation: a useful technique in peptide mass mapping. *Anal. Chem.* **72**, 2667-2670 (2000).

180. Laemmli U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685 (1970).

181. Sevchenko A., Wilm M., Vorm O, Mann M. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850-858 (1996).

182. Yates III J. R., Eng J. K., McCormack A. L., Schieltz D. M. Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal. Chem.* **67**, 1426-1436 (1995).

183. Gillam E. M., Guo Z., Guengerich F. P. Expression of modified human cytochrome P450 2E1 in Escherichia coli, purification, and spectral and catalytic properties. *Arch. Biochem. Biophys.* **312**, 59-66 (1994).

184. Benson D. A., Karsch-Mizrachi I., Lipman D. J., Ostell J., Wheeler D. L. GenBank. *Nucleic Acids Res.* **31**, 23-27 (2003).

185. Rodrigues A. D. Integrated cytochrome P450 reaction phenotyping; attempting to bridge the gap between cDNA-expressed cytochromes and native human liver microsomes. *Biochem. Pharmacol.* **57**, 465-480 (1999).

186. National Diagnostics catalogue 2003-2004. (2003).

187. Shevchenko A., Wilm M., Vorm O, Mann M. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850-858 (1996).

188. Simpson R. J. Proteins and proteomics: a laboratory manual. Cold Spring Harbor Laboratory Press (2002).

189. Sutter T. R., Tang Y. M., Hayes C. L., Wo Y. Y., Jabs E. W., Li X., Yin H., Cody C. W., Greenlee W. F. Complete cDNA sequence of a human dioxin-inducible mRNA identifies a new gene subfamily of cytochrome P450 that maps to chromosome 2. *J. Biol. Chem.* **269**, 13092-13099 (1994).

190. Murray G. I., Taylor M. C., McFadyen M. C. E., McKay J. A., Greenlee W. F., Burke M. D., Melvin W. T. Tumor-specific expression of cytochrome P450 CYP1B1. *Cancer Res.* **57**, 3026-3031 (1997).

191. McCarthy H. O., Yakkundi A., McErlane V., Hughes C. M., Keilty G., Murray M., Patterson L. H., Hirst D. G., McKeown S. R., Robson T. Bioreductive GDEPT using cytochrome P450 3A4 in combination with AQ4N. *Cancer Gene Ther.* **10**, 40-48 (2003).

192. Patterson L. H., McKeown S. R. AQ4N: a new approach to hypoxia-activated cancer chemotherapy. *Br. J. Cancer* **83**, 1589-1593 (2000).

193. Doherty M. M., Michael M. Tumoural drug metabolism: perspectives and therapeutic implications. *Curr. Drug. Metab.* **4**, 131-149 (2003).

194. Massaad L., de Waziers I., Ribrag V., Janot F., Beaune P. H., Morizet J., Gouyette A., Chabot G. G. Comparison of mouse and human colon tumors with regard to phase I and phase II drug-metabolizing enzyme systems. *Cancer Res.* **52**, 6567-6575 (1992).

195. Ducret A., Van Oostveen I., Eng J. K., Yates III J. R., Aebersold R. High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. *Protein Sci.* **7**, 706-719 (1998).

196. Yamreudeewong W., DeBisschop M., Martin L. G., Lower D. L. Potentially significant drug interactions of class III antiarrhythmic drugs. *Drug Safety* **26**, 421-438 (2003).

197. Naganuma M., Shiga T., Nishikata K., Tsuchiya T., Kasanuki H., Fujii E. Role of desethylamiodarone in the anticoagulant effect of concurrent amiodarone and warfarin therapy. *J. Cardiovasc. Pharmacol.* **6**, 363-367 (2001).

198. Chen X.-P., Tan Z.-R., Huang S.-L., Huang Z., Ou-Yang D.-S., Zhou H.-H. Isozyme-specific induction of low-dose aspirin on cytochrome P450 in healthy subjects. *Clin. Pharmacol. Ther.* **73**, 264-271 (2003).

199. Katoh M., Nakajima M., Yamazaki H., Yokoi T. Inhibitory potencies of 1,4-dihydropyridine calcium antagonists to P-glycoprotein-mediated transport: comparison with the effects on CYP3A4. *Eur. J. Clin. Pharmacol.* **55**, 843-852 (2000).

200. Simonsen U. Interactions between drugs for erectile dysfunction and drugs for cardiovascular disease. *Int. J. Impot. Res.* **14**, 178-188 (2002).

201. Thompson D. S., Kirshner M. A., Klug T. L., Kastango K. B., Pollock B. G. A preliminary study of the effect of fluoxetine treatment on the 2:16-alpha-hydroxyestrone ratio in young women. *Ther. Drug Monit.* **25**, 125-128 (2003).

202. Harvey A. T., Preskorn S. H. Fluoxetine pharmacokinetics and effect on CYP19 in young and elderly volunteers. *J. Clin. Psychopharmacol.* **21**, 161-166 (2001).

203. McDonough K. H. Antioxidant nutrients and alcohol. *Toxicology* **189**, 89-97 (2003).

204. Wolbold R., Klein K., Burk O., Nussler A. K., Neuhaus P., Eichelbaum M., Schwab M., Zanger U. M. Sex is a major determinant of CYP3A4 expression in human liver. *Hepatology* **38**, 978-988 (2003).

205. Silverman R. B. Mechanism-based enzyme inactivators. *Methods Enzymol.* **249**, 240-283 (1995).

206. Chun J., Kent U. M., Moss R. M., Sayre L. M., Hollenberg P. F. Mechanism-based inactivation of cytochromes P450 2B1 and P450 2B6 by 2-phenyl-2-(1-piperidinyl)propane. *Drug Metab Dispos.* **28**, 905-911 (2000).

207. Osawa Y, Pohl L. R. Covalent bonding of the prosthetic heme to protein: a potential mechanism for the suicide inactivation or activation of hemoproteins. *Chem. Res. Toxicol.* **2**, 131-141 (1995).

208. Kent U. M., Jushchyshyn M. I., Hollenberg P. F. Mechanism-based inactivators as probes of cytochrome P450 structure and function. *Curr. Drug. Metab.* **2**, 215-243 (2001).

209. Moreno R. L., Kent U. M., Hodge K., Hollenberg P. F. Inactivation of cytochrome P450 2E1 by benzyl isothiocyanate. *Chem. Res. Toxicol.* **12**, 582-587 (1999).

210. Blobaum A. L., Kent U. M., Alworth W. L., Hollenberg P. F. Novel reversible inactivation of P450 2E1 T303A by tert-butyl acetylene: The role of threonine 303 in proton delivery to the active site of cytochrome P450 2E1. *J. Pharmacol. Exp. Ther.* **310**, 281-290 (2004).

211. Goosen T. C., Mills D. E., Hollenberg P. F. Effects of benzyl isothiocyanate on rat and human cytochromes P450: identification of metabolites formed by P450 2B1. *J. Pharmacol. Exp. Ther.* **296**, 198-206 (2001).

212. Goosen T. C., Kent U. M., Brand L., Hollenberg P. F. Inactivation of cytochrome P450 2B1 by benzyl isothiocyanate, a chemopreventative agent from cruciferous vegetables. *Chem. Res. Toxicol.* **13**, 1349-1359 (2000).

213. Cai H., Guengerich F. P. Reaction of trichloroethylene and trichloroethylene oxide with cytochrome P450 enzymes: inactivation and sites of modification. *Chem. Res. Toxicol.* **14**, 451-458 (2001).

214. He K., Bornheim L. M., Falick A. M., Maltby D., Yin H., Correia M. A. Identification of the heme-modified peptides from cumene hydroperoxide-inactivated cytochrome P450 3A4. *Biochemistry* **37**, 17448-17457 (1998).

215. Yao K., Falick A. M., Patel N., Correia M. A. Cumene hydroperoxide-mediated inactivation of cytochrome P450 2B1. Identification of an active site heme-modified peptide. *J. Biol. Chem.* **268**, 59-65 (1993).

216. Koenigs L. L., Peter R. M., Hunter A. P., Haining R. L., Rettie A. E, Friedberg T., Pritchard M. P., Shou M., Rushmore T. H., Trager W. F. Electrospray ionization mass spectrometric analysis of intact cytochrome P450: identification of tienilic acid adducts to P450 2C9. *Biochemistry* **38**, 2312-2319 (1999).

217. Koenigs L. L., Trager W. F. Mechanism-based inactivation of cytochrome P450 2B1 by 8-methoxypsoralen and several other furanocoumarins. *Biochemistry* **37**, 13184-13193 (1998).

218. Koenigs L. L., Peter R. M., Thompson S. J., Rettie A. E, Trager W. F. Mechanism-based inactivation of human liver cytochrome P450 2A6 by 8-methoxypsoralen. *Drug Metab Dispos.* **25**, 1407-1415 (1997).

219. Koenigs L. L., Trager W. F. Mechanism-based inactivation of P450 2A6 by furanocoumarins. *Biochemistry* **37**, 10047-10061 (1998).

220. Kent U. M., Mills D. E., Rajnarayanan R. V., Alworth W. L., Hollenberg P. F. Effect of 17a-ethynylestradiol on activities of cytochrome P450 2B (P450 2B) enzymes: characterization of inactivation of P450s 2B1 and 2B6 and identification of metabolites. *J. Pharmacol. Exp. Ther.* **300,** 549-558 (2002).

221. Lin H.-L., Kent U. M., Hollenberg P. F. Mechanism-based inactivation of cytochrome P450 3A4 by 17a-ethynylestradiol: evidence for heme destruction and covalent binding to protein. *J. Pharmacol. Exp. Ther.* **301**, 160-167 (2002).

222. Yamasaki T., Izumi S., Ide H., Ohyama Y. Identification of a novel rat microsomal vitamin D3 25-hydroxylase. *J. Biol. Chem.* **279**, 22848-22856 (2004).

223. Fenyo D. Identifying the proteome: software tools. *Curr. Opin. Biotechnol.* **11**, 391-395 (2000).

224. Huang L., Baldwin M. A., Maltby D. A., Medzihradszky K. F., Baker P. R., Allen N., Rexach M., Edmondson R. D., Campbell J., Juhasz P., Martin S. A., Vestal M. L., Burlingame A. L. The identification of protein-protein interactions of the nuclear pore complex of Saccharomyces cerevisiae using high throughput matrix-assisted laser desorption ionization time-of-flight tandem mass spectrometry. *Mol. Cell Proteomics* **1**, 434-450 (2002).

225. Galeva N., Altermann M. Comparison of one-dimensional and two-dimensional gel electrophoresis as a separation tool for proteomic analysis of rat liver microsomes: cytochromes P450 and other membrane proteins. *Proteomics* **2**, 713-722 (2002).

226. Nisar S., Lane C. S., Wilderspin A. F., Welham K. J., Griffiths W. J., Patterson L. H. A proteomic approach to the identification of cytochrome P450 isoforms in male and female rat liver by nano-scale liquid chromatography electrospray ionisation tandem mass spectrometry. *Drug Metab Dispos.* **32**, 382-386 (2004).

227. Koenigs L. L., Peter R. M., Hunter A. P., Haining R. L., Rettie A. E, Friedberg T., Pritchard M. P., Shou M., Rushmore T. H., Trager W. F. Electrospray ionization mass spectrometric analysis of intact cytochrome P450: identification of tielenic acid adducts to P450 2C9. *Biochemistry* **38,** 2312-2319 (1999).

228. Bateman K. P, LeRiche T., Lee J., Baker J., Wilke M. Analysis of covalent cytochrome P-450 drug adducts using LC-MS. *16th International Mass Spectrometry Conference Edinburgh, UK*, (2003).

229. Kent U. M., Roberts-Kirchoff E. S., Moon N., Dunham W. R., Hollenberg P. F. Spectral studies of tert-butyl isothiocyanate-inactivated P450 2E1. *Biochemistry* **40**, 7253-7261 (2001).

230. Koenigs Lightening L., Jones J. P., Friedberg T., Pritchard M. P., Shou M., Rushmore T. H., Trager W. F. Mechanism-based inactivation of cytochrome P450 3A4 by L-754,394. *Biochemistry* **39**, 4276-4287 (2000).

231. Smyth R., Turton J. A., Clarke C. J., York M. J., Dare T. O., Lane C. S., Welham K. J., Munday M. R. The identification of superoxide dismutase in rat following carbon tetrachloride-induced hepatotoxicity. *Toxicology* **194**, 271-272 (2004).

Appendix 1

Supplementary tables

Table A1.1 Data used for the comparison of nano-LC-ES with nano-ES (Section 3.4.4).

Sample analysis [a]	P450	No. of peptides with Xcorr ≥2.5	Total no. of peptides	No. of repeat identifications	No. of incorrectly listed peptides on summary
Nano-LC-ES-1	1A2	5	9	2	0
	2E1	19	27	5	1
	3A4	9	12	8	2
Nano-LC-ES-2	1A2	7	11	4	0
	2E1	17	25	9	1
	3A4	9	13	6	2
Nano-LC-ES-3	1A2	5	10	4	1
	2E1	14	20	6	3
	3A4	6	11	4	1
Nano-LC-ES-4	1A2	3	8	3	0
	2E1	11	16	3	1
	3A4	7	10	1	0
Nano-ES-1	1A2	2	9	14	2
	2E1	4	13	22	3
	3A4	6	13	27	9
Nano-ES-2	1A2	6	11	38	2
	2E1	11	17	42	10
	3A4	11	12	35	8
Nano-ES-3	1A2	3	7	13	8
	2E1	2	9	13	16
	3A4	2	6	10	7
Nano-ES-4	1A2	4	11	28	5
	2E1	14	23	67	14
	3A4	11	12	72	13

[a] Analysis by nano-ES: the digest mixture was acidified to contain 0.1% TFA and desalted using ZipTips (Section 2.10.1). Analysis by nano-LC-ES: sample was not desalted; 1 µl of sample was injected via a 10 µl sample loop. CYP1A2, 2E1 and 3A4 were each present at approximately 290 fmol/µl. MS/MS data were searched with Sequest (Section 2.14.2) against the Finnigan Xcalibur human database (Section 2.14.1).

Table A1.2 Mean and standard deviation values from data used for the comparison of nano-LC-ES with nano-ES (Section 3.4.4). Figures shown are the mean values ± 1 standard deviation.

P450	Analysis method	No. of peptides with Xcorr ≥2.5	No. of repeat identifications	No. of incorrectly listed peptides on summary
1A2	Nano-LC-ES	5.0 ± 1.6	3.3 ± 1.0	0.3 ± 0.5
	Nano-ES	3.8 ± 1.7	23.3 ± 12.0	4.3 ± 2.9
2E1	Nano-LC-ES	15.3 ± 3.5	5.8 ± 2.5	1.5 ± 1.0
	Nano-ES	7.8 ± 5.7	36.0 ± 24.0	10.8 ± 5.7
3A4	Nano-LC-ES	7.8 ± 1.5	4.8 ± 3.0	1.3 ± 1.0
	Nano-ES	7.5 ± 4.4	36.0 ± 26.2	9.3 ± 2.6

Table A1.3 Data used to assess the effect of the use of ZipTips on the number of CYP1A2 and 3A4 peptides identified after in-gel tryptic digestion (Section 3.8.3).

P450 [a]	Sample no.	Desalted (Y/N)	Replicate analysis no.	No. of peptides with Xcorr ≥2.5	Total no. of peptides
CYP1A2	1	Y	1	5	6
			2	4	9
CYP1A2	2	Y	1	5	10
			2	3	8
CYP1A2	3	Y	1	2	3
			2	0	0
CYP1A2	4	N	1	5	9
			2	3	5
CYP1A2	5	N	1	5	9
			2	4	6
CYP1A2	6	N	1	6	8
			2	4	8
			3	5	8
CYP3A4	1	Y	1	0	3
			2	2	3
			3	0	0
CYP3A4	2	Y	1	0	1
			2	0	0
			3	0	0
CYP3A4	3	Y	1	1	5
			2	1	3
			3	1	3
CYP3A4	4	N	1	3	6
			2	1	3
			3	1	4
CYP3A4	5	N	1	5	6
			2	2	4
			3	4	7
CYP3A4	6	N	1	8	10
			2	3	6
			3	4	5

[a] Six replicate lanes each of CYP1A2 (2.2 pmol) and CYP3A4 (3.0 pmol) were analysed by SDS-PAGE. Gels were silver-stained (Section 2.7.2) and the protein bands and two blank gel bands subjected to in-gel tryptic digestion (Section 2.9). Three in-gel digests from each P450 and one blank in-gel digest were subjected to desalting using ZipTips (Section 2.10.1), whilst the remaining digests were not desalted, but stored at -80 °C. Desalted samples were dried and stored at -80 °C. Samples were reconstituted in 20 µl of 0.1% TFA and analysed by nano-LC-ES (Section 2.12.2): 1 µl of sample was injected via a 10 µl sample loop. Data was searched against the Finnigan Xcalibur human database (Section 2.14.1) using Sequest (parameters as described in Section 2.14.2, except the minimum TIC for the preparation of .dta files was set to 1 x 10^5 rather then 5 x 10^5).

Appendix 2

Publications resulting from the work presented in this thesis

C. S. Lane, W. J. Griffiths, L. H. Patterson. Identification of a covalently modified human cytochrome P450 CYP2E1 active site peptide after metabolism of benzyl isothiocyanate by tandem mass spectrometry. *American Association for Cancer Research Special Conference: Advances in Proteomics in Cancer Research,* Key Biscayne, Florida, USA, Oct 6-10, 2004.

C. S. Lane, W. J. Griffiths, L. H. Patterson. Identification of a covalently modified active site peptide after metabolism of benzyl isothiocyanate by human cytochrome P450 CYP2E1. *British Mass Spectrometry Society Annual Meeting*, Derby, UK, Sept 6-8, 2004.

C. S. Lane, S. Nisar, W. J. Griffiths, B. J. Fuller, B. R. Davidson, J. Hewes, K. J. Welham, L. H. Patterson. Identification of cytochrome P450 enzymes in human colorectal metastases and surrounding liver: a proteomic approach. *European Journal of Cancer*, **40**(14), 2127-2134 (2004).

S. Nisar, C. S. Lane, A. F. Wilderspin, K. J. Welham, W. J. Griffiths, L. H. Patterson. A proteomic approach to the identification of cytochrome P450 isoforms in male and female rat liver by nano-scale liquid chromatography electrospray ionisation tandem mass spectrometry. *Drug Metabolism and Disposition*, **32**(4), 382-386 (2004).

C. S. Lane, W. J. Griffiths, B. J. Fuller, B. R. Davidson, K. J. Welham, L. H. Patterson. The proteomic identification of cytochrome P450 isoforms in liver and tumors from patients with metastatic colorectal cancer. *American Association for Cancer Research 95th Annual Meeting*, Orlando, Florida, USA, March 27-31, 2004 (oral presentation).

R. Smyth, J. A. Turton, C. J. Clarke, M. J. York, T. O. Dare, C. S. Lane, K. J. Welham, M. R. Munday. The identification of superoxide dismutase in rat following carbon tetrachloride-induced hepatotoxicity. *Toxicology*, **194**(3), 271-272 (2004).

C. S. Lane, W. J. Griffiths, K. J. Welham, B. J. Fuller, B. R. Davidson, L. H. Patterson. A study of the cytochrome P450 isoforms expressed in human liver

metastases and 'normal' human liver. *16th International Mass Spectroscopy Conference*, Edinburgh, August 31-September 5, 2003.

S. Nisar, C. S. Lane, W. J. Griffiths, A. F. Wilderspin, K. J. Welham, S. Orr, L. H. Patterson. Nanoelectrospray ionisation tandem mass spectrometric identification of cytochrome P450s from human liver and hepatocytes. *16th International Mass Spectroscopy Conference,* Edinburgh, August 31-September 5, 2003.

C. S. Lane, L. H. Patterson, K. J. Welham. Analysis of human cytochrome P450 isoforms using nanoelectrospray ionization tandem mass spectrometry. *AAPS PharmSci Supplement*, **4**(4), R6013 (2002).

C. S. Lane, L. H. Patterson, C. Lenz, K. J. Welham. Analysis of human cytochrome P450 isoforms using nanoelectrospray ionization tandem mass spectrometry. *Journal of Pharmacy and Pharmacology*, **54**(supplement), S-1 (2002) (oral presentation).

Addendum

The identification of rat superoxide dismutase in rat urine following carbon-tetrachloride induced hepatotoxicity

The identification of rat Cu/Zn superoxide dismutase

A current study at the School of Pharmacy is concerned with the identification of novel protein markers in the rat by non-invasive methods following hepatic insult [231]. Existing methods for identifying hepatotoxicity involve invasive procedures such as liver biopsy and serum analysis. For this study, hepatotoxicity was induced in rats using carbon tetrachloride (CCl_4), and urine was collected. Urine samples were analysed by SDS-PAGE and a protein band at approximately 18.4 kDa was observed in the urine from CCl_4 treated rats that was not present in control urine. The protein was subjected to in-gel tryptic digestion, with desalting. The sample was analysed by nano-ES (Section 2.12.1), and searched using Sequest (Section 2.14.2) against the Finnigan Xcalibur rat protein database (Section 2.14.1).

The protein was identified as Cu/Zn superoxide dismutase (SOD). Four peptides were identified, all with Xcorr scores $\geq$2.5 and all unique to SOD. Figure 1 shows the full-scan spectrum, with identified peptides indicated, whilst Figure 2 shows the MS/MS spectrum of the doubly charged $[M+2H]^{2+}$ ion of the peptide with amino acid sequence DGVANVSIEDR.

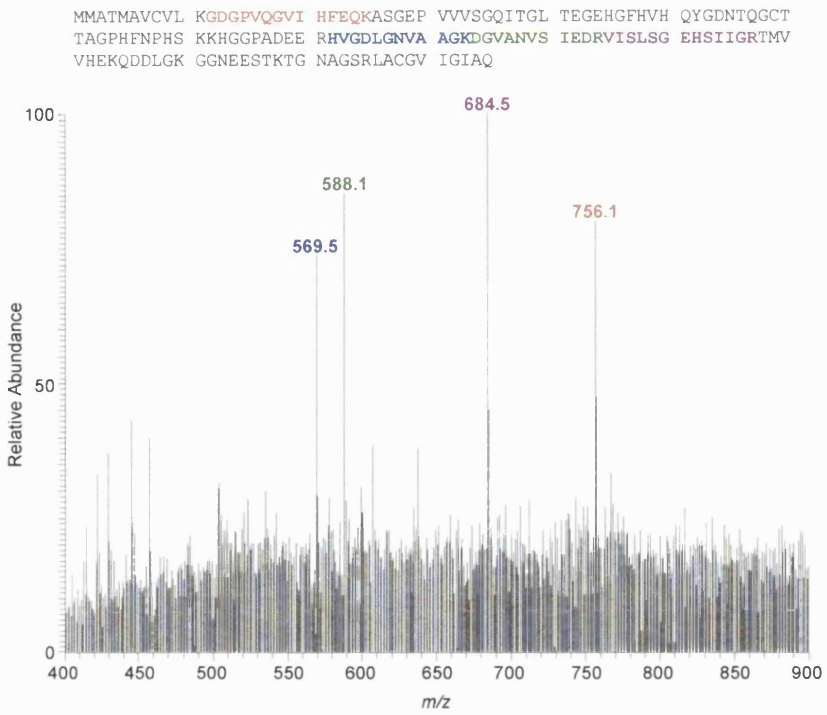


Figure 1 A full scan mass spectrum of SOD in-gel tryptic digest, acquired by nano-ES. The amino acid sequence of SOD is shown above the spectrum, with the identified peptides shown in red, blue, green and purple. Peaks corresponding to the identified peptides are labelled in the relevant colours. The spectrum is an average of 34 scans.

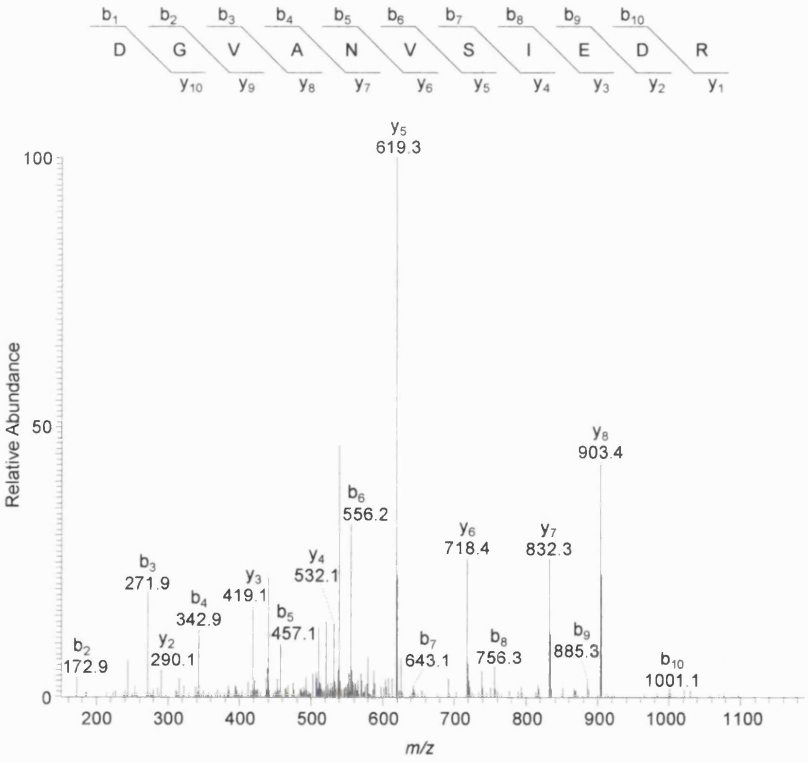


Figure 2 MS/MS spectrum of the tryptic peptide DGVANVSIEDR $[M+2H]^{2+}$ ion of *m/z* 588.4, identified to originate from SOD. The amino acid sequence of the peptide is shown above the spectrum. y and b ions are formed by peptide bond cleavage with charge retention on the C-terminus and N-terminus, respectively. Sample was subjected to in-gel tryptic digestion, with desalting, and analysed by nano-ES (Section 2.12.1).