Protein Adducts at Critical Protein Sites as Markers of Toxicological Risk

Presented by

Paul Getty

to

The University of Glasgow

for the degree of

Doctor of Philosophy

September 2012

College of Medical, Veterinary & Life Sciences

University of Glasgow

Abstract

The formation of conjugates between the electrophilic reactive metabolites of drugs and nucleophilic protein sites is known to be associated with toxicological risk. At present there is no low cost and high throughput means of reliably detecting the presence of drug-protein adducts in vitro or in vivo. The development of a reliable high throughput methodology would facilitate the study of underlying mechanisms of toxicity and prove useful in early screening of potential drug molecules. Assays using liver microsomes and trapping agents such as glutathione are used to produce and detect a wide range of drug reactive metabolites which are then characterised by mass spectrometry. The glutathione trapping is effective for metabolite identifications but, the modification of proteins by means of electrophilic attack on nucleophilic centres often occurs in an enzyme independent manner and is unlikely to be analogous to the glutathione model. In order to create a more suitable model system, three short polypeptides were designed and synthesised. These peptides were incubated with clozapine and human liver microsomes. The resulting metabolite-peptide conjugates were analysed by nanoLC-MS. Results indicated that a characteristic conjugate specific ion at 359.1 Da could be detected for each of the peptides. This data was used to create a precursor ion scan specific for the presence of this characteristic ion.

Protein separation techniques including SCX, Offgel IEF and 1d-gel electrophoresis, in conjunction with LC-MS (with the precursor 359 scan), were applied to microsome prep samples in order to identify modified proteins. Using these approaches some 1700 protein identifications were made, more than 1000 of these were unique hits. The precursor ion scan was found to have poor selectivity identifying roughly 1/3 as many proteins as the information dependant acquisition approach. No drug-protein adducts were identified.

Further to this a novel application of saturation DIGE was applied in order to enrich for the presence of protein adducts. The DiGE approach was used to identify some 15 proteins with apparent change in abundance (fluorescence intensity) between clozapine treated and untreated samples. Spots were excised from the 2d gel digested and analysed by reversed phase liquid chromatography mass spectrometry. The IDA scans identified some 147 unique protein hits, the precursor ion scans identified 18. Again no drug-protein adducts were found.

Biotinylated desmethyl clozapine was metabolised in the human liver microsome assay. Western blotting was carried out on a 2d gel run from an assay sample. The Western membrane was probed using an HRP-Streptavidin probe. Imaging of the membrane revealed the presence of several biotin bearing proteins, many of which were not present in the negative control sample. A print out of the image was used as a map for the excision of modified proteins from a duplicate gel. Digestion and LCMS analysis of the samples revealed the presence of several proteins but no protein-adducts were found.

Table of Contents

List of Tables

List of Figures

Academic acknowledgements

I would like to thank Professor Andrew Pitt, Dr. Nicholas Morrice and Dr. Richard Burchmore for their supervision of this project; Dr. Kathryn Gilroy for her support and insight; Dr. Karl Burgess for his help in maintaining and constructing various HPLC systems and related equipment; Dr. Sarah Cumming and Dr. Susan Horne for their instruction on molecular biology techniques and cheerful dispositions. Thanks to the DMPK and bioanalysis staff at Schering Plough/Merck: Dr. James Baker, Dr. Paul Scullion, Dr. Iain Martin and Dr. Stuart Best.

Personal acknowledgements

I owe thanks to my wife, Xiao Ling, and daughter, Scarlett, for their patience and support during these difficult years; to my friends Mark Crawford, Richard Crawford and Heather Henderson for their support and encouragement; and to Robert Kelly, Kshama Pansare and the other denizens of "the Pitt" for sharing in the joys of PhD studentship. If I have forgotten to mention you by name I apologise and cite Thesis Syndrome as the cause.

I owe special thanks to Dr. Sarah Cumming for making sure that this manuscript made it to the graduate school office whilst I was indisposed.

This project was funded by the EPSRC, BBSRC and a CASE award from Schering-Plough (Merck).

Declaration

I hereby declare that the thesis that follows is my own composition, that it is a record of the work done by myself, and that it has not been presented in any previous application for a higher degree.

Paul Getty

Abbreviations

A	Amperes
ACN	Acetonitrile
AC	Alternating Current
ADR	Adverse Drug Reaction
AmBic	Ammonium Bicarbonate
ANOVA	Analysis of Variance
APAP	Acetaminophen
b-	Biotinylated
BSA	Bovine Serum Albumin
BVA	Biological Variance Analysis
C18	Octadecyl Silica
CID	Collision Induced Dissociation
CyDye	Cyanine Dye
DC	Direct Current
DIA	Differential In-Gel Analysis
DiGE	Differential Gel Electrophoresis
DMC	Desmethyl Clozapine
DMSO	Dimethyl Sulfoxide
DTT	Dithiothreitol
ECD	Electron Capture Dissociation
ECL	Electrochemical Luminescence
EMS	Enhanced Mass Spectrum
ESI	Electrospray Ionisation
EPI	Enhanced Product Ion
ER	Endoplasmic Reticulum
ETD	Electron Transfer Dissociation
FA	Formic Acid
FASTA	FAST-ALL
FT-ICR	Fourier Transform Ion Cyclotron Resonance
GeLC	SDS-PAGE followed by digestion and LC-MS
GSH	Glutathione
GST	Glutathione-S-Transferase
HSA	Human Serum Albumin
HLM	Human Liver Microsomes
HPLC	High Performance Liquid Chromatography
IDA	Information Dependant Acquisition
IDR	Idiosyncratic Drug Reaction
IEF	Isoelectric Focussing
IEX	Ion Exchange
IPA	Isopropyl Alcohol
IPG	Immobilised pH Gradient
KC	Kupfer Cells
Kd	Dissociation Constant
kVh	Kilovolt hours
LC	Liquid Chromatography
LIT	Linear Ion Trap
MALDI	Matrix Assisted Laser Desorption Ionisation
MDF	Mass Defect Filtering
MeOH	Methanol
MeCN	Acetonitrile

MGF	Mascot Generic Format
MHCII	Major Histocompatibility Complex class II
MOWSE	Molecular Weight Search
MuDPIT	Multidimensional Protein Identification Technique
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
m/z	Mass to Charge Ratio
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NAPQI	N-Acetyl-P-Benzoquinone Imine
NCE	New Chemical Entity
NL	Neutral Loss
NSAID	Non-Steroidal Anti-inflammatory Drug
OTC	Over the Counter
P1	Synthetic Peptide 1
P2	Synthetic Peptide 2
P3	Synthetic Peptide 3
PAcIFIC	Precursor Acquisition Independent from Ion Count
PAGE	Poly-Acrylamide Gel Electrophoresis
PAMP	Pathogen Associated Molecular Pattern
PBS	Phosphate Buffered Saline
PEEK	Poly Ethyl Ethyl Ketone
PI	Precursor Ion
pKa	Acid Dissociation Constant
PMF	Peptide Mass Fingerprint
PTM	Post-Translational Modification
PVDF	Polyvinylidene fluoride
QqQ	Triple Quadrupole
RF	Radio Frequency
RP	Reversed Phase
SDS	Sodium Dodecyl Sulfate
SAX	Strong Anion Exchange
SCX	Strong Cation Exchange
SDS	Sodium Dodecyl Sulphate
SMX	Sulfamexazole
SNS	Self-Nonself
SRM	Single Reaction Monitoring
TFA	Trifluoroacetic Acid
TCA	Trichloroacetic Acid
TIC	Total Ion Chromatogram
ToF	Time of Flight
UGT	UDP-glucuronosyltransferase
UV	Ultraviolet
WAX	Weak Anion Exchange
WCX	Weak Cation Exchange
XIC	Extracted Ion Chromatogram

Chapter 1: Introduction

1.1 Drug Metabolism and Toxicity

The production of pharmaceuticals is central to modern healthcare and is an enormous industry in which company's annual revenues generally measure into the billions of pounds (Adams and Brantner, 2006). These companies generate and develop chemical compounds, so called new chemical entities (NCEs), which go on to become commercially available pharmaceuticals for global consumption. Compound generation and testing is formulaic in nature and is carried out in a series of discreet stages including identification of biological targets, mass screening of compounds versus targets, iterative refinement of compounds and preclinical/clinical trials.

Each of the stages represents an investment in time and money and at each stage compounds are eliminated. Classically, the elimination of compounds fits a pyramidal model with a steady loss of compounds and ultimately the emergence of very few successful drugs. The more advanced the stage at which a compound is eliminated, the higher the associated costs. Additionally, compounds eliminated during clinical trials are often flagged due to their toxic effects on human subjects.

The total costs involved in developing a new chemical entity (novel drug) from inception to market regularly exceed $500 million (Adams and Brantner, 2006) and can be compounded by litigation filed by victims of adverse reactions. Ideally, testing should identify unsuitable compounds at the earliest stage possible thereby reducing development costs, laboratory time and human/animal exposure.

In this short review, current methodologies for the early detection of potential drug molecules capable of causing toxicity in humans will be discussed. Particular attention will be given to techniques involving mass spectrometric detection of reactive metabolites of drug molecules.

1.1.1 Drug Development

Much of drug development involves the screening of a library of compounds against relevant biological targets. Compounds that show activity are then subjected to iterations of combinatorial chemistry in which they are subtly modified in order to maximise the efficiency of target interaction. Inevitably, this process often leads to the formation of molecules with detrimental characteristics.

Although structural knowledge can be used to guide compound development, we do not currently possess the knowledge to predict all possible associated toxicities. Careful testing is required in order to identify the effects of a novel drug in vitro, and in vivo.

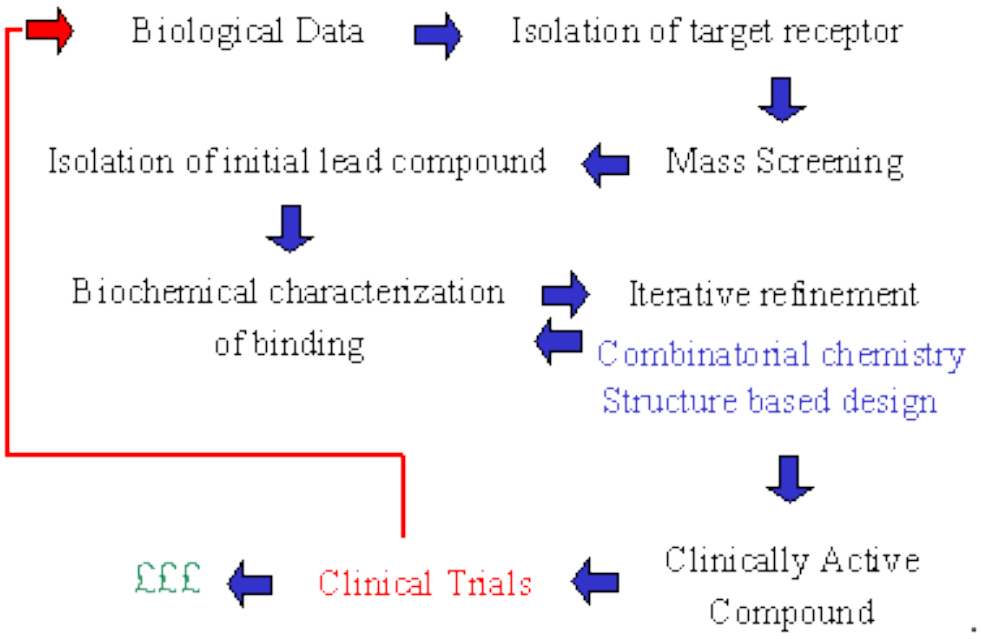


Figure 1. Compounds are selected for their activity against a biological target and are optimised for maximum effect. The clinically effective compounds are then put through pre-clinical and clinical testing in order to ensure their safety.

Adverse drug reactions (ADRs) have a variety of underlying causes; overdose, synergistic effects of drug treatment (polypharmacy) and genetic factors are commonly cited (Nguyen et al., 2006; Hersh et al., 2007). ADRs cover a wide spectrum of severity and can be very difficult to predict. In the United States ADRs are listed as the 4th most common cause of death (Lazarou et al., 1998). The identification of drugs capable of causing ADRs is paramount and begins early in the drug design process.

Typically, adverse reactions are not to the drug molecule itself but to its bioactivated metabolites, further compounding an already complex situation. Drug metabolism is a process by which the body can facilitate the removal of a xenobiotic from circulation. The process typically results in the inactivation/detoxification by way of enzymatic modification. Metabolites of drug molecules, often numerous, must be characterised and included when attempting to define mechanisms for ADRs.

1.1.2 Drug Metabolism

A vast array of xenobiotics can be found in the human body, these foreign molecules originate from sources such as dietary intake and the environment making their way into and through the respiratory tract, gastrointestinal tract and vascular system. These molecules, often with no nutritional value, must not be allowed to accumulate in the body, and therefore undergo elimination. The nature of xenobiotics dictates how they are distributed and partitioned within the body as well as their propensity for elimination. Lipid membranes form distinct compartments at the cellular and subcellular levels; lipid soluble molecules can pass freely through these membranes and gain access to cells and subcellular organelles making the job of regulating their location difficult. In order to combat this the body alters xenobiotics to a more hydrophilic state in which they cannot easily traverse lipid membranes without the aid of selective protein transporters. This allows a greater degree of selectivity regarding the location of the molecules, limiting their access to sensitive sites and making them more amenable to elimination. This process of chemical alteration is known as xenobiotic metabolism.

Metabolism of xenobiotics occurs in two discrete phases. Phase I, or bioactivation, occurs almost exclusively in the liver and is mediated by a range of enzymes, principally, the cytochrome P450 superfamily (CYP450). These monooxygenases can be found primarily in the endoplasmic reticulum of hepatocytes; they catalyze the oxidation of their substrates and require high energy electrons acquired from NADPH. Reactions catalyzed by these enzymes include hydroxylation, dealkylation, deamination, and epoxidation (Burka et al., 1983; Bellec et al., 1996; Boor et al., 1990; Kedderis et al., 1993). The CYP450 enzymes come in a variety of isoforms that are capable of reacting with various different drug types e.g. Zonisamide (1,2-benzisoxazole-3-methanesulfonamide) has been shown to be metabolized to SMAP (2-sulfamoylacetylphenol) by the CYP450 isoform 3A4 (Nakasa et al., 1993); CYP450 isoforms show interspecies variation, partially accounting for the disparity between animal and human drug trials (Jemnitz et al., 2008). Other enzymes including Flavin-containing monooxygenases, alcohol dehydrogenase, aldehydes dehydrogenase and monoamine oxidase are also involved in phase I reactions.

Phase I metabolism acts to convert lipophilic xenobiotics into a more hydrophilic state in order to enhance their clearance from the organism or to make them more susceptible to phase II metabolic processes. This is achieved primarily through oxidation, but reduction and hydrolysis also play important roles (Ahr et al., 1982; Amunom et al., 2011). Reduction, like oxidation, is handled by the cytochrome P450 enzymes, as well as various reductases (Matsunaga et al., 2006), but takes place under anaerobic conditions. Hydrolysis is catalyzed by esterases, amidases and epoxides hydrolases (Mentlein et al., 1980). No change to the oxidative state of the xenobiotic occurs, rather the molecule is cleaved via the uptake of a molecule of water. Hydrolytic reactions are not limited to the liver and occur in many other locations including skin, lung and blood (McCracken et al., 1993).

Phase II reactions comprise the conjugation of glutathione, glucuronic acid, sulfonates or amino acids with the xenobiotics and involve enzymes such as glutathione-S-transferase, UDP glucuronosyltransferase, methylransferase and N-acetyltransferase. Sites of conjugation include carboxyl (-COOH), hydroxyl (-OH), amino (NH2) and sulfhydral (-SH) groups (Booth et al., 1961; King et al., 2000; Lennard et al., 1997). Conjugation results in the production of more polar

molecules with increased amenability for elimination and is often carried out on species oxidized by phase I enzymes. The route or elimination is dependant on the molecular weight of the waste molecule. Higher molecular weights (glutathione conjugates and often glucuronide conjugates) are necessarily excreted in bile; lower molecular weight molecules are excreted in urine. Phase II metabolism also serves to lower the reactivity of metabolites and in some cases neutralises highly reactive metabolites generated during phase I (Dahlin et al., 1984).

It is known that metabolism of drug molecules can be complex and involve the production of many metabolite species. In some cases the metabolites of drugs can have enhanced or altered activity, this is known as bioactivation (Kalgutkar et al., 2005). Bioactivation can be taken advantage of when designing a new drug. A so called pro-drug form with enhanced ADME (absorption, distribution, metabolism and elimination) characteristics can be produced which then relies on the body's metabolic pathways for activation. However, it is also these same pathways that generate unexpected reactive metabolites that cause adverse effects to the organism (Attia, 2010). Highly reactive electrophiles arising from metabolism have been shown to covalently bind protein molecules. These protein-drug adducts, in comparison to native protein, can lose function and have altered routes of clearance (Ute et al., 2001; Jenkins et al., 2008; Crow et al., 2012). Although the products of both phase I and phase II reactions can be electrophilic in nature, phase I products have a greater tendency to be problematic.

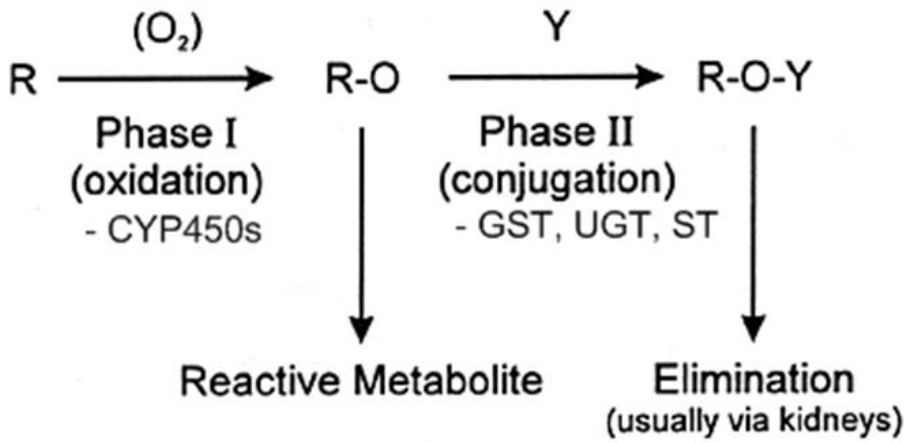


Figure 2. Metabolism of xenobiotics can lead to the formation of undesirable reactive metabolites.

1.1.3 Protein Modifications

Modification of proteins by reactive intermediates is a proposed mechanism in many cases of adverse drug reactions (ADRs). The metabolism of Xenobiotics is responsible for the generation of electrophilic reactive species known to target the nucleophilic thiol group of cysteines, heterocyclic nitrogen atoms of histidine, amino and guanidine groups of lysine/arginine and the phenolic ring of tyrosines (Rubino et al., 2007).

Adduct formation at critical sites can lead to the inactivation of enzymes or disruption of protein-protein interactions (Nelson and Pearson, 1990; Lin et al., 2008). The impairment of some critical proteins could lead to cellular damage and or death. Good candidates for critical target proteins would be any of the detoxification enzymes (Jenkins et al., 2008). Loss of function in these proteins could conceivably lead to a loss of suppression of oxidative stress in the cell and a scenario of runaway damage.

A large amount of work has been carried out on the subject and it has become increasingly obvious that routes of damage are complex and vary from drug to drug (Yukinaga et al., 2007). In many cases, levels of reactive metabolite in the cell dictate the extent of protein-adduct formation and as such the extent of physiological impairment.

1.1.3.1 Cellular Defences

It appears that cellular defences have been acquired to counteract the production of reactive electrophilic species. The highly nucleophilic nature of the cysteine sulfhydral side group makes it a prime target for electrophilic molecules. The cytosolic protein, KEAP1, is rich in cysteine residues (27 with no disulfide bridge formation) and forms a complex with CUL3 and NRF2. In this complex, KEAP1 acts as a sensor of cellular electrophile levels and can either allow NRF2 to, or prevent it from, initiating the production of detoxifying enzymes such as glutathione-s-transferase, heme oxygenase I and CYP450s. (Zhang et al., 2004; Hong et al., 2006; Liu et al., 2005; Satoh et al., 1985).

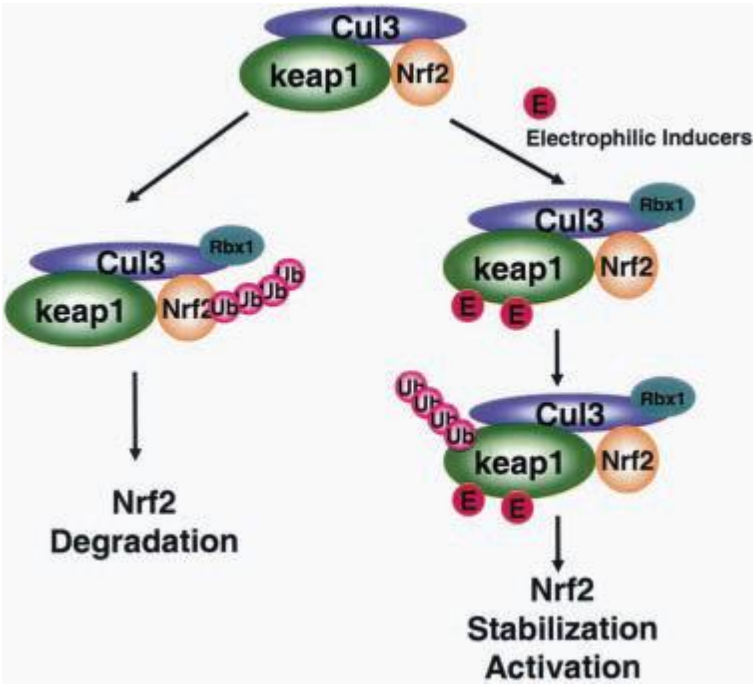


Figure 3. Binding of electrophilic species with keap1 prevents the degradation of Nrf2. Nrf2 can then go on to activate the production of detoxifying enzymes at the transcriptional level (Hong et al., 2006).

In addition to this intracellular defence mechanism is the role played by cells of the acquired immune system. Kupfer cells (KCs), a population of antigen presenting cells within the liver, are responsible for inducing tolerance to protein-drug adducts (Ju, 2009). Tolerance is mediated by KC cells acting as incompetent antigen-presenting cells and acting to suppress T cell activation

through the release of prostaglandins. Despite these measures drug toxicity continues to be problematic.

1.1.3.2 Dose Related Reactions

Adverse drug reactions (ADR), although poorly understood, can be attenuated through careful dosing. Indeed, dosing considerations are taken into account when deciding whether or not to progress a drug's development. A drug known to produce reactive metabolites but with a low therapeutic dose may be considered acceptable for further development (Evans et al., 2004). When considering dose however, it is necessary to take into account factors affecting the activity of Phase I enzymes such as the cyotochrome P450s. Increased activity, either through polypharmacy, genetic polymorphisms or physiological status can increase the formation of reactive metabolites and thus lower the level of dose required to cause toxicity (Sturgill and Lambert, 1997). The over the counter drug, N-acetyl-p-aminophenol (APAP), is a good example of this.

1.1.3.3 APAP metabolism

ADRs arising from APAP consumption are directly related to dose. At therapeutic doses APAP is detoxified mainly by glucuronidation (52-57%) and sulfation (30-44%) (Patel et al., 1990, 1992). An overdose leads to the saturation of the sulfation pathway, diverting more detoxification toward glucuronidation (66-75%) and resulting in a greater formation of an oxidised species known as N-acetyl-p-benzoquinoneimine (NAPQI) (7-15%)(Bessems and Vermeulen, 2001). NAPQI is electrophilic and readily reacts with cysteine sulfhydral groups; this metabolite is cleared from cells by its binding to glutathione and subsequent elimination in the urine. Upon depletion of cellular stores of glutathione, NAPQI begins to covalently bind to cellular protein and leading to severe disruption of normal calcium homeostasis (Tirmentstein and Nelson, 1989) and the subsequent associated necrosis of liver cells seen in APAP toxicity (Zhou et al., 2005; Rinaldi et al., 2002). APAP poisoning is mediated by several CP450 isoforms at low doses but at higher doses is mainly metabolised by CYP2A6 and CYP2E1 (Hazai, 2002).

Despite its hepatotoxicity, APAP remains available for OTC consumption due to its effectiveness and the disparity between its therapeutic dose and toxic dose. Unfortunately, for many other drugs this is not always the case. A very small yet significant number of patients show serious adverse effects with no apparent relation to dose.

1.1.3.4 Idiosyncratic Drug Reactions (IDR)

In contrast to the type of ADR mentioned previously, with a direct link between dose and toxicity and therefore a clear understanding of dose-risk, there exists what are known as idiosyncratic drug reactions (IDRs). The complexity of these often unpredictable adverse reactions is summarised in a review by Ulrich (Ulrich, 2007) in which many known risk factors including age, diet, genetic variation and repeated exposure are discussed. In some cases, the formation of a protein-drug adduct is capable of eliciting an immune response in the patient's body (Gardner et al., 2005; Roychowdhury et al., 2007). This specific response is mediated by antibodies raised when the peptide fragment with a drug adduct (acting as a hapten) is presented. The major antigenic determinant can be either the hapten (drug adduct) or part of the protein to which it is attached. As a consequence the immune system of the patient will begin to actively attack 'self' proteins (Martin and Weltzien, 1994; Kalish, 1995; Weltzien et al., 1996). In order for haptenation to occur however, it is necessary that the reactive electrophilic molecule covalently binds to a protein nucleophilic group (Park et al., 1987).

Generally hypersensitivity reactions involve the blood, liver and skin; presenting as signs such as rash, eosinophilia, fever and anaphylactic shock (Uetrecht, 1999; Smith and Schmid, 2006; Elahi et al., 2004). Agranulocytosis, depletion of granulocytes (basophils, neutrophils and eosinophils), is known to be caused by metabolites of the drugs Clozapine, Procainamide and Vesnarinone (Liu and Uetrecht., 1995). Each of these drugs yield different adduct profiles, although certain proteins are modified in all cases (Gardner et al., 2005). Major tissue targets of IDRs show a correlation to sites of reactive metabolite production

(Roychowdhury, 2007), likely due to the short lived presence of the highly reactive metabolites. There is strong evidence that bouts of inflammation play a major role in many cases of IDR. Exposure to an endotoxin or LPS during the course of treatment with an otherwise non-toxic drug can lead to liver toxicity (Roth et al., 1997).

Drugs known to induce idiosyncratic immune mediated toxicity include the tetracyclic antidepressant Mirtazapine, antiplatelet agent Ticlopidine, diuretic Tienilic acid and the sulfonamide Sulfamethoxazole (Zhou et al., 2005). Sulfamethoxazole (SMX) is an antimicrobial agent and it has been demonstrated that the hydroxylamine- (SMX-HA) and the nitroso- (SMX-NO) derivatives of this drug are capable of forming adducts with proteins. Both metabolites can do so at sub-toxic drug concentrations (Manchanda et al., 2001). Haptenation was shown to be inhibited by the presence of thiols and other antioxidants.

Phenytoin, an anticonvulsant, is known to cause idiosyncratic adverse reactions in 5-10 % of patients (Zhou et al., 2005). Lupus, Steven-Johnson syndrome and toxic epidermal necrolysis are adverse reactions associated with phenytoin. The generation of reactive metabolites and subsequent binding to cellular proteins, several isoforms of CYP450s in particular, leads to the raising of autoantibodies against CYP450s both modified and in their native states.

These examples are chosen to show the range of compounds and represent only a small number of drugs known to be problematic. It should be noted that a simple correlation between reactive metabolite production and pathology is insufficient. As seen previously in the cause of other types of ADR, the presence of drug-protein adducts does not always lead to toxicity or hypersensitivity (Gan et al., 2009; Obach et al., 2008). In the case of acetaminophen no immunotoxicity is encountered despite formation of protein-adducts (Nelson and Pearson, 1990).

Identification of drugs capable of eliciting immune response is compounded by the complexity of the immune system and by the physiological state of patients. An interesting explanation for the occurrence of IDRs has been posited and is known as the danger hypothesis.

1.1.3.5 The Danger Hypothesis (Model)

The danger model was put forward by Polly Matzinger in the early 1990s (Matzinger, 1994) and challenged the long standing SNS (self-nonself) model of immunology outlined by Burnet and Medawar in the 1960s. The SNS model asserts that the immune system actively engages any foreign, nonself, material whilst ignoring anything recognised as self. The danger model maintains that immune response is not mediated through this type of recognition but by activation of immune competent cells by a so called danger signal (Anderson and Matzinger, 2000) via toll-like receptors (Miyake, 2007). The mechanism results in the eliciting of an immune response in reply to the presentation of antigen (self or nonself) coupled with the presence of the danger signals. If an antigen is presented without the danger signal then tolerance to the antigen will occur.

Danger signals must be particular endogenous molecules present upon cell damage or death (Gallucci et al., 1999; Shi et al., 2000) whose presence may be elicited by exogenous molecules such as lipopolysaccharide associated with bacterial infection or so called PAMPs (Pathogen associated molecular patterns). Endogenous danger signal molecules identified so far include adenoside-5′-triphosphate (ATP), Uric acid, hyaluronan breakdown products, transcription factors such as high-mobility group box 1 (HMGB-1) , the S100 protein family and Heat shock proteins (Shi et al., 2003; Rovere-Querini et al., 2004; Melcher et al., 1998). The later 3 protein groups are collectively known as alarmins (Oppenheim, 2007) and are translocated from the nucleus or cytosol to the extracellular space in the event of cell damage or death whereupon they stimulate an immune response.

In the danger model, as applied to idiosyncratic drug reactions, a drug molecule or, more likely, a reactive metabolite acts as a hapten and is presented to helper T-cells via the MHCII receptor. Alarmins or other molecules representative of cellular damage then supply the danger signal and initiate a T-cell mediated immune response.

The capability of many drug molecules or their reactive metabolites to cause oxidative cell damage would make them potentially capable of eliciting an

immune response in line with the danger hypothesis. The oxidative damage and cell death coupled with drug-protein adducts could potentially supply both signals required. Immune tolerance when no danger signal is present would explain why many drugs known to form protein adducts do not go on to elicit immune response.

The question remains as to why immune response in patients to protein-drug adducts is idiosyncratic in nature given the fact that the danger hypothesis only requires that there be antigen fragments and cellular danger signals. In any drug capable of causing an ADR these criteria would be met and therefore should bring about an immune response. There is evidence of factors such as surgery and infection increasing the risk of IDRs, possible through production of danger signals in response to damage caused by physical trauma or there however there is insufficient evidence to suggest that this type of danger stimulation is commonly associated with an increased risk of IDR (Uetrecht, 1999). This may suggest that the immune system has some way of determining the cause of danger signals, limiting the direction of an immune response against molecules directly responsible for cellular damage.

1.1.3.6 Clearance of Protein-Drug Adducts

It has been suggested that a potential indicator of toxicity is the clearance time of drug-protein adducts from the body. A comparison of 1-biotinamido-4-(4′-[maleimidoethylcyclohexane]-carboxamido) butane (BMCC) and *N*-iodoacetyl-*N*-biotinylhexylenediamine (IAB), model electrophiles, was carried out by Lin et al (2008). IAB is known to cause apoptosis in HEK293 cells whereas BMCC does not (Wong and Liebler, 2008). Previous work had indicated that the two electrophiles had distinctly different adduction profiles with only 20% overlap; from this data the assumption was made that IAB must form an adduct with some critical protein in order to initiate apoptosis (Wong and Liebler, 2008).

Experiments revealed that BMCC levels decreases rapidly in cells after exposure, clearance occurs over a period of 4-6 hours. Additionally, the process occurs at a slower rate at lower temperatures suggesting a possible metabolic mechanism.

Enzymatic hydrolysis catalysed by an aminohydrolase is thought to either mediate the release of the adduct moiety or simply remove the means of detection.

It is clear however, that the same mechanism may not apply to adducts formed from other reactive metabolites.

1.1.4 Current Detection Methods

The complex nature of drug metabolism, adduct formation and subsequent toxicity makes identification of diagnostic markers difficult. The identification of metabolites and their interaction with proteins and cellular detoxification molecules provides a great challenge even before the consideration of autoimmune reactions.

Simply identifying proteins prone to adduct formation is a challenge in itself; the scarcity of modified relative to unmodified being a major barrier to detection (Zhou, 2003). Techniques such as radiolabelling or biotinylation of drug molecules, and where available, immunochemistry have been used in conjunction with mass spectrometry in order to identify the occurrence of drug-protein adduct formation. Mass spectrometry is used as a gel based approach is not sensitive enough to detect the level of changes occurring (Tirumalai et al., 2003).

1.1.4.1 Radiolabelling of Drugs and Total Protein Binding

Radiolabelling of drugs allows for a simple and sensitive method of adduct identification. A typical approach (Qiu et al., 1998), carried out in order to identify proteins targeted by reactive metabolites of APAP in liver cells, would follow the steps outlined in figure 4.

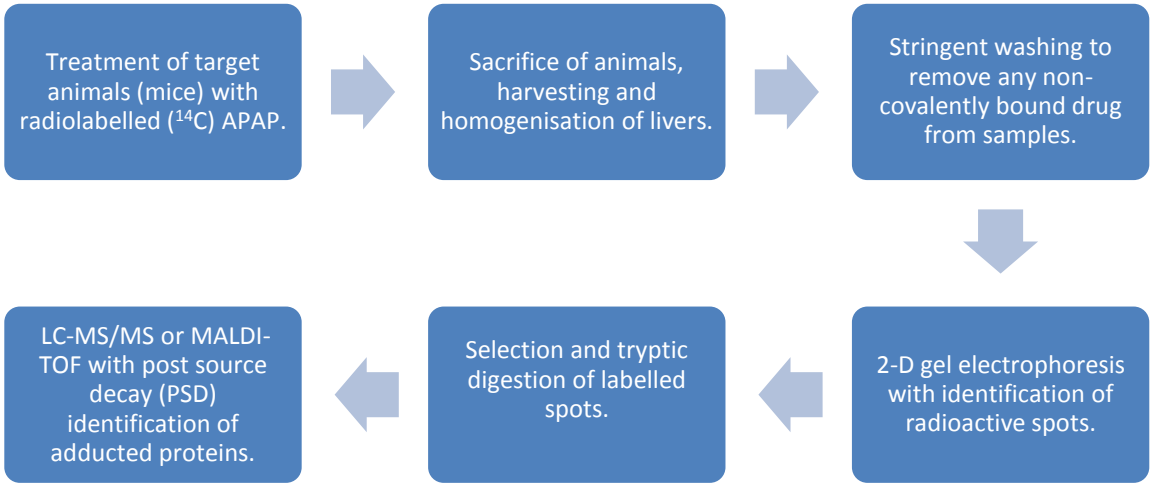


Figure 4. Identification of drug-protein adducts through the use of radio labelled drugs.

When used by Qiu et al (1998), this technique allowed for the identification of 23 adducted proteins but failed to identify others that were previously demonstrated to be present under these conditions (Qiu., 1998).

A major advantage associated with radiolabelling is the ability to quantify the extent of protein adduct formation (Noort et al., 1999).This approach is applied by Merck & Co., Inc. in order to determine whether or not to progress the development of a drug candidate. A carbon-14 labelled analogue of the drug is synthesised and in vitro and in vivo testing is carried out to identify the amount of covalent binding. An upper limit of 50 picomoles drug equivalent/milligram of protein is used to determine the suitability of drugs for progression. The figure comes from an analysis of covalent binding found in the livers of test animals subjected to prototypic hepatotoxic compounds (APAP, furosemide, bromobenzene or 4-ipomeanol) 50 picomoles/ milligram is 1/20th of the dose associated with hepatic necrosis. (Evans et al., 2004).

The limit is not a strict cut-off point however; considerations including the therapeutic dose, term of dosing, severity of adverse effects and the need to fill

an unmet clinical requirement must be weighed before a decision for progression is made.

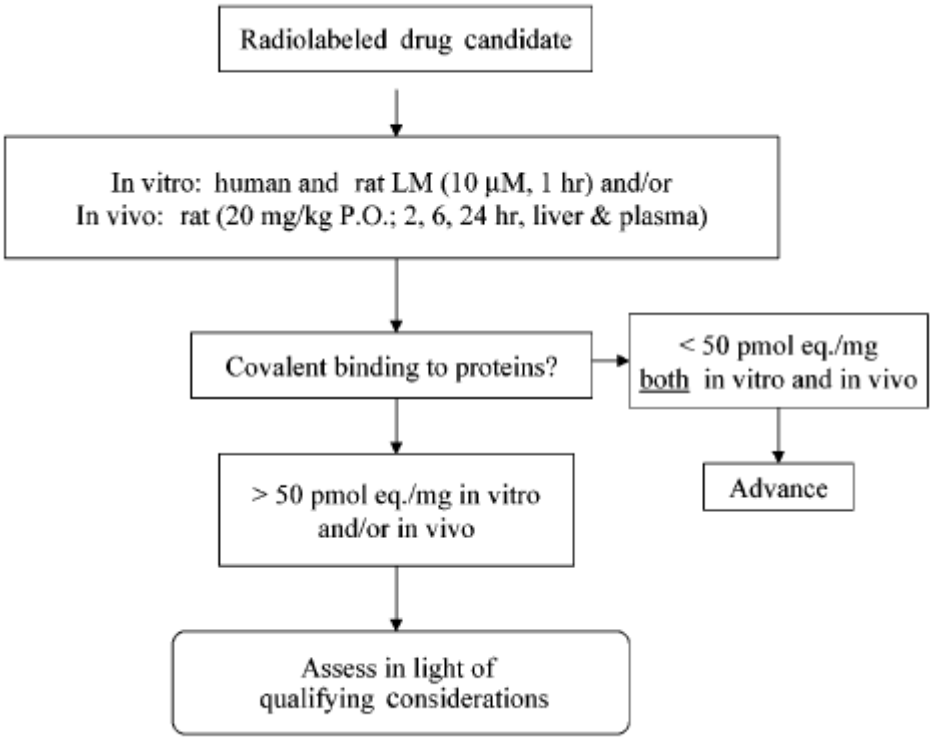


Figure 5. Decision tree regarding the progression of drug candidate as used by Merck. (Evans et al., 2004).

However, drawbacks such as the dangers inherent in radiation handling and the prohibitive cost of synthesising radiolabelled drugs make the technique less appealing (Evans et al.,2004). The technique also lacks in the ability to clearly identify adducted proteins. Gel spots with a radiolabel undoubtedly harbour these adducts but are likely to contain many more proteins besides. In gel digestion of spots and subsequent MS analysis will result in identification of many possible false positives. Depending on the level of modification present it may not be possible to directly identify modified peptides.

1.1.4.2 Biotinylation of Drugs

Affinity tagging has been used in xenobiotic covalent binding studies in order to enrich modified peptides from complex samples. A study by Shin et al used 1-biotinamido-4-(4'-[maleimidoethylcyclohexane]-carboxamido) butane (BMCC) and

N-iodoacetyl-N-biotinylhexylenediamine (IAB) labelled with biotin to identify electrophile sensitive proteins (Shin et al., 2007). A shotgun proteomic approach allowed for the identification of specific residues forming adducts. Protein targets included xenobiotic metabolising enzymes, enzymes of lipid metabolism, chaperones and ion transporter proteins.

Using this method it is possible to identify not only the proteins that are susceptible to modification but the site of adduct formation. By comparing the adduction profiles of BMCC (associated with toxicity) and IAB (no toxicity) we can begin to see that many different proteins are adducted in each case with a small overlap. From this the idea of so called 'Critical proteins' emerges; the premise being that adduction of specific proteins will determine the toxicity of a particular reactive metabolite. Data obtained from experiments like this one can single out protein targets for further investigation allowing for the characterisation of mechanisms of toxicity.

Additionally, work carried out by Dennehy et al demonstrated the affinity of cysteine thiol groups for electrophilic adduction using a biotin tagged electrophile system. They were able to identify 539 protein targets and 897 peptide targets using this method. However, only 20% of these proteins were adducted by both electrophiles (Dennehy et al., 2006). This seems to indicate that the nature of the electrophile is more important than the high reactivity of thiols. It is possible that these proteins are sensitive to adduction as they play a role in cellular sensing of oxidative stress.

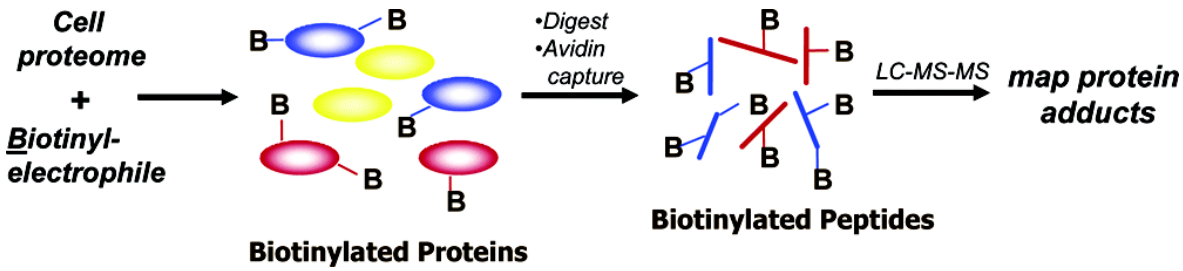


Figure 6. (Dennehy et al., 2006) Known electrophilic molecules were tagged with biotin and allowed to react with cellular proteins. These proteins were then enriched, digested and subjected to LC-MS-MS.

Biotin tagging and subsequent affinity purification provides a valuable tool for the characterisation of selected electrophiles and their protein binding partners. In contrast to radiolabelling, biotinylation is much simpler and comparatively inexpensive. The false positives detected as a by-product of 2d gel separation are eliminated in this affinity purification based technique. This method is useful in its ability to identify large numbers of protein targets which may help in the elucidation of mechanisms behind covalent binding of particular targets and toxic outcome. However, as a screening tool it is limited. The addition of a biotin tag to a small molecule is highly likely to alter its natural passage through a complex biological system. Altered penetration, metabolism and elimination are likely to create substantial differences between tagged and untagged molecules.

1.1.4.3 Immunoblotting of Protein-Drug Adducts

This method has been employed in the identification of protein adducts formed by the reactive metabolites of many xenobiotics including diclofenac, APAP and halothane (Satoh et al., 1985; Witzmann et al., 1994; Hargus et al., 1994). Targeting can be specific to particular drug-protein adducts or simply a means of concentrating a particular protein known to be susceptible to adduct formation (Hoos et al., 2007). Immunoblotting requires the availability of antibodies with sufficient specificity and sensitivity, limiting its usefulness in the identification of the many and varying modifications associated with adduct formation.

1.1.5 Model Systems

It is generally accepted that there is no animal model that can be used for humans and that current knowledge cannot accurately correlate covalent binding of reactive metabolites to toxicity. At present the best approach is to eliminate potentially problematic compounds from development as early as possible. New molecules are tested against trapping agents such as glutathione (GSH) and cyanide in order to identify reactive intermediates by subsequent LC-MS/MS or NMR (Evans et al., 2004). Modification of the chemical structure of the

molecule is made in the attempt to negate the production of these reactive metabolites.

The application of mass spectrometric analysis to the problem of reactive metabolite formation and protein adduction has yielded the development of various highly useful techniques (Wen and Fitch., 2009).

1.1.5.1 Chemical Oxidation of Drugs

It is possible to simulate the bioactivation of drug molecules using an extremely simple chemical oxidation step. Silver (i) oxide has been used to generate N-acetyl-p-benzoquinonimine (NAPQI), a reactive metabolite of APAP, in an in vitro setting which allowed for the subsequent detection of protein-drug adducts (Bessems et al., 1996; Damsten et al., 2007). Betalacotglobulin (BLG) was incubated with the NAPQI and the resulting adducts detected following tryptic digestion of the protein followed by liquid chromatographic separation and tandem mass spectrometric analysis. Adducts were identified by searching for known peptides associated with the tryptic digestion of BLG with the additional mass associated with the NAPQI adduct. This system provides a platform for basic study of adduct formation without the problems inherent in more complex biological systems. However, in order to be truly useful the complexities of a biological system must be incorporated into any model system.

1.1.5.2 Liver Microsome Based Assays

The liver carries out the vast majority of xenobiotic metabolism as well as vital functions including red blood cell degradation, glycogen storage and hormone production. It contains a wide range of enzymes responsible for drug metabolism which include the cytochrome p450 family, glutathione s-transferases, UDP-glucuronosyltransferases, sulfotransferases and N-acetyltransferases. The organ is found in all vertebrates and its functions cannot yet be fully emulated. Liver microsomes, both human and animal, are used as an in vitro means of metabolising drugs. These preparations consist primarily of ER with lesser

contributions from lysosome, nuclear membrane, cytoplasm, peroxisomes and plasma membranes. They contain high amounts of cytochrome P450s, UGT, GST and other xenobiotic metabolising enzymes. Microsomes represent a simple and effective system for the metabolism of xenobiotics in vitro and are often used to analyse the metabolism of drugs. Testing of microsomes for specific activities of CYP450 isoforms is carried out in order to maintain control between lots. Drugs are typically incubated in a microsome preparation which is spiked with the tripeptide glutathione. The highly nucleophilic sulfhydral group found in reduced glutathione acts as a trap for electrophilic species. Electrophiles that bind to GSH molecules can then be identified and the metabolites characterised.

The vast majority of work carried out currently on protein-drug adduct formation and reactive metabolites of drugs involves the use of human or animal liver microsomes for the metabolism of test compounds.

1.1.5.3 Hard and Soft Electrophiles

The nature of particular species of reactive metabolites effects their interactions with other molecules. Metabolites such as quinones, quinine imines, iminoquinone methides, epoxides, arene oxides, and nitrenium ions (Yan et al., 2007) are termed soft i.e. molecules with functional groups that are characterized as having a large radius and are easily polarized. Hard electrophiles have functional groups with a small radius and are difficult to polarize, aldehydes are the most common metabolites of this type. Based on the "hard and soft acid and bases" concept, hard electrophiles react more strongly with hard nucleophiles and soft electrophiles react more strongly with soft nucleophiles (Pearson, 1963).

Consideration must be given to this when attempting to identify reactive metabolites of a potential drug. Glutathione trapping preferentially identifies the production of soft metabolites as the sulfhydral group of cysteine, functional site, is a soft nucleophile. In an attempt to rectify this, and so allow for the detection of hard electrophiles, work was carried out using a "bifunctional" trapping agent *γ*-glutamylcysteinlysine (*γ* GSK) (Yan et al., 2007). The amine of

lysine in this molecules acts as the “hard” nucleophile to trap “hard” electrophilic metabolites. Using neutral loss scanning Yan et al., were able to demonstrate that this molecule was capable of simultaneously trapping both classes of reactive metabolites.

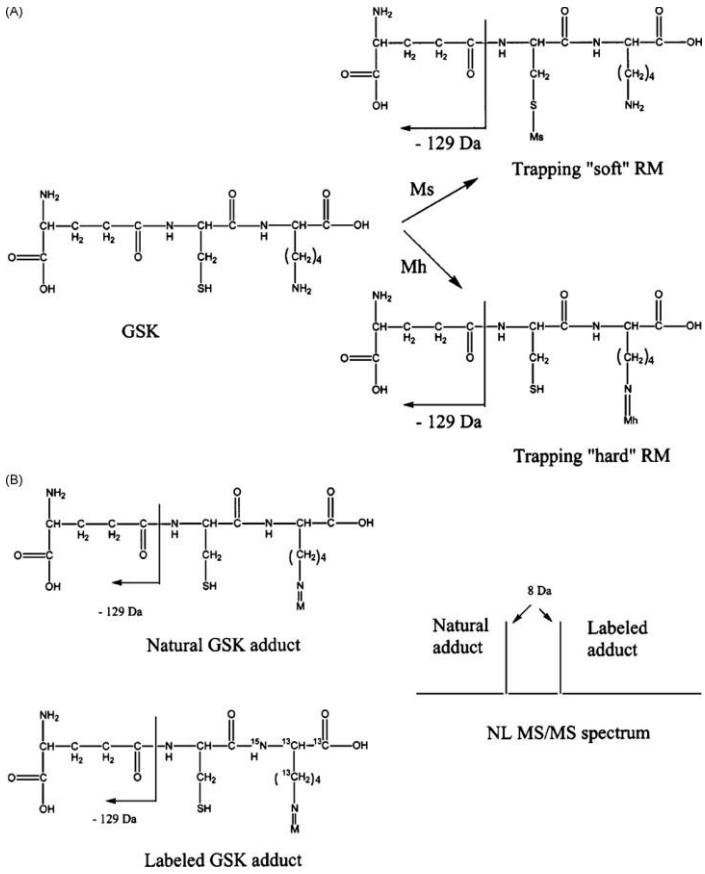


Figure 7. (A) Detection of hard and soft electrophiles using GSK as a trapping agent. (B) Verification of adduct identification through the use of GSK* (γ glutamylcystein-$^{13}C_6$-$^{15}N_2$-lysine) to rule out false positives (Yan et al., 2007).

The nucleophilic groups SH, NH and OH occur repeatedly in the biopolymers DNA and protein. These groups represent a spectrum of nucleophilicity ranging from soft SH to hard OH and intermediate NH. The terms soft and hard refer to the charge density of the nucleophiles but more specifically to their polarisability i.e. ability of their valence electron shells to deform. The rate of adduct formation

between hard/hard nucleophiles/electrophiles or soft/soft nucleophiles/electrophiles is greater than that of hard/soft nucleophiles/electrophiles. Bonding between similar types produces an intermediate state with a much lower potential energy MO than bonding between dissimilar species thereby favouring the reaction (Coles, 1984; Pearson 1963).

The attack of nucleophilic sites by electrophilic metabolites leads to the formation of drug-protein adducts by way of a substitution or addition mechanism. The nature of both electrophile and nucleophile are important in determining the formation of adducts. Protein modification is more likely to occur through attack by softer electrophilic species with favourable reactions with NH2 and SH groups (Parthasarathi, 2004).

Adduct formation at critical sites can lead to the inactivation of enzymes or disruption of protein-protein interactions (Nelson and Pearson, 1990; Lin et al., 2008). A large amount of work has been carried out on the subject and it has become increasingly obvious that routes of damage are complex and vary from drug to drug (Yukinaga et al., 2007). In many cases, levels of reactive metabolite in the cell dictate the extent of protein-adduct formation and as such the extent of physiological impairment.

1.1.5.4 Synthetic Peptides

Three short polypeptides were designed and synthesised, each peptide was N-terminally biotinylated. The design of each met with the following criteria:

i. Must contain a cysteine residue
ii. Must be a tryptic digest fragment of a protein of interest
iii. Must not contain a basic residue near to its midpoint
iv. Must contain at least 6 residues
v. At least one peptide should contain a lysine residue

Protein sequence information for Cytochrome P450s and KEAP1, proteins involved in metabolism and cellular defences against oxidative stress and documented targets of electrophilic species, were subjected to theoretical

tryptic digests. It was from this data that the synthetic peptides were selected. Biotinylation of these peptides should allow for their recovery from a complex background i.e. the liver microsome assay. The use of biologically accurate polypeptides is useful for several reasons. The proteins selected all have important roles in metabolism and cellular redox (reduction-oxidation) regulation. If metabolite-synthetic peptide adducts are formed it may indicate that these particular proteins are susceptible to attack. Additionally it will be possible to automatically identify the conjugates using the Mascot server in combination with a genomic protein database such as Swissprot.

1.2 Separation of Complex Protein Mixtures

1.2.1 Liquid Chromatography

The separation of molecules within mixtures based on their physicochemical properties is known as chromatography. A number of different techniques exist that allow for separation based on size (Dean, 1980; Dondi et al., 2002), hydrophobicity (Karger et al., 1976;Vailaya, 2005; Vailaya and Horvath, 1998), chiral conformation (Gholami et al., 2009; Narayana et al., 2003; Lipka et al., 2005) and affinity binding(Santucci et al., 1990; Tseng et al., 2004; Verdoliva et al., 2002). Each of these techniques requires the interaction of an analyte-containing mobile phase and an immiscible stationary phase with appropriate characteristics. Interactions with the stationary phase alters the time taken for molecules to traverse the column, molecules with favourable interactions with the stationary phase take longer to pass through. The time taken for molecules to elute from the column is known as retention time. Good chromatographic separation requires that molecules within the mixture elute with sufficiently different retention times and that their elution profiles (peak areas) are distinct.

Detection of analytes on elution from the column is routinely carried out by UV absorption measurements or mass spectrometry. Ideally, analytes should have sharp, symmetrical peaks. Height equivalent theoretical plates are an abstract means of evaluating a column's efficiency. Plates represent hypothetical regions in which the mobile phase and solid phase are in equilibrium, the greater the number of these plates, i.e. the smaller the plate height, the greater the

efficiency of separation. The number of theoretical plates can be calculated as follows:

$$N = \frac{L}{HETP}$$

N = number of theoretical plates
L = column length
HETP = Plate height

Column efficiency can be affected by factors including column length, particle size, packing quality, flow, dead volumes and retention factor.

Plate height can be calculated using the Van Deemter equation (van Deemter et al., 1956):

$$H = A + \frac{b}{u} + C.u$$

H = plate height (HETP)
A= eddy diffusion
B= longitudinal diffusion
C=resistance to mass transfer coefficient
u=linear velocity (flow)

The Gaussian curve is created by a distribution of retention time within a single species passing through a column. This variance is described in the terms of the Van Deemter equation with a higher H value being indicative of greater variances and as such broader peak widths.

Eddy diffusion (A)

Eddy diffusion describes the movement of molecules through the column along different paths through the stationary phase. Packing of the column, particle size and morphology are the major contributors to path length of analyte. The smaller the particle sizes the less variance in path length. Packing particles with smoother surfaces contribute less to differential path length than do those with rougher surfaces.

Longitudinal diffusion (B)

Analyte molecules diffuse throughout the mobile phase setting up a concentration gradient independent of the flow direction. Longitudinal diffusion is greatly affected by flow rate; as flow rate increases the effect of longitudinal diffusion (increases peak width) is diminished. Other factors affecting longitudinal diffusion include diffusion coefficient of the analyte in the mobile phase, mobile phase viscosity, temperature and the type of analyte (molecular mass).

Mass transfer (C)

Mass transfer occurs within each phase and between the two. Mass transfer in the mobile phase is effected by the differing velocities of analyte depending on their proximity to mobile phase or column wall. Analyte in close proximity to either of these moves with a lower velocity than analyte further away. In the stationary phase analyte is retained depending on its specific interactivity with that packing. As analyte travels the length of the column there is a constant exchange between mobile and stationary phase brought about by equilibration as the Gaussian profile of the analyte in the mobile phase.

Mass transfer is dependent on the speed of the partition coefficient (varies between molecules of the same analyte depending on their physical position). Using smaller packing particles results in decreasing the importance of mass transfer.

1.2.1.1 Reversed Phase Chromatography

Early chromatography columns, so called normal phase columns, were packed with unmodified silica or alumina resins, this type of stationary phase interacted strongly with hydrophilic molecules. In contrast, RP columns are packed with silica beads functionalised with alkyl chains of various lengths and separates molecules based on hydrophobic interactions. The more hydrophobic the molecule, the greater the retention time. Peptide separations are routinely carried out using octadecyl carbon chain (C18) bonded silica packed columns and a gradient of increasing polar mobile phase. The gradient of the mobile phase

can be tuned to enhance the separation of molecules with any given level of hydrophobicity. RP separation is typically used in direct conjunction with mass spectrometric analysis; samples loaded onto a high capacity C18 trap can be retained and thoroughly washed prior to separation and MS analysis, ionic salts in particular must be removed as they can cause a problematic level of ion suppression during electrospray ionisation (Annesley, 2003; Mallet et al., 2004). The compatibility of the mobile phases used in RP-chromatography with MS analysis confer a second advantage in the coupling of the techniques.

1.2.2 Difference Gel Electrophoresis (DiGE)

Differential gel electrophoresis allows for the direct comparison of multiple protein populations (samples) on a single 2 dimensional polyacrylamide gel. Up to three distinct samples can be loaded into a single gel; typically, a control sample, a treated sample and a pooled sample. The pooled sample contains an equal volume of both the control and treated and acts as a standard allowing for direct comparisons across multiple gels.

Differentiation between multiple samples in a single gel is made possible by dying proteins with 3 spectrally distinct fluors. Cy2, Cy3 and Cy5 (cyanine dyes) are used to label the separate samples which are then combined and run on standard 2D gels. Importantly, the dyes are both mass and charge matched to ensure that labelled proteins migration along the 1st dimensional pH gradient and their subsequent travel through the 2nd dimension of acrylamide gel do not differ dependant on which dye is applied.

Minimal labelling dyes are functionalised with an NHS ester group which reacts to form an amide linkage with the epsilon amino acid of lysine. As a consequence of the dye: protein ratio approximately 3% of proteins in the sample are labelled, each at a single lysine. The single positive charge of the dye replaces that of the lysine to which it binds; this ensures that the pI of the protein remains unaltered.

Saturation labelling dyes are functionalised with a maleimide group which reacts to form a covalent bond with the thiol group of cysteine. The saturation dye:

protein ratio is designed to allow complete labelling of reactive cysteines on all proteins. As currently only two of the dyes (Cy3 and Cy5) are available with the maleimide reactive group it is only possible to run two distinct samples on a single saturation DIGE gel. In order to create inter-gel consistency a pooled sample must be run on each gel along with either a treated or untreated sample.

In both cases the gels are then scanned using the appropriate wavelengths and the composite images subjected to software based spot matching. Differences in intensity between the dye pairs in each of the samples can then be normalised using the internal standards and a statistical analysis of changes can then be carried out.

1.2.3 Ion Exchange Chromatography (IEX)

Ion exchange chromatography can be divided into 4 categories; strong cation exchange (SCX), weak cation exchange (WCX), strong anion exchange (SAX) and weak anion exchange (WAX). Separation is based on Coulombic interactions between a charged mobile phase and oppositely charged stationary phase (Paull and Nesterenko, 2005). The cation exchange based columns utilise sulfonic acids (SCX) or carboxylic acids (WCX) functional groups to interact with positively charged proteins/peptides in a highly acidic (pH 2-3) and weakly ionic mobile phase. Anion exchange based columns utilise trimethylammonium groups (SAX) of primary, secondary or ternary amino groups (WAX) in a basic (pH 8.0) and weakly ionic solution. For both cation and anion exchange the elution of peptides/proteins is brought about by increasing the ionic strength of the mobile phase, molar amounts of sodium chloride, potassium chloride or ammonium sulphate are required. The extremely high levels of non-volatile ionic salts make the technique incompatible with mass spectrometry due to the potential for massive ion suppression.

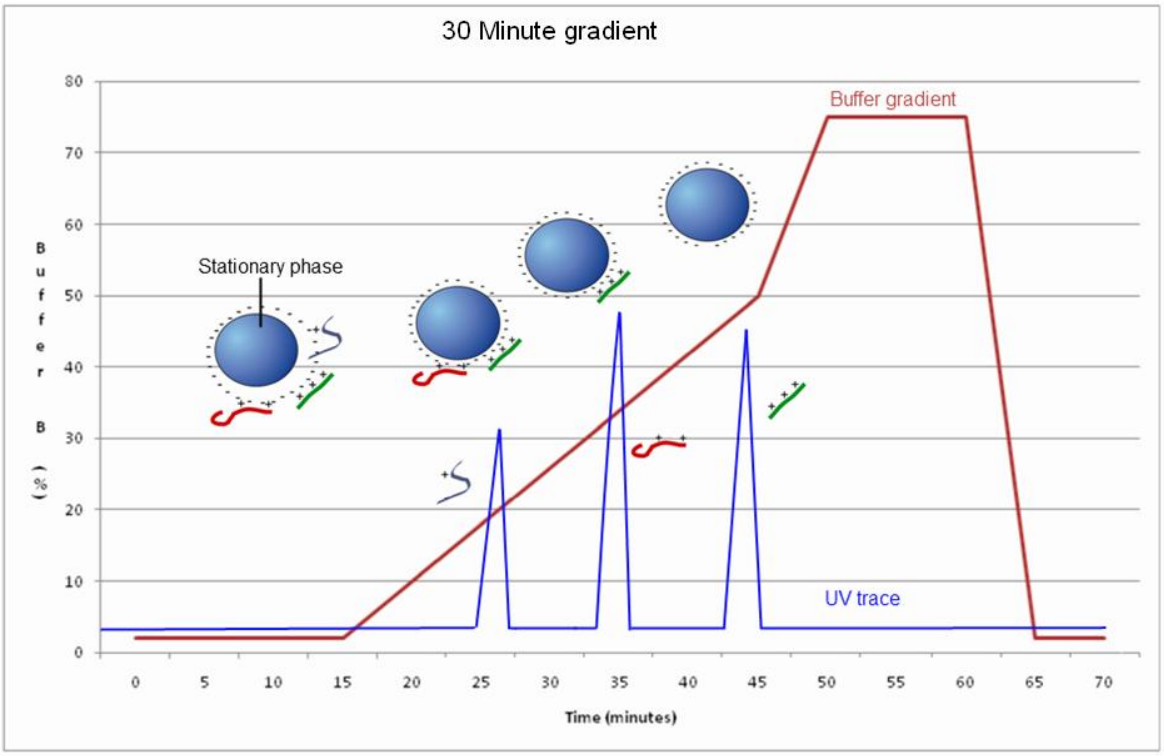


Figure 8. A simplified strong cation exchange gradient run involving 3 peptide species each with a different charge state. As the ionic strength of the buffer increases the Coulombic interactions between peptides and the stationary phase are disrupted. The peptides bearing fewer charges are eluted first.

1.2.4 MuDPIT (Multidimensional Protein Identification Technology)

Ion exchange chromatography is orthogonal to reversed phase liquid chromatography and as such the two techniques can be used together to produce a combined high resolution method for separation of analytes (Mohammed and Heck, 2011). MuDPIT (multidimensional protein identification technology) can be carried out online or offline with the reversed phase separation. In an online configuration analytes are eluted from the SCX column with multiple salt steps. At each step the analytes are loaded directly into the RP column and are separated by hydrophobic interactions before being introduced into the mass spectrometer for analysis. A two stage column with first stage SCX and second stage RP is often used in the online mode (Liu et al., 2006; Kang et al., 2005).

In the offline mode the SCX separation is carried out using a mobile phase with a gradually increasing ionic strength. Fractions are collected and undergo buffer exchange prior to loading on a RP column and subsequent MS analysis. Offline MuDPIT affords higher resolution separations and has been shown to provide a superior degree of protein identifications (Gokce et al., 2011).

1.2.5 Offgel Isoelectric Focussing

Isoelectric focussing allows for the separation of proteins or peptides based on their isoelectric points (pIs)(White and Cordwell, 2005). The technique is commonly used as the first dimension of separation in a 2d-PAGE experiment and is orthogonal to the size based separation of PAGE. The Offgel apparatus allows for the same degree of separation but with enhanced recovery capability. The IPG strip is separated into a number of discrete reservoirs covering portions of the IPG strip pH gradient. Separation proceeds with the application of an electric current along the length of the IPG strip. Proteins/peptides migrate along its length until reaching the point on the pH gradient at which they are in their neutral (uncharged) form. At this point the lack of charge prevents further electrophoretic migration of the molecules. After separation each well contains proteins/peptides with pIs relating to the underlying portion of the IPG strip. A small current is maintained post-separation in order to prevent the diffusion of proteins/peptides along the length of the strip.

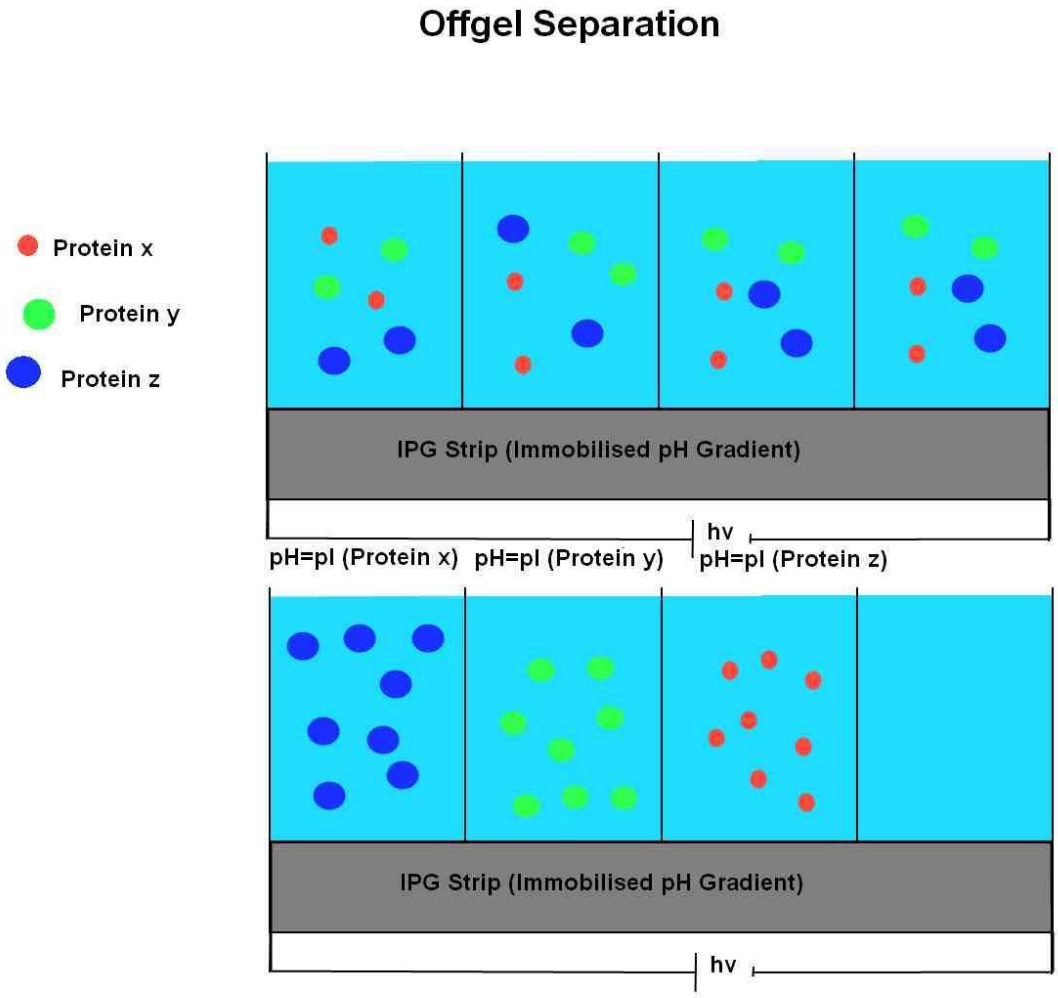


Figure 9. Offgel separation. Three protein species are shown before (top) and after (bottom) separation. Proteins (or peptides) can diffuse along the length of the IPG strip and accumulate in reservoirs above the point on the strip at which the pH causes the protein to lose all charge.

1.3 Mass Spectrometry and the Identification of Proteins

Mass spectrometry allows for the determination of a molecule's mass to charge ratio (m/z). All instruments share the same basic components; an ion source, a mass analyser and a detector. Ions are created and/or accelerated into the mass spectrometer via the ion source; ions with a particular m/z are selected by the mass analyser and accelerated towards the detector. In this study electrospray ionisation (ESI), pioneered by Fenn et al in the 1980s (Fenn et al., 1989) for the analysis of large macromolecules (e.g. proteins) was applied. This technique allows for the ionisation of macromolecules without necessarily causing their

fragmentation. In ESI, liquid phase analyte is converted to its gas phase by a process of desolvation and columbic explosion. Desolvation is driven by evaporation of solvent due to heating and exposure to a nitrogen gas stream. A large potential difference between the emitter or sample stream and a grounded counter-electrode is used to convert the sample stream into a fine aerosol directed towards the mass spectrometers inlet orifice. The fine droplets produced are then thought to undergo Coulombic explosion as desolvation leads to an increasing surface charge. Eventually, single gas phase ions are produced and accelerated into the mass spectrometer for analysis (figure 10).

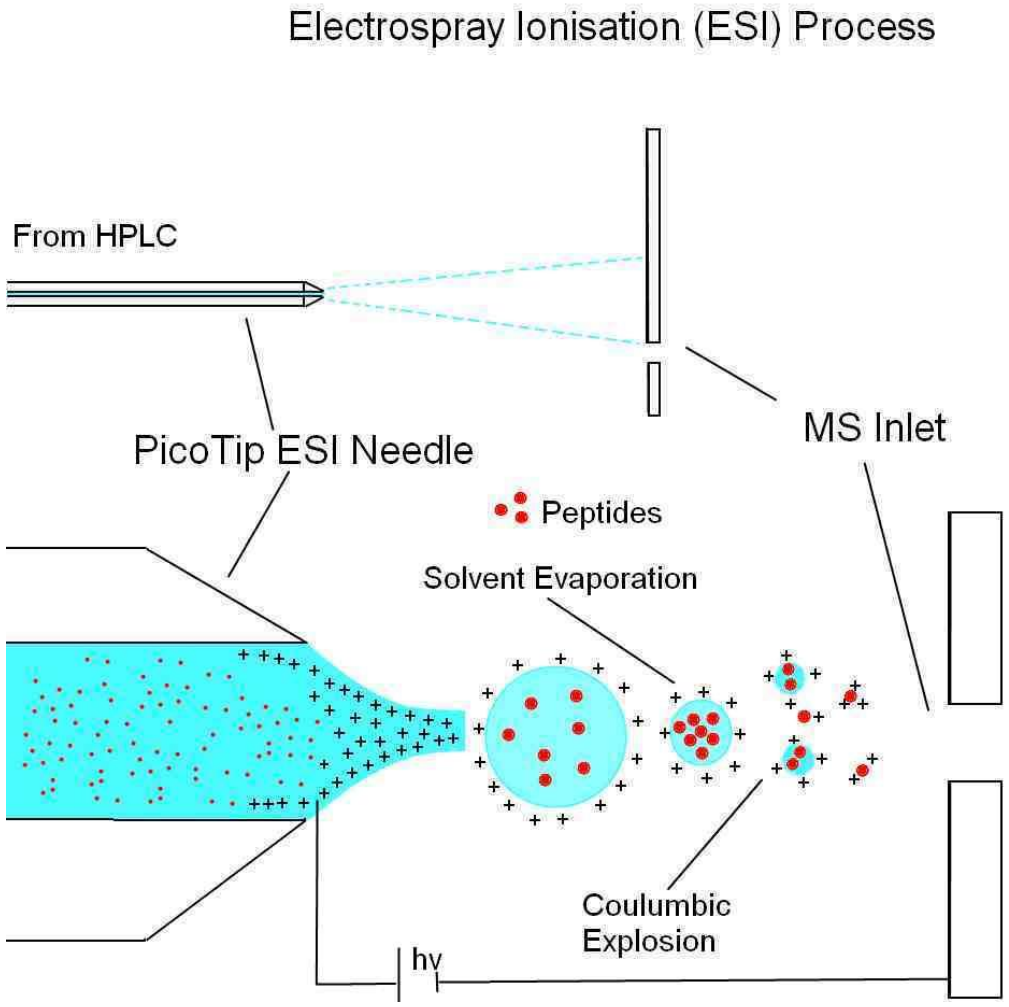


Figure 10. The ESI process. Relatively large droplets of sample (solvent and solute) form and rapidly dry. As the surface area decreases the building charge density reaches a critical state (The Rayleigh limit). A Coulombic explosion causes fission of the droplet. Further drying yields single solute molecules of various charge states.

1.3.1 Mass Spectrometry and the Fragmentation of Ions

The m/z of an intact protein holds useful information but in order to maximise the amount of data collected, fragmentation of the protein must be carried out. In this work, low energy (<100 eV) collision induced dissociation (CID) was used. CID generates so called b and y ions from the parent proteins (Johnson et al., 1987).

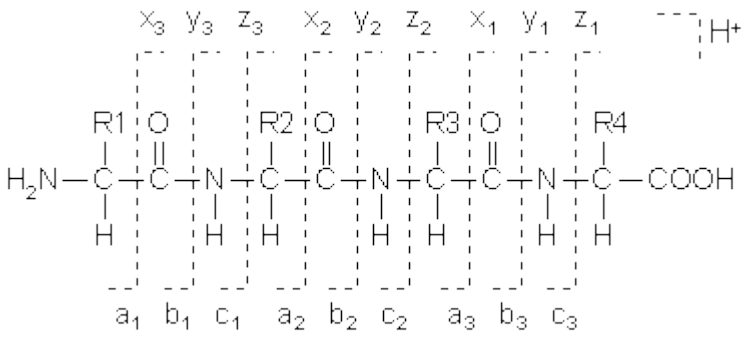


Figure 11. Fragmentation of parent ion (polypeptide) to form named daughter ions. Low energy CID produces b and y ions.

From the m/z data produced it is then possible to obtain protein amino acid sequence information.

1.3.2 Identification of proteins

The goal of a typical proteomics based mass spectrometry analysis is to acquire protein identifications or information regarding post translation modifications. Mass spectrometric analysis affords a fast and reliable method for protein identification. The raw data from mass spectrometers contain information about ion masses, intensities and charge states. This so called 'peak list' is submitted to a search engine for interrogation of genomic databases in order to identify which proteins best match the data. Commonly used search engines include Mascot (Perkins et al., 1999), SEQUEST (Eng et al., 1994) and OMSSA (Geer et al., 2004). Each search engine can search a variety of protein databases, the most

commonly used being Swissprot, MSDB and NCBInr. Swissprot is well curated and cross linked with many other databases; protein sequences in Swissprot are non-redundant rather than non-identical (MSDB and NCBInr) and as a consequence tandem MS searches may return fewer matches and fewer false positive errors.

1.3.2.1 Peptide mass fingerprinting

Proteins are typically subjected to a protease based digestion before being analysed by mass spectrometry. This process is highly specific, generating a set of peptide fragments based on cleavage rules known for each distinct protease. In the case of a trypsin based digestion, proteins are cleaved at the carboxyl side of the amino acids lysine and arginine, except when followed by a proline. Using this information an *in-silico* digestion of a protein database is carried out to yield all expected peptide fragments. It is then a matter of comparing experimental data with this modified database and matching observed masses with their predicted counterparts.

The identification of proteins from their peptide fragments is known as peptide mass fingerprinting (PMF). The matching process is not trivial and should give some indication of the statistical relevance of a match. The various search engines employee different approaches (1.3.3). Tandem MS information can be used to further distinguish between peptides with identical masses.

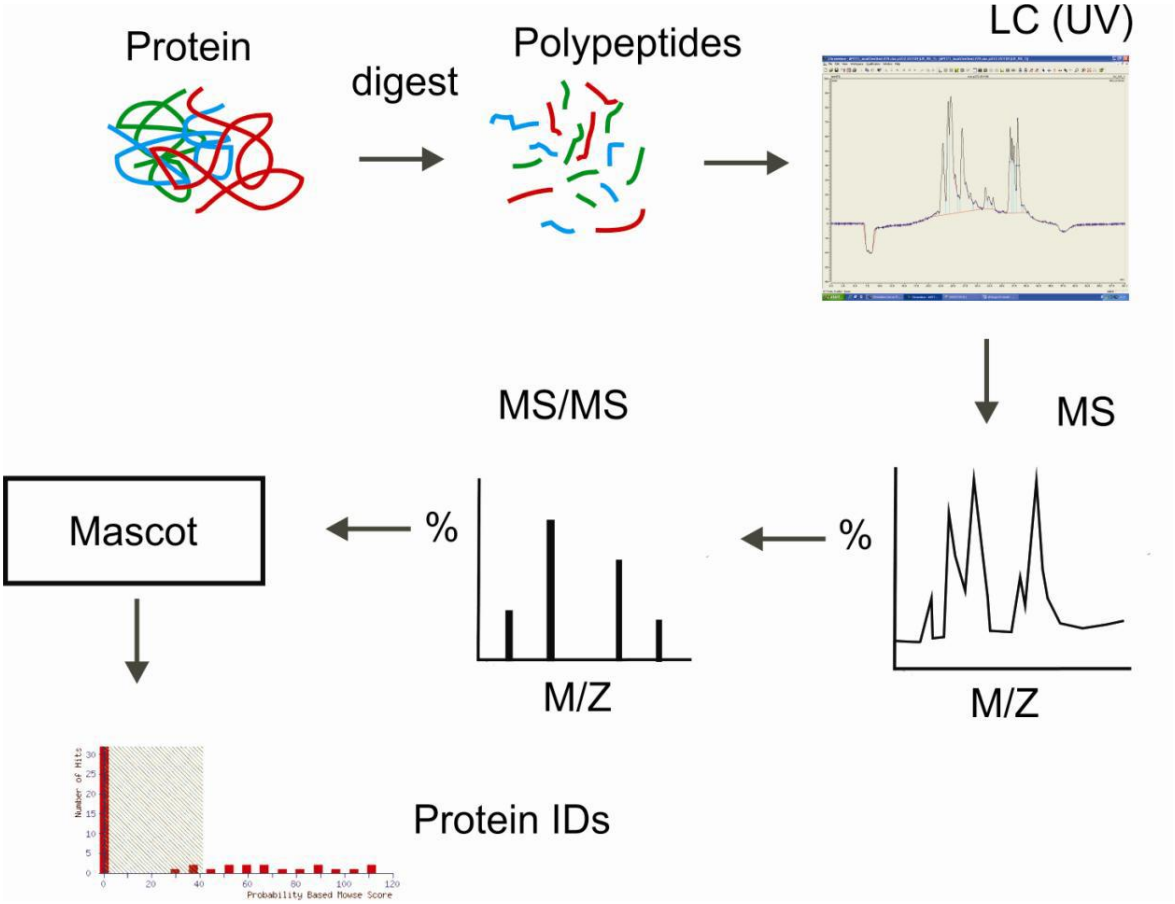


Figure 12. A typical bottom up (shotgun) proteomics approach. Proteins are digested and separated by liquid chromatography before mass spectrometric detection, analysis and database searching for identification. Here Mascot represents a proteomic search engine and is one of many such search engines.

1.3.3 Search Engines

Due to the amount of experimental data, spectra, gathered in a high throughput mass spectrometric investigation manual interpretation is impractical. Data formats vary depending on the software used for acquisition; this is often proprietary and varies by instrument manufacturer. Information regarding peptide m/z values, intensities, ms/ms fragments etc are encoded in these files and can be used for protein identification. The file containing the data of interest is submitted to a suitable search engine for searching against a protein sequence database. There are a variety of both protein search engines and protein sequence databases and a more in depth explanation of search engines follows (databases are more fully discussed in section 1.3.4).

Search engines are used to correlate mass spectrometric data to peptide fragments typically produced by the *in silico* digestion of a protein database (Pappin et al., 1993; Geer et al., 2004; Eng et al., 1994) . As each individual protein yields a unique profile of peptide fragments, this data can then be used to assign protein matches. Post translational modifications can greatly change the mass of peptides and therefore it is necessary to include the potential modifications as terms in the database search. Other information required for database searching includes the enzyme used for digestion, peptide and MS/MS tolerance of the instrument (to help narrow the window for potential matches),the type of instrument used (to determine which type of ions series should be detected) and the number of missed cleavages (to allow for inefficient digestion of proteins). Identification of proteins in this manner is not a trivial task, fragmentation of peptides rarely yields clean, fully realised spectra with complete series of b and y ion masses present. Instead spectra tend to be confounded by the presence of many peaks not related to the peptide of interest and the absence of peaks that would be expected. Chemical noise, instrument accuracy, electronic noise and poorly understood physico-chemical processes contributing to peptide fragmentation are some causes of this phenomenon.

Generally speaking search engines typically perform three general functions i) interpretation of data ii) filtering of data and iii) Scoring of matches.

1.3.3.1 Algorithms

There are many search engines currently available using a variety of search algorithms that can be grouped into the following four general categories, as suggested in the review by Sadygov et al (Sadygov et al., 2004): i) Descriptive ii) Statistical and probabilistic iii) Stochastic iv) Interpretive. Descriptive algorithms use predictions about how a given peptide would fragment on collision induced dissociation; this information is quantified and compared to experimental data using a correlation analysis to produce matches. Examples of programs that use descriptive algorithms include SEQUEST, SALSA and SONAR (Hansen et al., 2001; Colinge et al., 2003; Eng et al., 1994).

Statistical and probability based algorithms compare experimental data to peptides produced from a theoretical protein database. The statistical nature of

the dataset is taken into account when determining the significance to any matches made. Two of the most popular programs using this approach are known as Mascot and OMSSA (Yu et al., 2010); OMSSA has been shown to outperform Mascot in terms of identifications (Yu et al., 2010) but has since been discontinued due to a lack of funding (http://pubchem.ncbi.nlm.nih.gov//omssa/).

Stochastic models use empirical datasets to obtain statistical data on the fragmentation patterns of known peptides. This information allows probabilities to be assigned to each possible fragment ion pattern (a single fragmentation pattern would include all fragment ion series observed in a spectrum) this information along with known errors in the measurement of m/z associated with an instrument type are then used to generate theoretical spectra, for peptides in a given database, which are then compared to experimental data to find a best fit.

Interpretive models determine partial (contiguous) amino acid sequence from an experimental tandem MS spectrum. The data is used to generate a construct of three parts; i) the identified stretch of amino acids (tag) ii) The mass from the C terminus to the tag iii) the mass from the N terminus to the tag. The construct can then be searched against a protein database. The longer the tag sequence the more probable the match is correct. The determination of the amino acid tag sequence can be done manually or automatically.

More detailed examples for each of the algorithms are discussed in the follow sections.

1.3.3.2 Mascot

The search engine used in this work is known as Mascot. It uses a probability based scoring algorithm known as MOWSE (Molecular weight search) that assigns a score to each protein identified. The MOWSE score, detailed in the paper by Pappin et al (Pappin et al., 1993), is a calculation of the probability that a match is a random event. A database of proteins separated into their component peptides based on the known rules of enzymatic digestion is used as the basis for matching. Experimentally generated, mass spectrometric, data is then searched

against the theoretical database for any matches within a defined mass tolerance (Pappin et al., 1993):

$$\boldsymbol{DBMw - tolerance - 1} < PMw < DBMw + tolerance + 1$$

Where DBMw is database molecular weight and PMw is peptide molecular weight (query mass).

Proteins are sorted into 10kDa bins (e.g. 10 kDa to 20 kDa or 20 kDa to 30 kDa) and within each, theoretical peptide fragments are sorted into 100 Da bins. The frequency of occurrence for a given peptide mass within a particular protein bin is calculated by dividing the number of times it occurs by the total number of peptide fragments in the protein bin. These frequency values are then normalised to the largest value present within each 10 kDa protein bin. When an experimental fragment is matched to a theoretical fragment the normalised frequency value is looked up. In the case of multiple fragment matches to a single protein these values are multiplied together. The number is then inverted and normalised for protein mass of 50 kDa in order to control for score accumulation in larger proteins.

$$(\boldsymbol{P}) = \boldsymbol{50,000/(protein\ mass * W)}$$

Where: (P) = score
Protein mass = mass of matched protein
W = the inverted and normalised peptide frequency score

By scoring in this way the non-random distribution of peptide fragment masses in proteins of different sizes is taken into account.

MOWSE scores are expressed as $-10*\log_{10}(P)$ and therefore a probability of 10^{-5} (that the match is random) thus becomes a score of 50. The protein with the best match may or may not be relevant, depending on the size of the database. If the probability of a match by chance is for example 10^{-5} and the database contains 10^6 sequences then we would expect several of these matches to occur randomly and therefore we can deem the score insignificant. A significance threshold is set by defining first a significance level, typically <0.05. The software then calculates the threshold MOWSE score at which a match is likely

to occur by chance with a frequency of <5%. Any match with a higher score is deemed significant.

Tandem MS data (MS/MS ion search) is treated in much the same way; with the generation of expected fragment ions for each peptide fragment in the genomic database (or subset thereof e.g. species specific search) being compared with experimental data. The instrument type selected will define the type of ions generated with CID based instrumentation yielding mostly b and y ions. This approach has added accuracy as it can determine the difference between different peptides of the same mass.

1.3.3.3 OMSSA (Open Mass Spectrometry Search Algorithm)

An open source free search engine, OMSSA uses a statistical/probabilistic model for the interpretation of peptide matches. The basic assumption of OMSSA is that peptide matches follow a Poisson distribution. The results of a match are reported as an e-value which describes the chances of an equal or better quality match being made at random within the same database. A score of 1 would indicate that one other match of equal or better quality would be expected in a database of given size. In short, the lower the e-value the more statistically significant the match.

Firstly, the charge state of the precursor ion is determined as OMSSA selects peptides from the library based on the neutral mass of the precursor ion. If more than 95% of peaks in a spectrum are below the precursor mass then it is assumed that the precursor is singly charged. If the number is less than 95% then charge state is assumed to be 2+ or 3+ and the library is searched for both resulting neutral ion masses. Secondly, a noise reduction algorithm is applied to the data; any peaks with an intensity less than 2.5% of the maximum intensity are removed. Further noise reduction steps are applied depending on the charge state of the precursor and all with the intention of reducing the number of random matches and are detailed in the paper by Geer et al (Geer et al., 2004).

Comparison of experimental and theoretical data is carried out in two stages: i) A precursor mass is compared to theoretical peptide masses including any

relevant fixed or variable modifications. ii) A theoretical mass ladder, a list of ion fragments expected by CID fragmentation, is generated for the peptide of interest for comparison to the experimental data. If a match is made within a user defined mass tolerance in step one then the algorithm proceeds to the next step. If no match is made then the algorithm moves on to the next precursor mass. Matching is carried out within a user defined tolerance with each experimental ion being allowed to match only one theoretical ion in order to reduce random matching, particularly in low resolution data.

OMSSA reports expectation values (E-values) as its primary means of scoring matches (Geer et al., 2004). The E-values relate to the probability of the match being a false positive. This scoring method is also used for Blast local sequence alignment scoring. In detail:

$$E(y,\mu) = N(1 - (\sum_{x=0}^{y-1} P'(x,\mu_z))^N)$$

y = the number of successful product ion matches

z= 1 or 2 depending on the ion sequence searched

Like Mascot, OMSSA's probability based scoring is not based on the closeness of fit to a fragment model but on the probability that the match is a random event; the lower the E-score the more statistically relevant the result.

1.3.3.4 SEQUEST

SEQUEST is a commercially available search engine distributed by Thermo Scientific. It utilises a descriptive type algorithm. Initially, pre-processing of the MS/MS data is carried out; m/z values are converted to nominal masses, removal of low abundance ions and normalization of data is carried out along with the identification of immonium ions. These low m/z value ions are associated with particular amino acids (histidine, methionine, tryptophan and tyrosine) and are used in the identification process (Eng et al., 1994). Protein sequences present in the database are scanned for linear stretches of amino acids that match within a predefined tolerance the mass of the experimentally determined ions.

At this stage the masses of any modifications are also considered but are applied either to every occurrence of an amino acid or to none. The following formula is used to provide a preliminary score:

$$S_p = \left(\sum i_m\right) n_i(1+\beta)(1+\rho)/n_t$$

S_p : Preliminary score

n_i: Number of fragments that match experimental mass within tolerance

i_m: Abundances of ions matching the experimental mass

β: Incremental score for each ion present in the ion series (0.075)

ρ: Incremental score for each immonium ion present in the ion series (0.15)

n_t: Total number of ions in the theoretical ion series

The top 500 scoring matches are then analysed by cross correlation. The theoretical fragment spectrum for a given peptide is predicted (for b- and y-ions) and the main sequence ions assigned an abundance of 50, a window of 1 amu around these ions an abundance of 25 and water and ammonium losses an abundance of 10. The theoretical spectra are compared with experimental spectra using the following cross correlation function.

$$Corr(E,T) = \sum_{I=0}^{N-1} x_i\, y_{i+\tau}$$

$Corr(E,T)$: Correlation between theoretical data and experimental data

x: Theoretical spectrum construct

y: Experimental spectrum

The function serves to translate one spectrum across the other and measure the degree of similarity; the value τ is the degree of translation and is varied. If two spectra are the same then the correlation score should maximise at τ=0. To produce the final score (XCorr) the value at τ=0 minus the mean of the cross correlation in the region 75 < τ < 75 is calculated. The scores are then normalised to 1.0 (C_n). A further measure known as Δ C_n compares the top scoring peptide to its nearest scoring neighbours. This helps to indicate how unique any given match might be. The XCorr score does not give any statistical indication as to the correctness of a match but only the degree of correlation

between theoretical and experimental peptides and as such database size has no bearing on statistical significance. The second score is however dependant on the database and reflects the uniqueness of a match. A weakness of the XCorr method is higher degrees of matching between longer peptides or noisy experimental data; this can be corrected for however using appropriate normalisation methods (MacCoss et al., 2002; Sadygov et al., 2004).

1.3.3.5 Peptide Search

Peptide search (Mann et al., 1993) examines a tandem mass spectrum and calculates the mass differences between peaks in order to infer partial peptide sequences. Once a contiguous partial sequence is identified the spectrum is divided into three regions. Region two includes only the inferred peptide sequence; regions one and three represent peptides of unknown length but known mass. The direction of the sequence (which ions are b series and which are y series) is not known and as such it is treated firstly as a b series and then again as a y series. The program searches the database using the data from the three regions as well as the intact peptide mass, enzyme specificity and mass accuracy. Criterion considered when scoring a match include N terminal cleavage (N), region one mass, peptide sequence tag, region three mass and C terminal cleavage. Each of these criteria are assigned a discrete probability based on the chances that a match is random. The N and C terminal cleavage for a tryptic peptide is limited to either Arginine of Lysine or 2/20 amino acids simplified to 1/10 for each. The region one mass is considered at unit resolution using the average mass of the 20 amino acids or 1/110. The sequence tag probability is dependent on the number of amino acids in the tag, each weighted with a probability of 1/20 and cumulatively scored so that a tag of length two amino acids has a probability of 1/400 (1/20 * 1/20), three a probability of 1/8000 (1/20 *1/20 * 1/20) and so on. The region three mass is scored identically to the region one mass and as such is given the value 1/110 (Mann and Wilm, 1994). From these assumptions the probability of a match being made at random (false positive) is equal to:

$$P_{random} = P_{NtermCleavage} \times P_{R1} \times P_{1st\ tagposition} \times \dots P_{lasttagposition} \times P_{R3} \times P_{CtermCleavage}$$

The probability of a non-random match in a database with N peptides is then

$$P_{nonrandom} = (1 - 2 \times P_{random})^N$$

The size of a database relates directly to the likelihood of making a false positive match. As the size of a database increases the length of the peptide tag must also increase to keep the number of false positives below acceptable limits.

An advantage of this type of algorithm is its error tolerance. If there is some anomaly in the mass of a measured peptide e.g. a mutation or post translational modification, then a search of all three regions would not yield a true match. By searching for any combination of two of the three regions it is then possible to identify matches and locate the region of altered mass. Other search engines that use this type of algorithm include MS-Seq and Guten Tag (Clauser et al., 1999; Tabb et al., 2003).

1.3.3.6 Scope

Scope (Bafna et al., 2001) uses a two step stochastic process for identification. The first step uses, ideally, a large expertly curated empirical peptide database to predict which ions will be present in the MS/MS spectrum of a given peptide. The first part of the program computes the probability of a particular fragmentation pattern for a given peptide. The second part computes the probability of a particular spectrum for a given fragmentation pattern. The combining of these two steps allows for the probability of a spectrum being of any given peptide.

$$\varphi(S, p) = \sum_{F \subseteq \Im(p)} \varphi\,(S|F, p) Pr(F|p)$$

p: peptide

F: fragmentation pattern

S:Mass spectrum

Pr(F|p): Probability of fragmentation pattern F from peptide P

φ*(S|F,p)*: Probability density function, the probability of seeing spectrum S for fragmentation pattern F

$F \subseteq \Im(p)$: The fragment space containing all of the possible fragments for peptide p

The formula allows for the identification of the peptide that gives the maximum score for the spectrum being analysed along with the p-value. Another program that uses this kind of algorithm is known as OLAV (Colinge, 2003).

1.3.4 Protein Sequence Databases

Peptide and protein identifications from tandem MS data are realised in one of two ways. One option is to infer the sequence of a protein directly from the data i.e. measure the mass shifts between m/z values and use the data to identify particular amino acids (Liska and Shevchenko, 2003; Ma et al., 2005; Frank and Pevzner, 2005). This *de novo* sequencing approach is computationally expensive, requires high quality tandem MS data and the resulting peptide sequences must still be matched against some protein database using a modified version of the BLAST algorithm, MS-BLAST (Shevchenko et al., 2001). *De novo* sequencing difficulty is further compounded when applied to complex samples in which multiple peptides appear in the same tandem MS spectrum. Isobaric amino acids (Lysine and glutamine; leucine and isoleucine) offer further difficulties.

The second method involves trying to match the MS data directly against a database of protein sequences. The selection of an appropriate protein sequence database is important and results will vary between databases. At present there are three main types:

i) Non-identical and non-redundant manually curated databases such as Swiss-Prot and RefSeq (Partially): collapse together records with identical or near identical peptide sequences and have high quality manually reviewed information.
ii) Machine curated databases such as TrEMBL and RefSeq (X series): in which data is extracted predominantly from genomic databases with machine based analysis to assign information.
iii) Comprehensive databases such as NCBInr and OWL: contain a compilation of all publicly available sequences.

Curated databases, both human and machine, have annotated protein sequences. Information such as: taxonomy of the organism, functions, cellular location, polymorphisms, isoforms, PTMs, domains, molecular weight and pI is recorded

with each entry. Machine curation is of course much faster than human curation but is considered less reliable and less complete. Information about protein families, functional sites and domains can be inferred by searching for groups of amino acids that are conserved, signatures, (Sigrist et al., 2002) or statistically assigned profiles (Krogh et al., 1994; Durbin et al., 1998). There are many publicly available databases that can be used to search for these signatures (Hulo et al., 2006; Attwood et al., 2003; Finn et al., 2006; Letunic et al., 2006). InterPro (Mulder et al., 2007) combines all of these databases into one and allows for more comprehensive and unambiguous results. Manually curated databases offer higher reliability with data obtained from scientific publications which, importantly, offer solid evidence based assertions. Information is handled by experts and undergoes validation before addition to the database.

Data in comprehensive databases mainly comprises protein sequences directly translated from genomic data. In these types of database it is likely that a single gene will be represented by multiple gene products i.e. there is a degree of redundancy. Redundancies are introduced when compiling the primary databases when multiple records for a single protein are preserved. These databases are necessarily larger than the manually curated and automatically curated types that they comprise.

The choice of database is an important one and should reflect the nature of the experimental work being carried out. In most cases manually curated compact databases are the best choice, providing manageable datasets and accurate and full information on potential matches. However, as the objective of some experiments be to specifically identify mutated or novel alternatively spliced forms of proteins it may become necessary to use a more comprehensive database. Genomic databases such as EST (expressed sequence tag) databases contain information on polymorphisms and alternative splice forms and are therefore an option if protein sequence databases do not provide results. However, data from genomic databases is known to be prone to sequencing errors brought about by incorrectly predicted open reading frames and frame shifts. Another problem associated with larger databases is the increased risk of getting high scoring random matches to proteins (false positives) and the increase in computational time required to perform searches.

Ideally, protein databases represent the current state of our knowledge for a given protein. A description of some commonly used databases follows.

1.3.4.1 UniProt

Uniprot consists of Swiss-Prot- a section containing manually-annotated records with information extracted from literature and curator-evaluated computational analysis and TrEMBL- a section with computationally analyzed records that await full manual annotation.

1.3.4.2 Swiss-Prot

Swiss-Prot is a non-redundant database, combining sequences of near identical composition into a single entry with differences recorded in the annotations. With search engines that only search explicit database entries, not reading from annotations, potential matches can be missed. A potential problem is the alternative splicing of proteins. Entries in databases are typical a single isoform (the longest) of a protein and when compared to an experimentally detected variant isoforms matching is less than optimal. However, a program known as VarSplic (Kersey et al., 2000) can be used to generate variants from the sequences in the database and add them as entries thereby making optimal matching possible. The data contained in Swiss-Prot is reviewed, any discrepancies noted and duplicate information reviewed. There is an ongoing collaboration between Swiss-Prot, NCBI and DDBJ (DNA Databank of Japan) meaning that the databases share their data. Swiss-Prot details protein functions, known interactions, sub cellular locations, domains, PTMs and variants. It is well integrated with more than 50 other databases via cross-referencing. The Swiss-Prot database reflects the most up to date, manually curated collection of protein data available.

1.3.4.3 TrEMBL

TrEMBL is the automated counterpart to Swiss-Prot. This database contains translations of all of the coding nucleotide sequences in the DDBJ/EMBL (European molecular biology laboratory) /GenBank (Okubo et al., 2006; Kulikova et al., 2007; Benson et al., 2007) nucleotide databases along with sequences

found in literature and submitted to Uniprot. TrEMBL makes available these data for searching without adding them to Swiss-Prot. This prevents a lowering of the overall quality of notation in Swiss-Prot. TrEMBL entries are effectively queued for manual annotation, whilst this is pending relevant annotations from Swiss-Prot are applied to TrEMBL entries, with the new information superseding the old.

1.3.4.4 NCBI

The National Centre for Biotechnology Information is world class information hub containing some 39 literature and molecular biology databases with entries totalling in the hundreds of millions (Sayers et al., 2010). Two relevant databases including protein sequence information are known as RefSeq and NCBInr.

1.3.4.5 RefSeq

The RefSeq database contains information on an organism's genomic, transcript and protein sequences. Data is annotated and sourced from publicly archived databases including DDBJ, European nucleotide sequence database and GenBank. Initially, the data is produced by automated analysis of genomic information. These entries have accession numbers prefixed with an X. Manual reviewing of records is carried out and the reviewed record replaces the automatically modelled record. RefSeq pays particular focus to species of research significance and as of 2011 91.5% of all human protein entries were manually curated (Pruitt et al., 2011).

1.3.4.6 NCBInr

NCBInr is misleadingly named as it is not a truly non-redundant database. It contains multiple entries for proteins with sequences that vary by as little as a single residue. NCBInr is comprised of protein compiled from GenBank CDS translations, PIR (protein information resource, RefSeq, Swiss-Prot, PRF (protein research foundation), and PDB (protein data bank) (Kouranov et al., 2006). Using this type of database means potentially more matches but at the price of duplications in the search results and an increased risk of false positive matches.

1.3.4.7 MSDB

This database was run from Imperial College London. It was a compilation of Swiss-Prot, PIR, TrEMBL and GenBank. Sequences with 100% similarity were collapsed together to remove a degree of redundancy. The database was distributed with Mascot but as of 2006 it is no longer updated and should be considered obsolete.

1.3.4.8 EST databases

These databases look at the expressed sequence tags,single pass cDNA sequences, from organisms. The nucleic acid sequences are translated in all six reading frames to generate potential protein sequences. These databases contain a lot of information on polymorphisms and are typically very large. EMBL has 10 EST divisions including: Environmental_EST, Fungi_EST, Human_EST, Invertebrates_EST, Mammals_EST, Mus_EST, Plants_EST, Prokaryotes_EST, Rodents_EST, and Vertebrates_EST.

1.3.5 Mass Spectrometers

As previously stated, mass spectrometers at their most basic include only three basic components. An ion source, a mass analysed and a detector. In terms of mass analysers there are a wide range of options each with advantages and disadvantages that make them suitable for particular applications. The mass analysers used in this work, i.e. 3d and LIT ion trap and quadrupole, are all based on technology pioneered in the early 1950s by Woflgang Paul and his colleagues (Paul and Steinwedel., 1953).

1.3.5.1 Spherical (3d) Ion Trap

The spherical ion trap is composed of three hyperbolic electrodes designed to focus and trap ions through the formation of a 3 dimensional ion trapping field. Ions are allowed to enter the trap within which helium is present to a pressure of 1mTorr in order to reduce their kinetic energy. An oscillating potential, an AC voltage with periodicity in the RF range, is applied to the ring electrode and acts to focus and trap ions. Ions are affected by this so called fundamental RF voltage depending on their mass and charge. By altering the periodicity of the

fundamental RF it is possible to destabilise the trajectory of ions with a given mass and charge thereby eliminating them from the trap.

The application of low amplitude waveforms to the end cap electrodes can destabilise ion trajectories between these electrodes thereby increasing their collisions with the helium dampening gas. These collisions are capable of fragmenting ions, in the case of peptides fragmentation typically occurs along their backbones giving rise to b and y ions. Following the isolation and fragmentation of an ion species it is possible to then isolate and fragment one of the daughter ions. This process, so called MS(n), has been repeated up to MS(12). A limitation inherent to these instruments is the inability to retain ions with an m/z of less than 0.3 of the parents mass. A consequence of this that with peptides the first several b and y ions may not be detected .

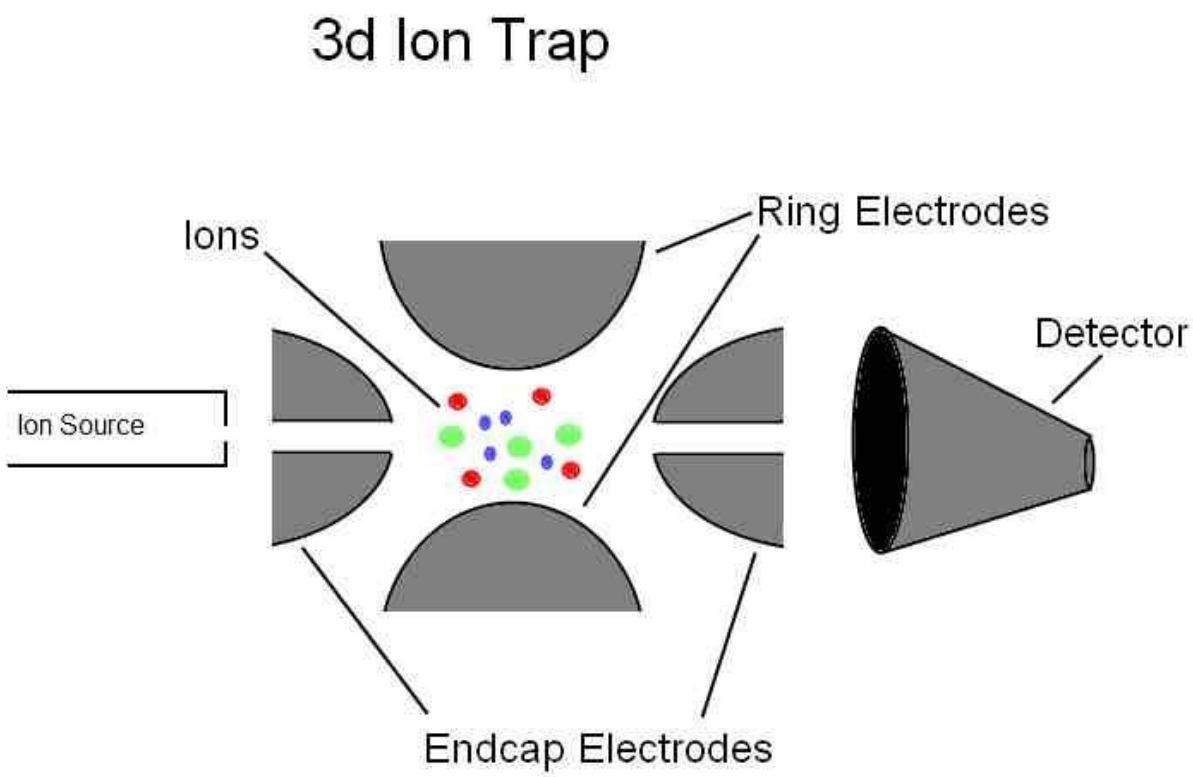


Figure 13. A 3d ion trap schematic. Ions enter through an endcap electrode. The ring electrode produces an RF voltage which acts to trap the ions. Ejection of the ions can be achieved by applying a supplementary RF voltage to the endcaps.

1.3.5.2 Linear Quadrupole Ion Trap

The linear quadrupole ion trap (LIT) comprises 4 parallel rods that use a combination of electrostatic DC fields to trap ions along their axis and an RF AC voltage to trap ions axially. Alteration of the RF field can lead to the destabilisation of ions causing them to collide with the quadrupole rods. Using this approach it is possible to eliminate all ions out with a range of interest from the trap. As with the spherical trap it is possible to apply low amplitude waveforms in order to bring about the fragmentation of ions and to perform MS(n) type experiments.

1.3.5.3 Quadrupole

This type of mass analyser comprises 4 parallel poles describing the corners of a square. Diagonally opposing corners are both either positively or negatively charged, this charge alternates at a predetermined frequency. Ions are accelerated along the length of the device by a DC field. The combination of the AC and DC fields cause ions to travel in a spiral along the device, a stable trajectory is held by ions of a particular m/z as determined by the AC frequency.

Triple quadruple instruments are a linear arrangement of quadrupoles. In this configuration it is possible to utilise neutral loss, multiple reaction monitoring and precursor ion scanning modes which greatly increase selectivity.

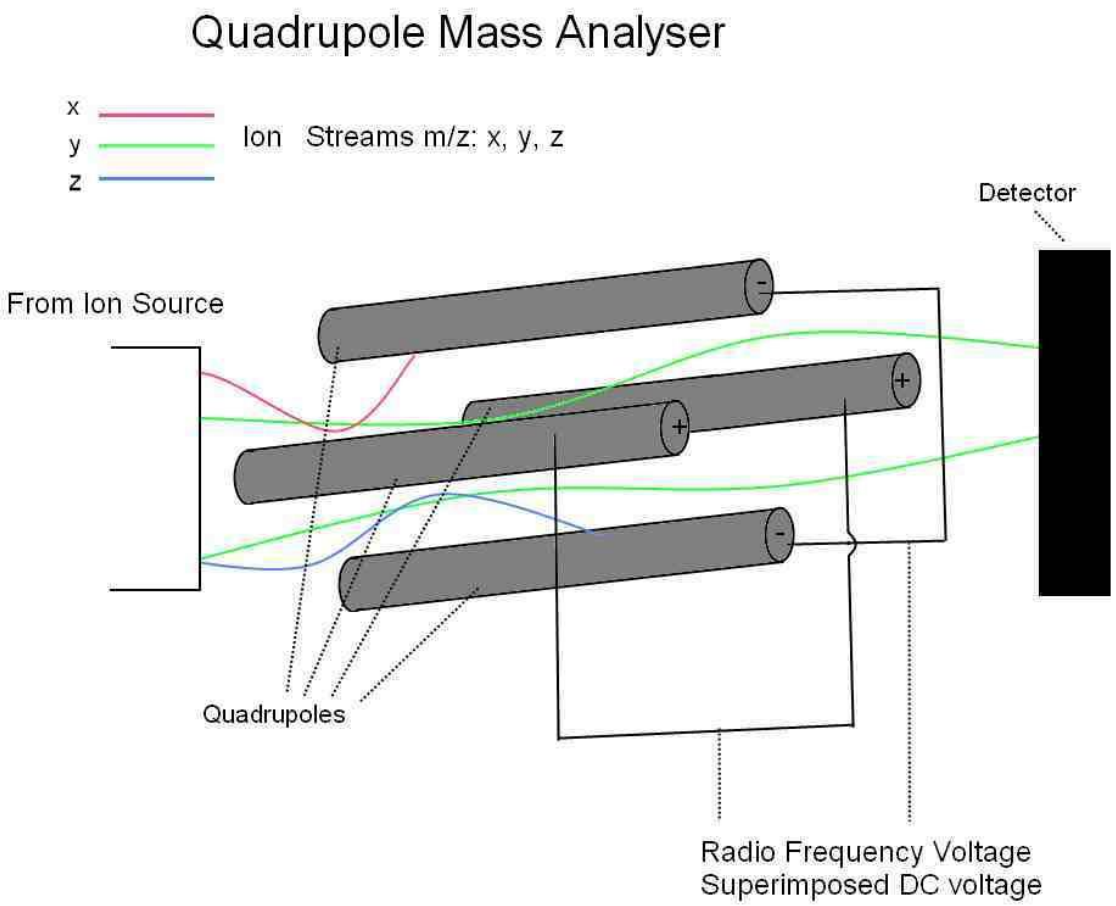


Figure 14. Quadrupole mass analyser. The RF voltage applied to the quadrupole determines the m/z of ions that have stable trajectories. In the above diagram only m/z: y has the stability to pass through to the detector. Both x and z are destabilised and lost. The supplementary DC voltage supplies ions with lateral acceleration toward the detector.

1.3.5.4 Hybrid Instruments

Hybrid instruments make use of two or more different types of mass analyser. These types of instrument combine the advantages of the different mass analysers in order to obtain higher quality data. Instruments such as the Qstar, a quadrupole TOF (time of flight) combination; Q-trap triple quadrupole LIT; LTQ Orbitrap a LIT Orbitrap; FT-ICR (Fourier transform ion cyclotron resonance) LIT ICR and TripleTOF triple quadrupole TOF.

1.3.6 Scanning Techniques

Owing to the various attributes of different instruments there are a wide variety of scanning techniques available. A brief summary of these follows.

1.3.6.1 Neutral Loss Detection

Detection of modified glutathione is routinely carried out using a constant neutral loss scan (NL) for 129 Da (Yu et al., 2005; Ma and Subramanian., 2006). Using a triple quadrupole instrument , quadrupole one is set to scan through a range of masses, quadrupole 2 fragments the selected ions and quadrupole three selects for a m/z of quadrupole 1 minus the 129 Da. This weight represents the loss of γ-pyroglutamic acid. This type of scan however is prone to 'false positives' as endogenous biological compounds can give rise to the same neutral loss.

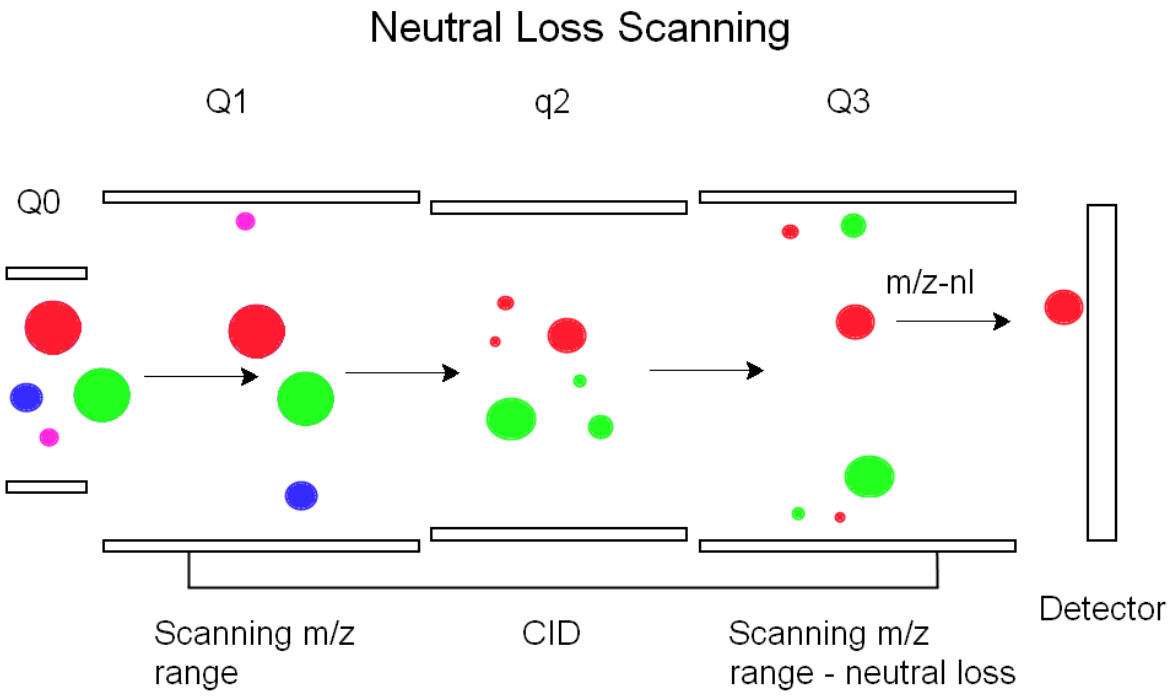


Figure 15. Ions enter the instrument and allowed through Q1 by mass. Q2 fragments ions and Q3 allows the detection of ions passed at Q1-129 Da. The mass of ions giving rise to neutral losses of 129 indicate GSH or GSH-RMs

In order to remove these results from the scan work by Yan and Caldwell focused on the use of heavy isotope labelled ($13C_2$-15 stable isotope) GSH. An equimolar ratio of labelled/unlabelled GSH was added to the drug/microsome preparations in order to react with electrophilic species. Using this method allows for the identification of doublet isotopic peaks with a 3 Da mass difference. These represent the labelled and unlabelled GSH conjugates.

Unfortunately not all test GSH adducts have the 129 Da neutral loss characteristic as part of a primary fragmentation pathway (Dieckhaus et al., 2005). In these cases adducts will escape detection and characterisation unless another method if detection is employed.

1.3.6.2 Precursor Ion Scanning

GSH can be detected by scanning in the negative ion mode for a 272 Da, a deprotonated ç-glutamyl-dehydroalanyl-glycine originating from the glutathionyl moiety. The abundance and uniqueness of this anion make it an excellent candidate for specific and sensitive detection of GSH-metabolite adducts. It has been shown capable of identifying previously unknown GSH conjugated metabolites such as those of meclofenamic acid (Wen et al., 2008).

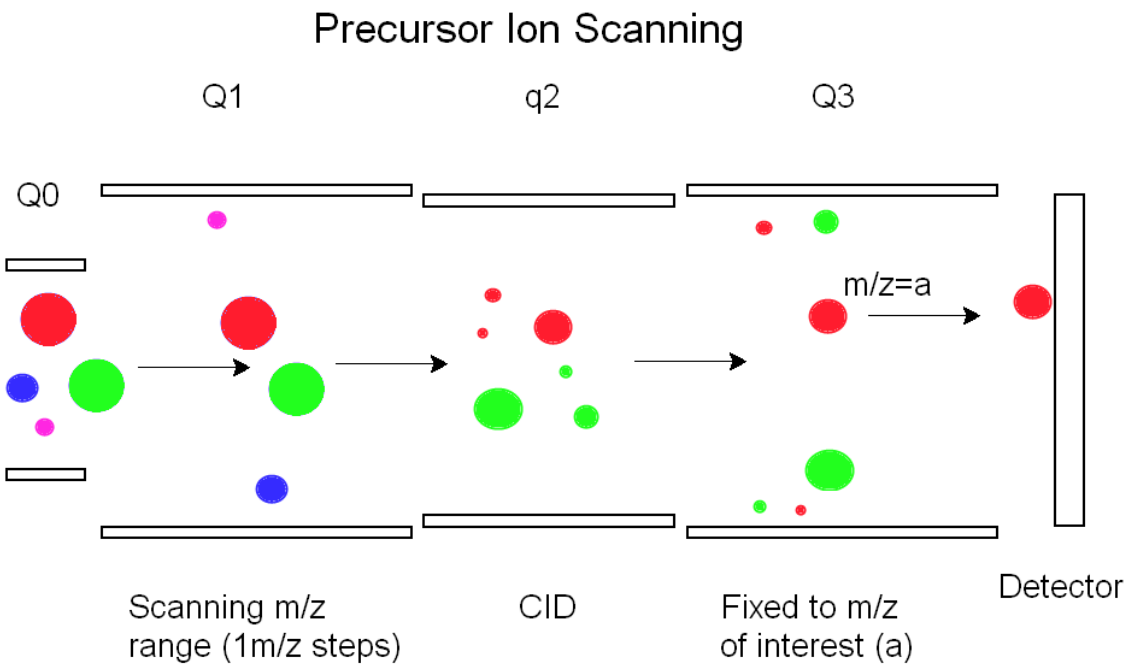


Figure 16. Precursor ion scans can identify ions giving rise to a characteristic fragment.

Detection of the 272 Da anion in negative mode has been used to trigger the acquisition of CID MS/MS of the precursor ion in positive ion mode. Data from this single run high throughput capable experiment can be used to selectively identify and characterise the structures of GSH bound reactive metabolites with superior selectivity, sensitivity and range in comparison to the standard NL 129 scan method.

1.3.6.3 Single Reaction Monitoring

Using a triple quadrupole mass spectrometer it is possible to operate in what is known as single reaction monitoring (SRM) mode. So called transitions, descriptions of ions present before and after CID, are used to select for particular molecules with a complex sample. SRM experiments produce unequalled sensitivity and with the advent of the Q-trap, can be used to initiate enhanced product ion spectra.

Transitions are obtained either through data gleaned from earlier experimentation or through theoretically expected changes to specific molecules e.g. biotransformation of drugs and subsequent binding to GSH. Previous work has shown that through the use of a list of some 114 SRM transitions calculated from common biotransformations of particular drug molecules, it was possible to detect the presence of GSH-reactive metabolite adducts (Zheng et al., 2007). In comparison to NL and PI scans SRM is more sensitive, provides fewer false positives and is capable of producing high quality MS/MS data on the same run (Zheng et al., 2007).

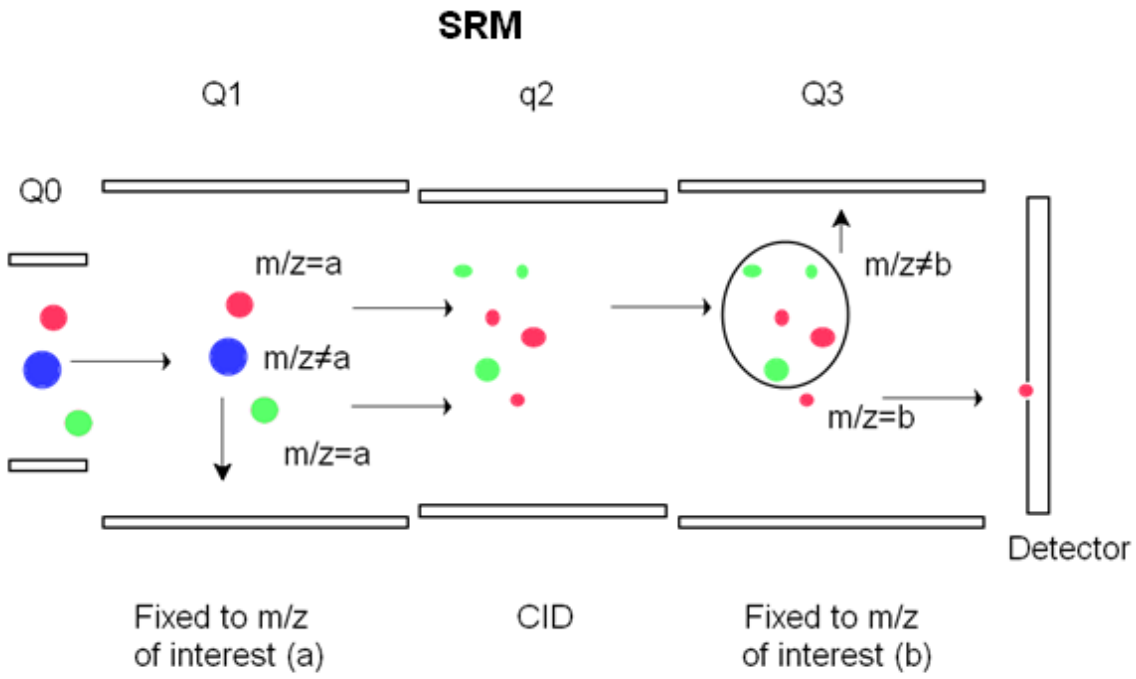


Figure 17. SRM experiments use known data to search for specific ions and products with a high degree of efficiency and sensitivity. Q1 is set to pass a specific mass, Q2 fragments them and q3 selects for a specific fragmentation product.

1.3.6.4 Post-Acquisition Data Mining

Techniques previously mentioned such as SRM, PI and NL scanning all technically require that the instrument used be a triple quadrupole. Alternative methods have been developed for use with higher mass accuracy machines such as Q-TOFs, Orbitraps and FT-ICRs. Software based approaches allow for emulation of PI and NL scanning through the format of precursor ion filtering (PIF) and neutral loss filtering (NLF). MS/MS data is collected and systematically searched using known PI or NL filters in order to identify ions of interest.

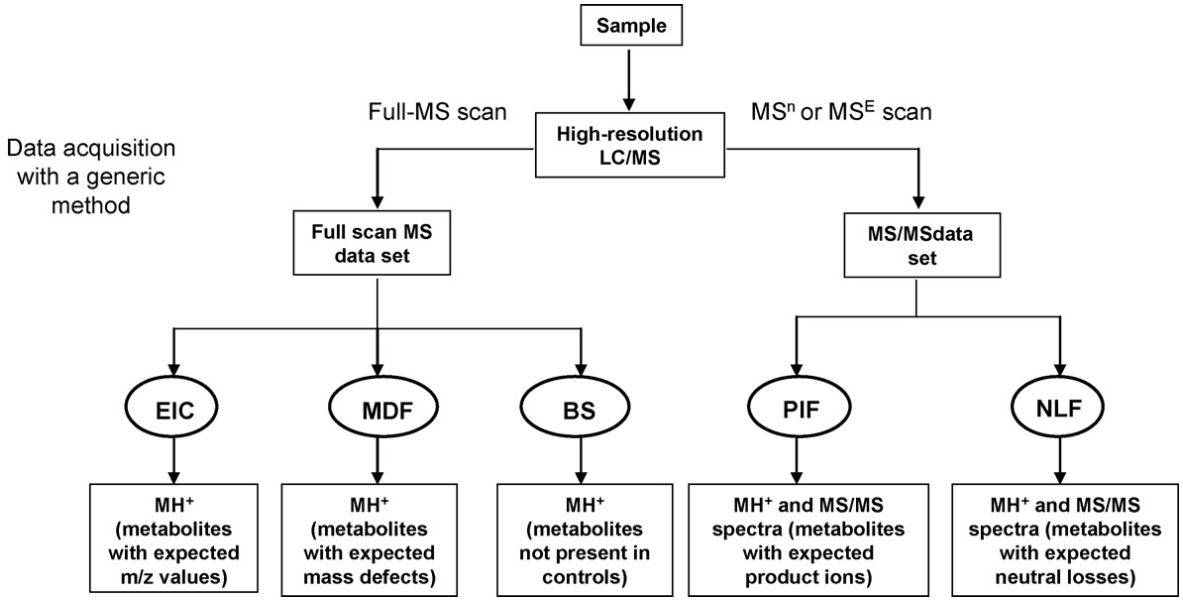


Figure 18. Schematic of common data mining techniques (Zhu and Ma, 2009).

Full scan data can be utilised to identify metabolites by MDF, extracted ion chromatograms (EIC) and background subtraction (BS).

Background subtraction involves the use of a control sample which undergoes an accurate full scan in order to obtain accurate masses of detected ions. Once the sample bearing reactive metabolites is scanned an algorithm is used to subtract all common ions from around the time frame of interest, with allowance made for differences in chromatography (inter run variation). The vast majority of remaining ions after subtraction of background should represent the presence of metabolites of interest.

1.4 The reactive metabolite target protein database

A web based database has been created with the intent of mapping the "Adductome". Results from many different groups involved in researching protein-drug adduct formation have been compiled into what's known as The Reactive Metabolite Target Protein Database (TPDB) (Hanzlik et al., 2008). This freely available resource is an attempt to identify relationships between the formation of particular protein adducts and toxicity. As of yet no clear relationship between adduction and physiological effects has been uncovered.
To date the database contains no information for any of the drugs used in the following work.

1.5 Statistics in Proteomics

Two of the workhorse techniques applied in the field or proteomics, 2D-PAGE/DiGE and Mass spectrometric analysis, are capable of generating vast datasets in a single experiment. The high dimensionality of the experiments necessitates a careful approach to data analysis. It has been observed that the kind of data gathered from 2D-PAGE and DiGE is very similar in nature to data gathered in genomic investigations using DNA microarray technology. As a consequence, many of the statistical tools developed for data analysis e.g. multiple hypothesis testing, classification methods and cluster analysis can be assimilated for use in protein expression analysis (Urfer et al., 2006). Univariate statistical tests, Student's t test and analysis of variation (ANOVA) are the most commonly used statistical tools for DIGE analysis (Meunier et al., 2007). Before data analysis takes place however, it is necessary to carry out some pre-processing of the DiGE data.

1.5.1 Data Pre-Processing

Two major assumptions of many statistical methods is that the dataset must fit a normal distribution and that variance in the data be homogenous. Assumptions that don't reflect the nature of DiGE raw data which is in the form of gel images. A typical first step in pre-processing involves a log transformation (most commonly $\log_2$) which acts to homogenise variance in the data. It has been observed that spots with higher mean intensities usual have larger variation

(Gion et al., 2005; Karp and Lilley, 2005). Additionally, it has been shown that the different CyDyes can introduce bias to the experiment due to their differing background signals (Karp et al., 2004; Karp and Lilley, 2005). This bias can however be controlled for by making sure that each of the dyes is used in both or all experimental groups (Timms and Cramer, 2008).

DiGE experiments include an internal standard, a mixture of all samples, which is run on each gel. A spot-wise division of each protein by its internal standard counterpart is carried out to control for experimentally introduced gel-gel variations. This normalisation across all gels, by accounting for technical variations, helps to identify biological variations. Technical replicates i.e. replicates produced using the same sample, must be included within any experiment for this purpose.

Biological replicates are necessary in order to account for variation in protein expression levels in protein populations (biological diversity between subjects, tissue samples, cell cultures etc) that have been subjected to the same conditions (to the best of our knowledge)(Karp et al., 2005). The extent of this biological variation must be taken into account when attempting to determine the changes of protein expression between different groups brought about by changes in experimental conditions (i.e. treatment with a drug vs. no treatment). In order to characterise this, multiple measurements for each test condition must be taken from biologically non-identical samples.

1.5.2 Type I and Type II Error

When dealing with large datasets it is necessary to statistically validate results. A typical DiGE experiment can consist of multiple gels with each allowing the visualisation of 2000-4000 proteins. As the size of a dataset increases the chances of reaching a false conclusion about any given data points also increases.

For a particular experimental question e.g. has protein expression changed? There can be four distinct outcomes.

i. True positive: A change has been detected and corresponds to a real biological change.

ii. False positive: A change has been detected but does not correspond to a biological change.

iii. True negative: No change has been detected and corresponds to no biological change.

iv. False negative: No change has been detected but a biological change has occurred.

From these four outcomes it is clear that ii and iv give erroneous results. The false positive, also known as a Type I error, can be controlled by choosing an appropriate significance level. Firstly, a so called null hypothesis is generated and is represented by $H_{0.}$ The null hypothesis is a statement of the question to be answered. In the case of a typical DiGE experiment, H_0 = there is no change between the mean expression of an given protein in the DiGE experiment. Now H_a, or the alternate hypothesis, is defined as a refutation of the null hypothesis i.e. that there is a change in the mean expression. Whether we reject or accept the null hypothesis is then dependant on the experimental data and a predefined significance level (α). The significance level defines the number of false positives that would be expected to occur by chance within a particular dataset. It is normally set at 0.05 or 0.01 (representative of 1/20 and 1/100).

The p-value is calculated from the experimental data and expressed as a real value between 0 and 1. The value represents the probability that the observed result (or more extreme result) would be observed given that the null hypothesis is correct. The p-value is directly compared to the significance level. A p-value lower than the predefined significance level compels us to reject the null hypothesis and thereby allows us to say that there is a significant difference in the mean protein expression between the two groups.

In an experiment with thousands of proteins probability dictates that false positives be expected. In order to reduce the number encountered the significance value must be made more stringent or modified in some way. This can be achieved by considering either the family wise error rate (FWER) or the false discovery rate (FDR).

1.5.2.1 FWER (Family Wise Error Rate)

FWER family wise error rate is defined as the probability of at least one false positive occurring amongst the whole dataset. A simple algorithm (the Bonferroni correction) is applied to alter p-values in such a way as to ensure that the probability of a false positive is kept below the previously defined significance level. The adjustment increases in severity as the number of hypotheses being tested increases. With the scale of many proteomics experiments it has been suggested that FWER based correction may not be the most useful approach as it introduces a higher probability for false negative, type II, errors.

The FDR (false discovery rate) approach can also be used to adjust p-values and control the likelihood of making a type I error (Cairns et al., 2009; Fodor et al., 2005; Dudoit et al., 2003). Two different algorithms are in use, a very conservative one which operates under the assumption that all hypotheses are independent (Benjamini and Yekutelli, 2001) and a less severe one that does not make this assumption (Benjamini and Hochberg, 1995).

False negatives, or Type II errors, are controlled for by designing experiments with a particular power. The power of an experiment is defined as the probability of not making a Type II error.

$$Power \propto \frac{ES\alpha\sqrt{n}}{\delta}$$

Where, ES: Effect size

α: significance level

δ: standard deviation

n: number of replicates

The effect size, the size of a particular difference that we would like to detect, must be determined experimentally by means of a small pilot study. From the equation we can see that a larger study is required to detect an effect of a smaller size. The power of the experiment is controlled by setting the sample

size. The power is the probability of rejecting the null hypothesis and is complimentary to type II error ($\boldsymbol{\beta}$) i.e.:

$$\boldsymbol{\beta = 1 - power}$$

An important consideration that must be taken into account is the fact that a DiGE experiment carried out in this thesis was simply a tool for identifying protein-drug adducts. Controlling the instance of type I errors is far less important than maintaining experimental power and not making type II errors. Following up on proteins identified as having altered expression did not bear a prohibitive cost in time or money, losing leads to type II errors however could very well render the experiment useless.

1.5.2.2 FDR (False Discovery Rate)

Due to the large number of protein spots being compared in a typical DiGE experiment (>2000) there is a considerable chance that type I, false positive, errors will occur. In order to control for this a false discovery rate (FDR) should be obtained. The false discovery rate indicates the number of random, incorrect, matches between experimental data and entries in the protein database and can be calculated as follows:

$$\boldsymbol{FDR = FP/(FP + TP)}$$

Where FP is false positive and TP is true positive.

There are several methods for obtaining a FDR and a commonly used method is known as the Benjamini-Hochberg approach. The Benjamini-Hochberg approach for controlling false discovery rates is much less conservative than FWER methods; rather than trying to control the chance of a single (i.e. any at all) false positive result, the Benjamin-Hochberg protocol (Benjamini and Hochberg, 1995) is used to limit the number of false positive results to a chosen proportion of all results.

By making the assumption that the p-values obtained follow a uniform distribution under the null hypothesis and arranging them into a sorted list it is possible to control for a desired false discovery rate. Starting from the lowest value on the list each p-value ($p_{(k)}$) is compared to a threshold value given by

multiplying the chosen false discovery rate by its location on the list (k) and dividing by the number of p-values being evaluated (m). For the following list of sorted p-values we accept a false discovery rate of 0.05 (α) (or 5%):

1) 0.0012		1) 0.0083
2) 0.0021		2) 0.0167
3) 0.04	k/m	3) 0.025
4) 0.071		4) 0.033
5) 0.11		5) 0.416
6) 0.36		6) 0.05

The list on the left represents the p-values and the list on the right represents the computed thresholds for a false discovery rate of 0.05. Starting from the top of the list (1), it can be seen that by the third term (0.04) the p-value exceeds the threshold, p-values from here on are then discarded as being outside of the desired false discovery rate.

1.5.3.3 FDR (Protein Identifications)

In the field of protein identifications, in which experimental data is searched against huge genomic databases, false positive matches are a considerable problem. An experiment carried out by one group (Cargile et al., 2004) demonstrated that it was possible to match experimental data to proteins from a mythical creature (in actuality, false positives). This serves to highlight the problem and makes clear that any attempt to follow up on these proteins would be not only futile but a waste of resources. To help combat this type of result FDRs are obtained.

A typical approach in protein identifications against a sequence database is to obtain an empirical measure of the FDR, in contrast to the calculation used in the Benjamini-Hochberg approach. This strategy would involve creating a copy of the database in which the protein sequences are reversed (Moore et al., 2002; Qian et al., 2005; Huttlin et al., 2007) or sequences retain the same frequency of amino acids but are randomised (Wang et al., 2009; Perkins et al., 1999; Higdon et al., 2005). The data is then searched against this new “decoy” database, as it is extremely improbable that proteins exist with these reversed

or randomised sequences it can be surmised that any match between the data and the sequences must be attributable to chance. The number of positive identifications received when searching against this database indicates the extent of type I present. This is assuming that the rate of false positives between the genomic database and decoy database is the same.

Yet another approach utilises tandem MS based peptide assignments along with their probabilities of being correct (Keller et al., 2002) to compute the probability that any given protein is present in the experimental data (Nesvizhskii et al., 2003; Sadygov and Yates, 2003). Software can be found that implements this statistical analysis and is known as Protein Prophet. However, assumptions made by the various models may not be applicable to all data sets nor be translatable to all proteomic instrument platforms.

The problem of false positive identifications is not one associated with current databases or instruments but with the mathematics of finite sets and is exacerbated by homology between proteins and the large number of proteins in a given database. Researchers continue to pursue this interesting field and tools for FDR characterisation are likely to continue to improve.

1.6 Future Work

The study of adverse drug reactions is an immensely complex field. Much remains to be discovered regarding the metabolism of xenobiotics, their conjugation to proteins and subsequent toxicity. The endeavour has been likened to that of cancer research; progress has been slow and hard come by. It has been previously suggested that a multidisciplinary approach will be necessary with contributions from proteomics, genomics and metabolomics based approaches (Merrick, 2008).

Work carried out has attempted to elucidate the mechanisms behind ADRs, to identify reactive metabolites with propensities to form covalent bonds with proteins, and to devise means with which to detect them at an early stage of drug development. Mass spectrometry will play an important role in unravelling these problems. Continued improvements in instruments leading to greater

accuracy, sensitivity and reduced scanning times as well as innovative ways to implement these technologies are key to future successes.

The development of sensitive methods for the detection of particular drug-protein adducts from complex biological backgrounds would provide an important step towards prevent human exposure to potentially toxic drugs. It has been noted that drugs withdrawn due to their toxicity are not often subject to further study (Park et al., 2006). This potentially rich source of data could be used to guide the design and development of future compounds.

Chapter 2: Methods

2.1 Methods

2.1.1 Proteomics

2.1.1.1 Protein concentration assay (Bradford)

Protein concentrations were determined using the Bradford assay. A kit was obtained from ThermoFisher Scientific, Loughborough, UK. Protein concentrations were measured as per the included instructions. Briefly, a series of known BSA concentrations (Final assay concentrations: 0.125, 0.25, 0.5, 0.75 and 1mg/ml) were spiked with Bradford reagent and their absorption at 595 nm was measured. The data were used to create a reference curve; curves with an R^2 value of at least 0.95 were accepted. Samples of unknown protein concentration were spiked with Bradford reagent and their absorptions measured at 595 nm. The reference curve was used to approximate a linear relationship between absorption and protein concentration. The data was used to interpolate the concentration of these unknown samples using the equation for a straight line:

$$y = mx + c$$

2.1.1.2 Protein precipitation

2.1.1.2.1 Acetone precipitation

Protein solutions were brought up to 80% v/v acetone and stored at -80˚C overnight. The samples were centrifuged at 14k rpm for 5 min and the supernatant discarded. Pellets were washed with 80% v/v acetone, 20% v/v ddH_2O. After subsequent centrifugation the pellets were reconstituted at the desired concentration in SDS-PAGE loading buffer.

2.1.1.2.2 TCA precipitation

4 parts of 100% w/v trichloroacetic acid (TCA) solution was added to 1 part protein solution. The mixture was incubated at 10˚C for 10 min then centrifuged at 14,000 rpm for min. The supernatant was discarded and the pellet washed in acetone. The centrifugation and washing steps were repeated. The pellet was then dried at 95 ˚C and reconstituted at the desired concentration in SDS-PAGE loading buffer.

2.1.1.3 In solution tryptic digestion

50 µg of protein was suspended in 25 µl 50mM ammonium bicarbonate solution (pH 8.0). 5 µl of 50 mM DTT was added, followed by a 30 min incubation at 60˚C in order to break disulfide bonds. 5 µl of Iodoacetamide was added with a subsequent incubation at room temperature in darkness for 15 min to prevent disulfide bond formation. 12.5 µl of 0.1 µg/µl trypsin solution was added along with 30 µl of acetonitrile followed by a one hour incubation at 37 ˚C. A final addition of 12.5 µl of trypsin solution was made followed by an overnight incubation (18 hours) at 37 ˚C. The reaction was stopped by the addition of 1 µl of 1% v/v formic acid solution. It should be noted that the proteins were not denatured using this method, this was an oversight and a more efficient digestion would have been possible had it been implemented for more details see section (4.7 Discussion).

2.1.1.4 1-dimensional polyacrylamide gel electrophoresis (1d-PAGE)

25 µg of protein dissolved in 24 µl of 50 mM AmBic was spiked with 6 µl of (5x) SDS-PAGE loading buffer (0.25 M Tris-HCL, pH6.8, 15% v/v SDS, 50% v/v glycerol, 25% v/v ß-mercaptoethanol and 0.01% w/v bromophenol blue) and loaded into a 4-12% gradient mini NuPAGE® polyacrylamide gel (Invitrogen, Paisley, UK). The gel was loaded into an XCell SureLock™ Mini-Cell Electrophoresis System (Invitrogen, Paisley, UK), submerged in NuPAGE ® tris-acetate SDS running buffer (Invitrogen, Paisley, UK) and run for 35 min at 200 V and 120 mA.

2.1.1.5 2-dimensional poly acrylamide gel electrophoresis (2d-PAGE)

2d-PAGE allows for the separation of proteins by both isoelectric point and mass. The isoelectric focusing is performed first followed by an SDS gel step for mass separation. A 24cm IPG strip with a pH gradient from 4-7 was selected and allowed to thaw at room temperature. Protein samples were re-dissolved in gel rehydration buffer solution (8M Urea, 4% w/v CHAPS, 0.0002% w/v bromophenol blue) spiked with DTT to a final concentration of 65 mM with a final volume of 500 µl. The solution was carefully pipetted along the length of an IEF strip holder. The plastic covering of the IEF strip was removed and the strip placed exposed side down into the protein solution, making sure that the gel is properly aligned to receive the cathode and anode of the Ettan™ IPGphor™ 3 system (GE healthcare, Little Chalfont, UK). Mineral oil was pipetted into the strip holder in sufficient quantity to immerse the gel strip and protein solution. The strip holder was then loaded into the Ettan™IPGphor™ 3 system and the appropriate program selected. The strips were allowed to accumulate 80,000 volt hours over a period of 24 hours.

Gels into which the IPG strips were to be loaded were then cast. Cleaned plates (25.5 x 20.5 cm) were loaded into an Ettan™ Dalt II gel caster frame (GE healthcare, Little Chalfont, UK) which was then filled with 500ml of acrylamide gel solution. DdH_2O saturated butanol was pipetted on top of the gel solution in order to ensure a level surface upon setting. After 1 hour the butanol was removed and replaced with running buffer and the top of the caster apparatus sealed with cling film to prevent evaporation. Gels were left overnight at room

temperature (around 21 ºC) to set, the following morning the plates containing the gels were removed from the caster washed with ddH_2O then placed upside down in a rack to dry.

The IPG strips were removed from the Ettan™IPGPhor™ 3 system and equilibrated in SDS equilibration buffer (SEB) containing 65 mM DTT. The strips were immersed in the solution and subjected to gentle rocking for 15 min. The solution was poured off and replaced with SEB containing 135 mM iodoacetamide then the samples gently rocked for a further 15 min. The IPG strips were loaded into the tops of the gels with the barcodes facing outwards and on the left hand side. The plates were placed in the ETTAN™ DALT II electrophoresis unit and the IPG strips fixed in place with molten agarose. About 8 litres of SDS running buffer was added to the tank in order to fill the bottom compartment. A further 2-3 litres of 2x SDS running buffer was added to ensure that the top compartment was full. The lid was fixed in place and the power pack set up to deliver 1 watt per gel for a period of 24 hours.

<u>SDS electrophoresis buffer 1x concentration:</u>

25 mM Tris pH 8.8
250 mM Glycine
0.1% SDS (w/v)

<u>Sample Equilibration buffer (SEB) 1x concentration</u>:

25 mM Tris-Cl pH 8.8
30% Glycerol (v/v)
1% SDS (w/v)
0.01 mg/ml bromophenol blue

<u>Acrylamide gel solution 1x concentration</u>:

12.5% acrylamide
375 mM tris pH 8.8
0.1% SDS (w/v)
1 mg/ml ammonium persulphate
0.14 µl/ml TEMED

2.1.1.5.1 Bind silane treatment

Gels to be stained were cast in bind silane treated plates. 4 ml of the bind silane solution was applied to the cleaned and dried plates. The solution was spread evenly over the surface using a lint free tissue lightly wetted with ethanol.

Bind silane solution 1x concentration:

20% ethanol (v/v)
1% acetic acid (v/v)
0.5% bind silane (v/v)

2.1.1.6 Agilent OFFGEL 3100 Fractionation

Samples from the liver microsome assay were tryptically digested (2.1.1.3) then reconstituted in 3.6 ml of OFFGEL buffer (Agilent technologies, Wokingham, UK). A 24 cm IPG strip with a gradient of pH 4-7 for each sample was thawed at room temperature, peeled then placed into the OFFGEL strip holder with the gel side face up. The 24 cm reservoirs were snapped into place and electrode pads wetted with OFFGEL buffer solution were put into place at the anode and cathode ends of the IPG strips. 40 µl of OFFGEL buffer solution added to each reservoir, once 15 min had elapsed the samples were loaded into their respective strips and spread equally among all reservoirs. The cover seals were put into place and 10 µl of distilled ddH_2O was applied to all electrode pads. Mineral oil was used to immerse the electrode pads in order to prevent dehydration. The fixed and movable electrodes were applied and the assembly seated in the OFFGEL 3100 fractionator apparatus (Agilent Technologies, Wokingham, UK). The samples were subjected to an electric field of 8000 V at a maximum of 50 µA until 50 kVh were accumulated (program 24-PE00). Upon completion of the program samples were recovered from the individual reservoirs by pipette , dried in a vacuum centrifuge then stored at -20 ˚C until needed.

2.1.1.7 SCX

Separation was performed using a POROS 10S column (inner diameter 300 µm, length 15 cm) (Dionex, Camberly, UK) and an UltiMate 3000® HPLC system (Dionex, Camberly , UK) in conjunction with an HTC PAL fraction collector (Bruker-Michrom, Auburn, CA, USA). A dual gradient pump supplied a 30 min salt

gradient (figure 23) from buffer A (5 mM KH_2PO_4, 5% v/v MeCN, pH 3.0) and Buffer B (5 mM KH_2PO_4, 5% v/v MeCN, 500mM KCl pH 3.0).

After tryptic digestion (2.1.1.3) samples were first purified using a C18 SPE cartridge (2.1.1.19) in order to remove salt then were reconstituted in 20 µl buffer A and loaded into a 96 well plate for injection into the HPLC system. Each sample was run through a 60 min gradient at a flow rate of 15 µl/min

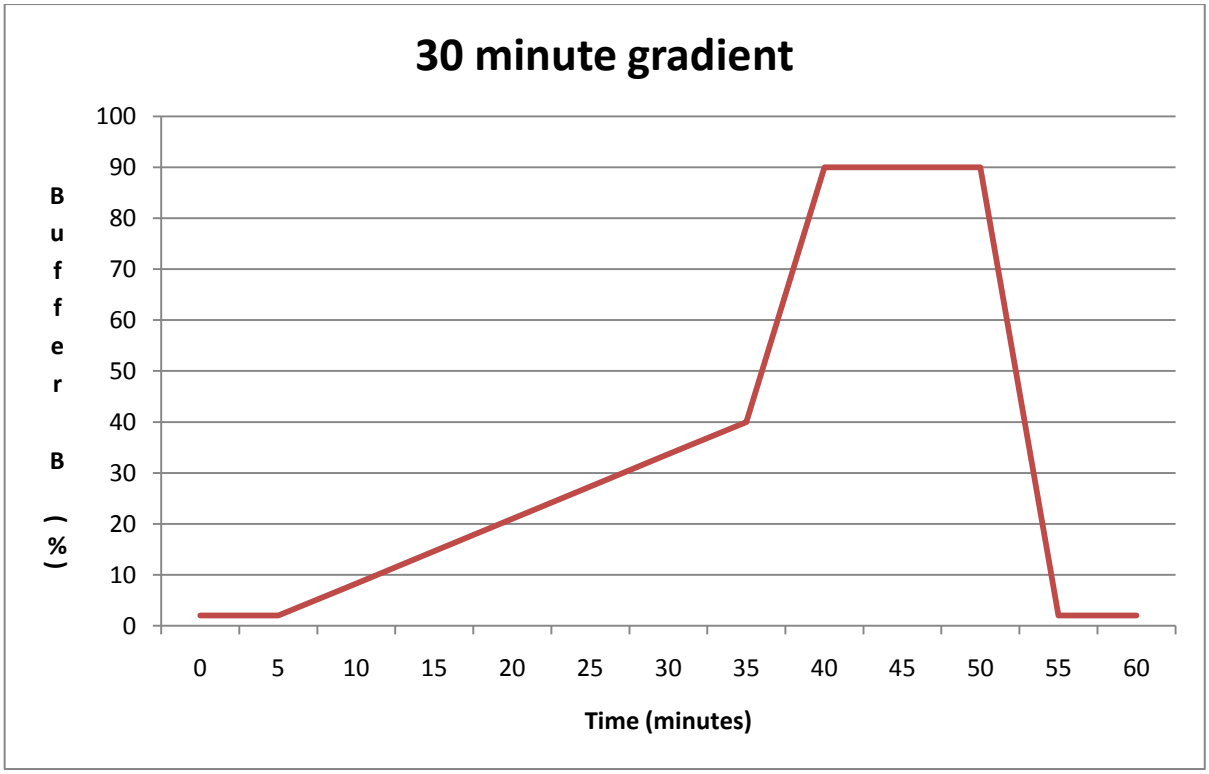


Figure 19. 30 minute SCX gradient.

Fractions were collected continuously into 96 well plates, limited to 3 min per well. UV data was collected at 214 nm, 280 nm and was used to characterise the complexity of fractions in order to select those amenable to pooling (low complexity).

2.1.1.8 Biotin affinity purification

A Softlink™ avidin column (Promega, Southampton, UK) was used for the recovery of biotinylated proteins from the microsome preparation. The column was first equilibrated with potassium phosphate buffer (pH 7.4). The sample was loaded into the column and allowed to flow through. Once loaded the column

was washed with several volumes of equilibration buffer, the washes were discarded. A 5 mM biotin solution (in equilibration buffer) was used to elute the bound proteins. Recovered fractions were dried down and reconstituted in 25 mM ammonium bicarbonate solution (pH 8.0) and subjected to a Bradford assay for protein concentration determination (2.1.1.1).

Potassium phosphate buffer (pH 7.4):

70 mM K_2HPO_4
30 mM KH_2PO_4

2.1.1.9 Delipidation

Samples from the liver microsome assay were centrifuged at 10,000 xg for 10 min and the supernatant stored. The pellets containing insoluble materials including proteins and lipids were resuspended in 1 ml of chloroform. The mixtures were shaken at 300 rpm for 1 hour at room temperature. 1 ml of 50% v/v methanol, 50% v/v ddH_2O was added to each sample followed by 30 min of vigorous vortexing at room temperature. Samples were centrifuged at 2,000 rpm for 1 min and the lipid enriched chloroform fractions discarded. 1 ml of chloroform was added to each sample followed by a 30 min sonication step in an ultrasonic bath filled with ice cold water to prevent overheating. Samples were again centrifuged at 10,000 rpm and the chloroform layer discarded. Samples were then subjected to an acetone precipitation (2.1.1.2.1) step for further cleanup then were reconstituted in 2 M urea and 250 mM ammonium bicarbonate prior to tryptic digestion.

2.1.1.10 In gel tryptic digestion and peptide extraction

Gel bands were excised and chopped into small pieces. Gel pieces were washed for 1 hour in 25 mM ammonium bicarbonate (AmBic) solution (pH 8), after washing the solution was discarded. A further 1 hour washing step was carried out using a 50% v/v acetonitrile, 50% v/v 100 mM AmBic solution, the solution was discarded after washing. Proteins were reduced with the addition of 10 µl of 45 mM DTT solution and a 30 min incubation at 60 ˚C. Alkylation was carried out

with the addition of 10 µl of 100 mM Iodoacetamide solution and a 1 hour incubation in darkness at room temperature. The solution was discarded and a 1 hour wash was carried out using 50% v/v acetonitrile, 100 mM AmBic solution. The wash was discarded and 50 µl of acetonitrile was added in order to shrink gel pieces. After 10 min the solvent was removed and the pieces dried in a vacuum centrifuge. 0.2 µg/µl sequencing grade modified porcine trypsin (Promega V111) in 25 mM AmBic solution was added in sufficient volume to rehydrate the gel pieces. 25 mM AmBic solution was added, ensuring that the gel pieces were fully immersed. Digestion was carried out overnight a 37 ˚C. Acetonitrile was added to the digest to 50% v/v and a 20 min incubation was carried out. Samples were centrifuged and the supernatants transferred to clean tubes. A further extraction step was carried out using 1% v/v formic acid, 99% v/v ddH_2O then a final extraction was carried out using acetonitrile. The supernatants from each extraction were added to those previously collected. Samples were dried down in a vacuum centrifuge and subsequently stored at -20 ˚C.

2.1.1.11 Western blotting

The samples of interest were run on 24 cm 2D gels using the protocol described (2.1.1.5). The gels were removed from the plates, washed in distilled ddH_2O. For each of the gels 6 pieces of appropriately sized (equal in size to the gel) blotting paper and one piece of PVDF (polyvinylidene fluoride) membrane were equilibrated in methanol for 5 min the transferred to distilled ddH_2O for 5 min. For each gel 3 pieces of blotting paper were placed inside the Amersham Pharmacia semi-dry blotter (GE healthcare, Little Chalfont, UK) followed by the PVDF membrane then the final 3 pieces of blotting paper. The lid of the transfer was locked in place and the power supply set to deliver 50 V at 400 mA for 2 hours.

Membranes were washed in PBST (0.05% v/v) solution for 5 min 3 times prior to blocking for 1 hour in 500 ml of a powdered milk solution (5% w/v). A further 3 PBST washes were carried out. The membrane was recovered and placed inside an A4 plastic pocket into which 70 ml of the blocking solution spiked with strep-

HRP (1000:1). The pocket was heat sealed and a 1 hour incubation at room temperature on a shaker was carried out. Following this a final 3 5 min washes were carried out with the PBST solution.

Pierce® enhanced chemiluminescence reagents (Thermo Scientific, Rockford, IL, USA)) were mixed at a 1:1 ratio, membranes were immersed in the solution for 3 min. Imaging was carried out in a G:box (Syngene, Cambridge, UK) using the "Chem blot" program. The settings allowed for the visualisation of each entire membrane and cumulative 30 second exposures (total exposure: 1 hour) were made in complete darkness.

Images of the gels were analysed using the Syngene software. Full size images of the gels were printed out for later use.

2.1.1.12 Colloidal Coomassie staining of 1d/2d gels

The bind silane treated gels had their front cover plates removed. The exposed gels were then each immersed in 500 ml of colloidal coomassie stain and placed on a shaker at 70 rpm for several days. The gels were then washed in distilled ddH_2O until the background staining had reduced enough to differentiate dyed protein spots.

The coomassie gels were visualised in the G:BOX (Syngene, Cambridge, UK)using the standard settings for coomassie stained gels.

Colloidal Coomassie dye stock:

0.1% v/v Coomassie brilliant blue G-250
76 mM ammonium sulphate
1.5 % w/v phosphoric acid

Colloidal Coomassie stain:
80 % v/v Colloidal Coomassie dye stock
20 % v/v Methanol

2.1.1.12.1Excision of Spots and Subsequent Tryptic Digestion

The gels were stained using a colloidal Coomassie stain (2.1.1.11) The print-outs from 2.1.1.11 were used to fix the location of proteins of interest on the stained gels i.e. the stained gels were placed on top of the 1:1 scale print outs and the regions on interest were highlighted. A round cutting tool was used to extract the spots of interest. These spots were then tryptically digested as described in 2.1.1.10.

2.1.1.13 Saturation DIGE (Analytical)

2.1.1.13.1HLM assay (Clozapine)

The assay consisted of a preparation of human liver microsomes (HLM), NADPH and suitable buffering system. Into this the drug of interest or an equivalent volume of DMSO was spiked. The mixture was incubated at 37 ˚C for 1 hour. Samples were cleaned up using 3 kDa spin filters. Proteins were recovered from the filter using pH 8.0 25 mM ammonium bicarbonate solution. The experimental design included 3 negative controls and 3 clozapine treatments with 3 biological replicates (batches) for a total of 6 samples. An equal aliquot from each of the 6 samples was taken and pooled to create a 7^{th} sample that would serve as the pooled internal standard.

2.1.1.13.2DIGE Labelling

Protein concentration was determined by Bradford assay (2.1.1.1) for each of the 7 samples. A volume equivalent to 5 µg of protein was taken from each sample and dried down in a SpeedVac concentrator. Each of these was then made up to a concentration of 1 µg/µl in 25 mM ammonium bicarbonate solution at pH 8.

Samples were labelled with 4 nmol of the appropriate CyDye™ solution, the aturation dyes used had maleimide reactive groups, (GE healthcare, Little Chalfont, UK) at 37 ˚C for 30 min in the dark. The reaction was stopped by the

addition of 2x sample buffer (7 M urea, 2 M thiourea, 4% w/v CHAPS, 2% v/v pharmalytes and 130 mM DTT) equal to the reaction volume.

2.1.1.13.3IEF

Samples were combined (Cy3 and Cy5 pairs), made up to volume in rehydration buffer and loaded onto the appropriate gel strips. The strips were loaded onto an Ettan™ IPGphor™ 3 system and a standard DIGE program was run.

Table 1. Identification of experimental conditions in relation to IPG gel number.

Gel No.	(Sample)	Pooled
62244	(1) Cy 3	Cy 5
62245	(2) Cy 3	Cy 5
62246	(3) Cy 3	Cy 5
62247	(4) Cy 3	Cy 5
62248	(5) Cy 3	Cy 5
62249	(6) Cy 3	Cy 5
62250	(1) Cy 5	Cy 3
62251	(2) Cy 5	Cy 3
62252	(3) Cy 5	Cy 3
62253	(4) Cy 5	Cy 3
62254	(5) Cy 5	Cy 3
62255	(6) Cy 5	Cy 3

2.1.1.13.4SDS-PAGE

The IPG strips were recovered from the Ettan™ IPGphor™ 3 system and treated with SDS equilibration buffer (SEB) spiked with 65 mM DTT for 15 min. Subsequently the buffer was emptied and replaced with SEB spiked with 135 mM Iodoacetamide and incubated for 15 min. The IPG strips were removed and

loaded onto prepared 2d gels (2.1.1.5), the strips were fixed in place with agarose. The gels were loaded and run for 16 hours (1W per gel).

2.1.1.13.5 Scanning of gels

Gels were imaged using a 9400 Typhoon scanner (GE healthcare, Little Chalfont, UK) . Each gel was scanned using the green laser (580 nm) for Cy3 and red laser (650 nm) for Cy5. The resolution was set to 100 microns/pixel.

2.1.1.13.6 Analysis of DIGE images

The 24 captured images (12 samples each with Cy3 and Cy5 images) were analysed using the proprietary DeCyder™ 7.0 software (GE healthcare, Little Chalfont, UK). Images were manually cropped in order to remove any obvious background noise at the extreme edges. The software matching algorithm was (DIA module) then applied in order to correlate spots across all of the different gel images. In order to ensure as complete matching as was possible some time was then spent in manually matching spots that had been missed by the software.

The biological variance analysis (BVA) module of the DeCyder™ program was used to assign statistical values to changes in protein concentrations across the gels with Clozapine treated images being compared to untreated images. A table was compiled of any statistically significant ($p<0.05$) decreases in intensity in the treated samples vs. untreated.

2.1.1.14 Preparative DIGE

2.1.1.14.1 HLM assay

The assay consisted of a preparation of human liver microsomes (HLM) (0.5ml Pooled human liver microsomes at a concentration of 20 mg/ml)(BD Biosciences, UK), NADPH and suitable buffering system. Into this the drug of interest was

spiked. The mixture was incubated at 37 ˚C for 1 hour. A sample was run with the inclusion of GSH as a positive control to ensure adduct formation, the sample to be run ion the preparative DiGE experiment did not contain GSH. The final assay concentrations were as follows: HLM 0.5g/ml, NADPH 1mM, MeOH 1.5% v/v, GSH 4mM (positive control only) and Clozapine 10µM. Samples were cleaned up using 3 kDa spin filters. Proteins were recovered from the filter using pH 8.0 25 mM ammonium bicarbonate solution.

2.1.1.14.2DiGE

The preparative CyDye™ (Cy3) was made up to a 20 mM working solution as specified in the supplied protocol (Amersham CyDye DIGE Fluor Libelling Kit for Scarce Samples). 250 µl of the 2 mg/ml sample was loaded into a fresh microfuge tube to which 20 µl of 20mM Cy3 saturation dye was added and mixed vigorously by pipetting. The sample was centrifuged briefly then incubated at 37 ˚C in the dark for 30 min. The reaction was stopped with the addition of 175.5 µl of 1x sample buffer (DTT/pharmalytes free) and vigorous mixing. 4.5 µl of pH 4-7 pharmalytes were added followed by mixing. 4.5 mg of DTT was added and a final mixing was administered. IEF, SDS-PAGE and scanning of gels was carried out.

2.1.1.14.8 Excision of spots from the preparatory DiGE gel

Scanning of the preparatory gel was carried out as described in 2.1.1.13.5. The image was loaded into DeCyder's DIA module as both a Cy3 and Cy5 channel and spot identification carried out. The resulting data was then entered into the BVA module and spot identification was carried out against the gel images loaded from the analytical DiGE experiments. Spots were added to a pick list which was then exported to the Spot Handling Workstation (Amersham Biosciences, UK). The gel was loaded into the robot and spots were then picked automatically. The large picking head (2.0mm) was used.

2.1.1.15 GSH trapping assay

Glutathione was used to trap the reactive metabolites produced from drugs using a human liver microsome (BD Biosciences) system. The reaction mixture comprised 200mM potassium phosphate buffer pH 7.4, 0.5mg/ml human hepatic microsomes, 10µM clozapine, 1mM NADPH, 4mM GSH, 1.5% v/v acetonitrile and 0.1% v/v DMSO. Negative controls were run each without either clozapine, NADPH or GSH. The reaction mixtures were pre-heated to 37 ˚C for 10 min in a shaking water bath prior to the addition of NADPH. After an hour at 37 ˚C the reactions were terminated with the addition of ice cold acetonitrile to 50% v/v. The reaction mixtures were cooled on ice for 15 min then centrifuged at 4000rpm for 10 min at 10 ˚C. The samples were cleaned up using a C18 solid phase extraction cartridge (detailed in 2.1.1.19) then dried down in a vacuum centrifuge for storage at -20˚C.

Prior to use samples were reconstituted in buffer A (2% v/v acetonitrile, 98% v/v ddH_2O, 0.1% v/v formic acid).

0.5 M Potassium phosphate buffer pH 7.4

359 mM K_2HPO_4
141 mM KH_2PO_4

2.1.1.16 Liver microsome assay with synthetic peptides

The assay is carried out as described in 2.1.1.15 without the acetonitrile precipitation (with the exception of the positive control). The reaction mixtures contained 200mM potassium phosphate buffer pH 7.4, 0.5 mg/ml human hepatic microsomes, 10µM clozapine, 1 mM NADPH, 1 nM Synthetic peptide (1, 2 or 3), 1.5 % v/v acetonitrile and 0.1 % v/v DMSO. The positive control was run with 4mM glutathione in place of the synthetic peptide.

The sample supernatants were subjected to affinity purification as detailed in (2.1.1.8).

2.1.1.17 Liver Microsome Assay for SCX, OFFGEL and GeLC

The assay is carried out as described in 2.1.1.16 without the acetonitrile precipitation, centrifugation and RP C18 cartridge clean up. Two reaction mixtures were run for each of the separation approaches one containing 200 mM potassium phosphate buffer pH 7.4, 0.5 mg/ml human hepatic microsomes, 10 µM clozapine, 1 mM NADPH, 1.5 % v/v acetonitrile and 0.1 % v/v DMSO. The other contained the same minus the clozapine. The samples for GeLC were stored at -20 ˚C. The others were subjected to the delipidation protocol detailed in 2.1.1.9.

2.1.1.18 Liver Microsome Assay With Other Drugs

The drugs clozapine, imipramine, tacrine, naproxen and acetaminophen were metabolised and their metabolites subsequently trapped using the assay described in 2.1.1.16. Two reaction mixtures were used for each drug, one containing 200 mM potassium phosphate buffer pH 7.4, 0.5 mg/ml human hepatic microsomes, 10 µM drug, 1 mM NADPH, 1.5 % v/v acetonitrile and 0.1 % v/v DMSO. The other contained the same minus the clozapine.

2.1.1.19 Solid phase extraction (SPE)

Sep-Pak reverse phased C18 cartridges (Waters, Hertfordshire, UK) were wetted using 6 cartridge volumes of acetonitrile. The cartridges were then equilibrated with a further 6 cartridge volumes of a 5% v/v acetonitrile, 95% v/v ddH_2O solution. Samples were then loaded onto the cartridge in buffer A (2% v/v MeCN, 98% v/v ddH_2O, 0.1% v/v Formic acid). Washing was achieved by flushing the cartridge with several volumes of the 5% v/v acetonitrile solution. 2 ml of an 80% v/v acetonitrile, 20% v/v ddH_2O was then injected to elute proteins from the cartridge. The elutions were collected and dried in a vacuum centrifuge. The samples were reconstituted in 200 µl of buffer A.

2.1.2 Mass Spectrometry and HPLC

2.1.2.1 Direct Injection Optimization of Collision Energy for Precursor Ion Scanning

Peptides 1-3 were suspended in a 50% v/v methanol, 49.9% v/v ddH2O solution spiked with 0.1% v/v formic acid at a final concentration of 1pg/ul. Approximately 10 µl of sample was pipetted into a Proxeon direct injection needle (Thermo Scientific, UK) loaded into a centrifuge fitting and briefly spun. The needle was then removed from the fitting, the end removed using a diamond edged cutting tool and mounted in the API 2000™ (AB SCIEX, Warrington, UK) direct injection assembly. Backpressure was applied to the needle via a syringe fitting; the needle tip was carefully broken in order to allow for electrospraying of the sample. Data was gathered for each sample in both +EMS mode and +EPI mode at a range of collision energies (30 eV, 40 eV and 50 eV). The intensity of the fragment ion at 359.1 m/z was monitored, collision energies were adjusted down from 50 eV to 45 eV in increments of 1 eV. It was found that for all 3 peptides tested a collision energy of 47 eV related to the highest intensity in the target ion at 359.1 m/z. Automatic optimisation of other parameters were carried out using the Analyst software. The following values were used:

CAD: -3, Curtain gas (CUR): 20, GS1: 10, GS2: 0, Interface heater temperature (IHT): 150, Collision cell exit potential (CXP): 12, Declustering potential (DP): 100, Entrance potential (EP): 10

2.1.2.2 Reversed phase liquid chromatography –UV-mass spectrometry

Samples were reconstituted in buffer A (97.9% v/v ddH_2O, 2% v/v MeCN, 0.1% v/v formic acid), loaded into 96 well plates or individual glass vials and mounted in the autosampler of an UltiMate® 3000 HPLC system. 2 µl of sample was injected into a 20 µl sample loop, from there onto a 300 µm i.d. x 5 mm C18 guard column (5 µm, 100 Å) at a flow rate of 30 µl/min using buffer C (97.9% v/v ddH_2O, 2% v/v MeCN, 0.1% v/v trifluoroacetic acid) as a loading buffer, after 5 min the valves were switched and a 30 min gradient was applied with a flow rate of 300 nl/min (a 1000:1 flow splitter was installed in the UltiMate® 3000 system). Buffer B composition was as follows: 80% v/v MeCN, 19.9% v/v ddH_2O, 0.1% v/v formic acid.

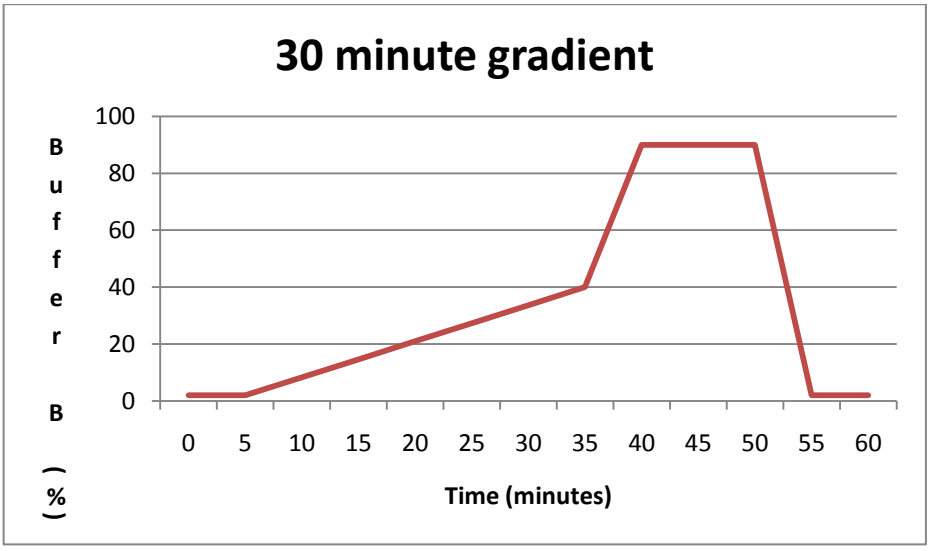


Figure 20. The reversed phase 30 minute gradient used in HPLC experiments.

A 75 µm i.d. x 15 cm Pepmap 100 C18 column (3 µm, 100 Å) was used for peptide separations. The column output was linked to an UltiMate® UV detector which in turn was connected to a PicoTip®(New Objective, Basingstoke, UK) fused silica emitter (i.d. 20 µm) connected to an API 2000™, 4000™ or 5500™ Q-trap via a nanospray source interface. The UltiMate® 3000 flow manager (FLM) unit 10 port switching valve was configured as shown in figure 25. It should be noted that the re-equilibration time (from 55-60 min; 5 min total (Fig 24)) was too short, and a length of at least 10 min should be applied.

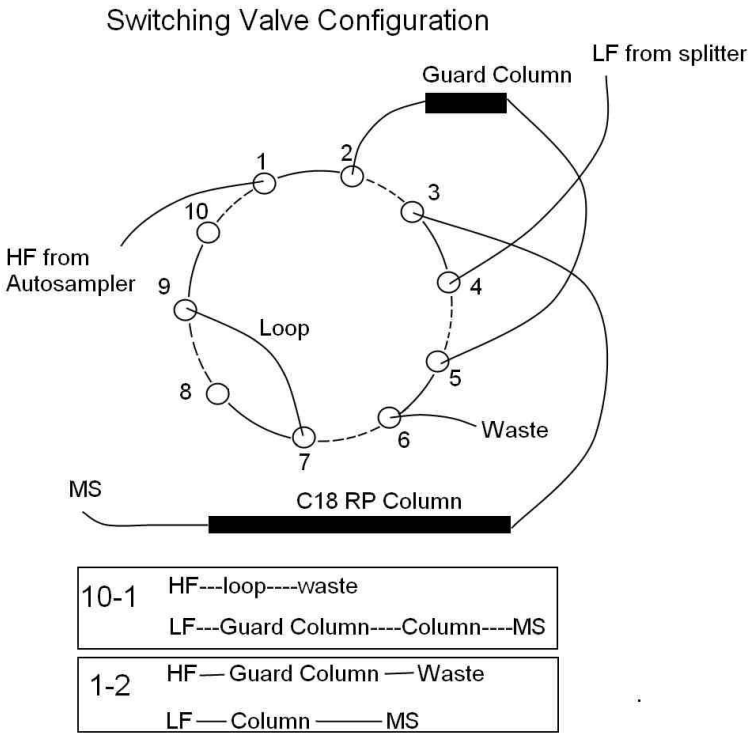


Figure 21. FLM 10 port switching valve configuration. In the 1-2 position samples is loaded onto the guard column. In the 10-1 position the sample is eluted from the guard column on to the column and subsequently the mass spectrometer. HF refers to high flow (30 µl/min), LF refers to low flow (300 nl/min).

UV data was collected on two channels at 214 nm and 280 nm. MS data was acquired using Analyst software package and a variety of acquisition methods detailed as follows.

2.1.2.3 Information dependant acquisition (IDA) of MS/MS (API 5500™)

An information dependant acquisition method was written to obtain data from samples based on the most intense peak as identified by a survey MS scan and fragmented by collision induced dissociation (CID) at a collision energy based on the mass and charge state of the ion based on data from an enhanced resolution (+ER) MS survey scan. The instrument's ion spray voltage was set to 2300 v, the collision gas to 20 and the interface heater to 150 °C. An EMS (enhanced mass spectrum) scan was set to analyse from 300-1000 m/z at a step size of 0.12 m/z.

An ER (enhanced resolution) scan was carried out with a step size of 0.02 m/z for the identification of charge states. An IDA (information dependant acquisition) step was added to allow for the exclusion of ions for 60 seconds after two subsequent MS/MS events. Three EPI (enhanced precursor ion) experiments were carried out, each scanning from 50-1000 m/z with a step size of 0.12 m/z. The total cycle time was 1.5 seconds.

2.1.2.4 NL129 scanning method (API 4000™)

The neutral loss scanning approach looks for a characteristic loss of mass between the first and third quadrupoles. The instrument's ion spray voltage was set to 2000 v, curtain gas to 20 and the interface heater to 150 ˚C. Experiment one was set to a neutral loss of 129 m/z with a step size of 1 m/z covering the mass range 300-650 m/z with a collision energy of 40 eV. The second experiment was set to ER (enhanced resolution) with a step size of 0.03 m/z. Experiments 3-5 were EPI (enhanced product ion) scans of ions identified in the ER experiment. The mass range covered was 50-1500 m/z with a step size of 0.12 m/z. The total cycle time was 4.3 seconds.

2.1.2.5 Selective precursor ion scanning (API 4000™ and API 5500™)

The precursor ion scanning approach is used to detect ions that upon CID produce a characteristic fragment of interest. Work was carried out using an API 4000™ Q-trap instrument coupled to an UltiMate® 3000 HPLC system. The flow rate was 300 nl/min over a period of 40 min with a linear gradient. 6µl of sample was loaded onto the guard column, equilibrated there and washed with buffer A (2% v/v MeCN, 98% v/v ddH_2O, 0.1% v/v Formic acid)for a period of 5 min prior to elution. Data acquisition was managed using the Analyst software (AB SCIEX, Warrington, UK). Briefly, the instrument was set to positive precursor scan for an m/z of 359 with a collision energy of 47 eV (2.1.2.1) and a curtain gas of 20; data was collected in peak hopping mode with a step size of 1 m/z unit over the range 450-1400 m/z. Subsequently the instrument was set to ER (enhanced resolution) mode to gather high resolution MS data at a step size of 0.03 m/z.

Three EPI (enhanced product ion) scans were then carried out with a step size of 0.12 m/z from 100-1400 m/z. Each cycle lasted for 4.3 seconds allowing for 480 cycles per LC run.

These experiments were also run using the API 5500™ Q-trap. Settings were modified to better suit the characteristics of the instrument. Briefly, the mass arrange scanned was changed to 400-1000 m/z and the EPI experiments were run at a faster scan rate allowing for a total cycle time of just 2.8 seconds.

2.1.2.6 Selective precursor scanning in the negative ion mode

Previous work has indicated that scanning for a precursor of 272 m/z in the negative ion mode is a sensitive method for the detection of glutathione-drug conjugates. In addition to the system specified in (2.1.2.2) an additional pump was used to introduce a 20% v/v methanol 80% v/v isopropyl alcohol (IPA) solution to the column output via a t-piece junction prior to electrospray ionisation. The IPA solution was set to flow at 100 nl/min and is required in order to ensure a stable current during negative ion mode electrospray, an ACCURATE® flow splitter(LC Packings/Dionex, Camberly, UK) was used to ensure a stable flow rate. A PicoTip™ fused silica emitter type FS360 75 xx 15 (New Objective, Basingstoke, UK) was used to reduce arcing and allow for an increased needle lifespan.

The instrument was set to –PI (negative precursor ion) mode with an ion spray voltage of -1750 v, curtain gas of 20 and a settle time of 700 ms. Subsequently a –ER (negative enhanced resolution) scan was carried out at a rate of 250 (m/z)/s at a resolution of 0.03 m/z. An IDA scan (information dependant acquisition) step was carried out to select ions for MS/MS experiments. Ions were allowed to be fragmented twice before being added to an exclusion list for 60 seconds. Three +EPI (enhanced product ion) experiments were carried out per cycle across the mass range 50-1500 m/z with a step size of 0.12 m/z and a scan rate of 4000 (m/z)/s. The total cycle length was 3 seconds, sufficient to sample data from a single peptide peak up to 10 times across a peak width of 30s (measured at 10% of full height.

2.1.2.7 Precursor ion scanning of 574 m/z (API 5500™)

The method was identical to that described in 2.1.2.5 but with the product mass set to 574 m/z.

2.1.3 Molecular biology

2.1.3.1 Transformation of E.coli with plasmid

E.coli BL21 (DE-3) cells were selected for transformation due to their high levels of protein expression. BL21 (DE-3) cells have an IPTG inducible T7 RNA polymerase gene which on induction leads to the processing of the plasmid gene of interest (His-ERK2). Cells were treated with 100 mM Calcium chloride ($CaCl_2$) to increase membrane permeability. 1 µl of plasmid (His-ERK2 with ampicillin resistance) was added to 100 µl of the cells and allowed to incubate for 30 min on ice. The cells underwent heatshock at 42 ˚C for 90 seconds then were cooled on ice for a further 5 min. 0.5 ml of broth was added with an incubation period of 1 hour at 37 ˚C. The cells were then spread onto an agar plate and stored at 37 ˚C until colonies became visible.

Broth (pH 7.0) :	Agar plate:
1 % w/v bacto-tryptone	Broth
0.5 % w/v bacto-yeast extract	0.0001% v/v ampicillin solution
171 mM NaCl	1.5 % w/v agar

2.1.3.2 Colony selection and protein expression

Plates were inspected for signs of contamination. A suitable colony was selected and added to 5 ml of LB broth spiked with 5 µl of 100 mg/ml ampicillin. A

replicate of the LB solution minus the addition of BL21 cells was used to act as a control. The samples were placed in a shaking incubator and stored overnight at 37 ˚C. The next morning the cultures were moved to the fridge if the control sample is negative, if it has growth then the work must be started over. The cells were stored in the fridge until noon then 1ml of the culture was extracted and used to inoculate 100 ml of overnight express™ medium containing 100 µl of ampicillin. A 4 hour incubation period at 37 ˚C in a shaking incubator was carried out. The temperature in the incubator was then lowered to 25 ˚C and the cells left to shake overnight. The culture was split into two 50 ml conical tubes and centrifuged at 4000 rpm at 4 ˚C for 30 min. The supernatant was discarded and the pellets stored at -80 ˚C. The Overnight Express™ system spontaneously induces protein expression in IPTG-inducible bacterial expression systems. This occurs after cells have grown to a high density and does not require the addition of any further inducers.

Overnight express™ medium:

6% w/v overnight express medium (Millipore
1 % w/v glycerol (sterile)
0.0001% w/v ampicillin

2.1.3.3 Recovery of protein

The pellets were removed from the -80 ˚C and allowed to thaw at room temperature then were resuspended in 5 ml of resuspension buffer (20 mg lyzosyme/20 ml of lysis buffer, 2 protease tablets). Samples were cooled on ice for 30 min prior to short bursts of sonication interspersed with cooling on ice for a total of around 15 min. The tubes were centrifuged at 4000 rpm and the lysate filtered through a 0.22 µm filter. The lysate was then loaded onto a His-select nickel affinity gel (Sigma, Dorset, UK) packed column prewashed with equilibration buffer. A 1 hour incubation period was carried out at room temperature then the lysate was allowed to flow through the column. 3 10 ml washes were carried out using the wash buffer solution followed by 5 1 ml elutions were carried out using the elution buffer solution. All fractions were

collected and samples analysed by 1d PAGE. Fractions bearing the His-ERK2 protein were pooled and protein concentration was determined by Bradford assay.

1x TBS (pH 7.4):

365 mM NaCl
27 mM KCl
248 mM tris Base

His-tag elution buffer (pH 8.0):

250 mM imidazole in 1xTBS

HIS-tag wash buffer (pH 8.0):

20 mM imidazole in 1xTBS

Lysis buffer (pH 8.0):

10 mM imidazole in 1xTBS
1mg/ml Lysozyme

2.1.4 Bioinformatics

2.1.4.1 In silico protein digestion

The MS-DIGEST tool found at http://prospector.ucsf.edu/prospector/mshome.htm was used to carry out in-silico digestion of proteins of interest.

2.1.4.2 In silico collision induced dissociation

The MS-Product tool at http://prospector.ucsf.edu/prospector/mshome.htm was used to simulate in-silico the collision induced dissociation fragments of the three synthetic peptides. The relevant amino acid sequences were entered along with an N-terminal biotinylation. Ion types b and y were selected and the program run. The theoretical peak table was recorded for each synthetic polypeptide.

2.1.4.3 Mascot

The Mascot search engine (http://www.matrixscience.com/) installed in-house on a Glasgow University server was used to identify proteins and modifications from mass spectrometric data. MS/MS ion searches were carried out against the Swissprot (version 56.6; 405506 sequences; 146166984 residues) genomic protein database against the human (homo sapiens) taxon (20413 sequences), the file of interest was selected along with the fixed modification of Carbamidomethyl (C) and the variable modification of Oxidation (M). When searching files with suspected protein-drug adducts no fixed modifications were selected, both Carbamidomethyl (C) and Oxidation (M) were selected as variable modifications along with relevant metabolite based modifications (Clozapine1 – 5 (C)). The enzyme used for digestion was set to trypsin and the possibility of 1 missed cleavage allowed for. Peptide tolerance was set to +/- 2 Da with MS/MS tolerance set to +/-0.6 Da. Peptide charge was set to 1+, 2+ and 3+ and monoisotopic mass was selected. The instrument type was set to ESI-QUAD.

Data files from the SCX and OFFGEL separation experiments were combined using the peaklist conversion tool (Proteomecommons.org IO framework 6.21) in order to improve protein identifications and sequence coverage.

2.1.4.4 3D protein analysis (DEEPVIEW)

The DeepView (Swiss-PdbViewer) software was downloaded from http://spdbv.vital-it.ch/ and installed. The proteins of interest were located in the Swissprot database and the most detailed x-ray crystallographic or NMR based 3d structural file (*.PDB) downloaded. The image was loaded into DeepView, the cysteine residues highlighted and those not involved in disulfide bridge formation identified. Those unpaired residues located at the surface of the protein, and as such potentially reactive, were noted.

2.1.4.5 Identification of membrane associated proteins

Protein identifications obtained from the Mascot search provided UniProtKB/Swiss-Prot accession numbers. An exhaustive list for each separation type was compiled and their associated FASTA files recovered from Uniprot (http://www.uniprot.org/). The FASTA files were submitted to the TMHMM server (http://www.cbs.dtu.dk/services/TMHMM-2.0/) for analysis.

2.1.4.6 Identification of potential electrophile binding motifs

FASTA files of proteins of interest were uploaded to the program motif_HUNTER (http://proteotools.pharmacy.arizona.edu/proteotools/motif.jsp). The KK, K?K, CH, HC, CR, RC, KC, CK were submitted as search terms. K represents lysine, C cysteine, H histidine and R arginine. The ? represents a wild card operator which allows for the presence of any amino acid.

2.1.5 Chemistry

2.1.5.1 Biotinylation of N-desmethyl clozapine

1 mg of N-desmethylclozapine and 6.6 mg of pentafluorophenyl biotin (PFP-biotin) was added to 100 µl of anhydrous DMSO. The solution was allowed to incubate overnight at room temperature. The unreacted PFP-biotin was quenched with the addition of lysine at an equimolar concentration. Anhydrous DMSO was added to the mixture to give a final concentration of biotinylated N-desmethylclozapine of 10 mM. The resulting solution was then purified by HPLC in order to separate biotinylated and unbiotinylated drug.

2.1.5.1 Purification of biotinylated desmethylclozapine (bDMC)

Products of the reaction were separate using an HPLC system equipped with a C18 reverse phase column and a UV detector. Fractions were collected as called

for by the UV trace (channels: 214nm, 254nm) and were analysed by mass spectrometry. The fraction containing biotinylated desmethylclozapine was retained (this fraction had a m/z value DMC plus that of biotin) and dried in a SpeedVac evaporator. The bDMC was reconstituted in DMSO and stored at 4 °C.

2.1.6 Materials

Unless stated otherwise all chemicals and reagents were acquired from Sigma Aldrich, UK.

Chapter 3: Trapping of Reactive Metabolites

3.1 Aims

An important step in determining the potential toxicity of a given new chemical entity (NCE) is to identify its metabolites. Triple quadrupole instruments are uniquely suited to this task with their highly selective neutral loss and precursor ion scanning modes being able to identify low abundance ions against the background of complex samples.

The following work focused on the identification of the metabolites of various drugs and the design of a selective precursor ion scan for the detection of said metabolites when conjugated with polypeptides.

The aims of the work carried out in this chapter were as follows:

1) The formation and trapping of drug metabolites in a liver microsome system spiked with glutathione.
2) Recovery of metabolite-glutathione adducts and their subsequent analysis by LC-MS for the identification of potential precursor ions through examination of CID fragmentation patterns.

3) Formation of metabolite-peptide adducts and their detection by LC-MS using the precursor ion scan.

3.2 Introduction

The simplest method for the generation of reactive metabolites is a direct chemical synthesis. Work carried out by Damsten et al (2007) demonstrated the ability to synthesise NAPQI through the treatment of APAP with freshly prepared silver oxide. They went on to demonstrate that incubating NAPQI with human serum albumin lead to the formation of NAPQI-HSA conjugates. Detection of these conjugations was possible using liquid chromatography mass spectrometry following enzymatic digestion of the modified HSA. This system was found to generate biologically accurate metabolites and protein adducts whilst maintaining an extremely simple chemical background suitable for study and further method development.

However it is not always the case that such a straightforward synthesis of reactive metabolites is possible. Phase I metabolism of drug molecules typically results in a variety of structurally distinct metabolites (Linnet and Olesen, 1997; Davis et al., 1995; Hinson, 1983; Lemoine et al., 1993; Zheng et al., 2011). A single chemical synthetic pathway is not capable of producing the wide variety of biologically mediated metabolites; therefore a range of different reactions must be used to ensure similar diversity. More importantly metabolism of a drug will produce the various metabolites at differing concentrations (Lemoine et al., 1993) which would most likely lead to particular protein modification profiles. In addition, the nature of a synthetic metabolite must be compared to a biological counterpart in order to verify its authenticity and usefulness as part of a model system.

Another approach involved the use of an electrochemical cell to mimic cytochrome P450 activity and lead to the successful detection of various drug metabolites (Jurva et al., 2003). However it was found that only one-electron oxidations could be produced resulting in hydroxylation of aromatics, oxidation of alcohols to aldehydes, S- and P- oxidation and N-dealkylation of amines. The strength of the technique is more limited to the identification of labile oxidation

sites of drug molecules rather than the accurate simulation of metabolism (Baumann and Karst, 2010). As such the system is not particularly useful for the identification of protein adducts.

A liver microsome system can be used to produce a wide range of metabolites from a given xenobiotic molecule. And although metabolism in a liver microsome system cannot produce the full range of metabolites produced in vivo (Rufer et al., 2007; Di et al., 2012) it can produce a comparable range of cytochrome P450 mediated metabolites (Di et al., 2012). The tripeptide glutathione (GSH) acts as a sink for a wide range of electrophilic metabolites (Jakoby, 1990; Boyland, 2006). The reduced form of the molecule (GSH) possesses a cysteine residue with a reactive sulfhydral side group allowing for the formation of glutathione-metabolite adduct formation. Glutathione adducts are readily formed in vitro using a simple assay and are easily recovered using an acetonitrile protein precipitation step. Using a triple quadrupole instrument, it is possible to detect metabolite-glutathione adducts using a highly selective neutral loss scan (Baillie and Davis, 1993; Yan and Caldwell 2004), precursor ion scan (Wen et al., 2008) or multiple reaction monitoring (SRM) (Zheng et al., 2007) . MS/MS scans of candidate ions reveal CID fragmentation information that can be used to characterise metabolites.

The formation of glutathione adducts is a highly efficient process mediated by the enzyme glutathione S-transferase (Coles, 1984; Booth et al., 1961).Consequently a glutathione based system can't accurately model the stochastic electrophilic attack of proteins . In order to more accurately model this process three synthetic peptides were designed (1.1.5.4). Data from the glutathione trapping assay was used to identify CID fragments of metabolite-glutathione adducts, based on metabolite structural information, that could be used to create a selective precursor ion scan. The selectivity of the precursor ion scan acts to distinguish modified from unmodified molecules and to isolate these from the highly complex background present in a liver microsome system (Annan et al., 2001; Zappacosta et al., 2002; Williamson et al., 2006).

The workflow thus comprised (i) glutathione trapping of reactive metabolites; (ii) MS/MS based characterisation of metabolites; (iii) design of suitable synthetic peptides; (iv) identification of characteristic metabolite ions; (v) liver

microsome assay spiked with synthetic peptides then searched using the precursor ion scan.

3.3 Methods and Materials

3.3.1 Glutathione Trapping Assay

Clozapine was incubated with glutathione and rat liver microsomes in order to generate and capture its reactive metabolites. After an incubation period of 1 hour the reaction was stopped with the addition of ice cold acetonitrile. The resultant solution was centrifuged and the supernatant retained. Cleaning of the sample was carried out using C18 reverse phase Sep-Pak cartridges (2.1.1.19). After cleaning, the sample was evaporated to dryness using a rotaevaporator. 400 µ l of buffer A was used to reconstitute the sample. For a more detailed account see (2.1.1.15).

3.3.2 Analysis of Assay Products by LC-UV-MS (NL129)

Samples were reconstituted in 1ml of buffer A (95% v/v ddH_2O, 4.9% MeCN, 0.1% FA) and 1 µl loaded into a 96 well plate with a further 19 µ l of buffer A. The plate was loaded into the autosampler of the UltiMate® 3000 HPLC system (Dionex) coupled to a Q-trap 4000™ (Applied Biosystems) mass spectrometer. A 60 min reverse phase gradient was run (2.1.2.2) and mass spectrometric data was gathered using a combination of a NL129 scan and MS/MS experiments(2.1.2.3 and 2.1.2.4). The data was manually inspected for the presence of metabolite glutathione adducts and the presence of ions with potential for use as precursor ion scanning targets.

3.3.3 Analysis of Assay Products by LC-UV-MS (PI272)

The drugs clozapine, tacrine, naproxen and imipramine were spiked into the glutathione trapping assay as described in (2.1.1.18). The products of the assay

were analysed by LC-UV-MS (PI272) running the Q-trap in the negative ion mode (2.1.2.6). UV data and MS data were manually analysed in order to identify possible metabolite glutathione adducts. Tandem mass spectrometric (MS/MS) data from these experiments was used to characterise adducts and examine fragment ions for potentially useful precursors.

3.3.4 Identification of Clozapine Glutathione Adducts Using a PI359 Scan

Clozapine modified glutathione was analysed by LC-UV-MS using an MS program designed to identify precursor ions giving rise to fragments with m/z 359 (2.1.2.5). The experiment was carried out on an API 4000™ model Q-trap.

3.3.5 Design of Synthetic Peptides

KEAP1 and two cytochrome P450 enzymes (isoforms A and B) were selected for the reasons previously detailed (1.1.5.4), the protein sequences were downloaded from the Swissprot database and subjected to an *in silico* tryptic digestion. The virtual digests were manually analysed and peptide fragments fitting the aforementioned criteria were selected for synthesis (Peptide Synthetics, Fareham, UK). In order to afford effective recovery from the complex background of the microsome assay the peptide sequences were N-terminally biotinylated.

3.3.6 Mass Spectrometric Characterisation of Synthetic Peptides

Peptides were reconstituted in a 50% v/v methanol, 50% v/v distilled water solution spiked with 0.1% v/v formic acid. 100 µl of the solution was loaded into a glass syringe that was then placed in a syringe drive and interfaced to an HCT ion trap (Bruker Daltonics, Bremen, Germany) mass spectrometer. A Proxeon steel needle was used for the electrospraying of the solution supplied at a rate of 5 µl/min by the syringe pump. Data was accumulated using Bruker's Compass software. Mass spectrometric data and subsequent CID fragmentation based

MS/MS data was recorded and analysed. Samples were again analysed using the API 4000™ Q-trap in direct injection mode (2.1.2.1) in order to obtain data on the low mass ions that could not be observed in the ion trap. For

3.3.7 Clozapine Synthetic Peptide Adducts Formation and Detection

A variation on the standard glutathione trapping assay was run with synthetic peptides acting as a replacement for glutathione (2.1.1.16). The peptides were recovered from the assay by affinity purification using an avidin functionalised column (2.1.1.8). Once reconstituted in buffer A the samples were analysed by LC-MS as previously described. An API 4000™ (Applied Biosystems) was used for mass spectrometric analysis; the previously established PI359 scan (2.1.2.5) was selected. Data obtained from these runs was manually analysed and de novo sequencing of the peptides was carried out and metabolite bearing fragments identified.

3.3.8 Reduction and Alkylation of Modified Peptides

Peptide samples recovered from the liver microsome assay by avidin purification were reconstituted into 50 µl of a 25 mM ammonium bicarbonate solution (pH 8.0). 5 µl of 50 mM DTT solution was added to 25 µl of each of the peptide solutions. After a 30 min incubation period at 60 °C 5 µl of 100 mM Iodoacetamide solution was added. A further incubation was carried out at room temperature for 15 min in the dark. Samples were dried in a vacuum centrifuge before being reconstituted in 1 ml of buffer A. 2 µl of each sample was loaded into a 96 well plate before undergoing LC-MS analysis as described in (2.1.2.2). The instrument used for MS analysis (2.1.2.3 and 2.1.2.5) was an API 4000™ Q-trap.

3.4 Results

3.4.1 Characterisation of Metabolites by GSH Trapping and the NL129 Scan

The Q-trap instrument offers the unique highly selective scanning modes, neutral loss/precursor ion/multiple reaction monitoring, of a triple quadrupole and combines them with the high sensitivity of an ion trap. This means that in a single duty cycle it is possible to identify a particular ion by a characteristic CID fragment or neutral loss, following this it is then possible to perform an enhanced resolution scan in order to gain a more accurate mass determination and thereby a more accurate charge state before initiating an MS/MS experiment.

Collision induced dissociation of glutathione or a glutathione adduct can be detected by the presence of a characteristic neutral loss 129 Da (NL129) representative of cleavage of the gamma-glutamyl moiety (Baillie and Davis, 1993). Ions producing this neutral loss were selected for and underwent CID in order to provide structural information on any metabolite adducts present.

The work was carried out at nanoflow rates (300 nl/min) consistent with typical proteomics based LC-MS experimentation. At such flow rates the usage of material is minimised and sensitivity maximised through concentration of analytes (Cutillas, 2005). Examination of a typical total ion chromatograph (TIC) from a liver microsome sample spiked with glutathione and clozapine reveals the presence of a single high intensity peak eluting at 25.91 min (figure 22). Collision induced dissociation (CID) was carried out on the ion responsible for the peak and its fragmentation pattern analysed. The parent ion, 632.3 m/z, (figure 23) represents a glutathione-clozapine conjugate with a range of its fragment ions being identified in figure 26. This previously discovered adduct is produced via the interaction of a glutathione molecule and a nitrenium ion of clozapine (Wen et al., 2008).

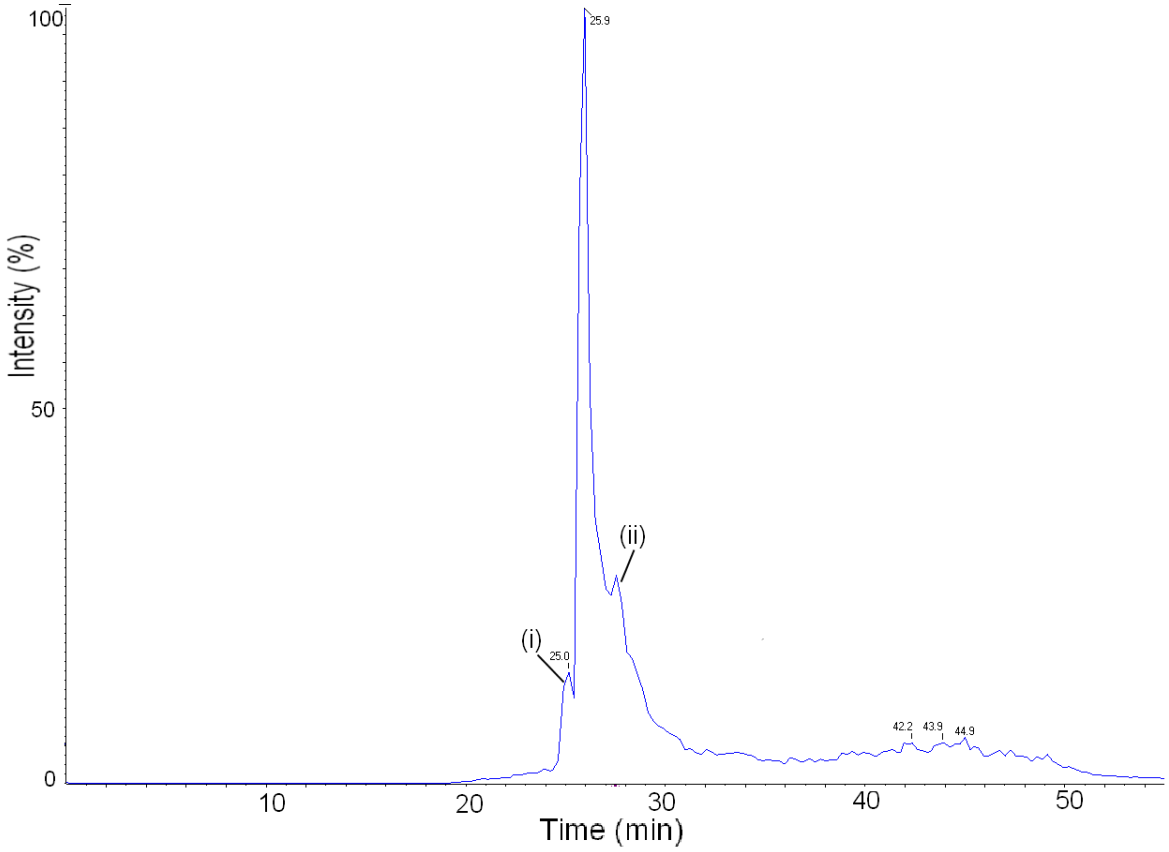


Figure 22. Total ion count chromatogram from a neutral loss of 129 (NL129) scan of clozapine treated GSH. A single intense peak is visible from around 25-30 min. The peak has obvious shouldering (i and ii).

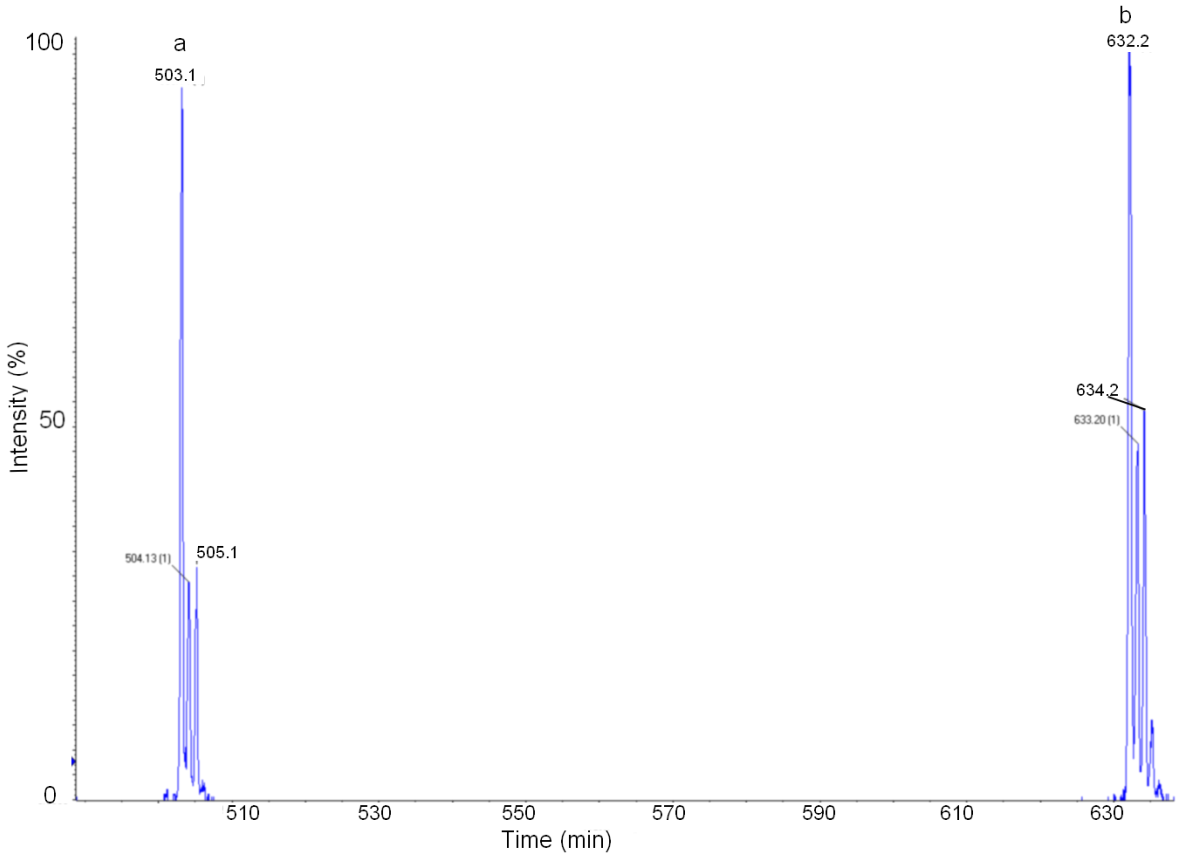


Figure 23. An enhanced resolution scan taken from the centre of the shouldered peak from figure 22. Two ions are visible, a and b. The mass of b is representative of a glutathione clozapine adduct (632.1 Da) as depicted in figure 26. The mass of "a" represents a fragment ion of "b", i.e. a glutathione clozapine conjugate, minus glutamic acid. In both peaks the isotopic distribution unique to chlorine is clearly visible.

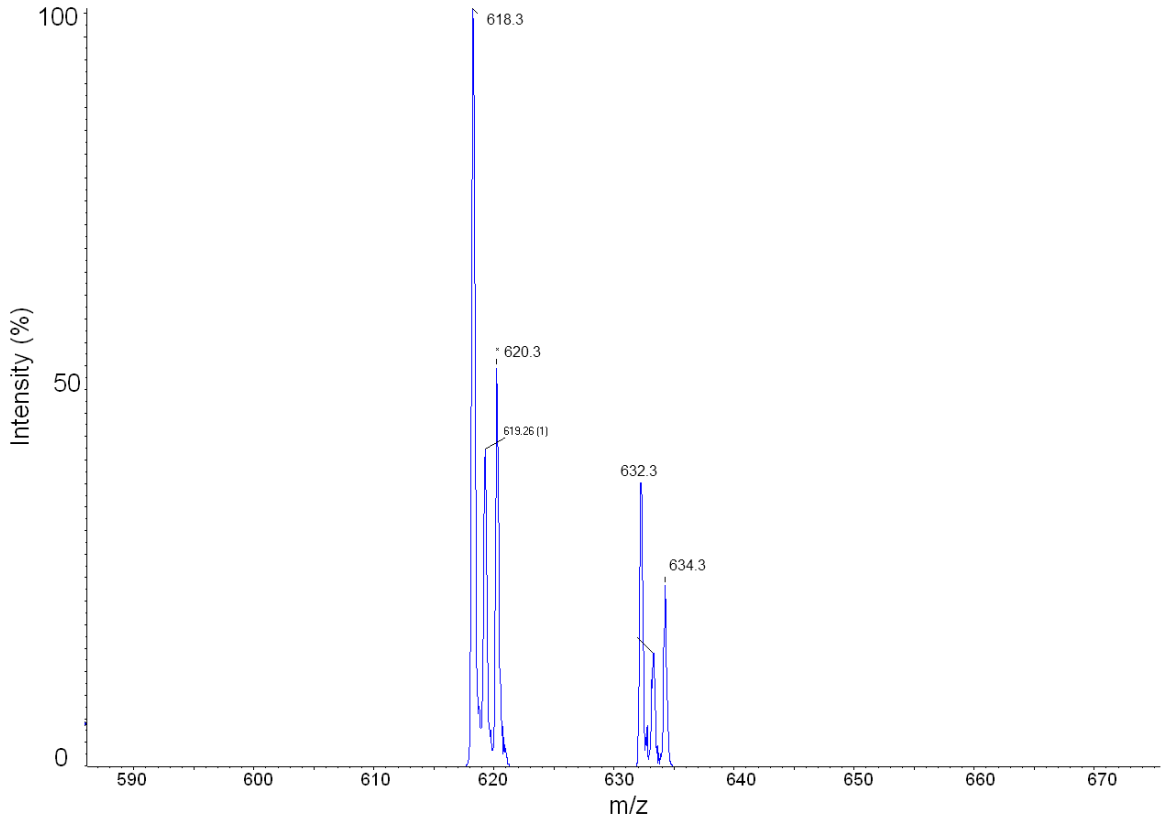


Figure 24. Enhanced resolution scan of shoulder (i) from figure 22. This ion with a mass of 618.3 Da is representative of a desmethylclozapine modification. The isotopic distribution is indicative of chlorine.

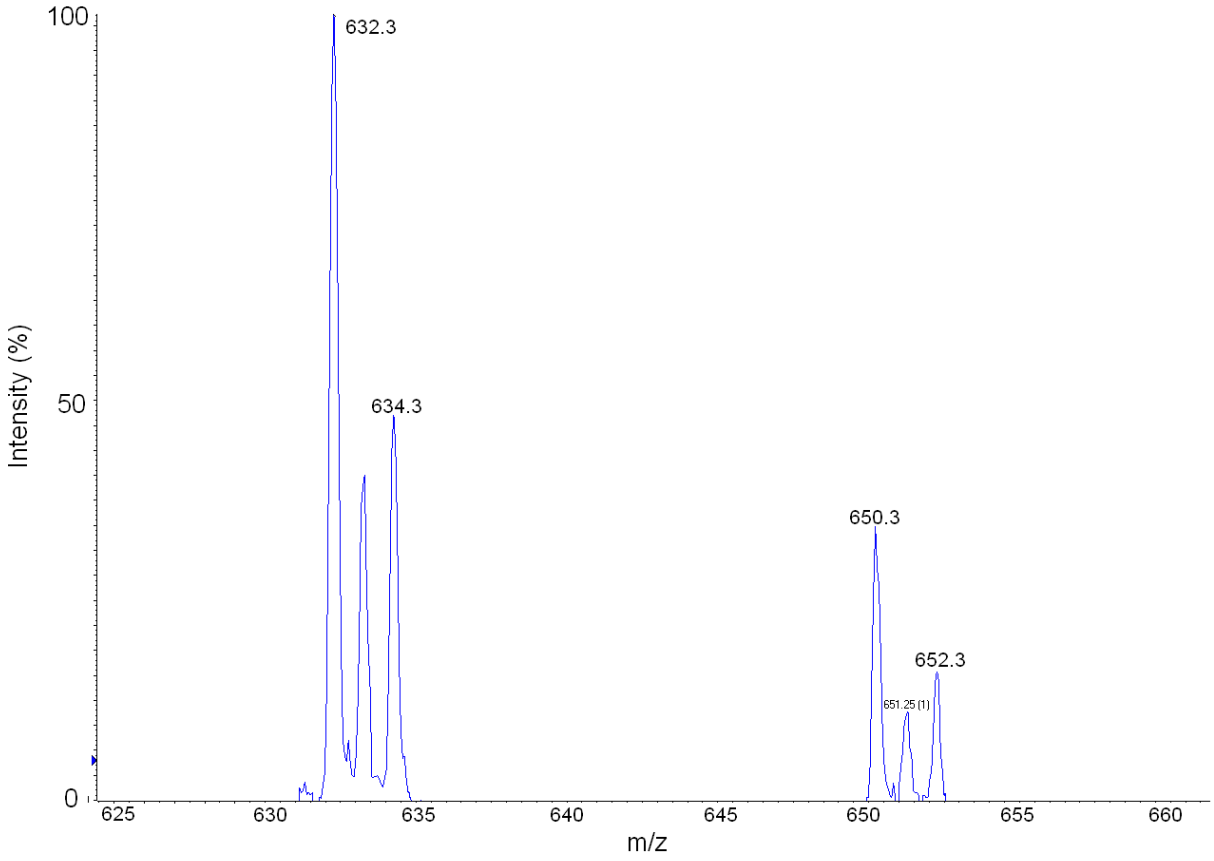


Figure 25. Enhanced resolution scan of the shoulder (ii) from figure 22. The ion at 650 Da likely represents the conjugation of a phase I metabolite of clozapine with glutathione (depicted in figure 28).

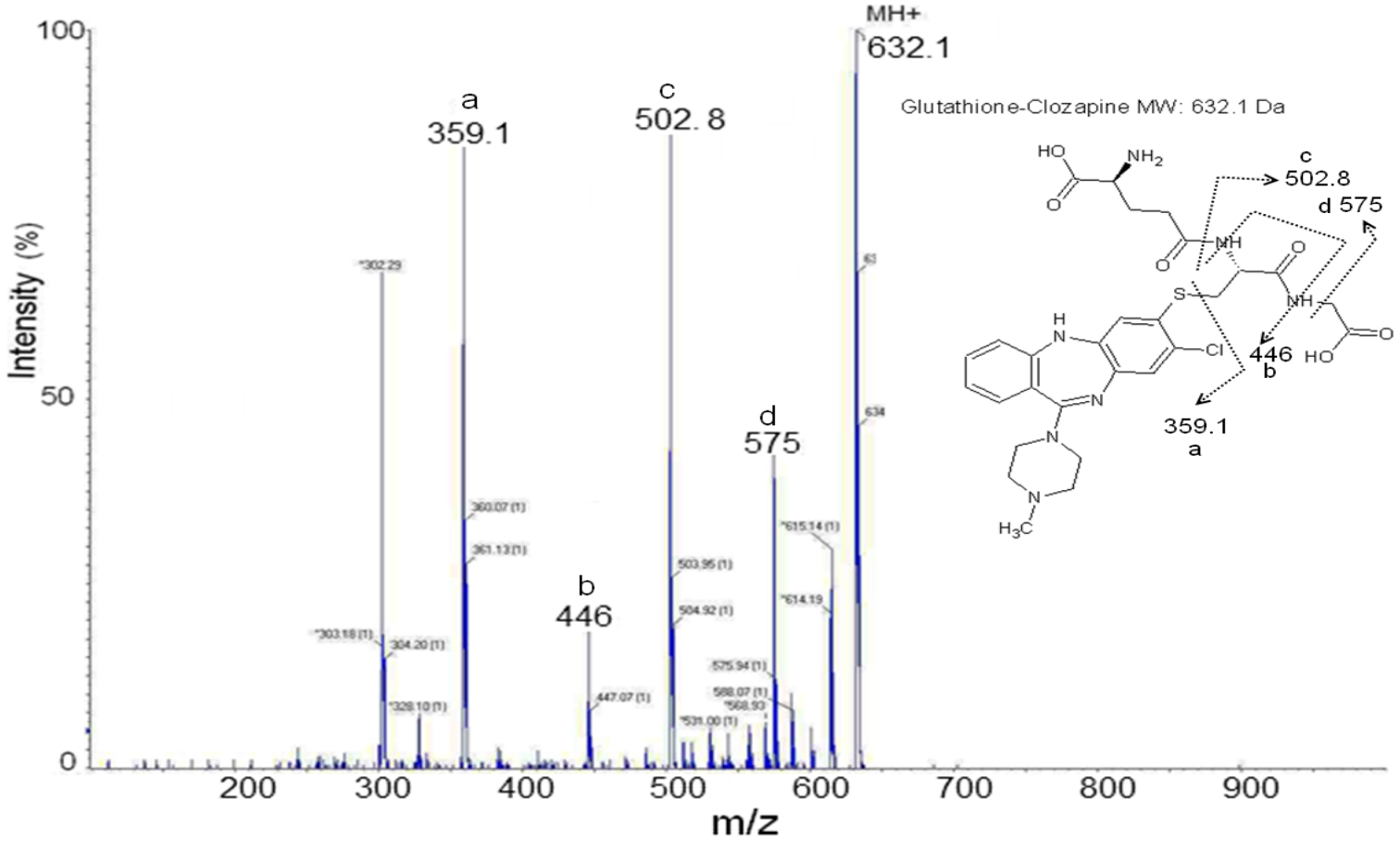


Figure 26. Tandem MS data from the 632.1 Da ion detected by the NL129 scan. The ion present at 359.3 Da is consistent with fragment "a" depicted in the molecular formula graphic.

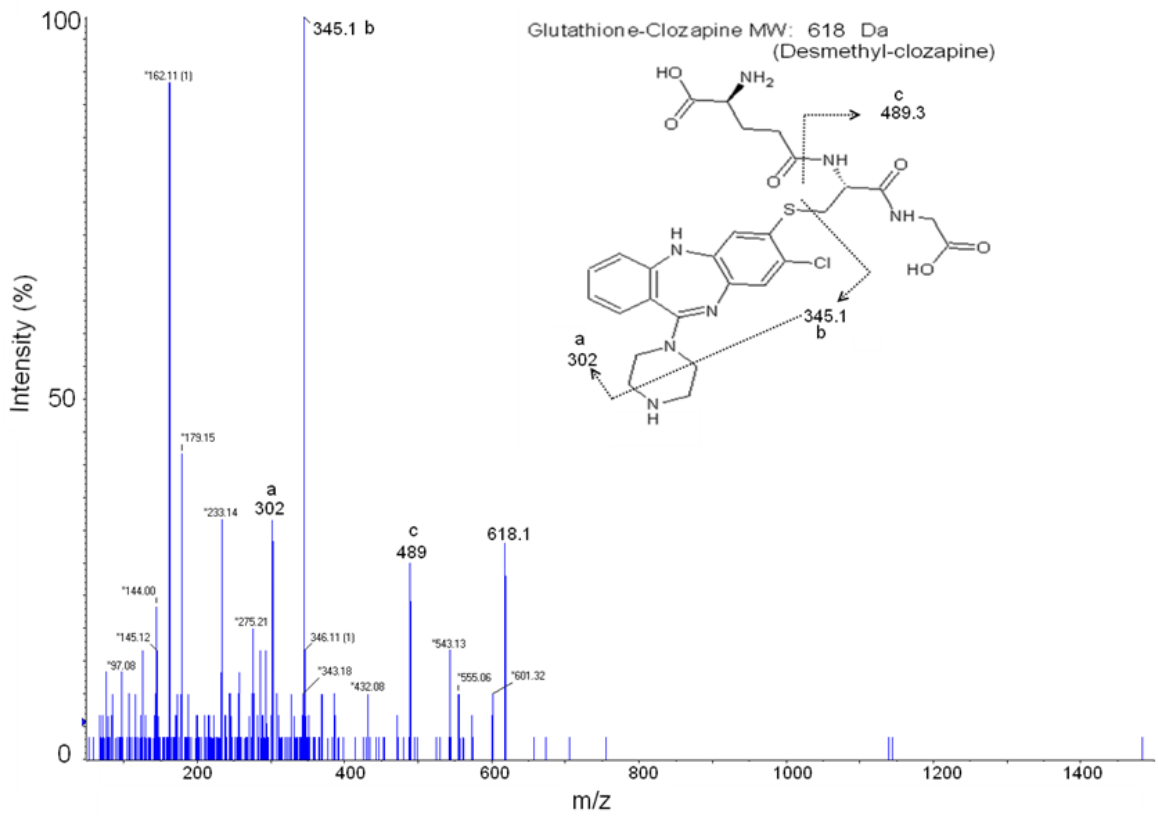


Figure 27. Tandem MS spectrum from the ion at 618 Da. The spectrum is consistent with the loss of a methyl group. Desmethylclozapine is a known metabolite of clozapine, with the methyl group lost from the piperazine ring.

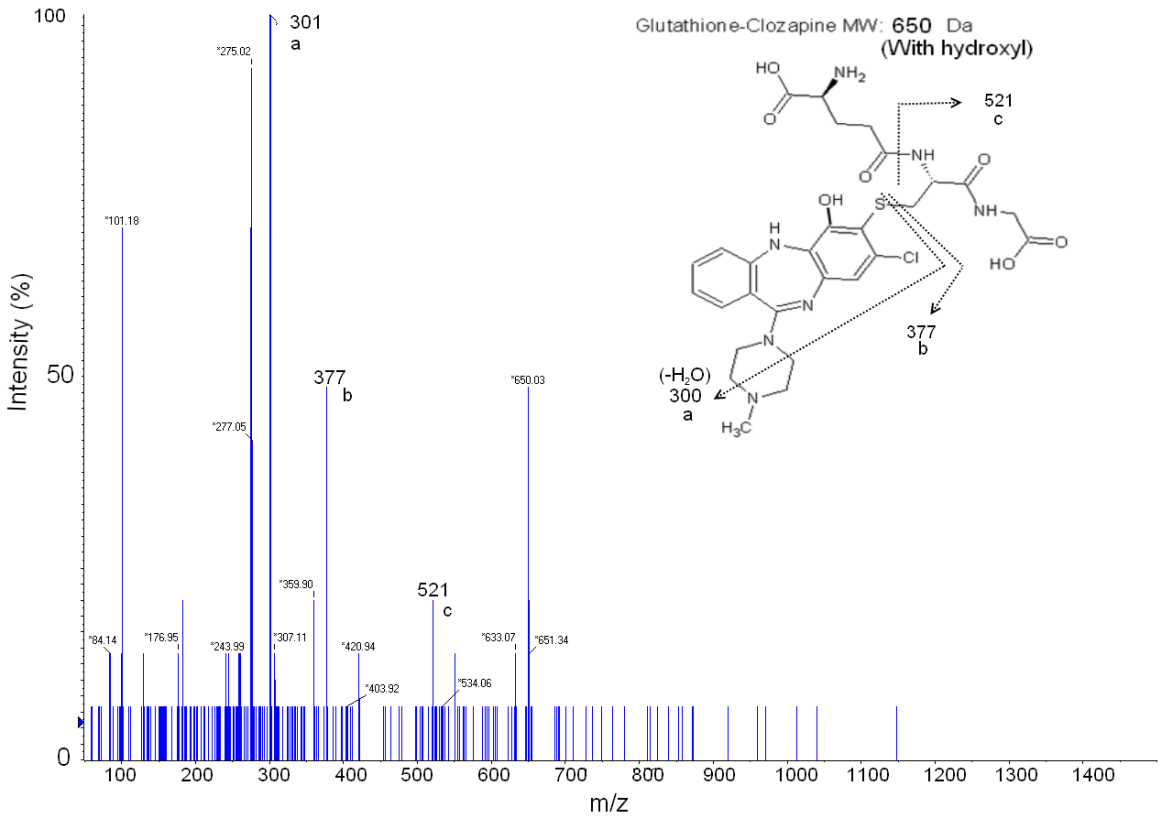


Figure 28. Tandem MS spectrum from the ion at 650 Da. The spectrum is consistent with a conjugate of glutathione and a hydroxylated clozapine metabolite.

The ion of 359 m/z identified in figure 26 is representative of the clozapine metabolite with the additional mass of a sulphur atom. Fragmentation of the clozapine-glutathione conjugate across the side group of cysteine is most likely to have created this ion. Using an instrument with higher resolution it would be possible to determine the exact mass for the fragment, this information could be used to work out the elemental composition and confirm the structure of the ion. In many of the spectra it is possible to see the distinct isotopic distribution associated with the presence of a chlorine (part of clozapine). The ^{35}Cl isotope makes up 75% whereas ^{37}Cl makes up about 25%, this fact makes it easy to spot molecules containing a chlorine from mass spectra as typically the second isotopic peak is greater than the third. This knowledge can be used to further confirm the identity of molecules.

3.4.2 UV Data for Clozapine Glutathione

UV data was gathered on two channels, 214 nm and 280 nm, downstream of chromatographic separation prior to MS analysis. Comparison of the UV chromatograph and the TIC chromatograph reveals a discrepancy in complexity. The UV data comprises several distinct peaks eluting from around 22-32 min (figure 29) in contrast to the single somewhat shouldered peak of the TIC data.

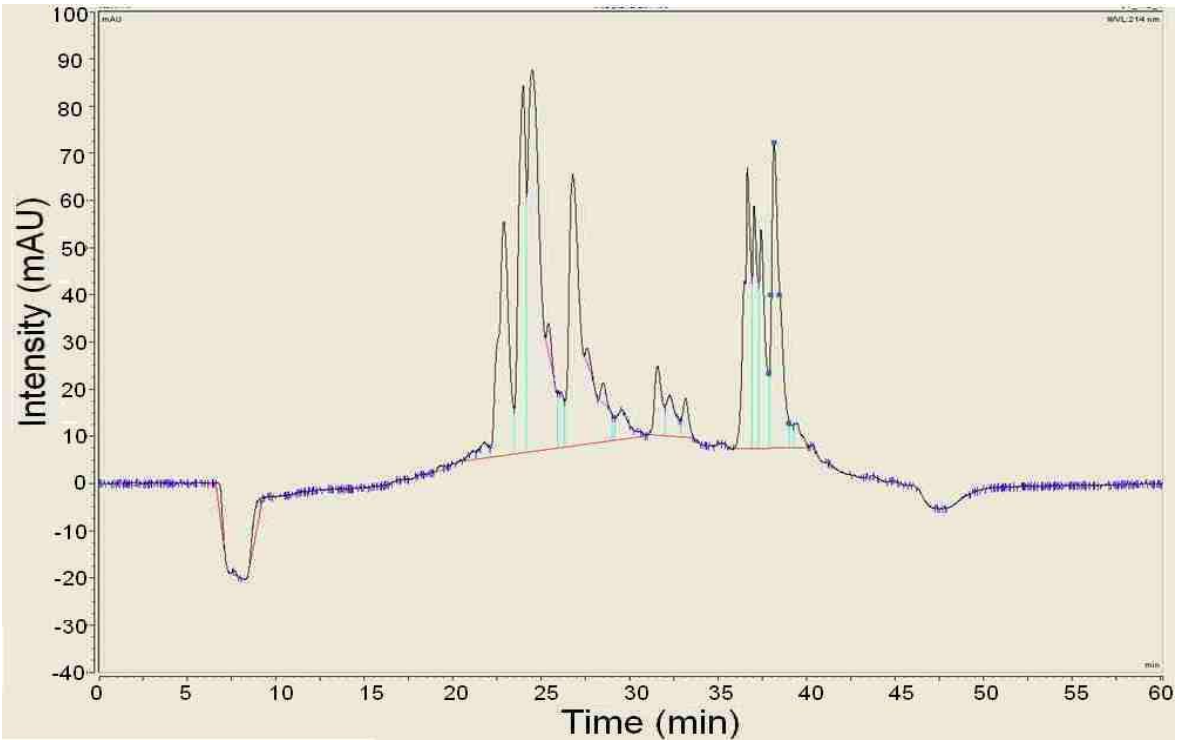


Figure 29. Clozapine GSH UV absorbance at 214 nm. Multiple peaks are easily observable in contrast to the single predominant peak observed in the TIC of the NL129 experiment.

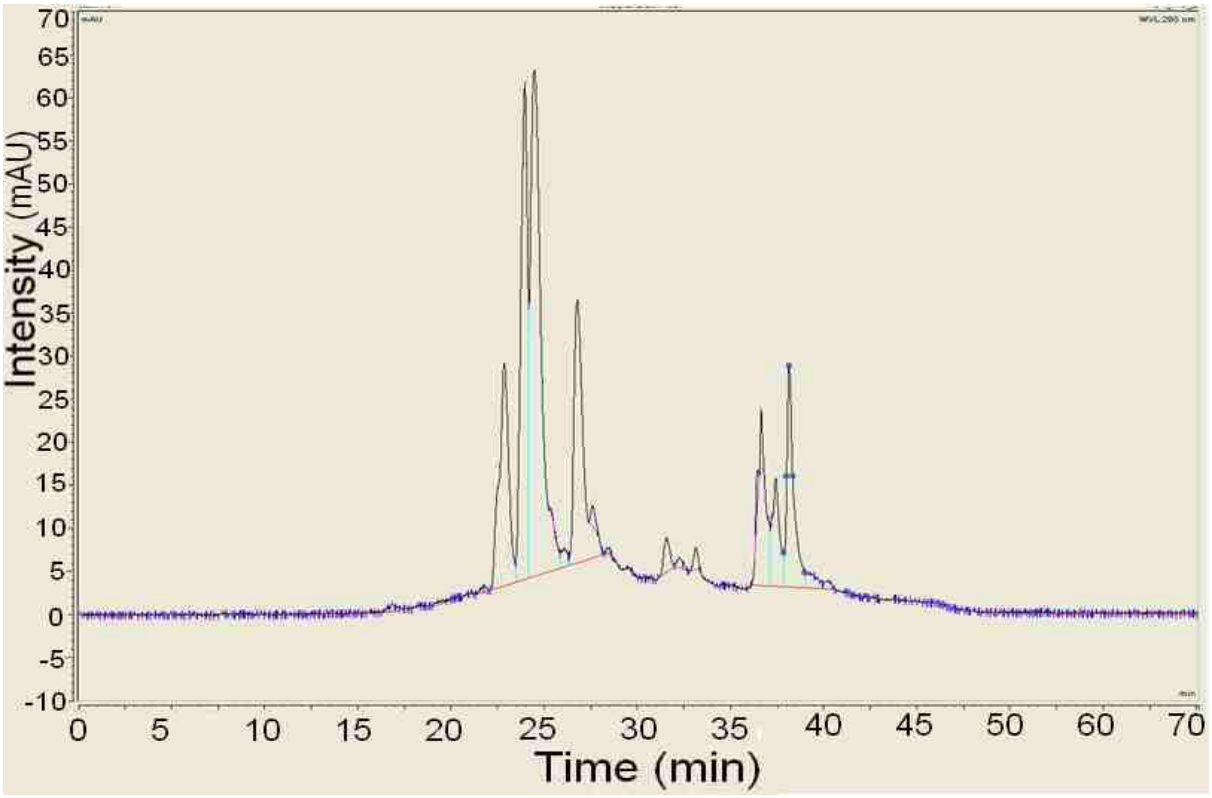


Figure 30: Clozapine GSH 280 nm. As with the spectrum at 214 nm multiple peaks are clearly visible.

Three potential adduct types were identified from the NL129 scanning method. It was noted however, that the UV data looked to have more features than did the TIC from the NL129 data. It is probable that these extra peaks represent unidentified adducts that did not appear in the NL129 data as the conjugates did not produce the neutral loss at 129 Da necessary for detection. Without this neutral loss these species would be absent from the TIC data and no tandem MS data would be gathered.

A review of the literature uncovered an alternative method of glutathione adduct detection involving the use of a precursor ion scan at 272 in the negative ion mode (Dieckhaus et al., 2005). Using this alternative approach several previously undetected adducts were characterised. Work carried out by Dieckhaus et al showed that MS/MS data was superior in the positive ion mode thereby necessitating a polarity switch between precursor scanning and MS/MS experiments.

3.4.3 PI272 Scan (Negative Ion Mode)

ESI instruments typically operate in the positive ion mode as maintaining a stable stream of negative ions requires additional care (Cech and Enke, 2001). Positive ion mode operation is well characterised and stable across typical gradients of organic solvent. In the negative ion mode the formation of corona discharge (Kebarle and Ho, 1997) and poor electrospraying of peptides is problematic (Williamson et al., 2006). It was found that low concentrations of organic solvent, such as those found at the early stages of a RP-LC gradient, were associated with these negative effects. In order to avoid these problems it was necessary to increase the concentration of organic solvent subsequent to elution from the column. The approach taken by Williamson et al was to infuse a solution of 80% v/v propan-2-ol, 20% v/v ddH_2O at a flow rate of 100 nl/min, using a secondary pump, into the post-column (300 nl/min) stream via a microtee connection. The effect was to increase the organic solvent of the electrospray stream without affecting column separation of peptides.

Scanning for a loss of 272 m/z is representative of the glutathione molecule minus the sulfhydral group. The propensity for aromatic thioesters, undergoing CID fragmentation, to cleave at the sulfhydral group of cysteine with the liberation of the xenobiotic and sulfhydral group has been previously noted (Baillie et al., 1993). This suggests interesting possibilities; firstly that rather than using the glutathione backbone as a product ion mass it would be possible to use the mass of the metabolite plus the sulphur. Obviously this would only be useful for downstream applications after which the primary metabolites have been identified. Secondly this same mechanism feasibly applies to the CID fragmentation of metabolites of other drugs. The prerequisites being that the metabolite be highly electrophilic in nature and possess the characteristics of a soft electrophile (tendency to react with soft nucleophiles i.e. SH group).

The mass of the SH group could be added to the mass of the known metabolites and used as selective precursor ion scans for the identification of peptide-metabolite conjugates.

3.4.3.1 PI272 Scan with Clozapine

Using the PI272 scan a more complete range of metabolite-glutathione adducts were detected and characterised (figures 31-38). In addition to those shown a further two adducts were identified and confirmed against independent studies. The masses of all five adducts were calculated in order to determine the mass shift associated with their conjugation to glutathione. These data along with the chemical composition and description of each adduct has been compiled in table 2.

Table 2. Clozapine metabolites.

Modification	Conjugate description	Chemical formula	Modification m/z (monoisotopic)	Retention time (min)
Clozapine	Clozapine+GSH	$C_{18}H_{18}ClN_4$	325.1219	25.5
Clozapine 2	Clozapine+GSH+O	$C_{18}H_{18}ClN_4O$	341.1169	21.3
Clozapine 3	Clozapine+GSH+2O	$C_{18}H_{16}ClN_4O_2$	355.1118	34.2
Clozapine 4	Clozapine+GSH-HCl	$C_{18}H_{17}N_4$	289.1453	Not detected
Clozapine 5	Clozapine+GSH-CH2	$C_{17}H_{16}ClN_4$	311.1142	24.2

It has been shown that by using the PI272 scan it is possible to identify a greater range of metabolites captured in the trapping assay. With knowledge of the range of metabolites it is possible to design more effective strategies for their detection. Moving away from using the glutathione as means of identification to using the actual metabolite means that not only glutathione conjugates can be discovered.

Modified peptides can then be detected based on analysis of their fragmentation patterns and identification of known metabolite fragment ions. From the work it was possible to identify a potential precursor ion, at 359 m/z, which could possibly be used to detected modified peptides.

In total five distinct metabolite adducts were identified and their associated masses added to the Mascot search engine to be available in the list of potential modifications.

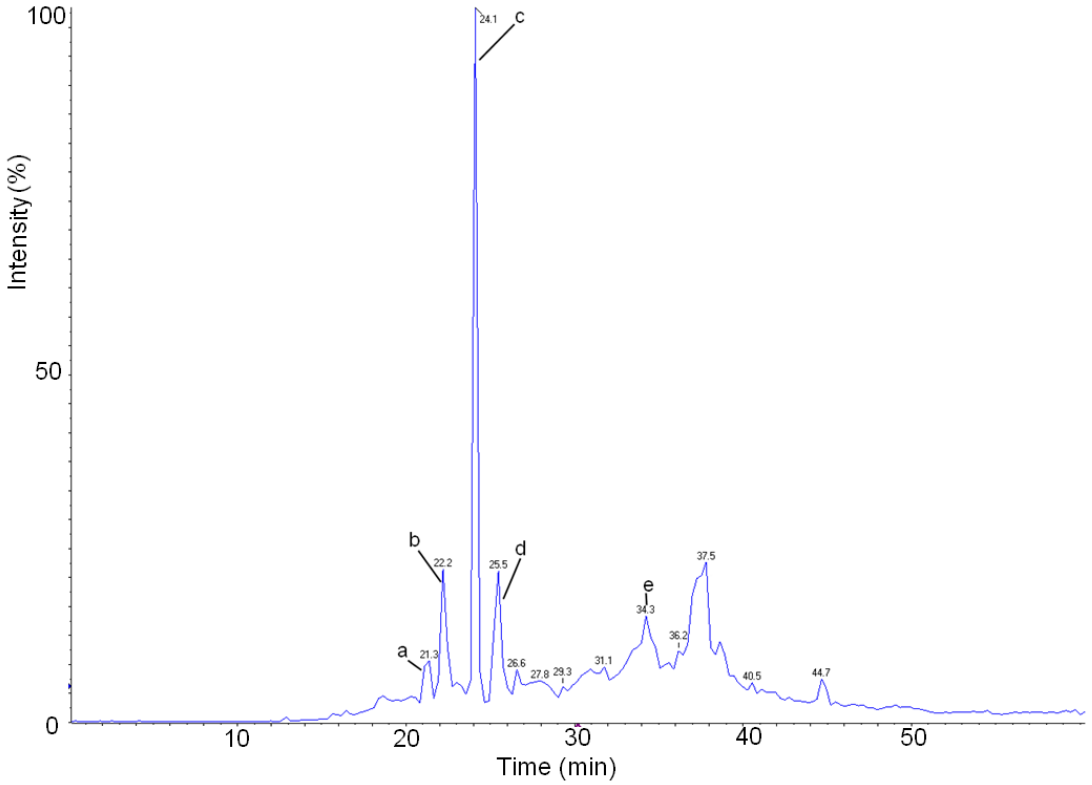


Figure 31. Precursor ion TIC spectrum from the PI272 experiment with clozapine-glutathione. All of these peaks present have been previously identified as clozapine metabolites conjugated to glutathione. (Wen et al., 2008). The spectrum matches closely with that obtained in the UV analysis with the major peaks b, c, d, and e appearing in the same locations.

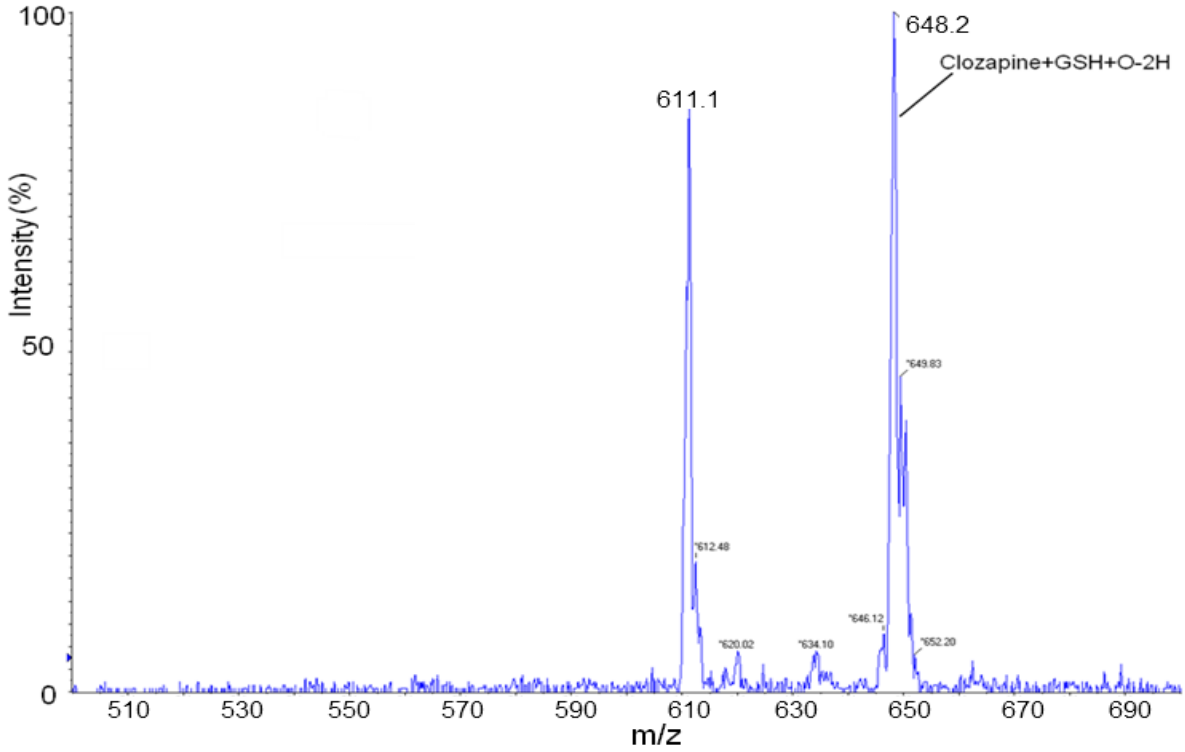


Figure 32. Precursor ion 272 scan for peak "a" from figure 31. The ion at 648.2 Da has previously been identified (Wen et al., 2008). The ion at 611 Da appears to be some kind of contaminant.

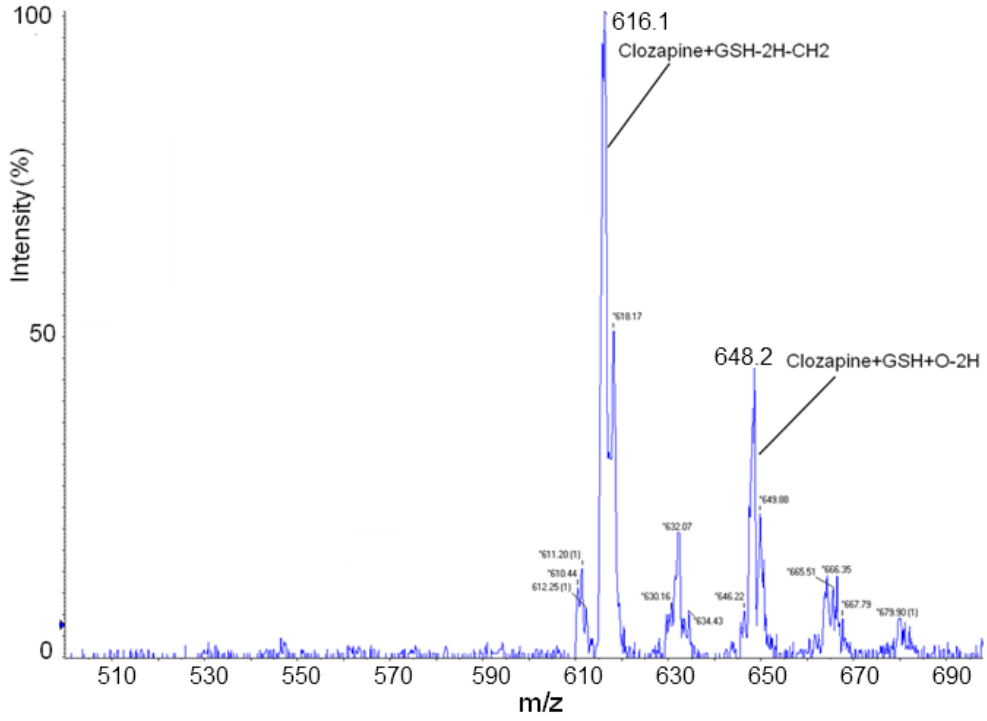


Figure 33. Precursor ion 272 scan for peak "c" from figure 3. The ion at 616.2 Da has previously been identified (Wen et al., 2008).

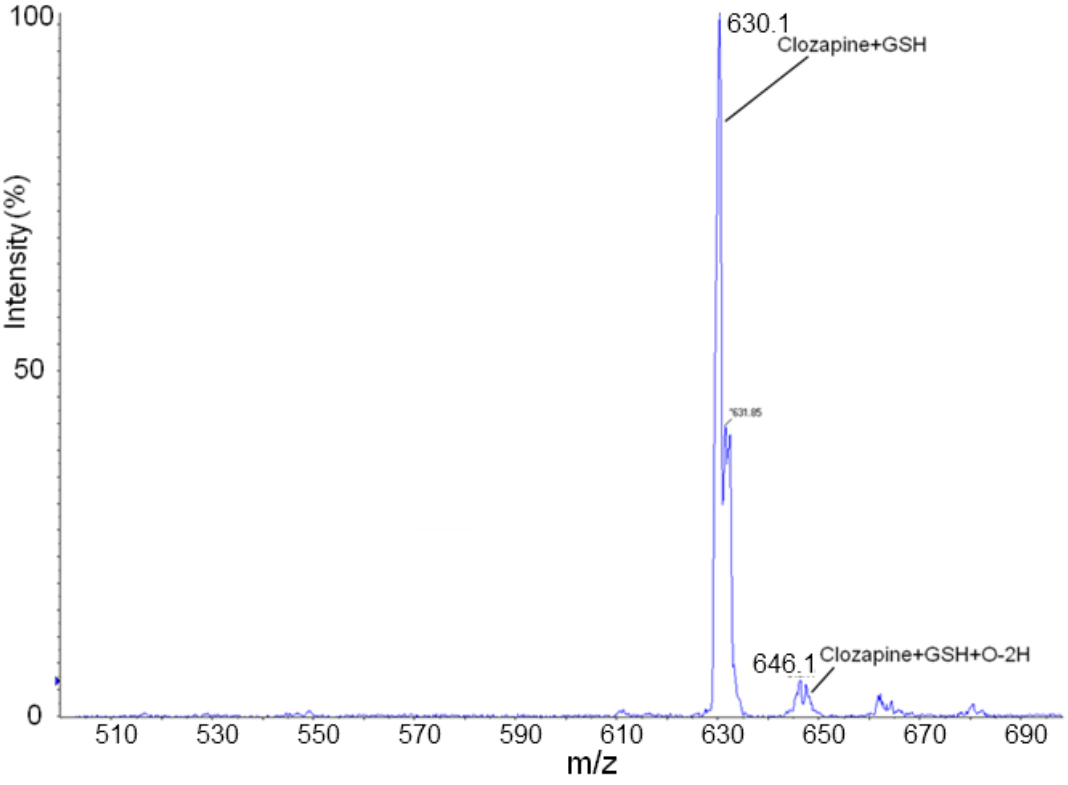


Figure 34. Precursor ion 272 scan for peak "d" from figure 3. The ion at 630.2 Da has previously been identified (Wen et al., 2008).

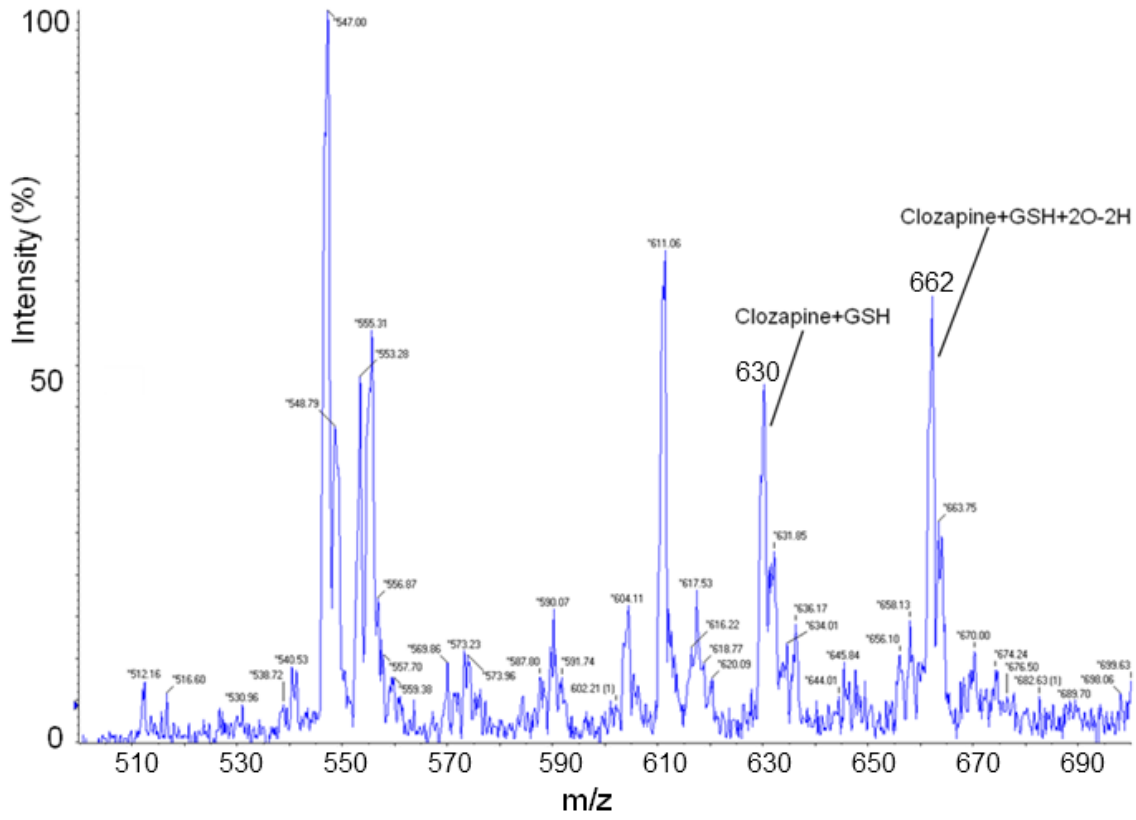


Figure 35. Precursor ion 272 scan for peak "e" from figure 3. The ion at 662.2 Da has previously been identified (Wen et al., 2008).

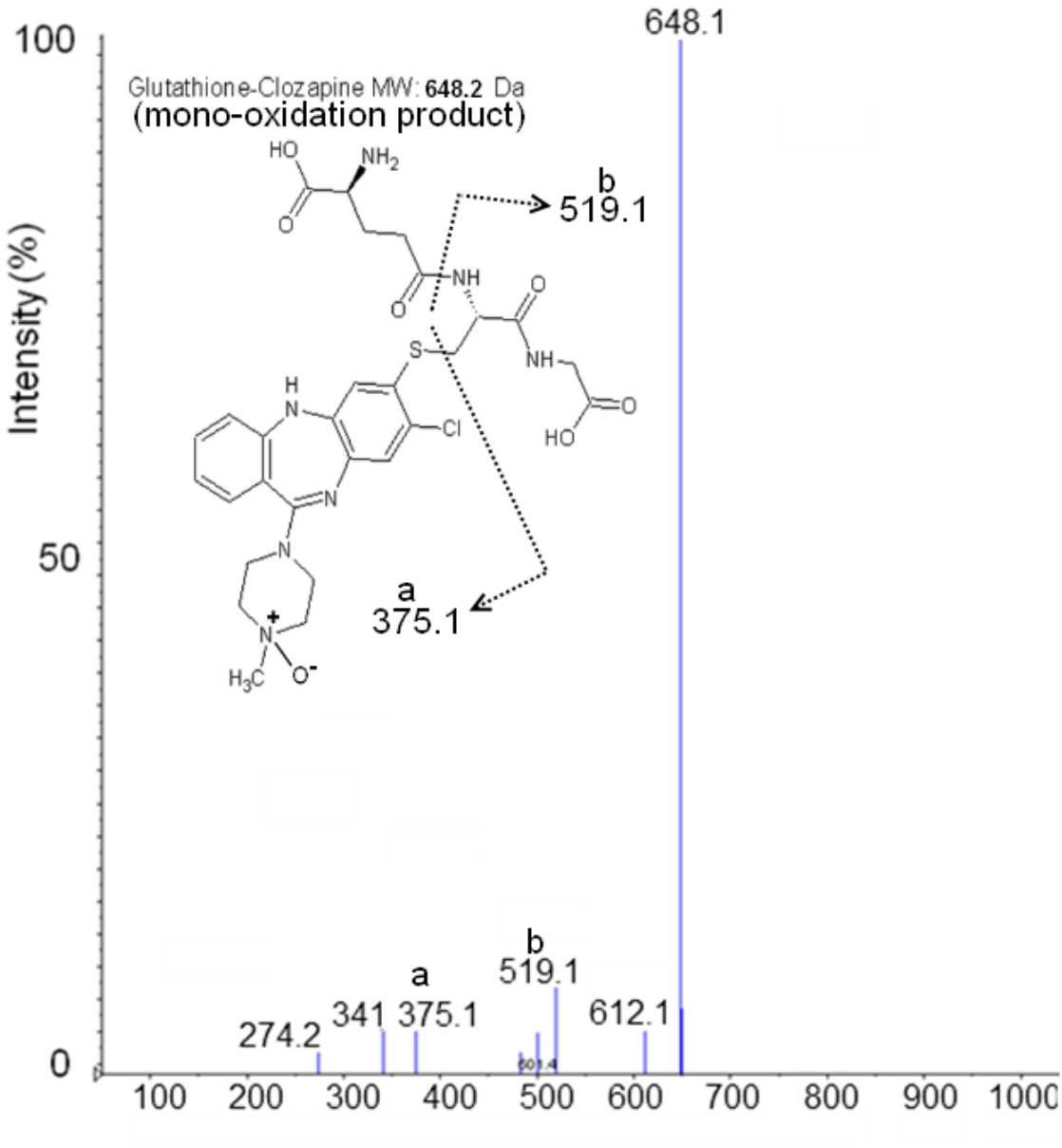


Figure 36. EPI scan of ion at 648 m/z from clozapine treated glutathione sample. The peak at 648.1 m/z is consistent with a mono-oxidation product (Parent ion – 2H + O) of clozapine. The parent ion and the ions at 375.1 and 519.1 match up with the data in the paper by Wen et al (Wen et al., 2008).

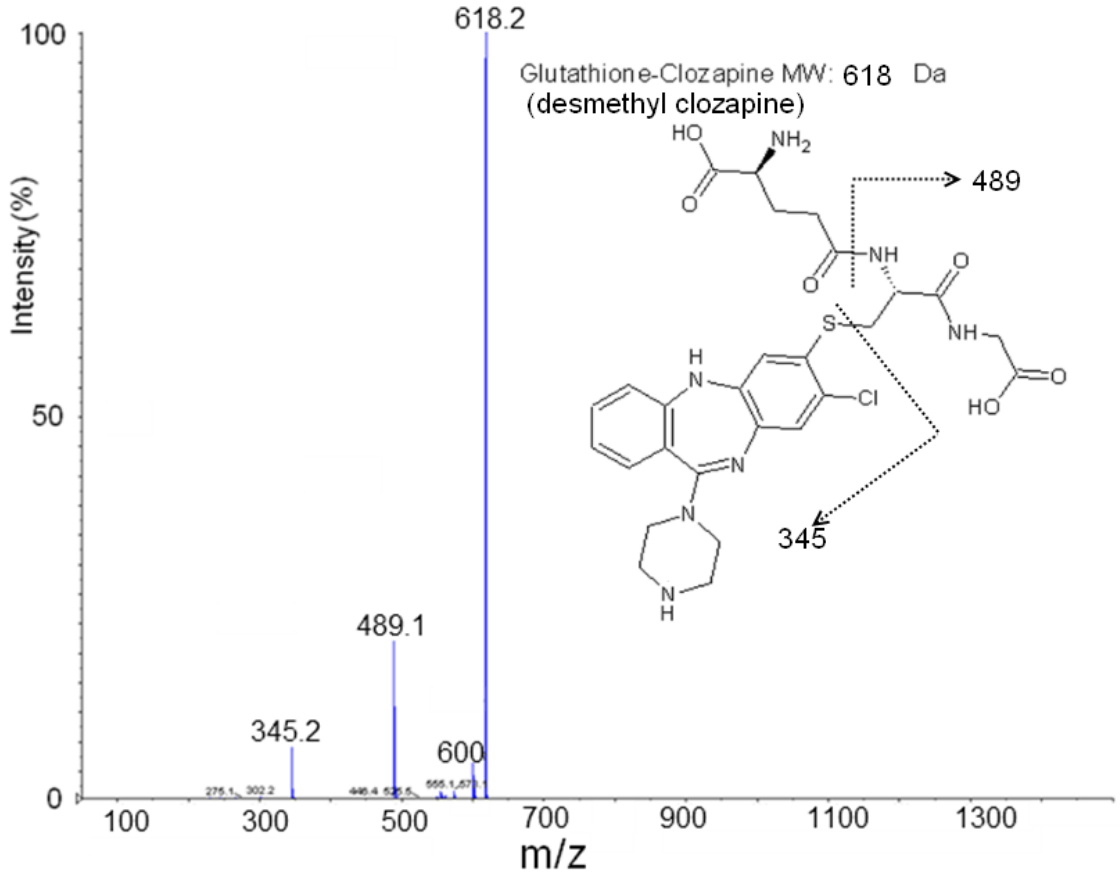


Figure 37. EPI of ion at 618 m/z consistent with a clozapine metabolite conjugate with glutathione. This mass indicates that the metabolite is the desmethyl form of clozapine. Additionally the ion at 345.2 represents the mass of the metabolite plus the SH group procured from a cysteine. The ions detected here fit well with those detected by Wen et al (Wen et al., 2008).

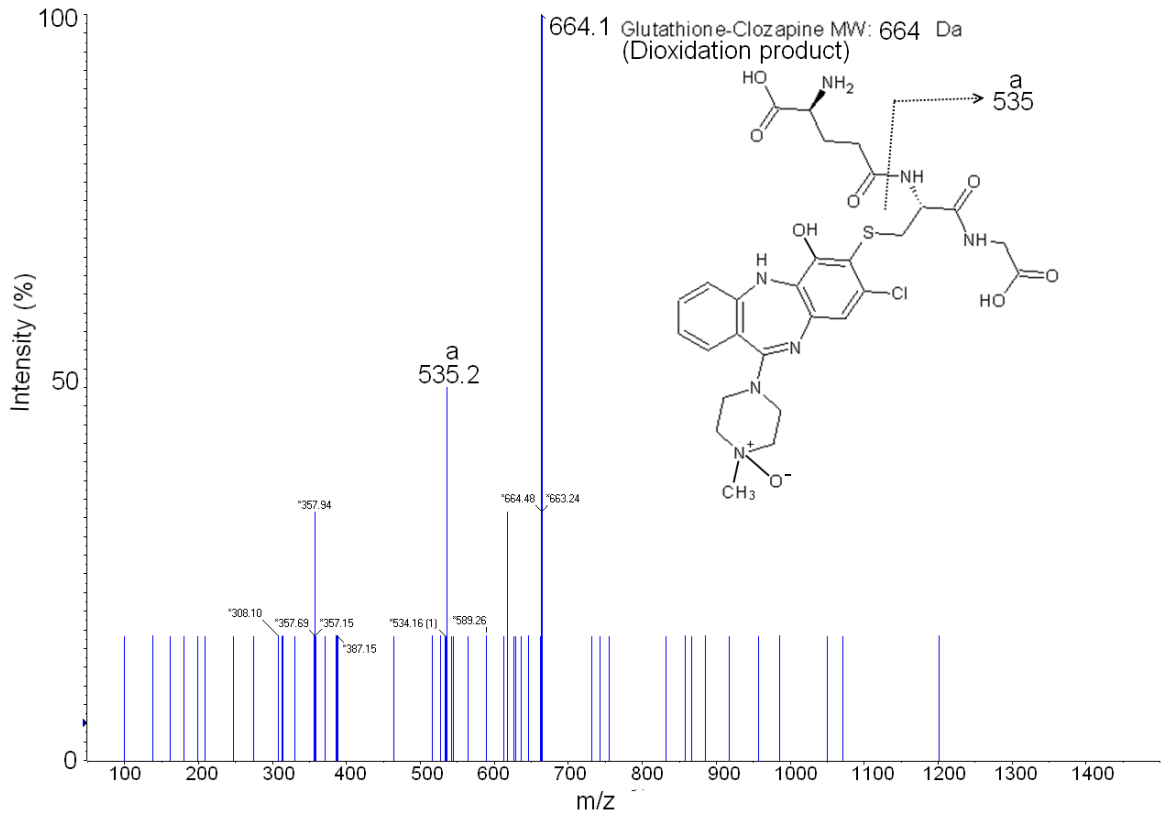


Figure 38. EPI spectrum of the ion at 664 Da (662 Da negative ion). 'the poor signal is evident yet the ion at 535.2 is clear to see and correlates with what was observed by Wen et al.

The MS/MS data gathered is of a very low signal due to an error in the experimental design. The MS/MS data was gathered for the peaks detected in the negative ion mode, meaning that the actual mass was off by 2 Da (the difference between [M-H]- and [M+H]+). The ion transmission window used for the precursor scan was wide enough that some ions were subjected to MS/MS but this number was extremely limited. This means that a minimum of fragments ions were observed. It is noted that the distinctive chlorine isotopic distribution cannot be determined from the tandem MS data. This is because the enhanced product ion scan (EPI) responsible for obtaining the tandem MS data has an narrow isolation window (0.1 m/z) this means that only a single isotopic peak is fragmented; this results in the production of a spectrum in which m/z values do not carry isotopic information beyond that of the product ion.

3.4.3.2 Negative Ion Mode Scanning of Other Drugs

The negative precursor of 272 scan was applied to several other drugs. The drugs were metabolised in the human liver microsome assay with glutathione trapping as previously described (2.1.1.19). Samples were then analysed by RP-LCMS as previously described (2.1.2.2 and 2.1.2.6). The resulting data was then analysed manually with potential metabolite glutathione conjugates identified and their MS/MS spectra examined.

3.4.3.2.1 Imipramine (3-(10,11-dihydro-5*H*-dibenzo[*b,f*]azepin-5-yl)- *N*,*N*-dimethylpropan-1-amine)

Imipramine is a tricyclic antidepressant used in the treatment of depression and of enuresis. Phase I metabolism of imipramine is carried out by various CYP450 enzymes including 2D6, 3A4, 2B6 and 3A7 (Koyama et al.,. 1997; Preissner et al.,. 2010). Imipramine is thought to form a reactive arene-oxide through its hydroxylation. The arene oxide is highly reactive and can go on to form a protein adduct, in some cases causing inhibition of CYP450 (Masubuchi et al., 1996).

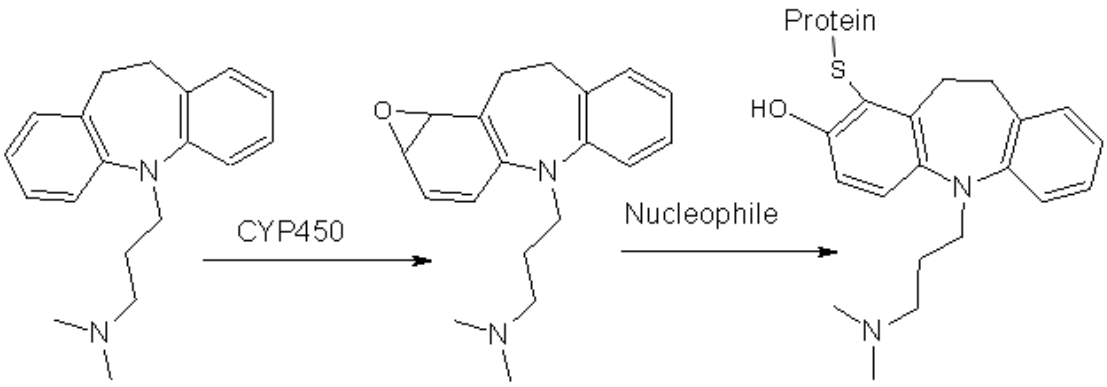


Figure 39. Metabolism of imipramine to hydroxyimipramine via a highly reactive arene oxide intermediate. The product is the most likely isomer with the opening of the arene oxide directed by the presence of the nitrogen. However, an isomer in which the location of the protein and the OH group are swapped is also possible.

Three distinct metabolite-glutathione adducts were detected and characterised including a hydroxy imipramine adduct (figure 41) and a didesmethyl hydroxyimipramine adduct (figure 42).

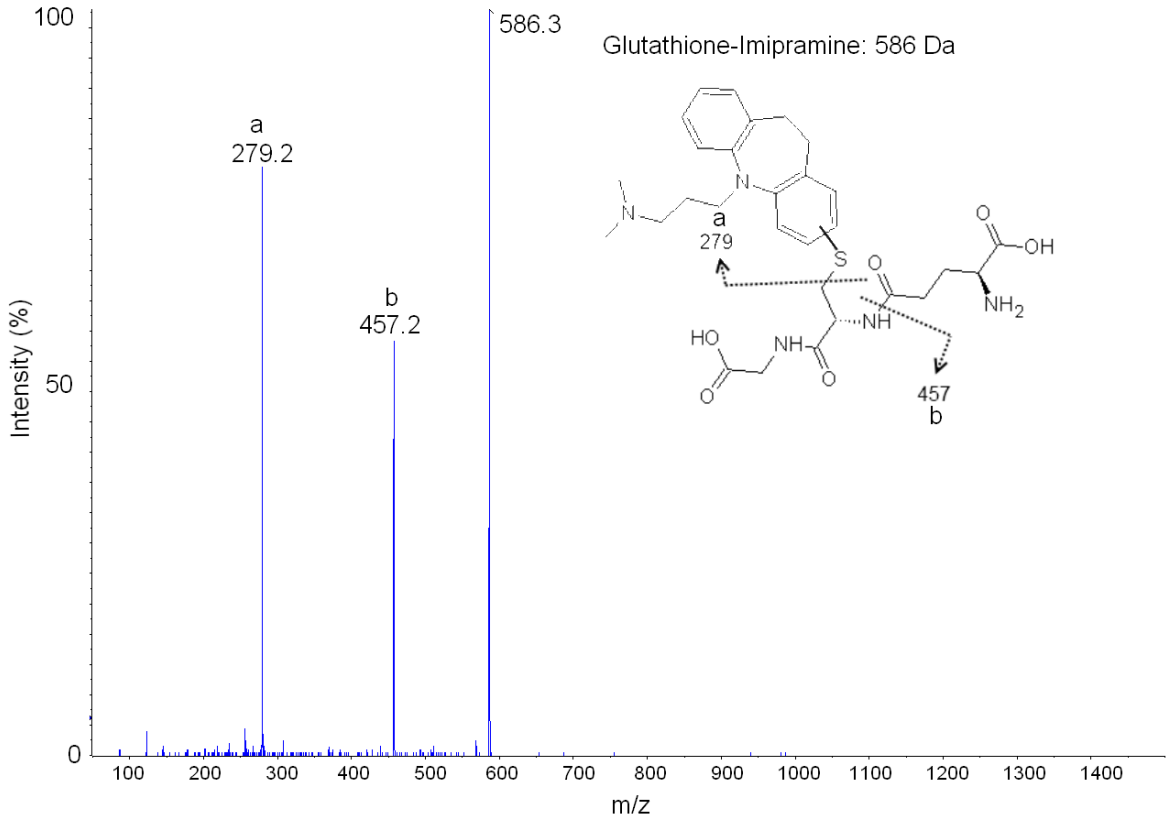


Figure 40. Enhanced product ion spectrum of m/z 586.2. The spectrum represents an imipramine metabolite-glutathione conjugate. Inset is the chemical structure of the molecule along with 2 characteristic fragment ions present in the spectra. All fragment ions were also detected by Wen et al.

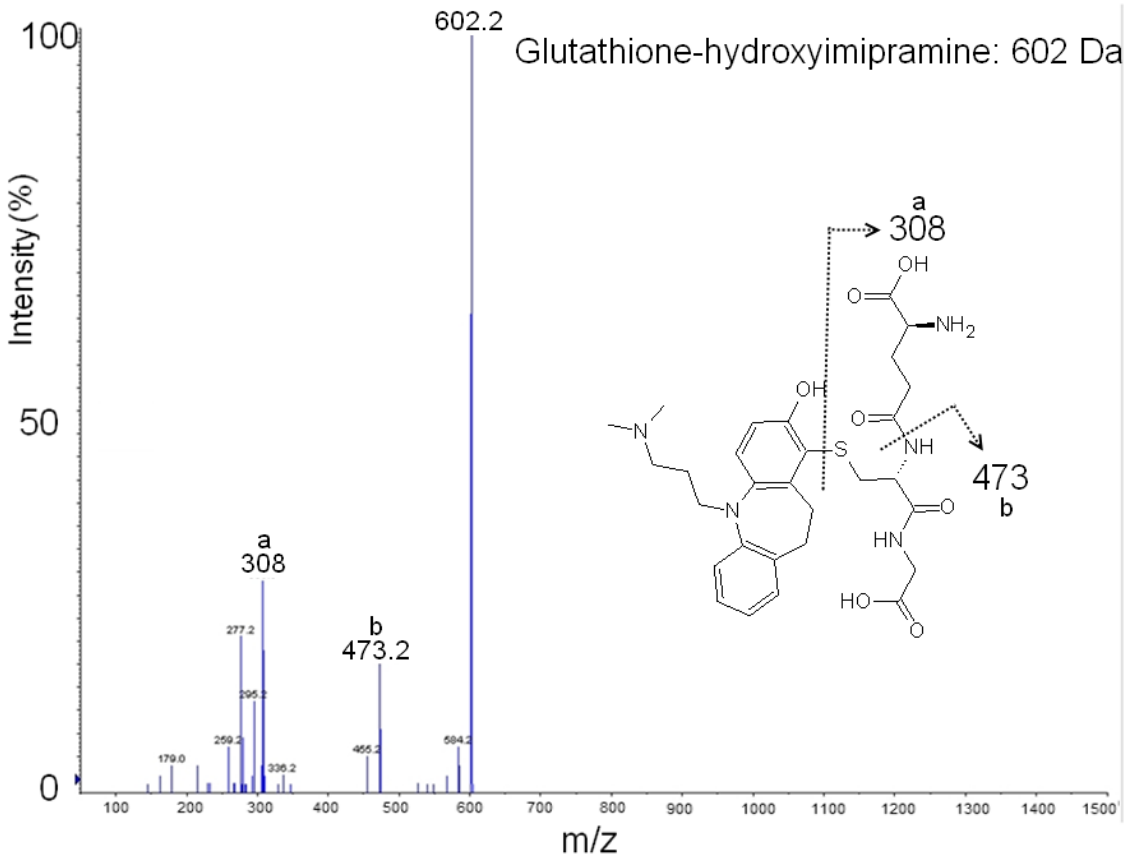


Figure 41. Hydroxy-imipramine-glutathione conjugate with m/z 602. The graphic depicts 2 characteristic fragments visible in the mass spectrum. These and an ion at 329 m/z (characteristic of the metabolite with the sulphur from glutathione) were detected by Wen et al.

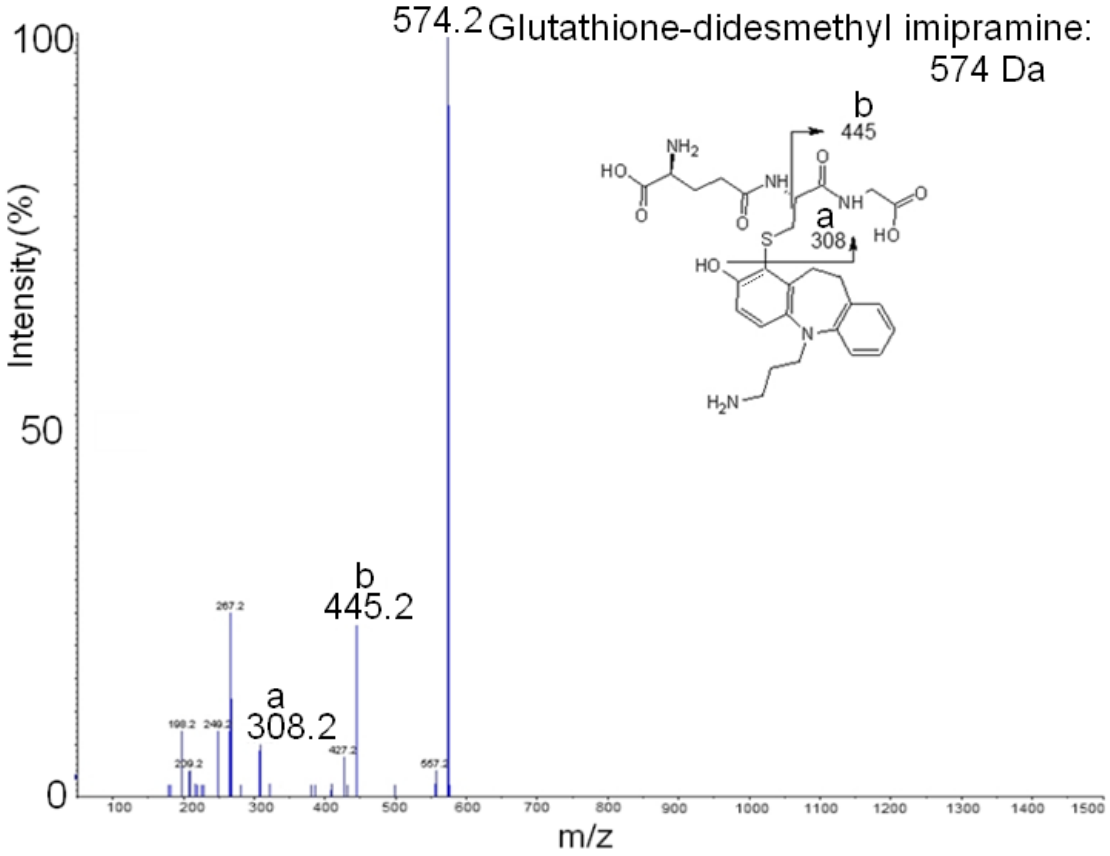


Figure 42. Didesmethyl hydroxyimipramine-glutathione conjugate with m/z 574.2. Characteristic ions from the mass spectrum are depicted in inset graphic. These ions as well as one at 301 m/z were detected by Wen et al.

3.4.3.2.2 Naproxen (Propanoic Acid)

Naproxen is a non steroidal anti inflammatory drug (NSAID) that provides mild pain relief as well as a reduction of inflammation. It has been established that carboxylic acid drugs form acyl-Coenzyme A and acyl-glucuronide thioesters on metabolic activation (Olsen et al., 2002). These thioesters can go on to react with the nucleophilic groups of proteins; the acyl-coenzyme A molecules being some 70 times more reactive than the acyl-glucuronides (Olsen et al., 2002).

The adduct O-desmethyl naproxen can be seen in figure 43. Two more potential conjugates with parent masses of 602 and 618 m/z units were observed; it was not possible to assign these adducts as the masses were too great to be

explained by any typical phase I reaction. All of these fragments are consistent with glutathione conjugate molecules.

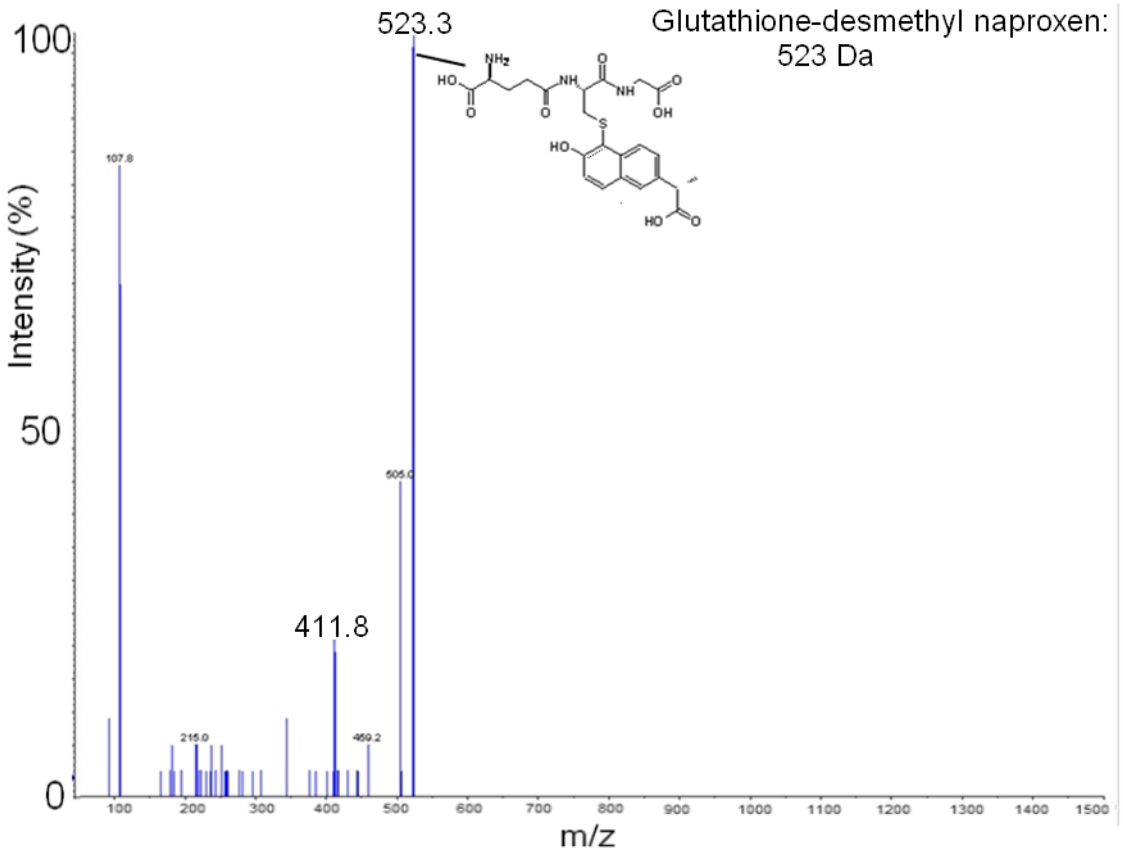


Figure 43. Desmethyl naproxen-glutathione conjugate shown at 523.2 m/z. The signal was poor suggesting a low abundance ion. The ion at 411.8 m/z remains unassigned.

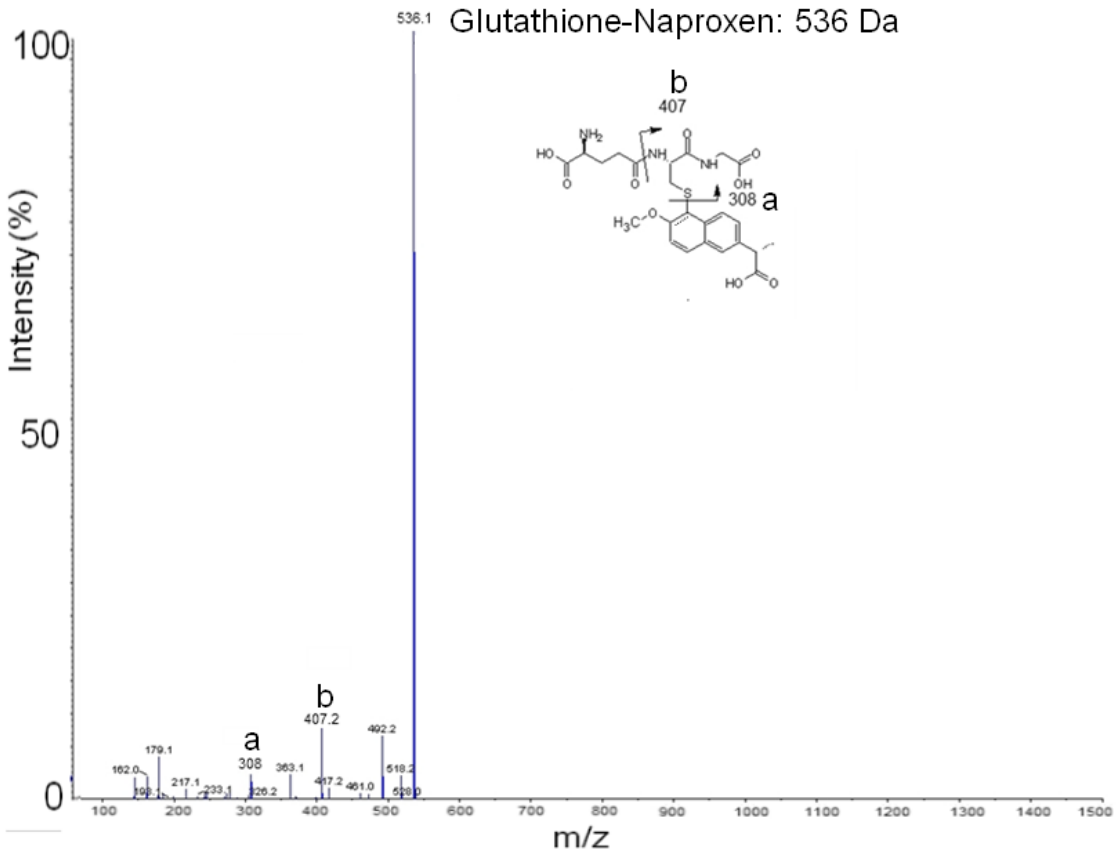


Figure 44. Naproxen glutathione conjugate 536 m/z EPI. Inset graphic shows 2 characteristic fragment ions from the mass spectrum.

3.4.3.2.3 PI272 Tacrine (1,2,3,4-tetrahydroacridin-9-amine)

Tacrine is an anticholinesterase used in the treatment of Alzheimer's disease; due to adverse drug reactions possibly exacerbated by the high doses required for effectiveness the drug is no longer commonly used (Qizilbash et al., 1998). The formation of reactive metabolites has been shown to follow a two step process involving a 7 hydroxylation followed by a two electron oxidation resulting in a quinone methide (Madden et al., 1995; Park et al., 1994) mediated by CYP450 1A2 (Obach and Reed-Hagen, 2002).

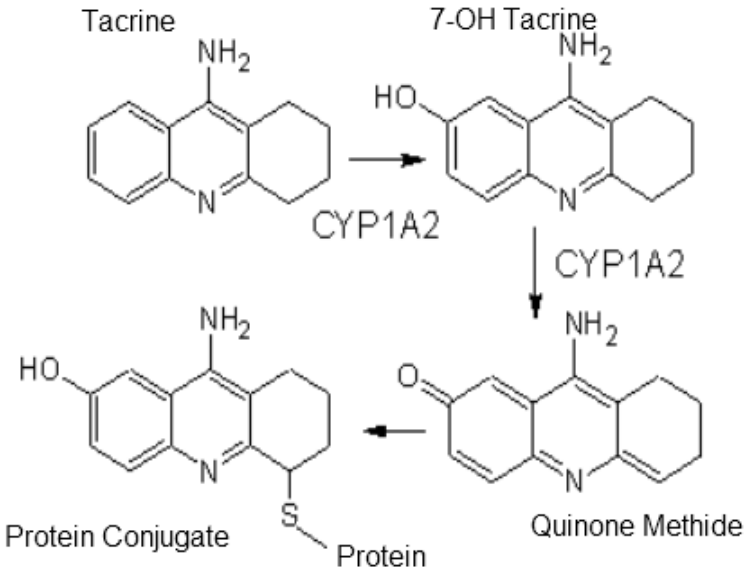


Figure 45. Formation of tacrine-protein conjugates by way of the reactive quinone methide intermediate.

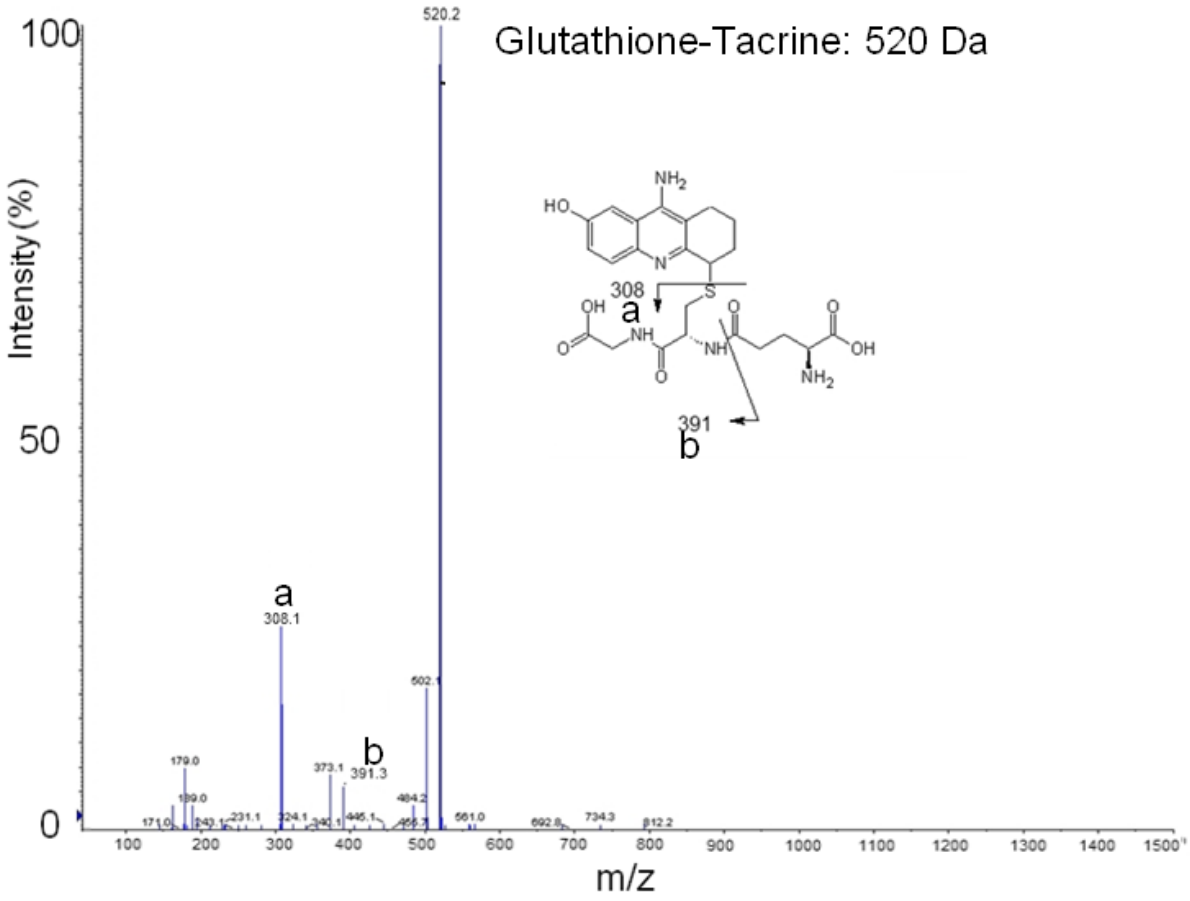


Figure 46. Tacrine glutathione adduct with m/z 520.2. 2 characteristic ions from the mass spectrum are identified in the inset graphic.

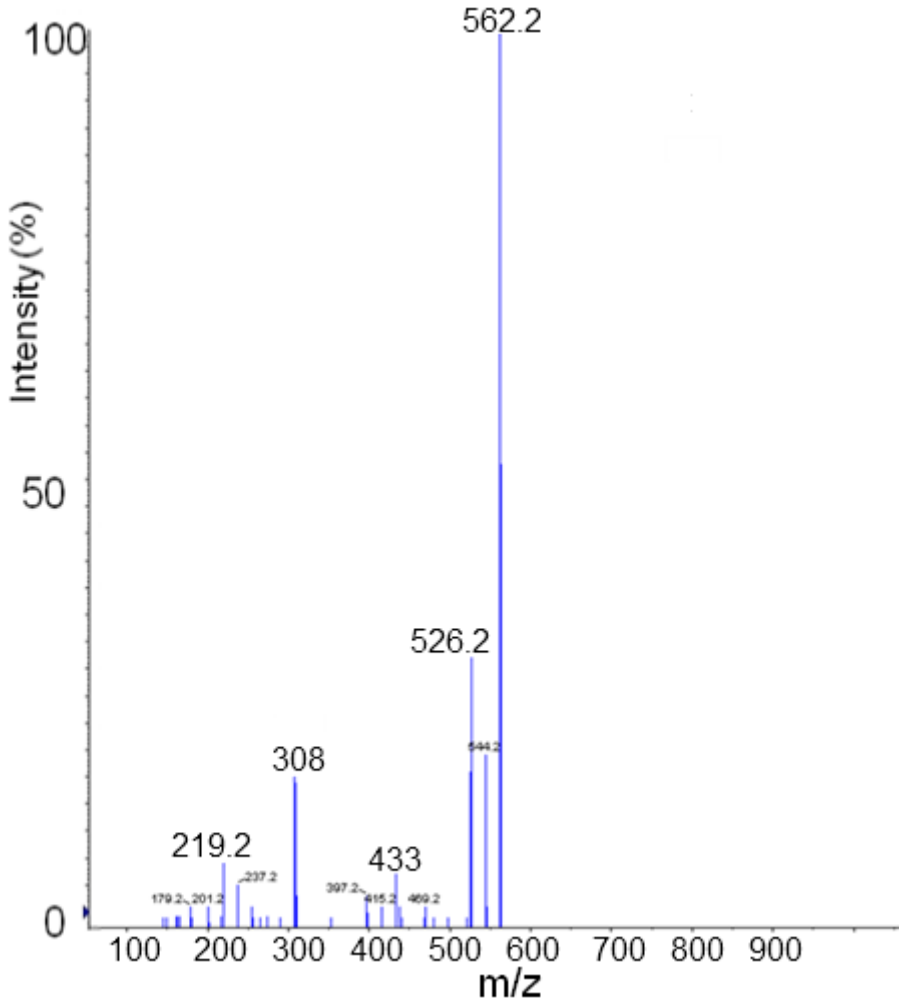


Figure 47. EPI scan of tacrine glutathione conjugate with m/z 562.2. It was not possible to assign a specific adduct to this spectrum. The ion at 308 m/z does however indicate the presence of glutathione, meaning that this is indeed some kind of conjugate. The ion at 433 further supports this as it represents a loss of 129 from the parent ion (562-433=129) a common fragmentation route of glutathione.

3.4.3.2.4 PI272 Summary

Using the PI272 method it was possible to identify metabolite-glutathione adducts for all of the drugs tested. At least two metabolites were identified for each of the drugs and their characteristic CID fragmentation patterns were analysed. It was possible to identify ions in the tandem mass spectrometric data

relating to the fragmentation of metabolite glutathione adducts. An ion at 308 Daltons, the mass of reduced glutathione, was identified in almost all samples.

The data suggests that once the metabolite mass is known it can be used to create selective neutral loss scans for metabolite adducts. Interestingly, the CID fragmentation pattern of glutathione-hydroxyimipramine yielded an ion comparable to the one at 359 Daltons i.e. a drug metabolite plus the added mass of a sulphur acquired from glutathione. The precursor ion at 272 Da used for the triggering of MS/MS experiments, was not observed in any of the samples. This is explained by the fact that the fragment at 272 Da is in fact an anion whereas the MS/MS scan was performed in positive ion mode.

On closer inspection of the results and comparison with molecular formulae it was noted that the adducts discovered were all mass shifted by 2 Da. This is consistent with the mass shift from a negative ion ([M-H]-)to a positive ion([M+H]+) and is due to the selection for this negative mass in the precursor scan. The subsequent EPI experiments targeting these masses resulted in the relatively low signals observed. It is highly probable that the data gathered is not optimal. Indeed, work by Wen et al., showed that the CID fragmentation patterns of many drugs including APAP, imipramine and meclofenamic acid actually do yield an equivalent ion to the one at 359 m/z seen in clozapine. In order to correct for this it would simply be necessary to correct for the 2 Da mass shift when performing the MS/MS experiments.

Additionally, the instrument used in these experiments was an API 2000™ Q-trap a now fairly outdated machine. When performing at optimum capacity the instrument is still at least an order of magnitude less sensitive than the newer API 4000™ and two orders of magnitude less sensitive than the latest API 5500™ model.

3.4.4 Characterisation of Synthetic Peptides

Three N-terminally biotinylated synthetic peptides were reconstituted in buffer A at a concentration of 25 nM. 20 µl of each (500 fmol) was injected into the UltiMate HPLC system for subsequent MS analysis on the API 2000™ (Applied

biosystems). The peptides were subjected to CID and their fragmentation patterns compared to the theoretical fragmentation patterns predicted using the protein prospector MS-Product software (http://prospector.ucsf.edu/prospector/mshome.htm). The peptides analysed were as follows:

Peptide 1: biotin-LNSAECYYPER

Peptide 2: biotin-LCVIPR

Peptide 3: biotin-CIGEVLAK

The primary structures of each of the three peptides were validated. A wide range of both b and y ions characteristic of CID fragmentation were observed in all cases (figures 43-45). The intensity of fragment ions depends on their prevalence and at present is not possible to accurately determine by means other than empirical observation. Some of the predicted fragments were not observed, this could suggest that the CID fragments were not capable of holding a charge (ion formation) and as such were invisible to the detector. In all 3 cases fragment ions bearing the amino acid of interest, cysteine, were visible. This indicates that if they were to be modified they would likely be detectable by mass spectrometry.

3.4.4.1 Synthetic Peptide 1

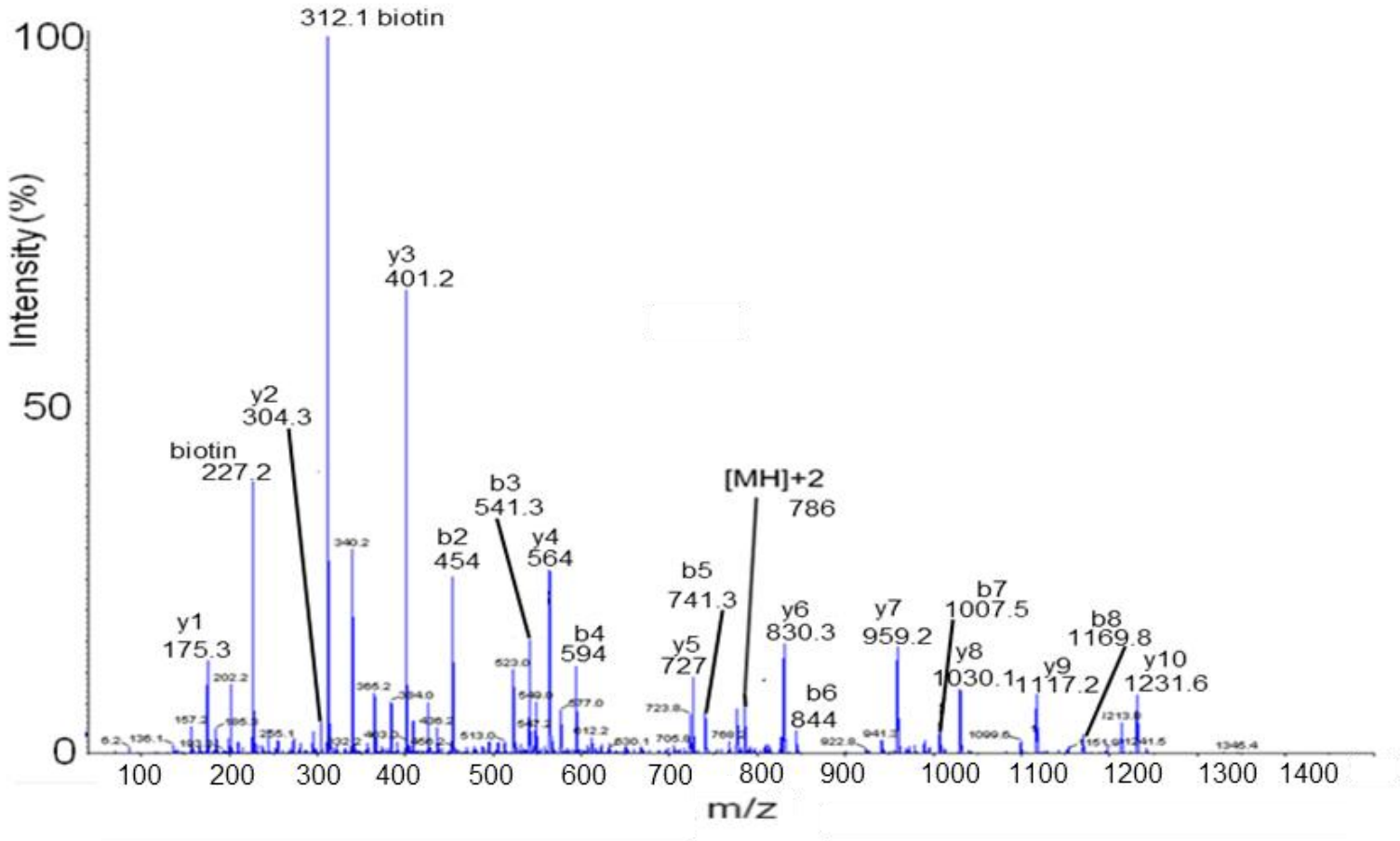


Figure 48. CID fragmentation of synthetic peptide 1 reveals the presence of most theoretical fragments along with ions at m/z 227.2 and 312.1 that are characteristic of the biotin tag.

Table 3. List of theoretical CID fragmentations ions (b and y) for synthetic peptide 1. Ions in green were experimentally confirmed.

Ion	mass $[MH]^{+}$	mass $[MH]^{+2}$	Sequence
Parent	1570.67	785.833	b-LNSAE**C**YYPER
b1	---		---
b2	454.212	227.606	b-LN
b3	541.244	271.122	b-LNS
fb4	612.281	306.641	b-LNSA
b5	741.324	371.162	b-LNSAE
b6	844.333	422.666	b-LNSAE**C**
b7	1007.4	504.198	b-LNSAE**C**Y
b8	1170.46	585.73	b-LNSAE**C**YY
b9	1267.51	634.256	b-LNSAE**C**YYP
b10	1396.55	698.777	b-LNSAE**C**YYPE
b11	---		---
y11	---		---
y10	1231.5	616.252	NSAE**C**YYPER
y9	1117.46	559.231	SAE**C**YYPER
y8	1030.43	515.715	AE**C**YYPER
y7	959.393	480.196	E**C**YYPER
y6	830.35	415.675	**C**YYPER
y5	727.341	364.171	YYPER
y4	564.278	282.639	YPER
y3	401.214	201.107	PER
y2	304.162	152.581	ER
y1	175.119	88.0595	R

3.4.4.2 Synthetic Peptide 2

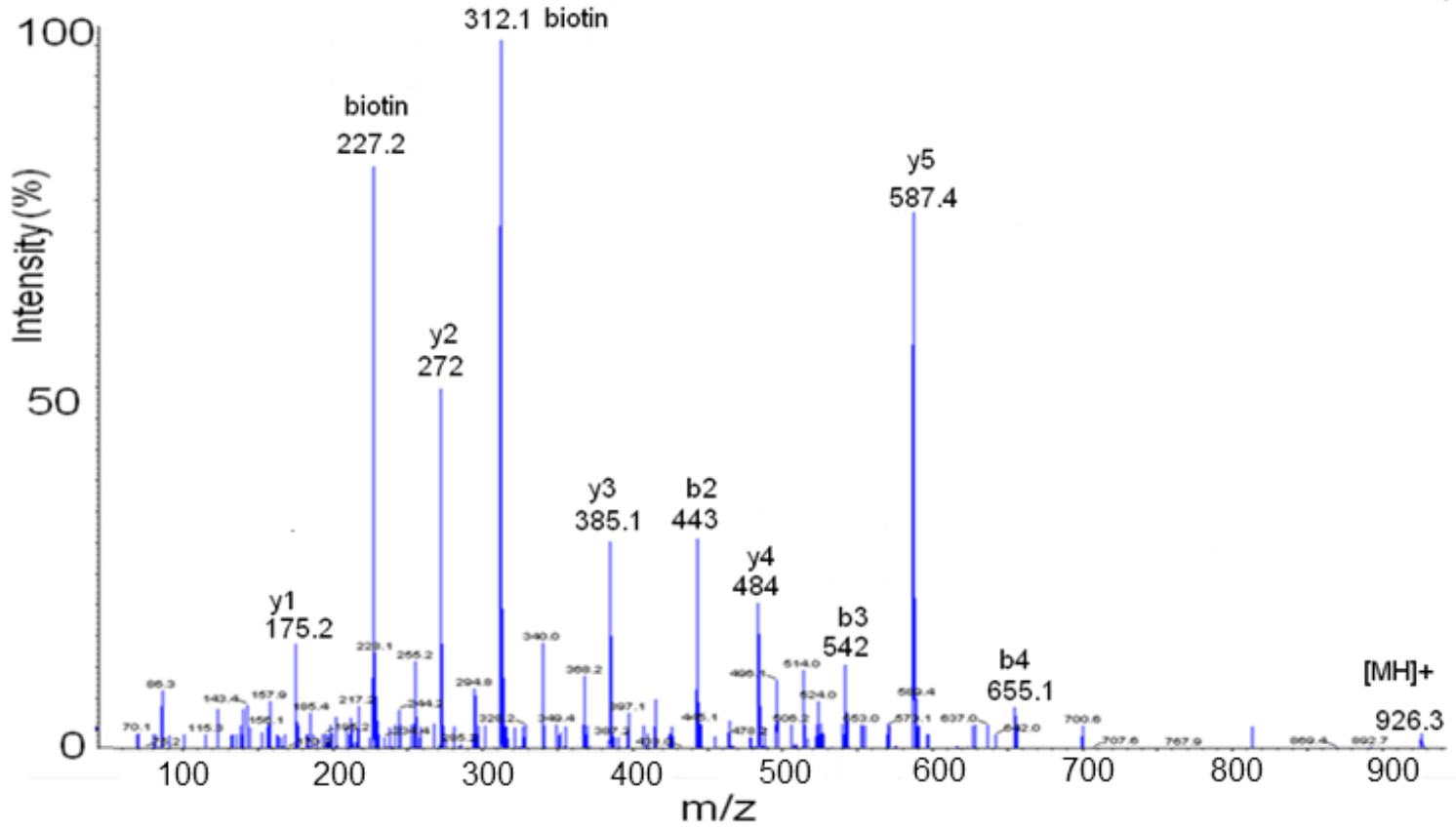


Figure 49. CID fragmentation of synthetic peptide 2 reveals the presence of most theoretical fragments along with an ion at m/z 312.1 that are characteristic of the biotin tag.

Table 4. List of theoretical CID fragmentation ions (b and y) for synthetic peptide 2. Ions in green were experimentally confirmed.

Ion	mass $[MH]^{+}$	mass $[MH]^{+2}$	Sequence
Parent	926.495	463.7475	b-LCVIPR
b1	---		---
b2	443.178	222.089	b-LC
b3	542.247	271.623	b-LCV
b4	655.331	328.165	b-LCVI
b5	752.383	376.692	b-LCVIP
b6	---		---
y6	---		---
y5	587.333	294.167	CVIPR
y4	484.324	242.662	VIPR
y3	385.256	193.128	IPR
y2	272.172	136.586	PR
y1	175.119	88.0595	R

3.4.4.3 Synthetic Peptide 3

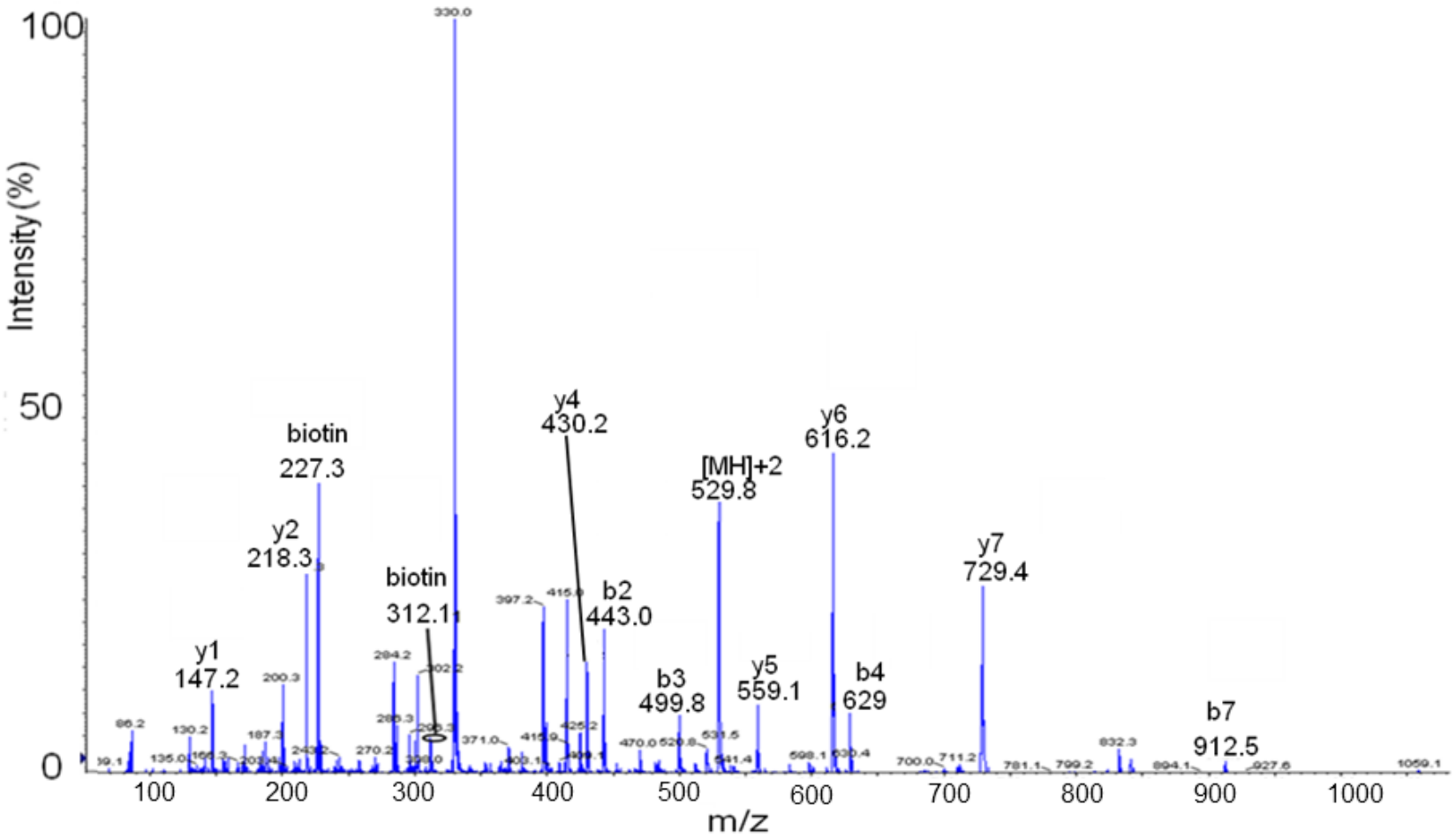


Figure 50. CID fragmentation of synthetic peptide 3 reveals the presence of most theoretical fragments along with ions at m/z 227.3 and 312.1 that are characteristic of the biotin tag.

Table 5. List of theoretical CID fragmentation ions (b and y) for synthetic peptide 3. Ions in green were experimentally confirmed.

Ion	mass $[MH]^{+}$	mass $[MH]^{+2}$	Sequence
Parent	1058.54	529.7687	b-CIGEVLAK
b1	---		---
b2	443.178	222.0891	b-CI
b3	500.2	250.5998	b-CIG
b4	629.242	315.121	b-CIGE
b5	728.311	364.655	b-CIGEV
b6	841.395	421.1974	b-CIGEVL
b7	912.432	456.7159	b-CIGEVLA
b8	---		---
y8	---		---
y7	729.451	365.2253	IGEVLAK
y6	616.367	308.6833	GEVLAK
y5	559.345	280.1725	EVLAK
y4	430.302	215.651	VLAK
y3	331.234	166.117	LAK
y2	218.15	109.575	AK
y1	147.113	74.0564	K

3.4.5 PI359 Based Detection of Synthetic Peptide Conjugates

The synthetic peptides were spiked into the liver microsome assay with clozapine. The peptides were recovered from the complex mixture of the assay by avidin based affinity purification (2.1.1.8). Recovered peptides were then analysed by LC-MS using the previously designed precursor ion scan (PI359) for the identification of clozapine-peptide adducts.

The results were a number of intense well defined peaks as seen in figure 52. An IDA based scan was carried out for each of the samples in order to compare its effectiveness with the precursor ion method. The IDA method results in a TIC without any definable individual peaks (Figure 51). It was possible to identify the modified peptides by searching fthe MS/MS data for ions with the calculated mass but this data was obscured by a much greater amount of irrelevant ions. For the purposes of these experiments all non-modified peptides and other ions are essentially background noise. The modified ions are competeing against this noisy background for detection and discovery. Whereas being able to identify modified peptides in a sample of this complexity is trivial, the next step would require that peptides be identified in a sample of far greater complexity and therefore greater background noise.

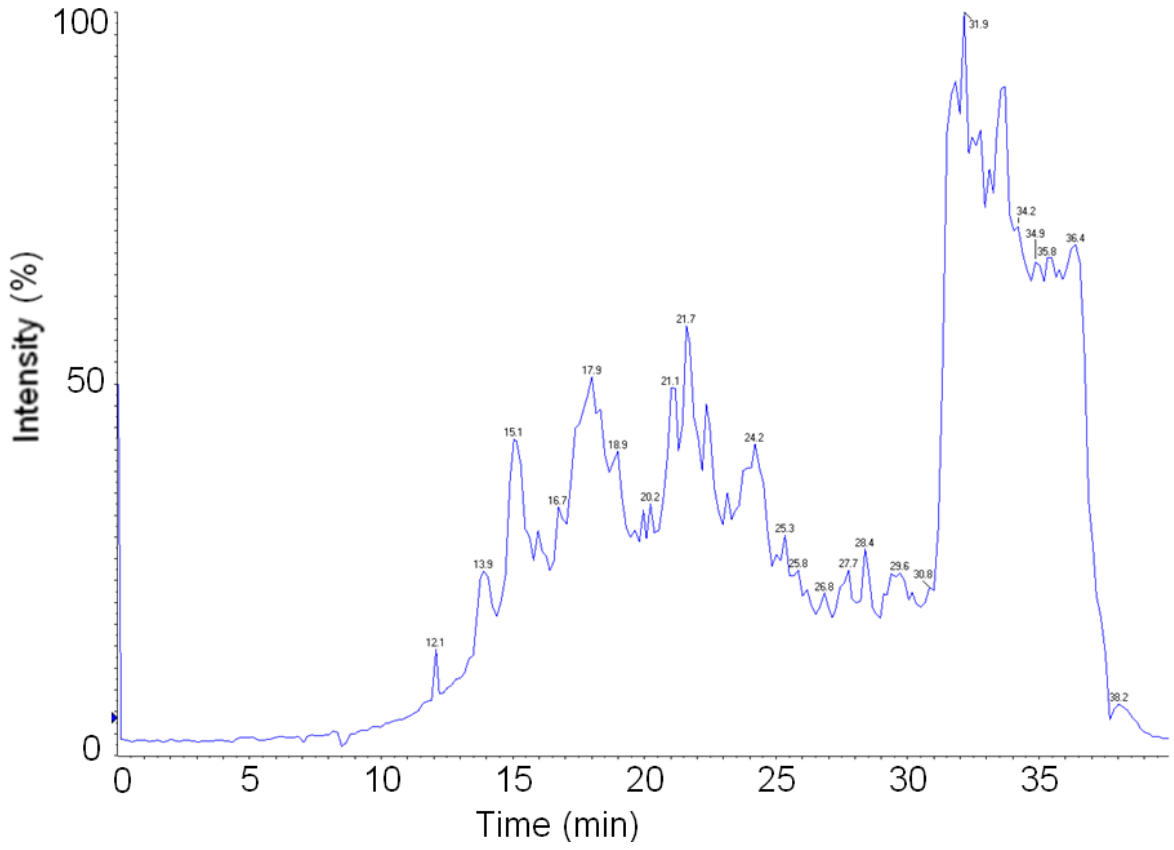


Figure 51. Clozapine treated b-P3 TIC from an IDA MS/MS experiment. The complexity of the chromatogram is evident with no distinguishable or dominant peaks visible.

3.4.5.1 PI359 Scan for Peptide 1

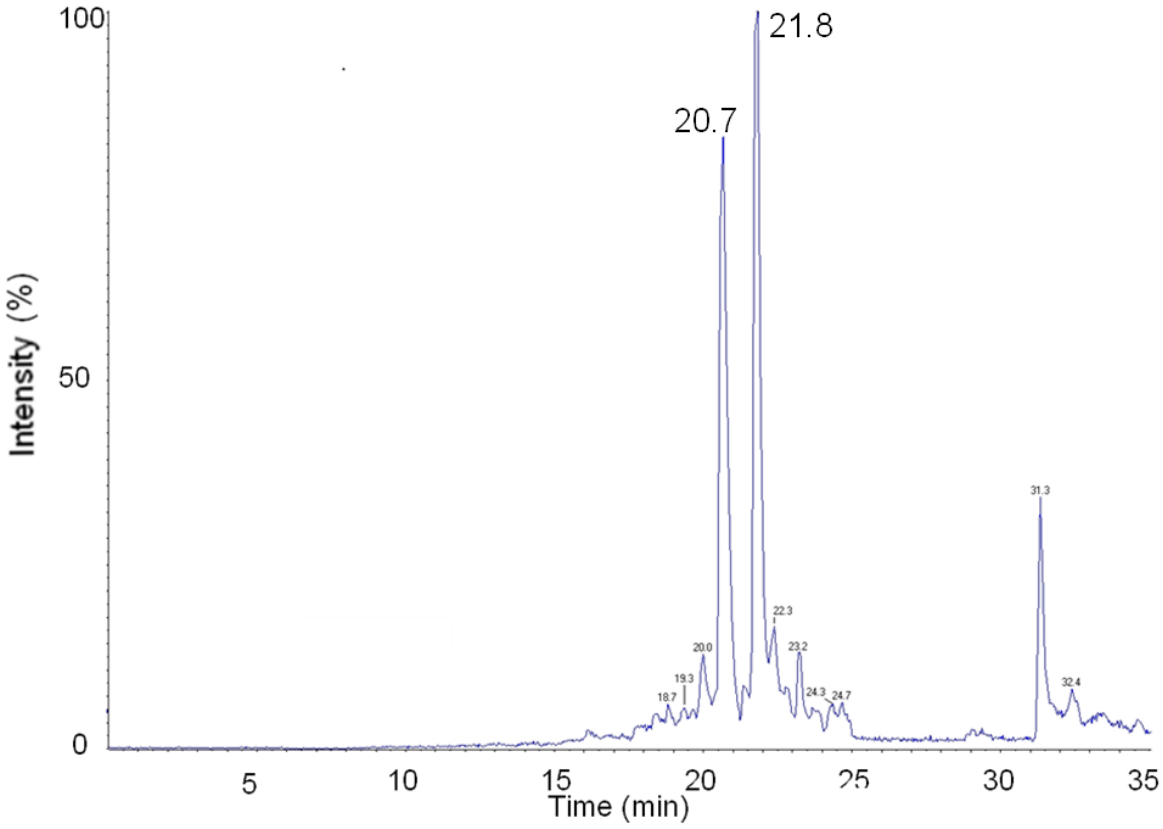


Figure 52. TIC of PI 359 scan of clozapine-P1. The complexity of the chromatogram is relatively low with 2 dominant peaks clearly visible.

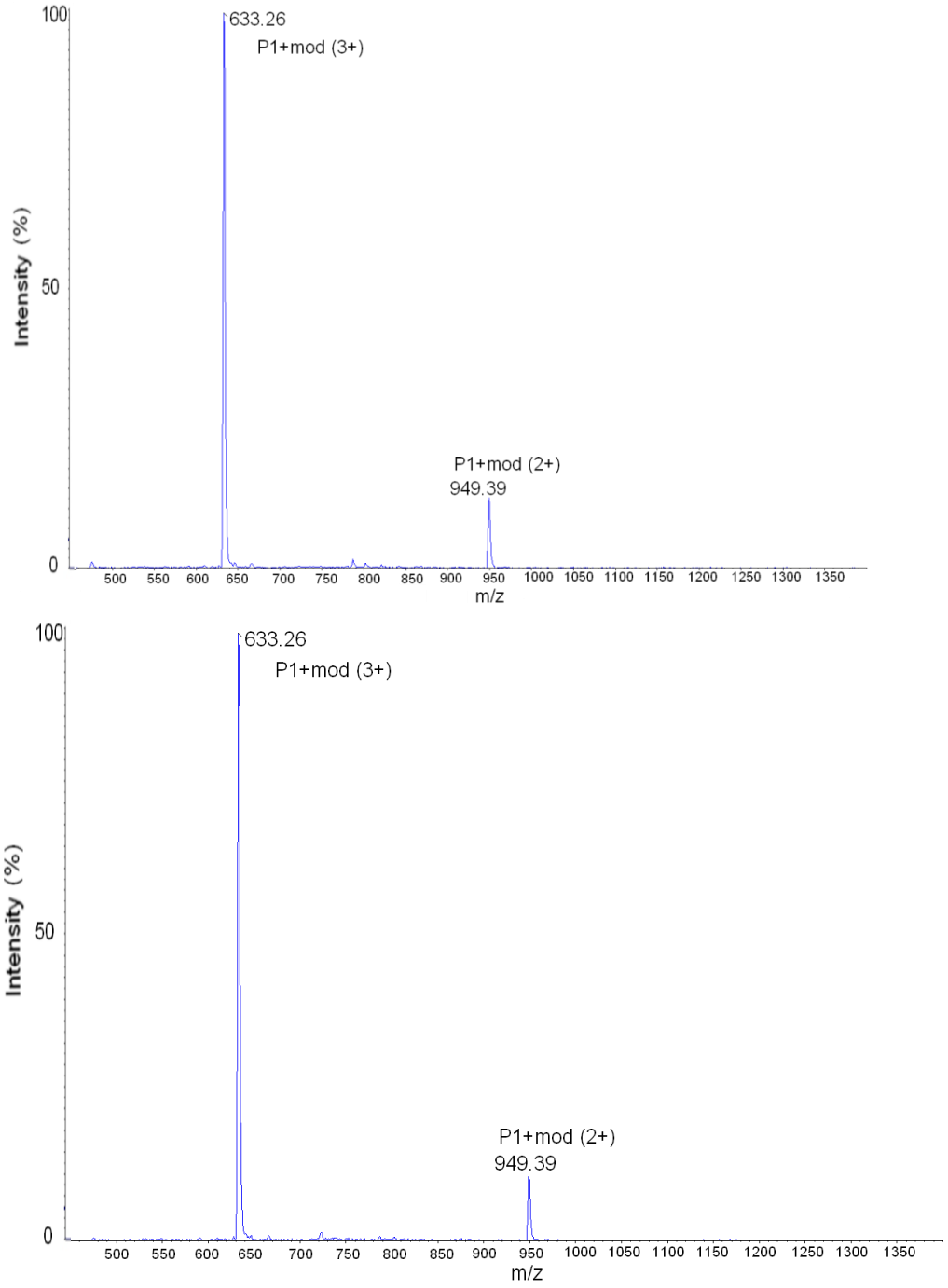


Figure 53. Precursor ion scan of the major peaks seen in figure 52. Clozapine modified peptide is visible at 633.3 m/z ([M+H]3+) and 949.4 m/z ([M+H)2+]. The top image is from the peak at 20.7 min; the bottom from the peak at 21.8 min.

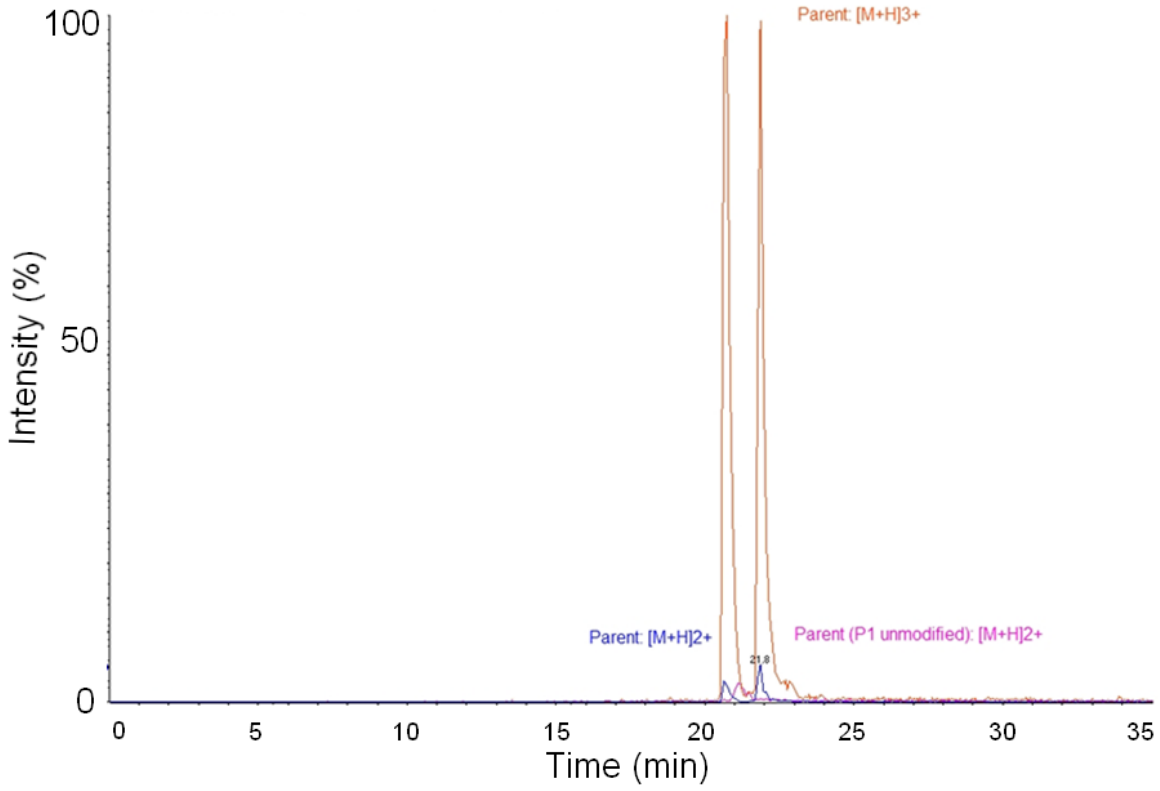


Figure 54. XIC of some of the ions previously detailed in figure (m/z 633.3 and 949.4) 53 and of the unmodified peptide (m/z 786). The triply charged modified ions dominates the spectrum. The doubly charged unmodified peptide is visible, the singly charged ion would fall outside the detection range used.

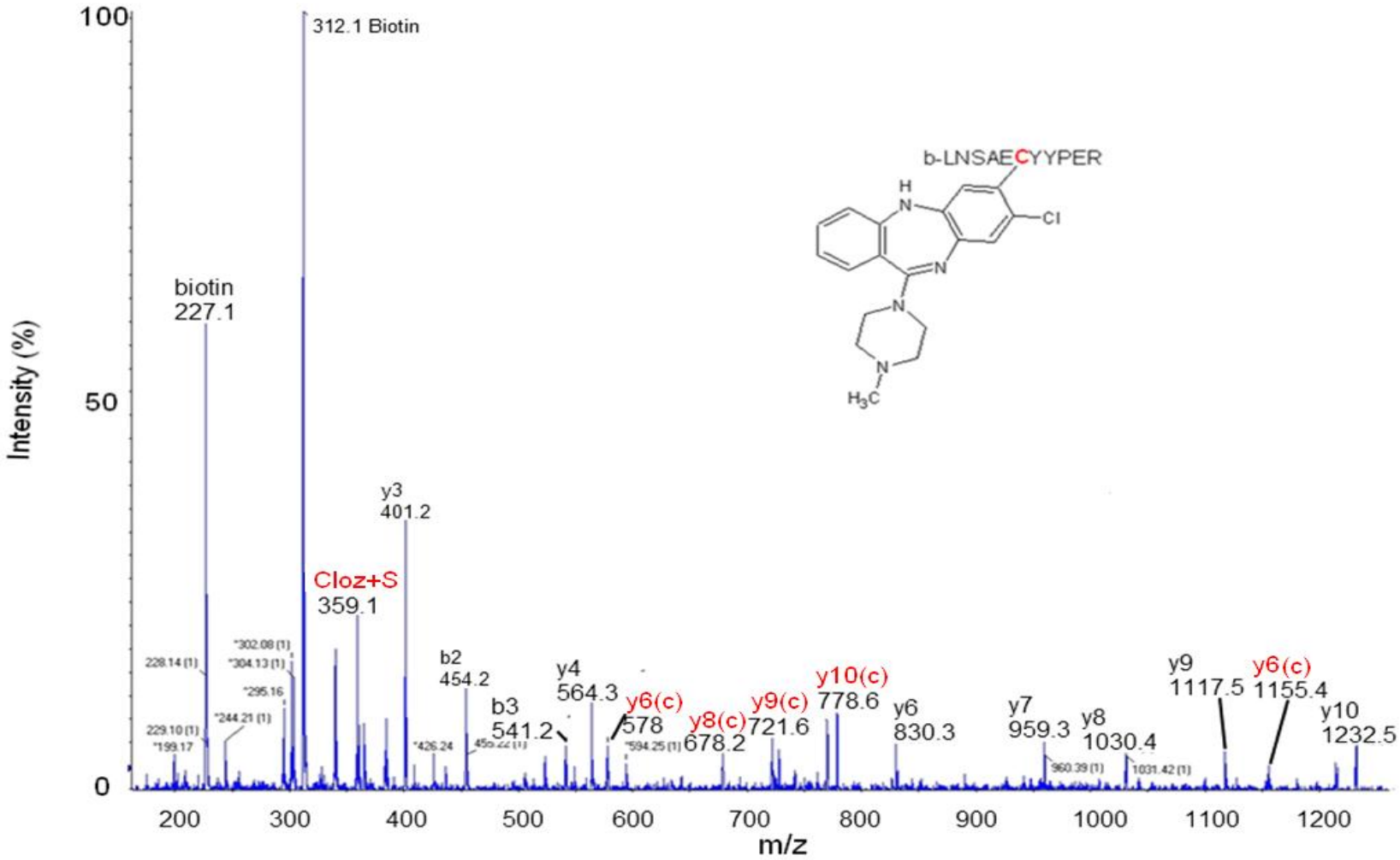


Figure 55. CID fragmentation of clozapine-P1. B and y ions bearing the modification are visible in red and marked with (c).

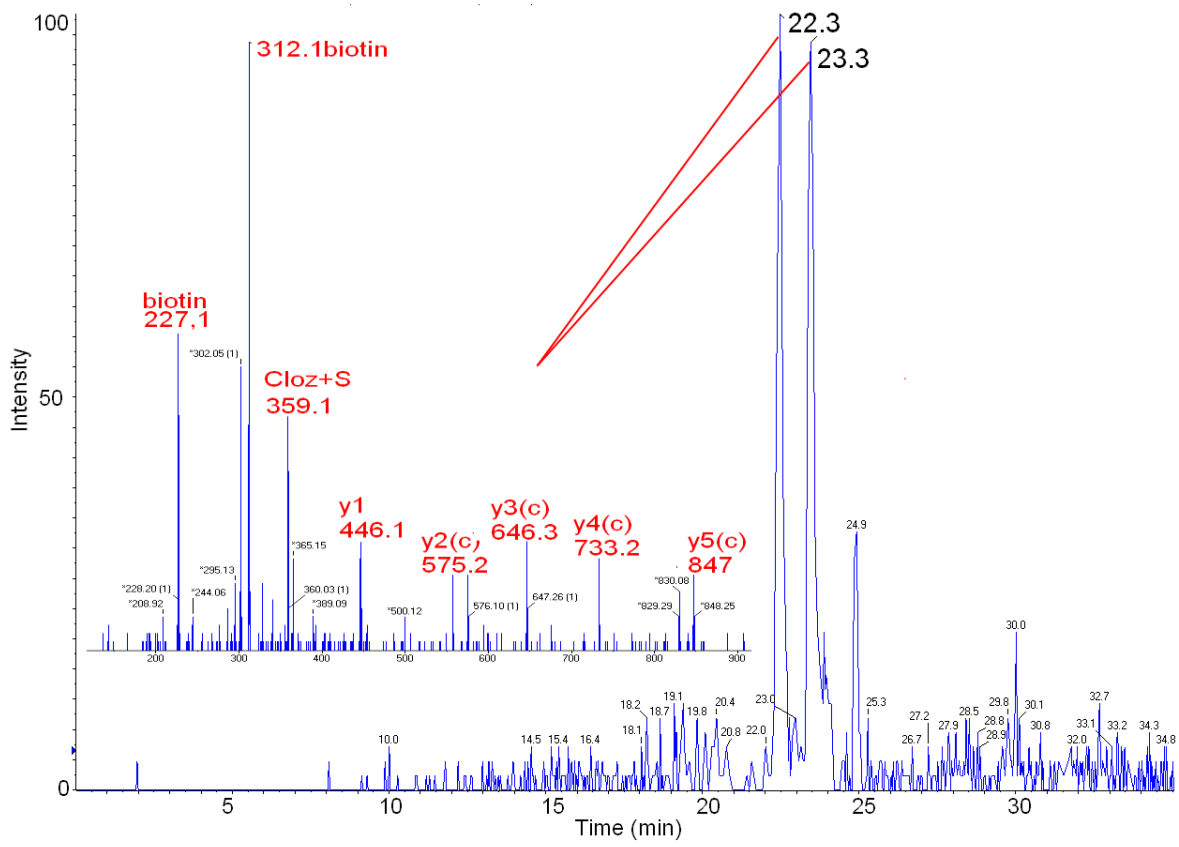


Figure 56. XIC of m/z 593.8 with a corresponding MS/MS spectrum. The ion is representative of a peptide with the sequence b-LNSAEC, a truncated version of the peptide b-LNSAECYYPER.

Table 6. Ions that were identified in MS/MS experiments are highlighted in red or green. Red signifies a clozapine modification bearing ion, green signifies an ion without an associated modification.

Ion	mass [MH]+	mass [MH]+2	Adduct [MH]+3	Sequence
Parent	1894.795	947.8975	632.265	b-LNSAECYYPER
b1	---	---	---	---
b2	454.2119	227.606	---	b-LN
b3	541.2439	271.122	---	b-LNS
b4	612.281	306.6405	---	b-LNSA
b5	741.3236	371.1618	---	b-LNSAE
b6	1168.462	584.7309	---	b-LNSAEC
b7	1331.525	666.2626	---	b-LNSAECY
b8	1494.588	747.7942	---	b-LNSAECYY
b9	1591.641	796.3206	---	b-LNSAECYYP
b10	1720.684	860.8419	---	b-LNSAECYYPE
b11	---	---	---	---
y11	---	---	---	---
y10	1555.634	778.3169	---	NSAECYYPER
y9	1441.591	721.2955	---	SAECYYPER
y8	1354.559	677.7795	---	AECYYPER
y7	1283.522	642.2609	---	ECYYPER
y6	1154.479	577.7396	---	CYYPER
y5	727.341	364.1705	---	YYPER
y4	564.278	282.6388	---	YPER
y3	401.214	201.1072	---	PER
y2	304.162	152.5808	---	ER
y1	175.119	88.0595	---	R

3.4.5.2 PI359 Scan of Peptide 2

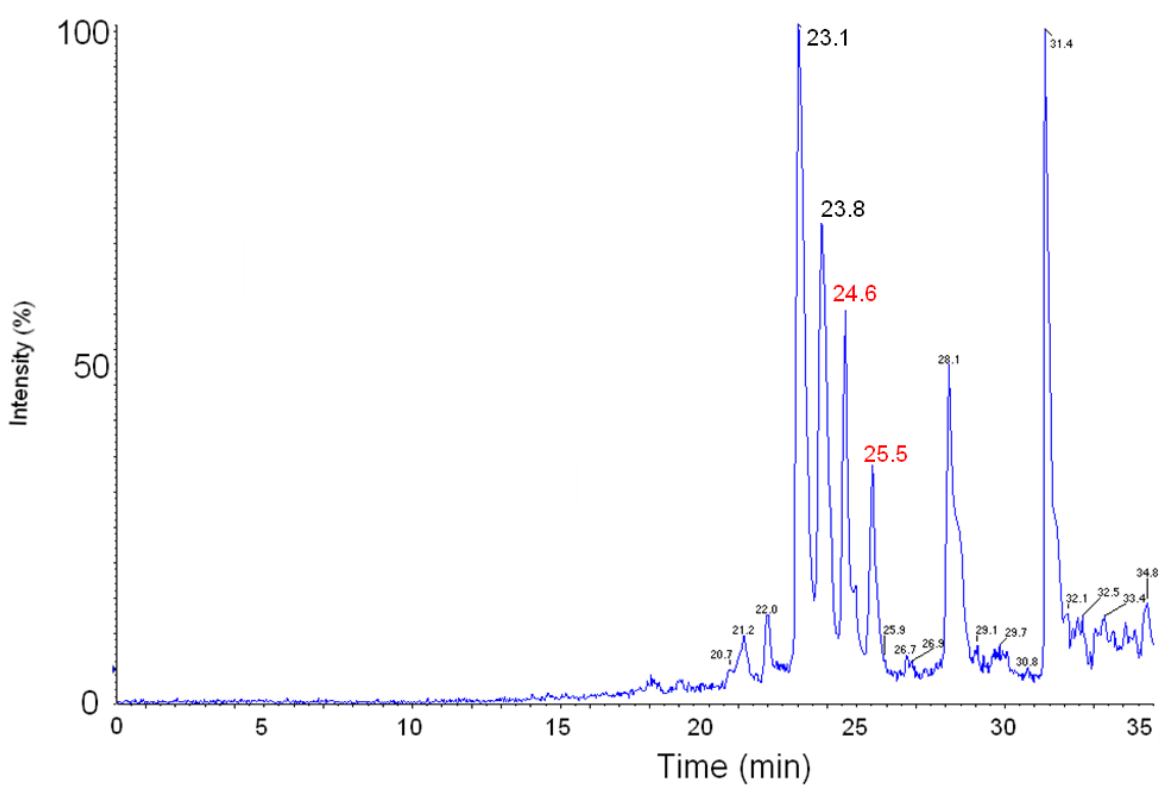


Figure 57. TIC of PI 359 scan of clozapine-P2. A higher number of peaks are visible in this sample. Some of which were identified as carry over contamination from P1 (24.6, 25.5).

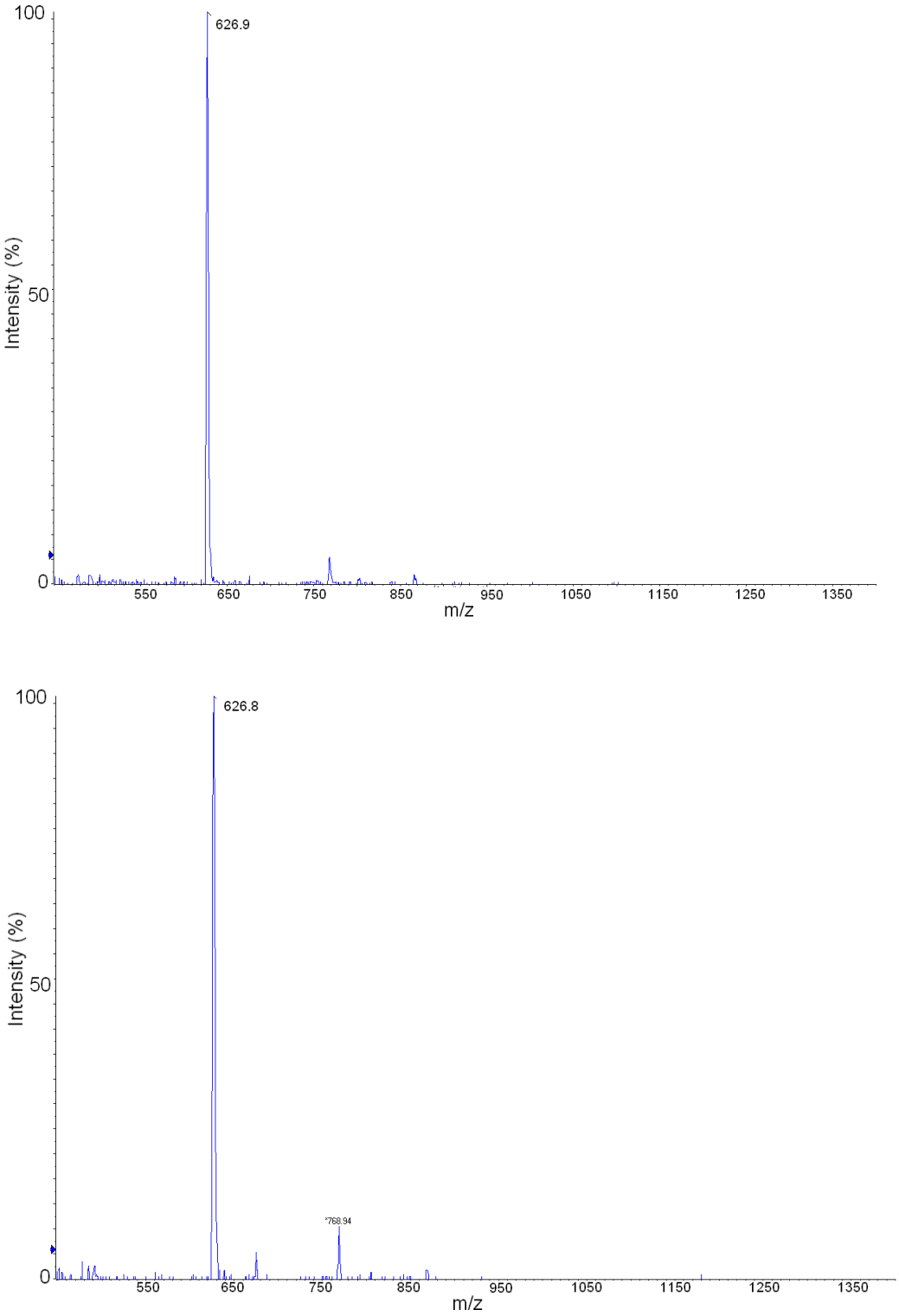


Figure 58. PI359 scan data from the peaks at (top) 23.1 min and (bottom) 23.8 min from figure 57. The ion at 626 m/z is dominant in both.

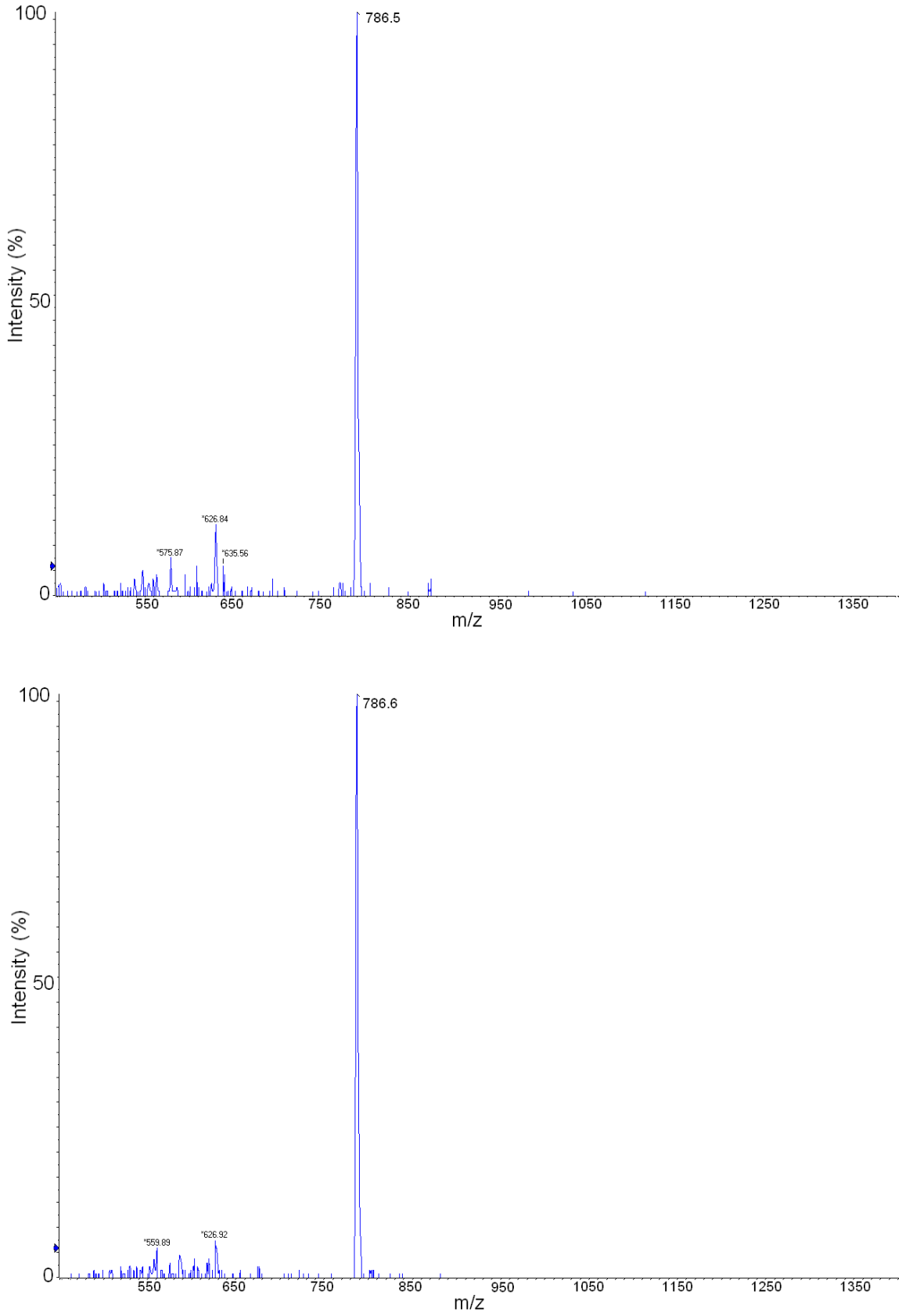


Figure 59. PI359 scan data from the peaks at (top) 24.6 min and (bottom) 25.5 min from figure 57. The ion at 786.6 m/z is dominant.

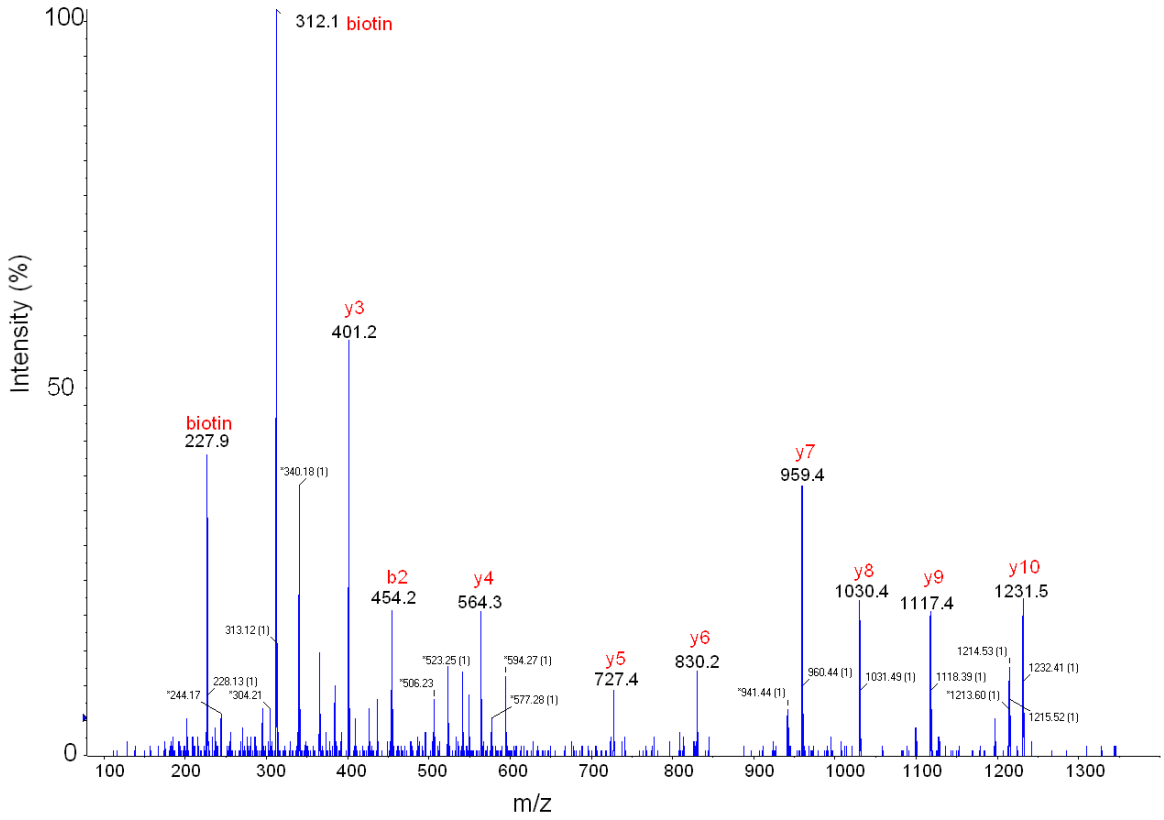


Figure 60. Tandem MS data from the peak at 786.6 m/z seen in figure 59. The spectrum matches up extremely well with that of unmodified peptide 1. The b and y ion series matching peptide 1 is marked in red.

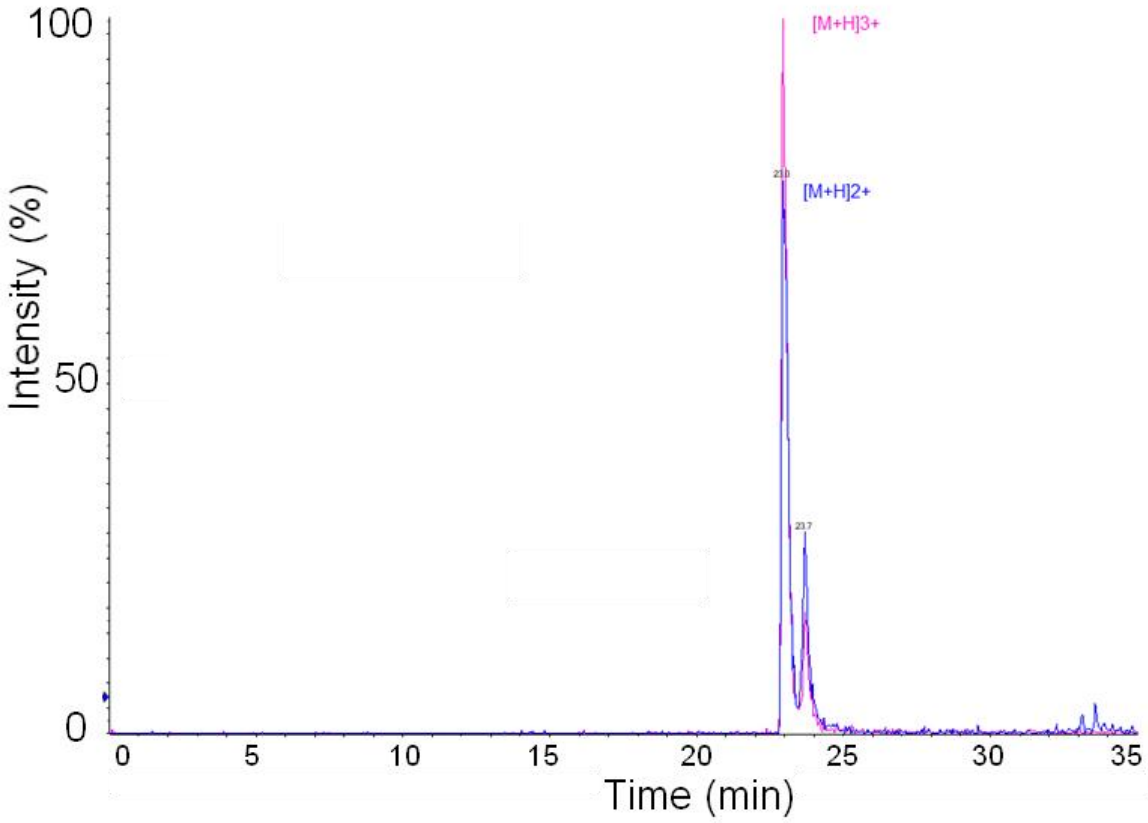


Figure 61. XIC of clozapine-P2 in the 2+ (625.8 m/z) and 3+ (417.5 m/z) charge states. These ions have exactly the same elution profile as would be expected.

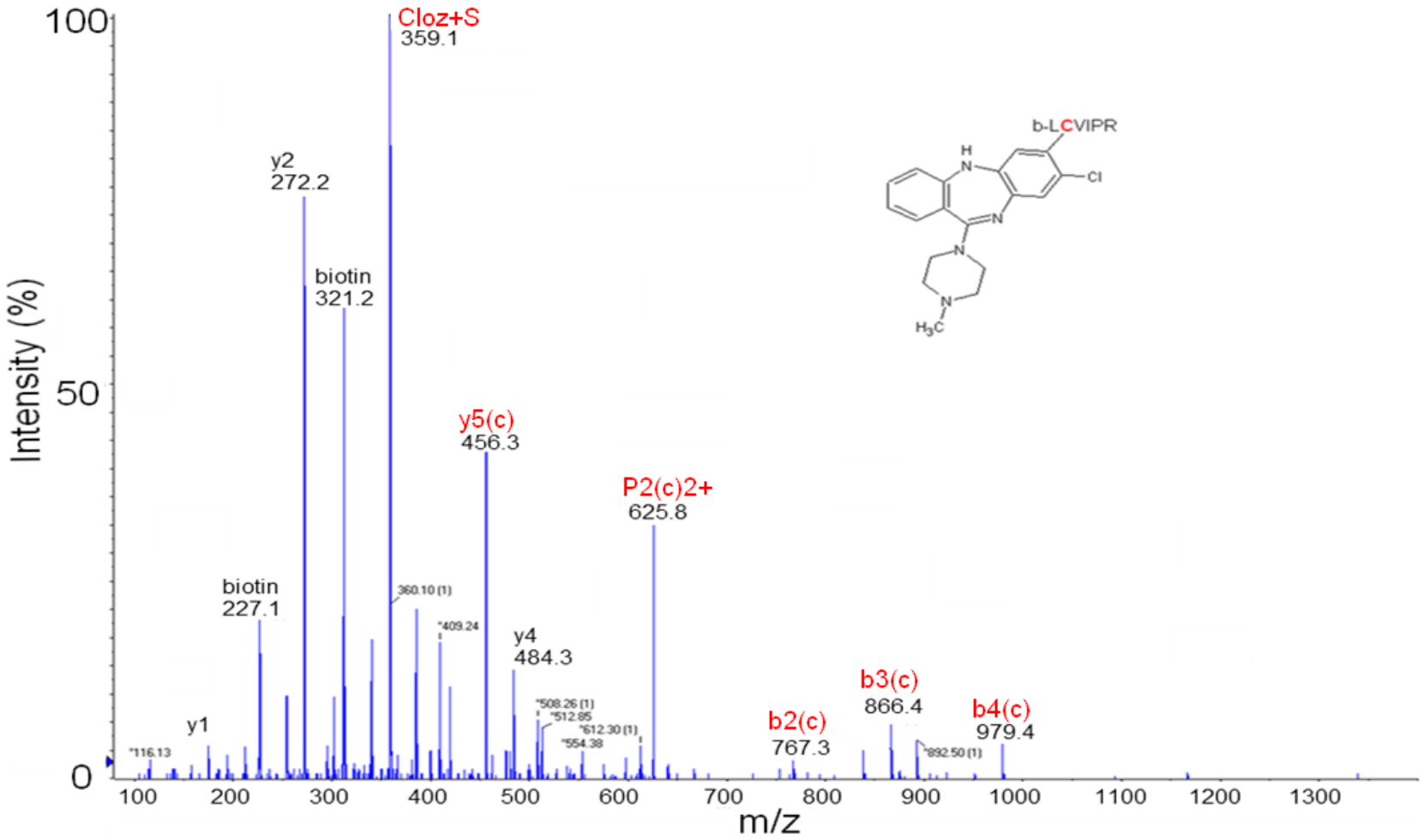


Figure 62. CID fragmentation of clozapine-P2 conjugate at m/z 625.8. Fragments bearing the metabolite adduct are marked in red with a (c).

Table 7. Detected ions are highlighted either red or green. Red signifies the detection of a modification bearing ion.

Ion	Adduct $[MH]^{+}$	Adduct $[MH]^{+2}$	Adduct $[MH]^{+3}$	Sequence
Parent	1250.624	625.812	417.54	b-LCVIPR
b1	---	---	---	---
b2	**767.3071**	**384.1536**	---	b-LC
b3	**866.3755**	**433.6878**	---	b-LCV
b4	**979.4596**	**490.2298**	---	b-LCVI
b5	**1076.512**	**538.7562**	---	b-LCVIP
b6	---	---	---	---
y6	---	---	---	---
y5	**911.4624**	**456.2312**	---	CVIPR
y4	484.324	242.6621	---	VIPR
y3	385.256	193.1279	---	IPR
y2	272.172	136.5859	---	PR
y1	175.119	88.0595	---	R

3.4.5.3 PI359 Scan of Peptide 3

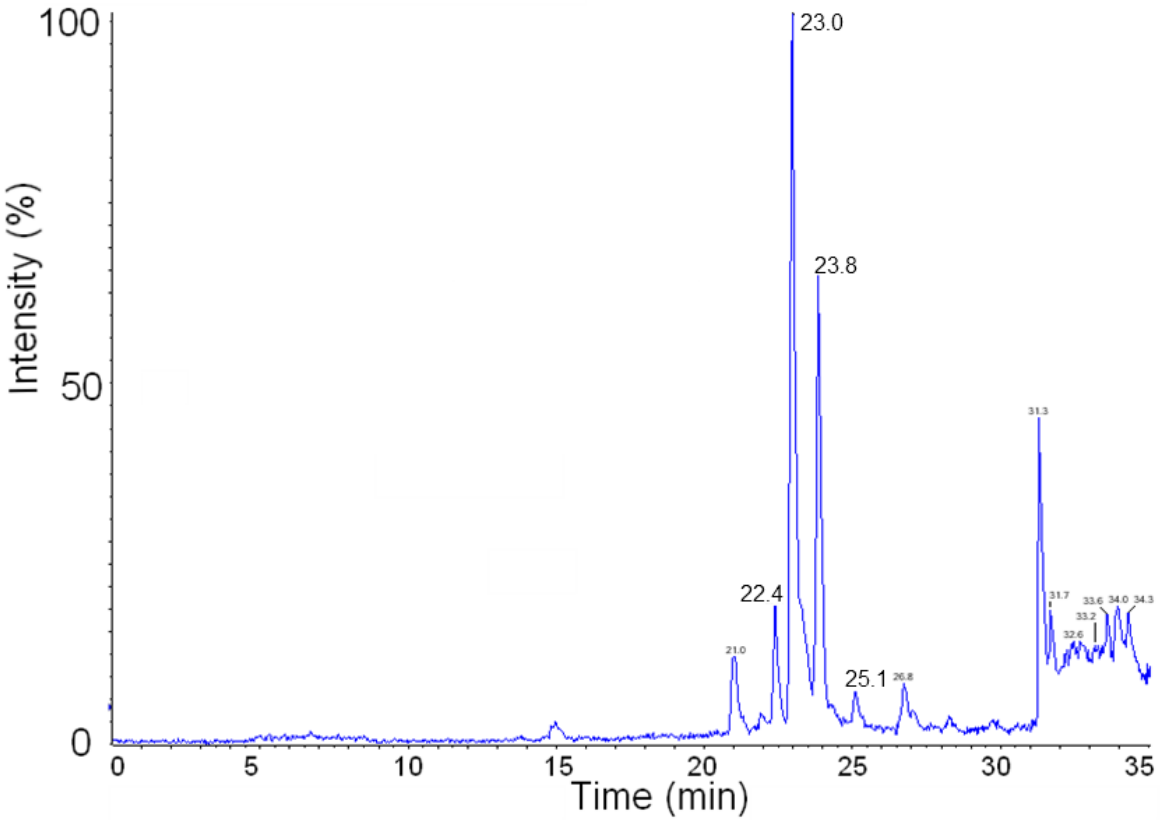


Figure 63. TIC of PI 359 scan of clozapine-P3.

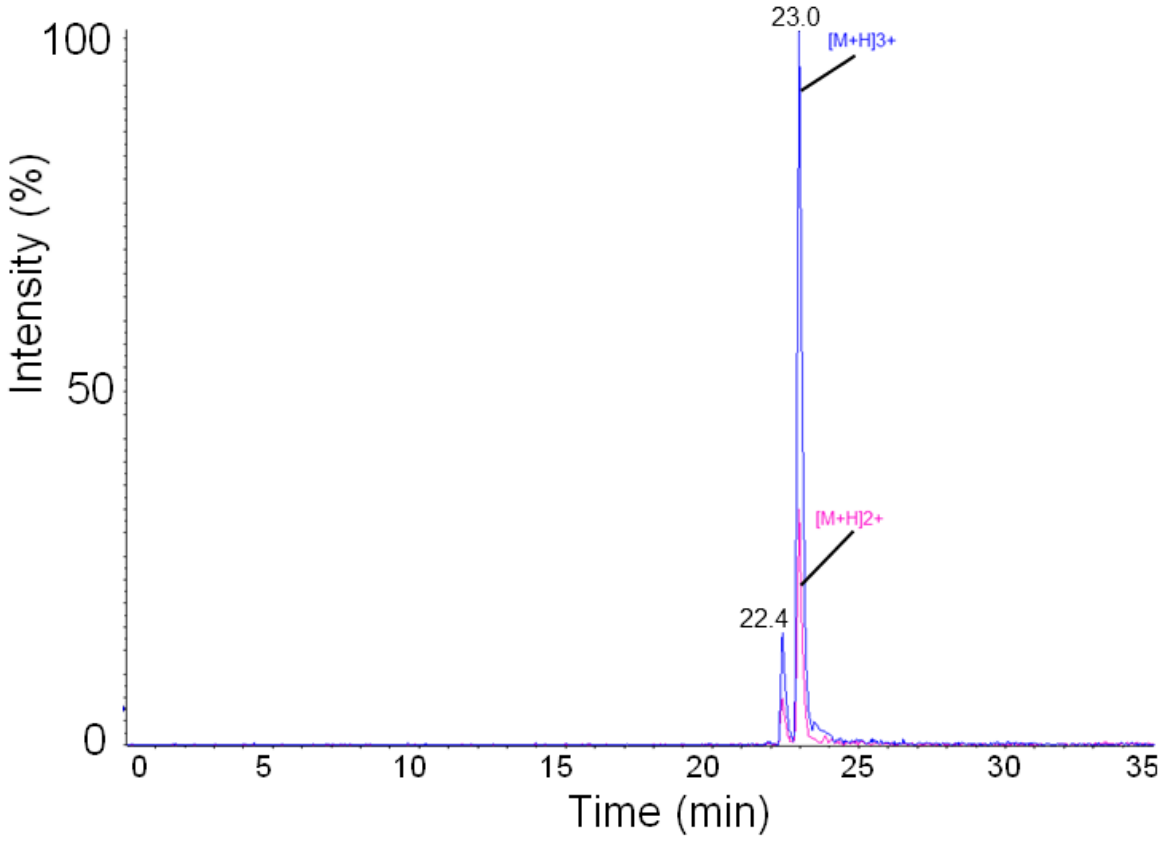


Figure 64. XIC of clozapine-P3 in the 2+ (691.8m/z) and 3+ (461.5m/z) charge states.

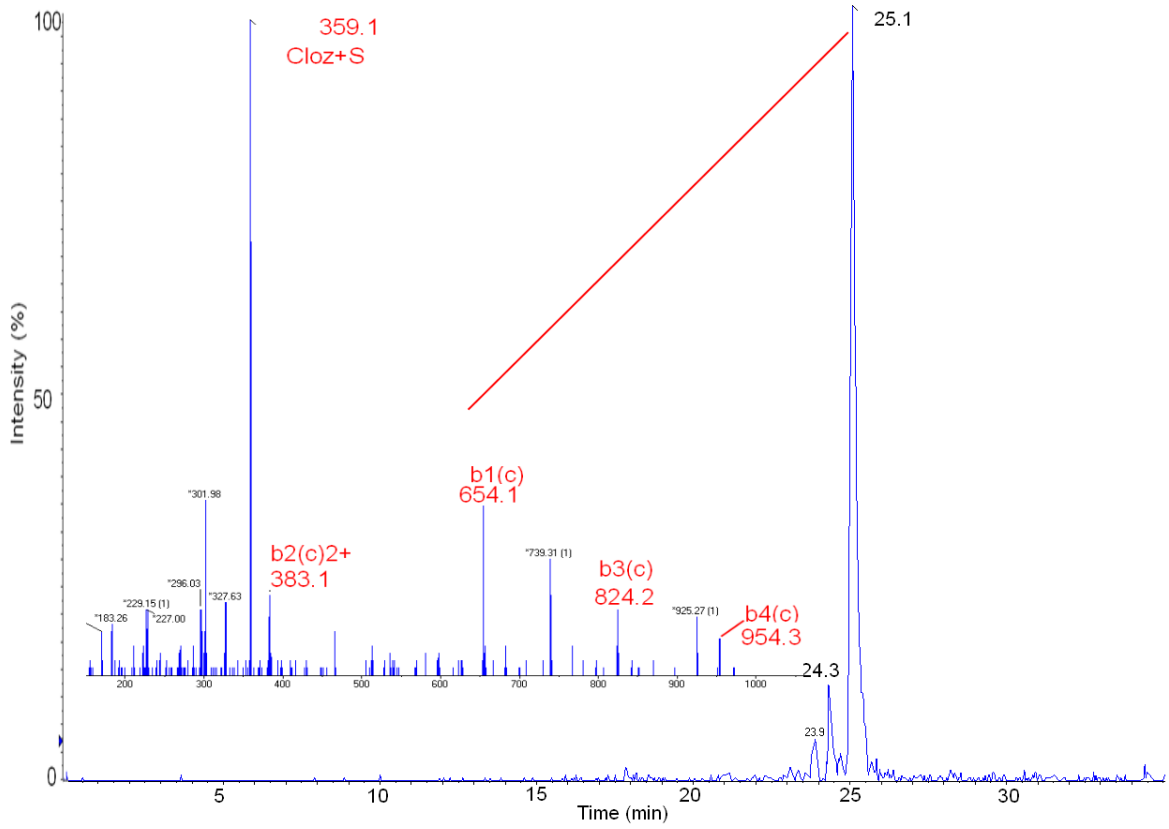


Figure 65. XIC of m/z 536.7 with an MS/MS spectrum inset. This mass is consistent with a fragment of P3 with the sequence b-CIGEV. It represents a truncated form of the peptide b-CIGEVLAK. The ions identified in red represent clozapine modified fragments of the peptide b-CIGEV.

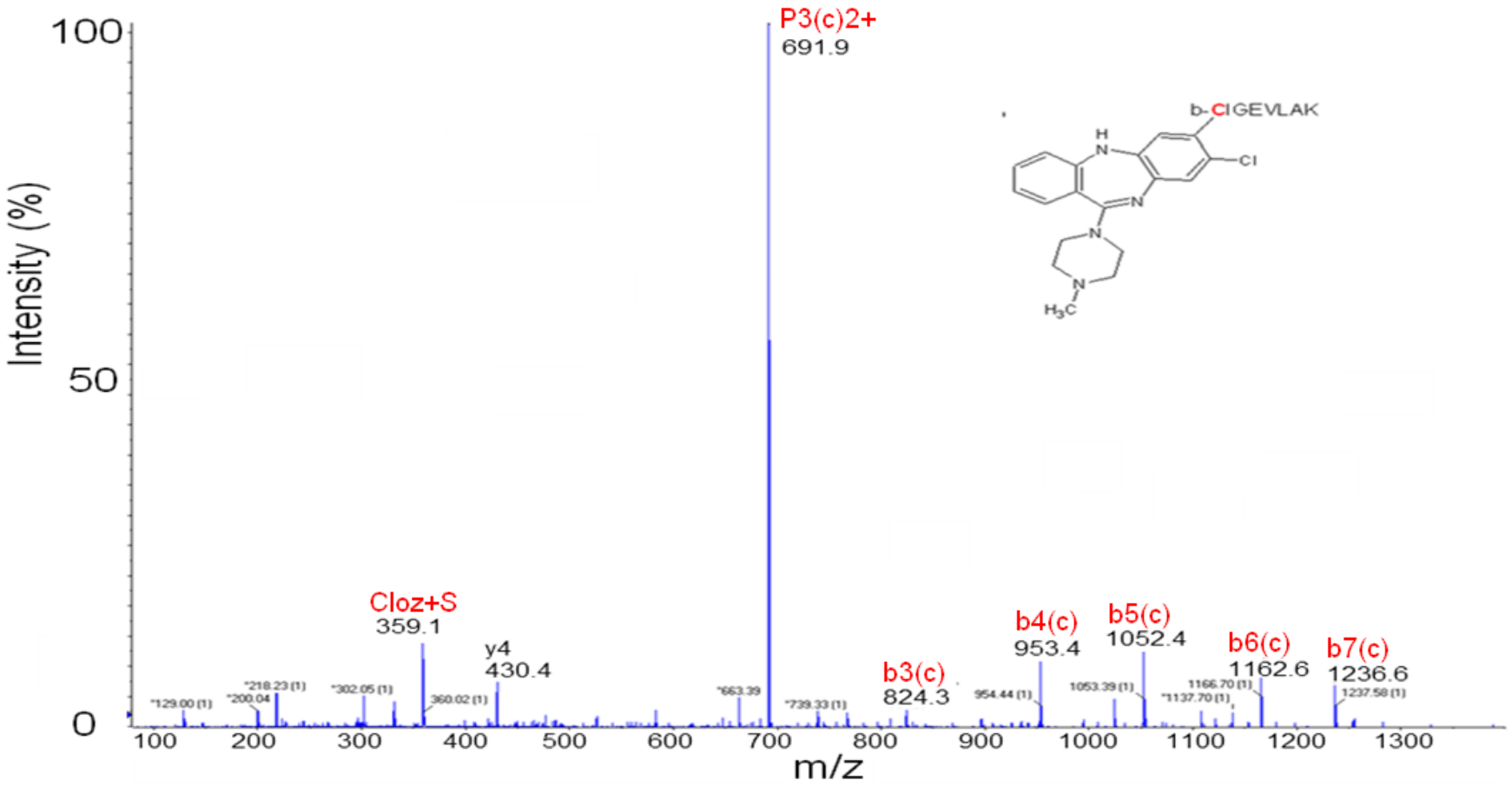


Figure 66. CID fragmentation of clozapine-peptide3 conjugate at m/z 691.9. Fragments bearing metabolite adducts are marked in red with a (c).

Table 8. Detected ions are highlighted either red or green. Red signifies the detection of a modification bearing ion.

Ion	Adduct $[MH]^{+}$	Adduct $[MH]^{+2}$	Adduct $[MH]^{+3}$	Sequence
Parent	**1382.666**	**691.8332**	**461.55**	b-CIGEVLAK
b1	---	---	---	---
b2	**767.3071**	**384.1536**	---	b-CI
b3	**824.3286**	**412.6643**	---	b-CIG
b4	**953.3712**	**477.1856**	---	b-CIGE
b5	**1052.44**	**526.7198**	---	b-CIGEV
b6	**1165.524**	**583.2619**	---	b-CIGEVL
b7	**1236.561**	**618.7804**	---	b-CIGEVLA
b8	---	---	---	---
y8	---	---	---	---
y7	729.451	365.2253	---	IGEVLAK
y6	616.367	308.6833	---	GEVLAK
y5	559.345	280.1725	---	EVLAK
y4	430.302	215.6512	---	VLAK
y3	331.234	166.117	---	LAK
y2	218.15	109.575	---	AK
y1	147.113	74.0564	---	K

3.4.5.4 Synthetic Peptides

The ability to detected modified synthetic peptides has been demonstrated. With all three of the synthetic peptides the modified peptides appeared as split peaks. MS/MS analysis revealed the ions in each split pair to have the same CID fragmentation patterns. Interestingly the unmodified peptide was detected between the split peak. This seems to suggest that the modification can alter the hydrophobicity of the peptide making it both more or less hydrophobic. Contamination of the C18 column was considered as a possible cause of the split

peak, testing of the column with other samples however did not produce similar effects. The column was cleaned with elevated levels of solvent for an extended period of time to remove any contamination. From figures 57, 59-60 it is clear that a contaminant is present that was identified as synthetic peptide 1. The presence of an analyte from a previously injected sample detected in subsequent runs is known as carryover. It is a complex problem with many underlying causes including interaction of analytes with surfaces within the HPLC system, poorly plumbed HPLC systems and scratches in the rotor/stator system of the injector or switching valves. Void volumes in a nanoflow LC system are problematic even at very low volumes. Poor connections between tubing and other components and by the volume of scratches on switching or injection valves provide sufficient voids to allow carry over. A method for eliminating void volumes within nanoflow HPLC systems is discussed in the paper by Mitulovic et al (Mitulovic et al, 2003). Additionally, TFE (2,2,2 Trifluoroethanol) was shown to be effective in washing out a nanoflow HPLC system when added to running buffers and used as a flushing solvent (Mitulovic et al, 2009). It was demonstrated that the addition of the solvent increased the number of peptide and protein Ids and that a user define program for flushing the HPLC system prevented carry over between runs. TFE is used to dissolve proteins prior to enzymatic digestion and is known to enhance solubility of peptides (Polverino et al, 1995; Craig et al, 2008). TFE is compatible with the materials that are used in typical HPLC systems and has even been shown to extend the useful lifespan of reversed phase columns (Bidlingmeyer and Wang, 2006). Application of these techniques to the nanoflow HPLC system used in these experiments could have eliminated the presence of carry over and should be utilised in any follow up work.

3.4.6 Mascot Searching of Synthetic Peptides

The Mascot search engine can be used to screen raw datafiles against large genomic databases. It is possible to allow for the presence of post translational modifications when submitting datafiles for searching. Typically carbamidomethylation of cysteine, representative of iodoacetamide based alkylation, is selected as a fixed (always present) modification and oxidation of

methionine as a variable modification. For the following searches carbamidomethylation was not selected as no alkylation of the peptides was performed. The five previously characterised metabolites of clozapine (3.4.3.1) were added to the Mascot servers modification list. In all cases searching was carried out with all of these selected as variable modifications. Finally, a biotin modification (N-terminal) was selected as a fixed modification in order to allow for the N terminal biotin tags present on all of the synthetic peptides.

3.4.6.1 Mascot Results

The detection of a clozapine metabolite modified synthetic peptides using the selective PI359 scan has been clearly demonstrated. The data from the mass spectrometric analysis was then successfully searched and the clozapine metabolite modification identified using the Mascot search engine. Examination of the matched ion lists revealed the presence of many significantly scoring, high quality ions representative of clozapine modifications. Positive identifications were made for all 3 of the synthetic peptides, each with MOWSE scores well beyond the minimum required for statistical matching.

3.4.6.1.1 Peptide 1

Mascot Score Histogram

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 33 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.

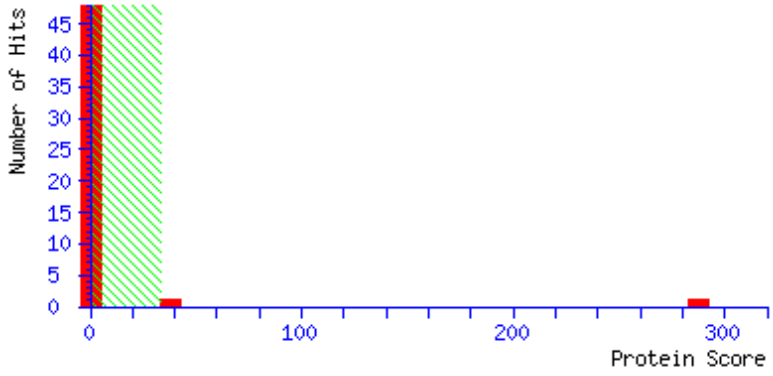
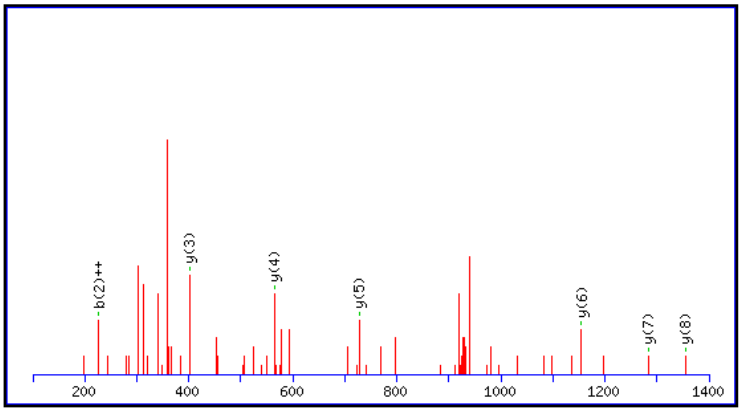


Figure 67. HLM P1 Mascot results MOWSE Score. A score of 288 is of high statistical significance.

1. Q14145 **Mass:** 69847 **Score:** 288 **Matches:** 20(7) **Sequences:** 1(1) **emPAI:** 0.12
Kelch-like ECH-associated protein 1 OS=Homo sapiens GN=KEAP1 PE=1 SV=2
Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
61	785.9245	1569.8344	1569.6591	0.1753	0	(54)	0.00053	1	U	R.LNSAECYYPER.N
62	785.9807	1569.9468	1569.6591	0.2877	0	(67)	2.6e-05	1	U	R.LNSAECYYPER.N
63	785.9866	1569.9586	1569.6591	0.2995	0	(56)	0.00029	1	U	R.LNSAECYYPER.N
64	785.9998	1569.9850	1569.6591	0.3259	0	68	1.9e-05	1	U	R.LNSAECYYPER.N
65	786.1183	1570.2220	1569.6591	0.5629	0	(58)	0.00014	1	U	R.LNSAECYYPER.N
66	786.1526	1570.2906	1569.6591	0.6315	0	(59)	0.00012	1	U	R.LNSAECYYPER.N
67	786.1969	1570.3792	1569.6591	0.7201	0	(50)	0.001	1	U	R.LNSAECYYPER.N
77	632.2769	1893.8089	1893.7733	0.0355	0	(13)	6.1	1	U	R.LNSAECYYPER.N + Clozapine (C)
78	632.2894	1893.8464	1893.7733	0.0730	0	(9)	13	2	U	R.LNSAECYYPER.N + Clozapine (C)
79	632.3185	1893.9337	1893.7733	0.1603	0	(4)	41	10	U	R.LNSAECYYPER.N + Clozapine (C)
80	948.0183	1894.0220	1893.7733	0.2487	0	(32)	0.067	1	U	R.LNSAECYYPER.N + Clozapine (C)
81	632.3537	1894.0393	1893.7733	0.2659	0	(11)	8	1	U	R.LNSAECYYPER.N + Clozapine (C)
83	632.3761	1894.1065	1893.7733	0.3331	0	(0)	92	10	U	R.LNSAECYYPER.N + Clozapine (C)
9	632.3830	1894.1272	1893.7733	0.3538	0	(17)	4.3	1	U	R.LNSAECYYPER.N + Clozapine (C)
84	948.0734	1894.1322	1893.7733	0.3589	0	(13)	4.6	1	U	R.LNSAECYYPER.N + Clozapine (C)
85	632.3943	1894.1611	1893.7733	0.3877	0	(1)	71	7	U	R.LNSAECYYPER.N + Clozapine (C)
86	948.1139	1894.2132	1893.7733	0.4399	0	(23)	0.42	1	U	R.LNSAECYYPER.N + Clozapine (C)
87	948.1141	1894.2136	1893.7733	0.4403	0	(28)	0.13	1	U	R.LNSAECYYPER.N + Clozapine (C)
88	632.4241	1894.2505	1893.7733	0.4771	0	(15)	2.2	1	U	R.LNSAECYYPER.N + Clozapine (C)
89	632.4576	1894.3510	1893.7733	0.5776	0	(14)	2.9	1	U	R.LNSAECYYPER.N + Clozapine (C)

Figure 68. HLM P1 Mascot results protein hits. The modified peptide sequence was identified in both its 2+ and 3+ charge states.



Monoisotopic mass of neutral peptide Mr(calc): 1893.7733
Fixed modifications: Biotin (N-term) (apply to specified residues or termini only)
Variable modifications:
C6 : Clozapine (C)
Ions Score: 32 **Expect:** 0.067
Matches : 7/116 fragment ions using 11 most intense peaks (help)

#	a	a^{++}	a^{*}	a^{*++}	b	b^{++}	b^{*}	b^{*++}	Seq.	y	y^{++}	y^{*}	y^{*++}	#
1	312.1740	156.5906			340.1689	170.5881			L					11
2	426.2169	213.6121	409.1904	205.0988	454.2119	227.6096	437.1853	219.0963	N	1555.6190	778.3131	1538.5924	769.7999	10
3	513.2490	257.1281	496.2224	248.6149	541.2439	271.1256	524.2173	262.6123	S	1441.5761	721.2917	1424.5495	712.7784	9
4	584.2861	292.6467	567.2595	284.1334	612.2810	306.6441	595.2545	298.1309	A	***1354.5440***	677.7757	1337.5175	669.2624	8
5	713.3287	357.1680	696.3021	348.6547	741.3236	371.1654	724.2971	362.6522	E	***1283.5069***	642.2571	1266.4804	633.7438	7
6	1140.4520	570.7297	1123.4255	562.2164	1168.4470	584.7271	1151.4204	576.2138	C	***1154.4643***	577.7358	1137.4378	569.2225	6
7	1303.5154	652.2613	1286.4888	643.7481	1331.5103	666.2588	1314.4837	657.7455	Y	***727.3410***	364.1741	710.3144	355.6608	5
8	1466.5787	733.7930	1449.5522	725.2797	1494.5736	747.7904	1477.5471	739.2772	Y	***564.2776***	282.6425	547.2511	274.1292	4
9	1563.6315	782.3194	1546.6049	773.8061	1591.6264	796.3168	1574.5998	787.8036	P	***401.2143***	201.1108	384.1878	192.5975	3
10	1692.6741	846.8407	1675.6475	838.3274	1720.6690	860.8381	1703.6424	852.3249	E	304.1615	152.5844	287.1350	144.0711	2
11									R	175.1190	88.0631	158.0924	79.5498	1

Figure 69. Ion 80. -.LNSAECYYPER.- + Clozapine (C). A good quality spectrum is visible with a good representation of y ions and b ions.

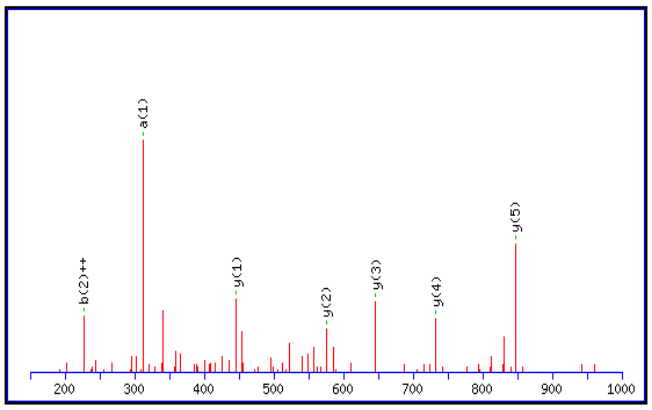
1. in|PAG001|PAG001 **Mass:** 1570 **Score:** 306 **Queries matched:** 25 **emPAI:** 252.39

- HLM_P1

☐ Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
☑ 29	593.2400	1184.4654	1185.4502	-0.9848	0	(32)	0.49	1	-.LNSAEC.Y + Clozapine (C)
☑ 30	593.2684	1184.5222	1185.4502	-0.9280	0	(29)	0.8	1	-.LNSAEC.Y + Clozapine (C)
31	593.2719	1184.5292	1185.4502	-0.9210	0	(6)	1.6e+02	4	-.LNSAEC.Y + Clozapine (C)
☑ 32	593.7651	1185.5156	1185.4502	0.0654	0	(33)	0.36	1	-.LNSAEC.Y + Clozapine (C)
☑ 33	593.7768	1185.5390	1185.4502	0.0888	0	40	0.071	1	-.LNSAEC.Y + Clozapine (C)
☑ 34	593.7870	1185.5594	1185.4502	0.1092	0	(30)	0.67	1	-.LNSAEC.Y + Clozapine (C)
☑ 35	593.7994	1185.5842	1185.4502	0.1340	0	(34)	0.25	1	-.LNSAEC.Y + Clozapine (C)
☑ 36	593.8077	1185.6008	1185.4502	0.1506	0	(35)	0.21	1	-.LNSAEC.Y + Clozapine (C)
☑ 51	728.9075	1455.8004	1456.5751	-0.7746	0	43	0.027	1	L.NSAECYYPER.-
☑ 57	785.9245	1569.8344	1569.6591	0.1753	0	(71)	5e-05	1	-.LNSAECYYPER.-
☑ 58	785.9807	1569.9468	1569.6591	0.2877	0	(70)	5.1e-05	1	-.LNSAECYYPER.-
☑ 59	785.9866	1569.9586	1569.6591	0.2995	0	(70)	4.7e-05	1	-.LNSAECYYPER.-
☑ 60	785.9998	1569.9850	1569.6591	0.3259	0	87	9.9e-07	1	-.LNSAECYYPER.-
☑ 61	786.1183	1570.2220	1569.6591	0.5629	0	(78)	6.9e-06	1	-.LNSAECYYPER.-
☑ 62	786.1526	1570.2906	1569.6591	0.6315	0	(67)	8.2e-05	1	-.LNSAECYYPER.-
☑ 63	786.1969	1570.3792	1569.6591	0.7201	0	(64)	0.00016	1	-.LNSAECYYPER.-
75	632.2769	1893.8089	1893.7733	0.0355	0	(14)	21	3	-.LNSAECYYPER.- + Clozapine (C)
76	632.2894	1893.8464	1893.7733	0.0730	0	(10)	54	5	-.LNSAECYYPER.- + Clozapine (C)
☑ 78	948.0183	1894.0220	1893.7733	0.2487	0	(35)	0.15	1	-.LNSAECYYPER.- + Clozapine (C)
79	632.3537	1894.0393	1893.7733	0.2659	0	(13)	25	6	-.LNSAECYYPER.- + Clozapine (C)
☑ 82	948.0734	1894.1322	1893.7733	0.3589	0	(16)	9.6	1	-.LNSAECYYPER.- + Clozapine (C)
☑ 84	948.1139	1894.2132	1893.7733	0.4399	0	(26)	0.97	1	-.LNSAECYYPER.- + Clozapine (C)
☑ 85	948.1141	1894.2136	1893.7733	0.4403	0	(28)	0.68	1	-.LNSAECYYPER.- + Clozapine (C)
☑ 86	632.4241	1894.2505	1893.7733	0.4771	0	(18)	6.5	1	-.LNSAECYYPER.- + Clozapine (C)
☑ 87	632.4576	1894.3510	1893.7733	0.5776	0	(17)	6.9	1	-.LNSAECYYPER.- + Clozapine (C)

Figure 70. HLM P1 Mascot results protein hits. The modified peptide sequence was identified in all 3 charge states. A truncated peptide LNSAEC.Y+Clozapine was identified.



Monoisotopic mass of neutral peptide Mr(calc): 1185.4502
Fixed modifications: Biotin (N-term)
Variable modifications:
C6 : Clozapine (C)
Ions Score: 40 **Expect:** 0.071
Matches (Bold Red): 7/60 fragment ions using 9 most intense peaks

#	a	a^{++}	b	b^{++}	b*	b^{*++}	b^{0}	b^{0++}	Seq.	y	y^{++}	y*	y^{*++}	y^{0}	y^{0++}	#
1	**312.1740**	156.5906	340.1689	170.5881					L							6
2	426.2169	213.6121	454.2119	**227.6096**	437.1853	219.0963			N	**847.2959**	424.1516	830.2693	415.6383	829.2853	415.1463	5
3	513.2490	257.1281	541.2439	271.1256	524.2173	262.6123	523.2333	262.1203	S	**733.2529**	367.1301			715.2424	358.1248	4
4	584.2861	292.6467	612.2810	306.6441	595.2545	298.1309	594.2704	297.6389	A	**646.2209**	323.6141			628.2104	314.6088	3
5	713.3287	357.1680	741.3236	371.1654	724.2971	362.6522	723.3130	362.1602	E	**575.1838**	288.0955			557.1732	279.0903	2
6									C	**446.1412**	223.5742					1

Figure 71. Ion 33. -.LNSAEC.Y + Clozapine (C). The figure depicts a full set of y ions consistent with the peptide fragment described.

3.4.6.1.2 Peptide 2

Mascot Score Histogram

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 36 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.

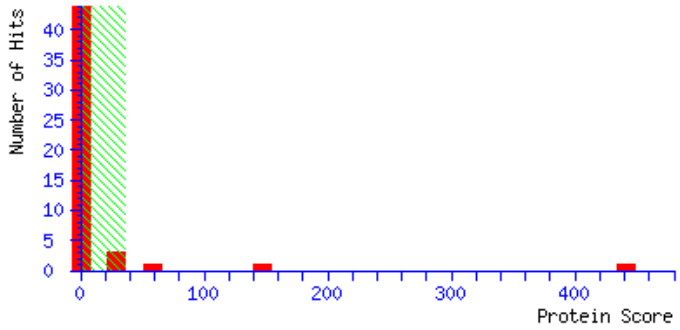
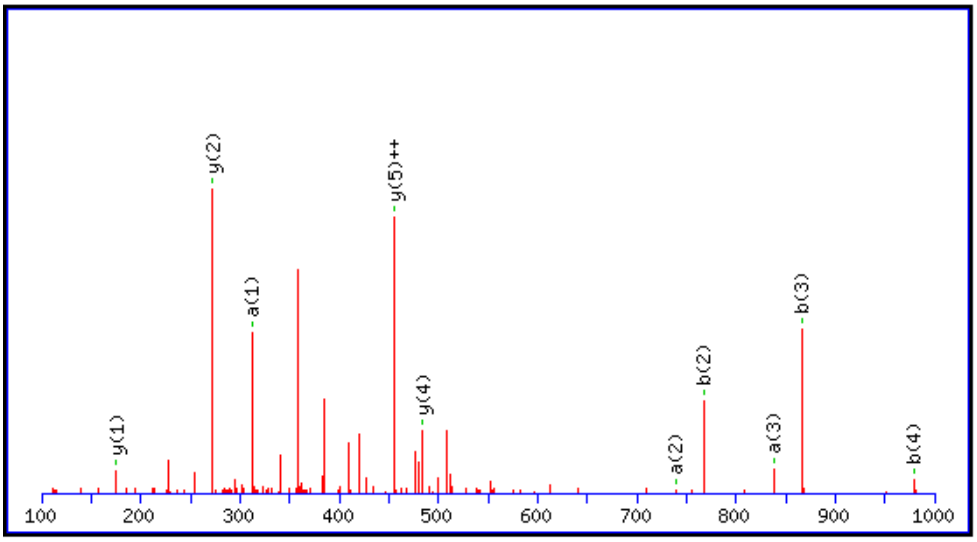


Figure 72. HLM P2 Mascot results MOWSE score of 442, a highly significant statistical match.

1. P05181 **Mass:** 57039 **Score:** 442 **Matches:** 58(40) **Sequences:** 1(1) **emPAI:** 0.07
Cytochrome P450 2E1 OS=Homo sapiens GN=CYP2E1 PE=1 SV=1
Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
260	463.2011	924.3876	925.4877	-1.1001	0	(32)	0.2	1	U	K.LCVIPR.S
262	463.2221	924.4296	925.4877	-1.0581	0	(32)	0.2	1	U	K.LCVIPR.S
498	463.2357	924.4568	925.4877	-1.0309	0	(45)	0.005	1	U	K.LCVIPR.S
499	463.2420	924.4694	925.4877	-1.0183	0	(7)	29	3	U	K.LCVIPR.S
507	463.2781	924.5416	925.4877	-0.9461	0	(42)	0.01	1	U	K.LCVIPR.S
508	463.2788	924.5430	925.4877	-0.9447	0	(44)	0.0052	1	U	K.LCVIPR.S
264	463.2917	924.5688	925.4877	-0.9189	0	(36)	0.074	1	U	K.LCVIPR.S
510	463.2965	924.5784	925.4877	-0.9093	0	(42)	0.0092	1	U	K.LCVIPR.S
265	463.2970	924.5794	925.4877	-0.9083	0	(35)	0.089	1	U	K.LCVIPR.S
511	463.2972	924.5798	925.4877	-0.9079	0	(42)	0.0089	1	U	K.LCVIPR.S
512	463.3195	924.6244	925.4877	-0.8633	0	(42)	0.008	1	U	K.LCVIPR.S
513	463.3241	924.6336	925.4877	-0.8541	0	(42)	0.0088	1	U	K.LCVIPR.S
514	463.3293	924.6440	925.4877	-0.8437	0	(35)	0.044	1	U	K.LCVIPR.S
515	463.3308	924.6470	925.4877	-0.8407	0	(37)	0.022	1	U	K.LCVIPR.S
516	925.7700	924.7627	925.4877	-0.7250	0	(21)	2.3	1	U	K.LCVIPR.S
552	463.8656	925.7166	925.4877	0.2289	0	(47)	0.003	1	U	K.LCVIPR.S
280	463.8900	925.7654	925.4877	0.2777	0	(26)	0.83	1	U	K.LCVIPR.S
283	464.0361	926.0576	925.4877	0.5699	0	(36)	0.061	1	U	K.LCVIPR.S
613	417.1686	1248.4840	1249.6019	-1.1180	0	(14)	6.4	8	U	K.LCVIPR.S + Clozapine (C)
64	417.4200	1249.2382	1249.6019	-0.3638	0	(8)	30	6	U	K.LCVIPR.S + Clozapine (C)
69	417.5100	1249.5082	1249.6019	-0.0938	0	(15)	7.7	4	U	K.LCVIPR.S + Clozapine (C)
70	417.5127	1249.5163	1249.6019	-0.0857	0	(15)	7.9	4	U	K.LCVIPR.S + Clozapine (C)
623	417.5310	1249.5712	1249.6019	-0.0308	0	(17)	3	5	U	K.LCVIPR.S + Clozapine (C)
625	417.5388	1249.5946	1249.6019	-0.0074	0	(14)	5.8	6	U	K.LCVIPR.S + Clozapine (C)
626	625.8116	1249.6086	1249.6019	0.0067	0	(12)	8.7	3	U	K.LCVIPR.S + Clozapine (C)
71	417.5460	1249.6162	1249.6019	0.0142	0	(10)	23	6	U	K.LCVIPR.S + Clozapine (C)
77	417.6300	1249.8682	1249.6019	0.2662	0	(5)	63	8	U	K.LCVIPR.S + Clozapine (C)
636	417.9293	1250.7661	1249.6019	1.1641	0	(14)	5.2	7	U	K.LCVIPR.S + Clozapine (C)

Figure 73. A view of a selection of ions detected and matched. The drug modified matches are not considered the best match for the data (they are not highlighted in red). This is a mistake caused by Mascot wrongly matching these ions to another peptide fragment not present in the sample.



Monoisotopic mass of neutral peptide Mr(calc): 1249.6019
Fixed modifications: Biotin (N-term) (apply to specified residues or termini only)
Variable modifications:
C2 : Clozapine (C)
Ions Score: 17 **Expect:** 3
Matches : 10/40 fragment ions using 18 most intense peaks (help)

#	a	a^{++}	b	b^{++}	Seq.	y	y^{++}	y^{*}	y^{*++}	#
1	***312.1740***	156.5906	340.1689	170.5881	L					6
2	***739.2974***	370.1523	***767.2923***	384.1498	C	911.4476	456.2274	894.4210	447.7141	5
3	***838.3658***	419.6865	***866.3607***	433.6840	V	***484.3242***	242.6657	467.2976	234.1525	4
4	951.4499	476.2286	***979.4448***	490.2260	I	385.2558	193.1315	368.2292	184.6183	3
5	1048.5026	524.7550	1076.4975	538.7524	P	***272.1717***	136.5895	255.1452	128.0762	2
6					R	***175.1190***	88.0631	158.0924	79.5498	1

Figure 74. Ion 623. -.LCVIPR.- + Clozapine (C). Several adduct bearing fragments are present in the data.

3.4.6.1.3 Peptide 3

Mascot Score Histogram

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 34 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.

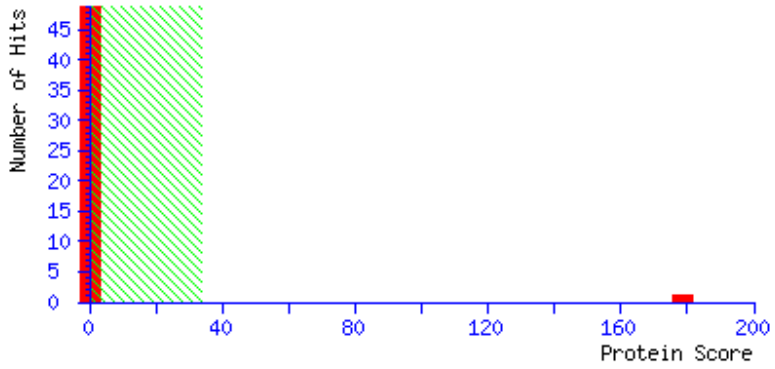
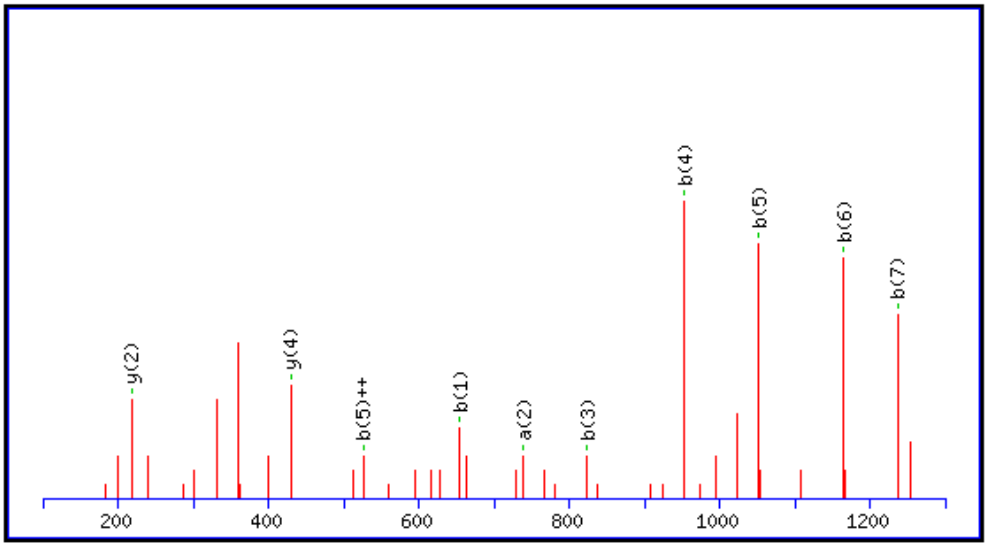


Figure 75. HLM P3 Mascot MOWSE score.

1. P05177 **Mass:** 58483 **Score:** 179 **Matches:** 12(8) **Sequences:** 1(1)
Cytochrome P450 1A2 OS=Homo sapiens GN=CYP1A2 PE=1 SV=3
☐ Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
☑ 22	530.0225	1058.0304	1057.5300	0.5004	0	(49)	0.0016	1	U	R.CIGEVLAK.W
☑ 23	530.1245	1058.2344	1057.5300	0.7044	0	(48)	0.0019	1	U	R.CIGEVLAK.W
☑ 24	530.1750	1058.3354	1057.5300	0.8054	0	58	0.00028	1	U	R.CIGEVLAK.W
☑ 37	691.8001	1381.5856	1381.6442	-0.0585	0	(35)	0.049	1	U	R.CIGEVLAK.W + Clozapine (C)
☑ 38	691.8230	1381.6314	1381.6442	-0.0127	0	(26)	0.43	1	U	R.CIGEVLAK.W + Clozapine (C)
☑ 39	691.8275	1381.6404	1381.6442	-0.0037	0	(47)	0.0034	1	U	R.CIGEVLAK.W + Clozapine (C)
☑ 40	691.8363	1381.6580	1381.6442	0.0139	0	(32)	0.093	1	U	R.CIGEVLAK.W + Clozapine (C)
☑ 41	691.8587	1381.7028	1381.6442	0.0587	0	(34)	0.064	1	U	R.CIGEVLAK.W + Clozapine (C)
☑ 42	691.8751	1381.7356	1381.6442	0.0915	0	(46)	0.0034	1	U	R.CIGEVLAK.W + Clozapine (C)
☑ 43	691.8868	1381.7590	1381.6442	0.1149	0	(32)	0.095	1	U	R.CIGEVLAK.W + Clozapine (C)
☑ 6	691.9029	1381.7912	1381.6442	0.1471	0	(47)	0.0046	1	U	R.CIGEVLAK.W + Clozapine (C)
☑ 44	691.9154	1381.8162	1381.6442	0.1721	0	(46)	0.0034	1	U	R.CIGEVLAK.W + Clozapine (C)

Figure 76. Ions matched to peptide fragments. A selection of modified and unmodified peptide is visible. These ion scores indicate very high quality spectra.



Monoisotopic mass of neutral peptide Mr(calc): 1381.6442
Fixed modifications: Biotin (N-term) (apply to specified residues or termini only)
Variable modifications:
C1 : Clozapine (C)
Ions Score: 47 **Expect:** 0.0034
Matches : 10/56 fragment ions using 11 most intense peaks (help)

#	a	a^{++}	b	b^{++}	Seq.	y	y^{++}	y^*	y^{*++}	#
1	626.2133	313.6103	***654.2082***	327.6078	C					8
2	739.2974	370.1523	767.2923	384.1498	I	729.4505	365.2289	712.4240	356.7156	7
3	796.3188	398.6631	***824.3138***	412.6605	G	616.3665	308.6869	599.3399	300.1736	6
4	925.3614	463.1844	***953.3564***	477.1818	E	559.3450	280.1761	542.3184	271.6629	5
5	1024.4299	512.7186	***1052.4248***	526.7160	V	**430.3024**	215.6548	413.2758	207.1416	4
6	1137.5139	569.2606	***1165.5088***	583.2581	L	331.2340	166.1206	314.2074	157.6074	3
7	1208.5510	604.7792	***1236.5459***	618.7766	A	**218.1499**	109.5786	201.1234	101.0653	2
8					K	147.1128	74.0600	130.0863	65.5468	1

Figure 77. Ion 39. -.CIGEVLAK.- + Clozapine (C). The data contains several fragments bearing the clozapine modification.

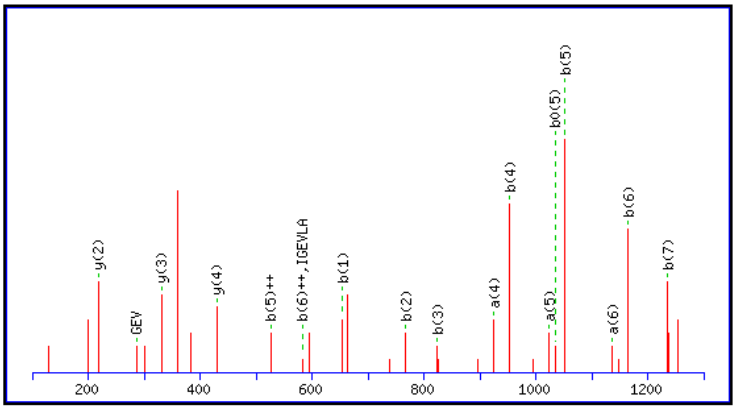
1. in|PAG003|PAG003 **Mass:** 1058 **Score:** 187 **Queries matched:** 16

- HLM_P3

Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
18	530.0225	1058.0304	1057.5300	0.5004	0	(55)	0.0012	1	-.CIGEVLAK.-
19	530.1245	1058.2344	1057.5300	0.7044	0	(57)	0.00094	1	-.CIGEVLAK.-
20	530.1750	1058.3354	1057.5300	0.8054	0	58	0.0009	1	-.CIGEVLAK.-
21	535.7323	1069.4500	1069.4281	0.0220	0	(21)	4.9	1	-.CIGEV.L + Clozapine (C)
22	535.7339	1069.4532	1069.4281	0.0252	0	24	2.7	2	-.CIGEV.L + Clozapine (C)
23	535.7389	1069.4632	1069.4281	0.0352	0	(20)	7	2	-.CIGEV.L + Clozapine (C)
32	627.7834	1253.5522	1253.5492	0.0030	0	40	0.051	1	-.CIGEVLA.K + Clozapine (C)
33	691.8001	1381.5856	1381.6442	-0.0585	0	(35)	0.17	1	-.CIGEVLAK.- + Clozapine (C)
34	691.8230	1381.6314	1381.6442	-0.0127	0	(28)	0.88	1	-.CIGEVLAK.- + Clozapine (C)
35	691.8275	1381.6404	1381.6442	-0.0037	0	(47)	0.012	1	-.CIGEVLAK.- + Clozapine (C)
36	691.8363	1381.6580	1381.6442	0.0139	0	(32)	0.33	1	-.CIGEVLAK.- + Clozapine (C)
37	691.8587	1381.7028	1381.6442	0.0587	0	(36)	0.14	1	-.CIGEVLAK.- + Clozapine (C)
38	691.8751	1381.7356	1381.6442	0.0915	0	(49)	0.007	1	-.CIGEVLAK.- + Clozapine (C)
39	691.8868	1381.7590	1381.6442	0.1149	0	(32)	0.32	1	-.CIGEVLAK.- + Clozapine (C)
40	691.9029	1381.7912	1381.6442	0.1471	0	(51)	0.0037	1	-.CIGEVLAK.- + Clozapine (C)
41	691.9154	1381.8162	1381.6442	0.1721	0	(46)	0.01	1	-.CIGEVLAK.- + Clozapine (C)

Figure 78. HLM P3 Mascot protein hits. A truncated form of the peptide is detected.



Monoisotopic mass of neutral peptide Mr(calc): 1381.6442
Fixed modifications: Biotin (N-term)
Variable modifications:
C1 : Clozapine (C)
Ions Score: 51 **Expect:** 0.0037
Matches (Bold Red): 18/85 fragment ions using 25 most intense peaks

#	a	a^{++}	b	b^{++}	b^{0}	b^{0++}	Seq.	y	y^{++}	y^{*}	y^{*++}	y^{0}	y^{0++}	#
1	626.2133	313.6103	**654.2082**	327.6078			C							8
2	739.2974	370.1523	**767.2923**	384.1498			I	729.4505	365.2289	712.4240	356.7156	711.4400	356.2236	7
3	796.3189	398.6631	**824.3138**	412.6605			G	616.3665	308.6869	599.3399	300.1736	598.3559	299.6816	6
4	**925.3614**	463.1844	**953.3564**	477.1818	935.3458	468.1765	E	559.3450	280.1761	542.3184	271.6629	541.3344	271.1709	5
5	**1024.4299**	512.7186	**1052.4248**	**526.7160**	**1034.4142**	517.7107	V	**430.3024**	215.6548	413.2758	207.1416			4
6	**1137.5139**	569.2606	**1165.5088**	**583.2581**	1147.4983	574.2528	L	**331.2340**	166.1206	314.2074	157.6074			3
7	1208.5510	604.7792	**1236.5460**	618.7766	1218.5354	609.7713	A	**218.1499**	109.5786	201.1234	101.0653			2
8							K	147.1128	74.0600	130.0863	65.5468			1

Figure 79. Ion 40. -.CIGEVLAK.- + Clozapine (C).

3.4.7 DTT and Iodoacetamide Treated Human Liver Microsome Peptide 3

The next step in the project will be the detection of modified proteins the identification of which will be carried out using a bottom up shotgun proteomics approach. This type of approach necessitates an enzymatic digestion step. Tryptic digestion is a typical part of many proteomics workflows; in order to maximise digestion efficiency it is necessary to disrupt protein tertiary structure and to alkylate cysteine residues to prevent disulfide bridge formation. Reduction of already present disulfide bridges is carried out using dithiothreitol (DTT) and can potentially reduce metabolite bearing cysteine residues.

In order to examine this, a sample containing modified P3 was subjected to the standard reduction and alkylation protocol used when carrying out tryptic digestion. Results indicate that the treatment does not prevent the detection of peptide adducts (Figure 80). It is possible however that any change in the level of modification may result in peptides falling out with the limit of detection. It would be appropriate that any further investigation to include experiments to determine i) the absolute level of modification and ii) any quantative change in modification caused by reduction and alkylation.

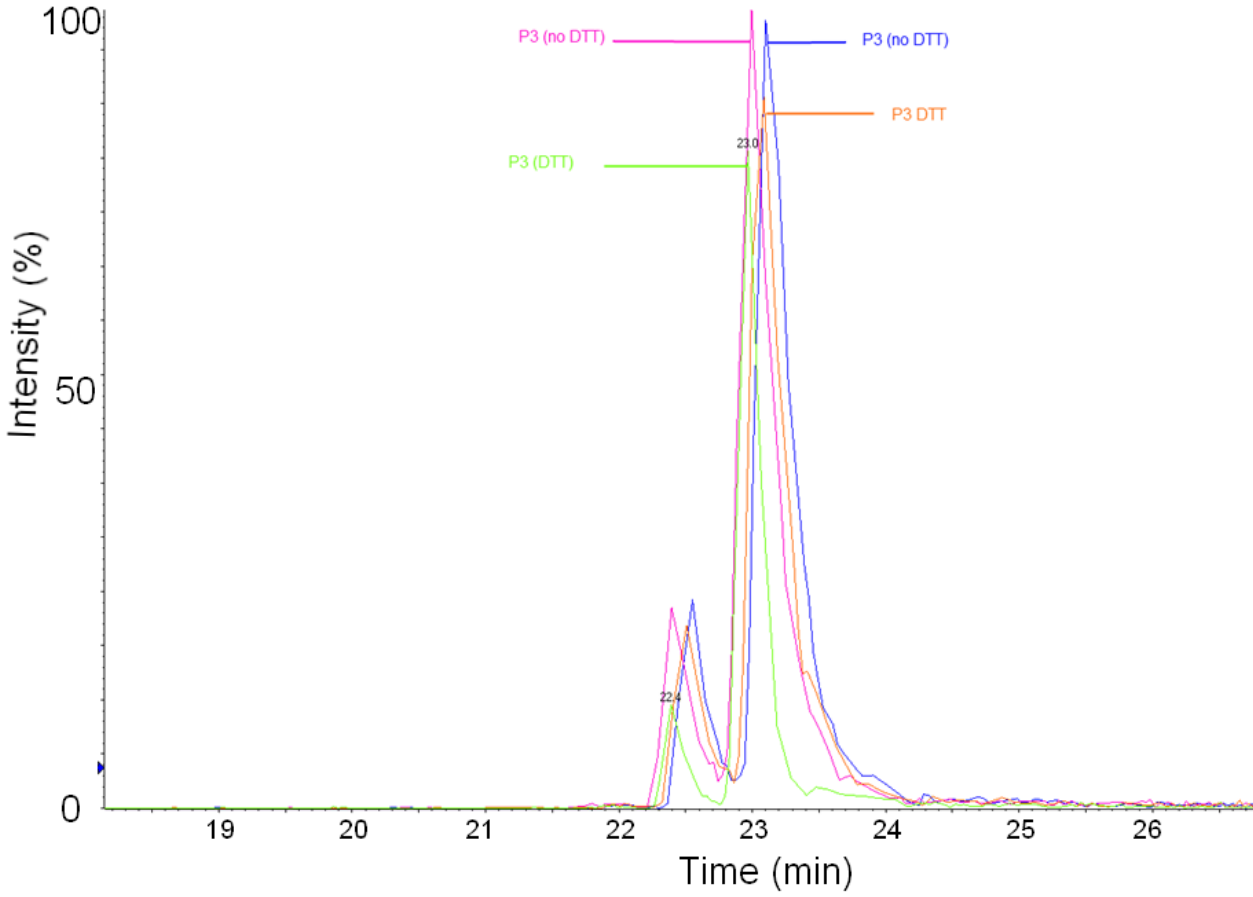


Figure 80. P3 exposed to clozapine in a human liver microsome assay followed by DTT and iodoacetamide treatment. The XIC of 461.5-462.5 is representative of the triply charged adduct of bCIGEVLAK. The DTT and iodoacetamide treated samples (Orange and Green) show marginally less intensity than the samples that were not exposed to DTT and iodoacetamide (Pink and Blue).

1. in|PAG003|PAG003 **Mass:** 1058 **Score:** 797 **Queries matched:** 70 **emPAI:** 2645.96
- HLM_P3

Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
9	459.0066	915.9986	915.4194	0.5792	0	(22)	11	1	-.CIGEVL.A + Carbamidomethyl (C)
106	478.2155	954.4164	954.5208	-0.1044	0	8	1e+02	1	C.IGEVLAK.-
133	529.3175	1056.6204	1057.5300	-0.9096	0	(5)	2.2e+02	7	-.CIGEVLAK.-
134	529.7057	1057.3968	1057.5300	-0.1332	0	(24)	2.8	1	-.CIGEVLAK.-
135	529.7409	1057.4672	1057.5300	-0.0628	0	(29)	0.92	1	-.CIGEVLAK.-
138	529.7700	1057.5254	1057.5300	-0.0046	0	(50)	0.0069	1	-.CIGEVLAK.-
139	529.8455	1057.6764	1057.5300	0.1464	0	(30)	0.59	1	-.CIGEVLAK.-
140	529.8743	1057.7340	1057.5300	0.2040	0	60	0.00055	1	-.CIGEVLAK.-
142	535.7220	1069.4294	1069.4281	0.0014	0	20	7.2	1	-.CIGEV.L + Clozapine (C)
143	535.7745	1069.5344	1069.4281	0.1064	0	(19)	8.9	1	-.CIGEV.L + Clozapine (C)
149	558.2436	1114.4726	1114.5515	-0.0788	0	(38)	0.13	1	-.CIGEVLAK.- + Carbamidomethyl (C)
150	558.2444	1114.4742	1114.5515	-0.0772	0	(43)	0.043	1	-.CIGEVLAK.- + Carbamidomethyl (C)
151	558.2635	1114.5124	1114.5515	-0.0390	0	(40)	0.079	1	-.CIGEVLAK.- + Carbamidomethyl (C)
152	558.2644	1114.5142	1114.5515	-0.0372	0	(38)	0.13	1	-.CIGEVLAK.- + Carbamidomethyl (C)
153	558.2691	1114.5236	1114.5515	-0.0278	0	(51)	0.006	1	-.CIGEVLAK.- + Carbamidomethyl (C)
154	558.2727	1114.5308	1114.5515	-0.0206	0	(34)	0.32	1	-.CIGEVLAK.- + Carbamidomethyl (C)
155	558.2732	1114.5318	1114.5515	-0.0196	0	(42)	0.053	1	-.CIGEVLAK.- + Carbamidomethyl (C)
156	558.2799	1114.5452	1114.5515	-0.0062	0	(49)	0.0096	1	-.CIGEVLAK.- + Carbamidomethyl (C)
157	558.2814	1114.5482	1114.5515	-0.0032	0	(49)	0.0086	1	-.CIGEVLAK.- + Carbamidomethyl (C)
158	558.2842	1114.5538	1114.5515	0.0024	0	(44)	0.031	1	-.CIGEVLAK.- + Carbamidomethyl (C)
159	558.2853	1114.5560	1114.5515	0.0046	0	(48)	0.012	1	-.CIGEVLAK.- + Carbamidomethyl (C)
160	558.2879	1114.5612	1114.5515	0.0098	0	(25)	2.2	1	-.CIGEVLAK.- + Carbamidomethyl (C)
161	558.2897	1114.5648	1114.5515	0.0134	0	(25)	2.1	1	-.CIGEVLAK.- + Carbamidomethyl (C)
162	558.2898	1114.5650	1114.5515	0.0136	0	(33)	0.41	1	-.CIGEVLAK.- + Carbamidomethyl (C)
163	558.2905	1114.5664	1114.5515	0.0150	0	(38)	0.12	1	-.CIGEVLAK.- + Carbamidomethyl (C)
164	558.2942	1114.5738	1114.5515	0.0224	0	(38)	0.11	1	-.CIGEVLAK.- + Carbamidomethyl (C)
165	558.2943	1114.5740	1114.5515	0.0226	0	(45)	0.023	1	-.CIGEVLAK.- + Carbamidomethyl (C)
166	558.2947	1114.5748	1114.5515	0.0234	0	(38)	0.11	1	-.CIGEVLAK.- + Carbamidomethyl (C)
167	558.3025	1114.5904	1114.5515	0.0390	0	(24)	3.1	1	-.CIGEVLAK.- + Carbamidomethyl (C)
168	558.3033	1114.5920	1114.5515	0.0406	0	(34)	0.3	1	-.CIGEVLAK.- + Carbamidomethyl (C)
169	558.3036	1114.5926	1114.5515	0.0412	0	(46)	0.018	1	-.CIGEVLAK.- + Carbamidomethyl (C)
170	558.3061	1114.5976	1114.5515	0.0462	0	(58)	0.0011	1	-.CIGEVLAK.- + Carbamidomethyl (C)
171	558.3066	1114.5986	1114.5515	0.0472	0	(34)	0.28	1	-.CIGEVLAK.- + Carbamidomethyl (C)
172	558.3080	1114.6014	1114.5515	0.0500	0	(38)	0.13	1	-.CIGEVLAK.- + Carbamidomethyl (C)
173	558.3151	1114.6156	1114.5515	0.0642	0	(49)	0.0097	1	-.CIGEVLAK.- + Carbamidomethyl (C)
174	558.3183	1114.6220	1114.5515	0.0706	0	(47)	0.013	1	-.CIGEVLAK.- + Carbamidomethyl (C)
175	558.3222	1114.6298	1114.5515	0.0784	0	(49)	0.009	1	-.CIGEVLAK.- + Carbamidomethyl (C)
228	592.3094	1182.6042	1182.5121	0.0921	0	25	2.4	1	-.CIGEVL.A + Clozapine (C)
257	461.6156	1381.8250	1381.6442	0.1808	0	(30)	0.53	1	-.CIGEVLAK.- + Clozapine (C)
258	461.6240	1381.8502	1381.6442	0.2060	0	(35)	0.16	1	-.CIGEVLAK.- + Clozapine (C)
259	691.9345	1381.8544	1381.6442	0.2103	0	(28)	0.94	1	-.CIGEVLAK.- + Clozapine (C)
260	691.9430	1381.8714	1381.6442	0.2273	0	(31)	0.39	1	-.CIGEVLAK.- + Clozapine (C)
261	691.9436	1381.8726	1381.6442	0.2285	0	(28)	0.87	1	-.CIGEVLAK.- + Clozapine (C)
262	691.9612	1381.9078	1381.6442	0.2637	0	(46)	0.012	1	-.CIGEVLAK.- + Clozapine (C)
20	461.6524	1381.9354	1381.6442	0.2912	0	(43)	0.07	1	-.CIGEVLAK.- + Clozapine (C)
263	691.9776	1381.9406	1381.6442	0.2965	0	(24)	2.2	1	-.CIGEVLAK.- + Clozapine (C)
264	691.9853	1381.9560	1381.6442	0.3119	0	(43)	0.025	1	-.CIGEVLAK.- + Clozapine (C)
265	461.6846	1382.0320	1381.6442	0.3878	0	(35)	0.14	1	-.CIGEVLAK.- + Clozapine (C)

Figure 81. DTT and Iodoacetamide treated HLM P3.

3.5 Discussion

The detection and characterisation of drug metabolites has been the subject of much study and numerous reviews (Zhang et al., 2011; Ma et al., 2006; Prakash et al., 2007; Kostiainen et al., 2003; Holcapek et al., 2008), many approaches involved the use of the physiologically abundant tripeptide glutathione (Baillie and Davis, 1993; Dieckhaus et al., 2005, Zheng et al., 2007; Mutlib et al., 2005; Zhu et al., 2007; Gan et al., 2005). The formation of glutathione metabolite adducts is an efficient and directed mechanism catalysed by the enzyme glutathione transferase which serves to deprotonated reduced glutathione molecules thereby rendering them reactive (Atkins et al., 1993). The highly nucleophilic sulfhydral group makes glutathione a particularly effective trapping agent for the so called soft electrophiles; this class of drugs includes epoxides, quinone imines, quinone methides, quinones and imine methides as well as others (Tang and Lu, 2010; Ma and Subramanian, 2006). In addition to these soft electrophiles glutathione has been shown to form adducts with hard electrophiles which include nitrenium ions and carboxylic acids (Sidenius et al., 2004).

The drugs chosen in this study were metabolised to form both hard and soft electrophiles which were trapped with glutathione and characterised by mass spectrometry. Multiple metabolites were identified for each of the drugs tested, a more complete characterisation could have been carried out by including a hard nucleophilic target such as cyanide anion (Gorrod et al., 1991; Argoti et al., 2005) or by using the more sensitive glutathione monoethyl ester (Wen and Fitch, 2008) as trapping agents. The scope of the study was to identify proteins that form covalent adducts with proteins and as such glutathione was thought to be a more appropriate initial model system.

The detection of modified synthetic peptides showed that the metabolites of clozapine were capable of forming conjugates with polypeptides without the intervention of the GST enzyme. Both glutathione and synthetic peptide conjugates were detected using the precursor ion, neutral loss and basic IDA type scans. The precursor ion scan was deliberately designed to be drug specific, the precursor ion was shown to be produced both by glutathione trapped

metabolites and synthetic peptide adducts. This should allow for the detection of previously uncharacterised adducts. i.e. modified proteins.

The relatively simple chemical composition of the metabolites limit the degree of selectivity that they can show. A similar principle applies to both glutathione and the synthetic peptides, as both short polypeptides are not complex enough to exhibit either secondary or tertiary structures. The interaction of metabolites and these short polypeptides would most likely proceed as simple stochiometric chemical reactions. The interaction between metabolite and a fully formed protein with secondary, tertiary and potentially quaternary structure would presumably involve a more complex dynamic. It is well known that the interaction between enzyme and substrate is focused at the so called active site; typically the active site operates based on a highly specific physic-chemical interaction. Proteins are most likely attacked at sites with favourable physico-chemical characteristics; important factors include the solvent accessibility of the cysteine residue and a favourable local pKA. As previously mentioned enzymes are often targets of reactive metabolites, particularly those enzymes involved in metabolism. The active site can be the site of adduct formation when particularly reactive species are produced, leading to so called mechanism based inactivation of said enzyme (Massey et al., 1970; Almira et al., 2005).

The pKA of a given cysteine sidechain influences its reactivity. The cysteine at the active site of the enzyme protein tyrosine phosphatase (PTP) was observed to have a pKa of 4.67, considerably lower than the 8.5 of a typically cysteine residue (Lundblad and Noyes, 1984). The stability of the thiolate anion at this lowered pH was shown to be dependent on neighbouring peptides (Zhang and Dixon, 1993), in this case a histidine residue and an arginine residue. Point mutations in either of these two flanking residues to non-basic alternatives lead to a notable increase in the pKa of the cysteine residue.

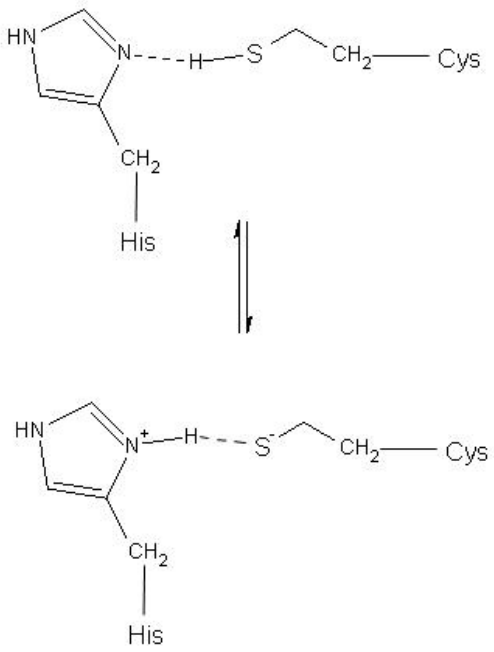


Figure 82. Proposed mechanism for the stabilisation of the thiolate anion (cysteine) by a neighbouring imidazole ring (histidine) (Zhang and Dixon, 1993). The interaction involves the formation of a zwitterionic ion pair. This relationship leads to a drastic lowering of the apparent pKa of the thiol.

In addition to the lowered pKa, the localised area of positive charge will affect the interaction of electrophiles via electrostatic interactions. These chemical and physical forces presumably come into play when considering the interaction of any given electrophile and protein. Obviously the pH of the environment and by extension the localisation of both protein and electrophile within a cell or organ will also affect the probability of adduct formation. Work carried out by Fisher et al managed to identify the presence of a putative electrophile binding motif in many proteins known to form adducts by nucleophilic/electrophilic attack (Fisher et al., 2007). Proteins with a higher than normal lysine content were reportedly at higher risk of adduct formation, in particular the motif KxK, KKx or xKK where K is lysine and x represents a nucleophilic amino acid was shown to further enhance the risk of adduct formation (Labenski et al., 2009).

The motifs described represent local interactions based on protein primary structure, similar effects could feasibly be produced through higher order structure in native proteins. Identification of such proteins is much more

difficult, it should be possible to use a combination of previously determined structural conformation and homology modelling (Pitman and Menz, 2006). However direct structural information from NMR or x-ray crystallography would be required for confirmation.

The primary structure of the synthetic peptides did not carry any such motifs but were still subjected to modification. These short linear peptide sequences do not possess the secondary or tertiary structure of native proteins and are not subjected to the same accessibility problems. In this respect the limitation of a synthetic peptide model becomes apparent.

Mascot assigns scores to peptides based on the probability that the match could occur at random when a search is carried out against a particular database; the higher the score the lower the match was made by chance. Peptides are grouped according to whichever protein they belong and protein score is calculated based on the contributions of its assigned peptides. A protein's score is the sum of its peptide scores, the highest scoring peptide is chosen in the case of duplicate ions, with a small correction to account for the contribution of multiple low scoring ions. A significance level is defined, p=0.05 by default, and an equivalent MOWSE score threshold is computed (Mascot is discussed in much greater detail in section 1.3.3.2). Proteins with a score greater than the threshold are reported as statistically significant matches ($p>0.05$). Peptides that do not contribute to proteins with a score exceeding the threshold are reported as unassigned. As the score for a protein is basically the sum total of its peptide scores this means that peptides that are sole matches for a given protein become statistically less significant. Often the spectra associated with these peptides is of poor quality i.e. there is high background noise or uncharacteristic ions comprise most of the peaks; in a CID based tandem mass spectrum one does not typical expect to find c or z ions which are typically associated with fragmentation by ETD. However, it can be the case that even relatively high scoring peptides, which may come close to passing the significance threshold alone, are included in a list of unmatched peptides at the end of the mascot report. These peptides, upon manual inspection, may derive from a high quality spectrum with several of the b and y ion series identified. To ensure completeness any of these unassigned ions that potentially contained a drug adduct (as identified by mascot) were manually examined and the quality of the data assessed. It was found that none

of the potentially modified peptides were of sufficient quality for further investigation.

It was noted that peptides 1 and 3 had undergone fragmentation prior to MS analysis. Figures 52 and 56 show that there is a difference of 1.5 min in the elution HPLC elution of intact peptide 1 and the fragmented form. Mascot search data shown in figures 70 and 71 initially called attention to this anomaly. Similar evidence of this was found in the experiments with peptide 3 (figures 65 and 78). It is possible that the fragmentation is a product of the drug derivatisation of the peptides. The clozapine adduct contains a diazepine ring which acts as a strong base and may be responsible for intramolecularly catalysing hydrolysis of the peptide bond. The fragmentation is not seen in peptide 2 suggesting that the structure of the peptide plays a role in this fragmentation process and the peptide 2 forms a more stable adduct. A review of literature uncovered little to shed light on this phenomenon and further work would need to be carried out in order to determine the underlying processes.

DTT based reduction and Iodoacetamide based alkylation of modified synthetic peptides did not have a pronounced effect on the conjugation of metabolite and peptide. These experiments verify that these treatments, critical for efficient and effective tryptic digestion of proteins, can be carried out without undue loss of protein-drug adducts. This is important for the next stage in testing that relies on this method of digestion for a bottom up shotgun proteomics based search of the human liver microsome for protein-drug adducts.

For the protein based work to follow the latest model of the Q-trap (API 5500™) will be used, this instrument has a sensitivity at least an order of magnitude greater than the API 4000™ model owing to enhanced ion optics and greatly improved linear ion trap. The increased performance does however come at the cost of a reduced m/z range. Predominantly designed for metabolomics based workflows the 5500 has a m/z maximum of 1000. This will effectively narrow the mass range of detectable peptides.

It was demonstrated that the precursor ion scan at 359 m/z can be used to detected both clozapine-glutathione adducts and clozapine-synthetic peptide adducts. Reduction and alkylation did not markedly reduce the levels of adduct formation and using the Mascot search engine it was possible to automatically all

three clozapine-synthetic peptide adducts. All three adducts were correctly matched with their parent proteins indicating that even with limited sequence information it is possible to get an accurate protein identification.

Chapter 4: Protein Separations

4.1 Aims

The liver microsome fraction comprises a huge number of proteins (Peng et al., 2012; Huang et al 2011). Proteolytic digestion of such a sample further increases complexity and thereby increases the challenge of identifying individual molecules. In order to maximise protein identifications, sequence coverage and the identification of post translational modifications it is necessary to separate the samples prior to mass spectrometric (MS) analysis. Three types of separation methodologies, orthogonal in nature to reversed phase liquid chromatography, were employed:

1) Protein separation by 1d PAGE followed by tryptic digestion of small sections of the protein ladder.
2) Separation of peptides, produced by the tryptic digestion of the liver microsome fraction, using the Offgel isoelectric focusing system.

3) Ion exchange (IEX) liquid chromatographic separation of peptides.

Subsequent to these separation methods all samples were subjected to RP-LCMS and Mascot searches against the Swissprot protein database.

4.2 Introduction

The usefulness of reversed phase LC-MS for analysis of highly complex samples is limited by the peak capacity of a given chromatographic column (Giddings, 1967). Increasing the length of a single gradient run can improve the protein identifications attainable, peak capacity increases as the solvent gradient becomes more shallow (Liu et al., 2007; Wang et al., 2006). Instrument availability sets limits on the maximum length of separation gradients that can be applied; overly long gradients result in a decrease in detection sensitivity due to a widening of chromatographic peaks. Analytes elute from the column over a particular range of physical conditions based on interactions between the analytes, the mobile phase and the stationary phase, in the case of RP separation, hydrophobic/hydrophilic interactions. A particularly shallow gradient leads to an increase in the time at which conditions favour analyte elution i.e. the hydrophobic conditions of the mobile phase are suitable for a longer period. The immediate effect is that analytes elute over a longer period, leading to peak widening and necessarily, a decrease in the concentration of analyte entering the mass spectrometer at any given time. Simply, the analyte elution occurs more gradually with the same total amount being eluted with negation of the concentrating effect normally observed during chromatographic separation.

In order to maximise dynamic range and proteome coverage it is necessary to use orthogonal methods of separation (Issaq et al 2002; Righetti et al., 2003) in so doing peak capacity can be increased dramatically (Giddings, 1987). Approaches including 1d PAGE, ion exchange chromatography and isoelectric focusing are both mature and suitably orthogonal technologies. Spreading the separation over two dimensions maximises the opportunity to detect poorly represented or poorly ionised peptide species otherwise lost using a single separation dimension. That is to say that superior chromatographic separation works to increase the dynamic range of molecules detected by reducing the complexity of the ESI stream.

The combination of ion exchange and RP-LC-MS is known as MuDPIT (Multidimensional protein identification technology) and has the added advantage of being amenable to automation (Bailey et al., 2007; Jiang et al., 2007). HPLC systems such as the UltiMate™ 3000 (Dionex) can be configured to operate both RP and IEX columns in serial, allowing for so called on-line 2D LC-MS analysis (Washburn et al., 2001; Mohammed and Heck, 2011), the nature of the technique necessitates the presence of salt in droplets formed during ESI causing ion suppression and reducing the sensitivity of the mass spectrometric analysis (Annesley, 2003). This is not the case with the offline method in which the eluent from the ion exchange column can be captured in a guard column and washed free of salt prior to RP separation giving superior sensitivity (Peng et al., 2002). Additionally a superior peak capacity is attributed to the offline mode due to its use of a gradient separation in comparison to the step-wise elution of the online method (Wagner et al., 2003). As sensitivity takes precedent over automation at this stage of the project all work was carried out in the offline configuration.

1d-PAGE (1 dimensional polyacrylamide gel electrophoresis) is a workhorse proteomics tool and as such is well characterised and robust. Non-native SDS-PAGE (sodium dodecyl sulphate) allows for the processing of hydrophobic proteins and primarily separates proteins based on their molecular masses (Laemmli, 1970). Migration of proteins through the gel depends on their molecular mass, the applied electric field and the gel matrix density. Within a single gel a concentration gradient can be used to vary the density of the gel along its length. The gradient allows for improved resolution due to the sieving effect created by the decreasing pore size of the gel (Rodbard and Chrambach, 1970).

Proteins are easily digested and recovered from gel bands (Rosenfeld et al., 1992) the resultant solution is then amenable to analysis by LC-MS. This coupling of 1d-PAGE and LC-MS is most commonly described as GeLC-MS (Gel Liquid Chromatography Mass Spectrometry). Using this simple combination it is possible to identify hundreds to thousands of proteins; a group investigating rat pancreatic cells were able to identify some 1350 non-redundant proteins by GeLC-MS (Paulo et al., 2011) using only 10 gel sections per lane.

Offgel fractionation makes recovery of proteins or peptides from the IEF stage simpler by eliminating the need for gel extraction and the associated loss of proteins. Using a 24cm pH 4-7 IPG strip with a 24 well holder it has been shown that a resolution of 0.15 pH is attainable (Michel et al., 2006). The performance of Offgel fractionation is comparable to that of a MuDPIT based approach (Elschenbroich et al., 2009).

The three techniques chosen represent different mechanisms of separation and are likely to give a synergistic overview of the human liver microsome proteome. An increase in the detectable peptide fragments clearly leads to an increase in the likelihood of detecting a metabolite-peptide adduct.

4.3 Methods and Materials

4.3.1 Metabolism of Drugs and Formation of Drug-Protein Adducts

Clozapine was incubated with human liver microsomes as described in (2.1.1.17) without the addition of glutathione. After a 1 hour incubation samples, except for those to be separated by 1d-PAGE, were spun at 4000 rpm for 10 min, the acetonitrile precipitation step was foregone. The pellet fraction was subjected to delipidation (2.1.1.9). The supernatant fraction was stored at -20 °C until tryptic digestion could be carried out.

4.3.2 1d SDS-PAGE

For 1d PAGE analysis 24 µl of the assay solution (12.5 µg of protein) was spiked with 6µl of (5x) SDS-PAGE loading buffer and loaded into a 12 cm gradient polyacrylamide gel. The gel was submerged in SDS-PAGE running buffer and run for 35 min at 35 V and 120 mA.

4.3.3 In solution tryptic digestion of proteins

Proteins were digested in solution as described in (2.1.1.3) prior to separation by either IEX liquid chromatography or the Offgel system. Briefly, samples were reduced using DTT and alkylated using Iodoacetamide in order to break down disulfide bonds and maximise digestion. It has been shown (3.4.7) that clozapine metabolite binding to peptides was not reversed by this particular treatment.

4.3.4 In Gel Tryptic Digestion of Proteins

Each gel lane was cut into 12 equally sized sections along their lengths. Each of these sections was chopped into smaller pieces before undergoing washing, tryptic digestion and recovery. For details see (2.1.1.10). Samples were stored at -20 ˚C until LCMS was carried out.

4.3.5 Offgel Separation of Peptides

After tryptic digestion samples were reconstituted in Offgel buffer and separated by their isoelectric points along a pH gradient gel strip. Recovery of separated peptides from each of the 24 regularly sized reservoirs along the strip consisted of a simple pipetting step. Samples were dried in a vacuum centrifuge and stored at -20 ˚C until analysis by LC-MS. Full details can be found at (2.1.1.6).

4.3.6 Ion Exchange Liquid Chromatography

Peptides from the tryptic digestion step were loaded into an appropriate buffer and injected into an LC system equipped with a strong cation exchange column. An increasing salt gradient was applied over the course of 1 hour and fractions of equal length were collected using an HTC pal robotic fraction collector. For full details see (2.1.1.7).

4.3.7 Reversed Phase Liquid Chromatography

All samples were reconstituted in buffer and loaded into 96 well plates or glass vials for mounting in the UltiMate® 3000 HPLC system and autosampler. Separation was carried out by applying a gradient with increasing concentration of organic solvent (acetonitrile) over a period of 60 min. Full details at (2.1.2.2).

4.3.8 Mass Spectrometric Analysis of Peptides

All analysis in this chapter was carried out using an API 5500™ series Q-trap (AB SCIEX) data was acquired using a standardised information dependant acquisition (IDA) approach (2.1.2.3) and a more selective precursor ion of 359 (PI359) based approach (2.1.2.5).

4.3.9 Identification of Peptides Modified by Clozapine Metabolites

The five known clozapine metabolites were added to the Mascot database. Data from each of the files obtained from both the IDA and PI359 scans of all samples, including the negative controls, were uploaded to the Mascot server and searched against the Swissprot database (The version of Swissprot used was not noted; searching was carried out in 2012) against the human taxon Swissprot was used at it is a high quality, manually curated non-redundant database. No fixed modifications were selected, oxidation of methionine, carbamidomethylation of cysteine and the five clozapine metabolite adducts of cysteine were selected as possible variable modifications. False discovery rates were automatically calculated by Mascot and are based on searching the mass spectrometric data against a decoy database in order to quantify the extent of matches. The decoy database was generated to have the same average amino acid composition, and number of proteins (of the same lengths) as those in the target database being searched. The number of hits detected in the target database is compared to the number of hits from the decoy database (assumed to be false positives) in order to give the false discovery rate for the experiment. This process is more fully explained in section 1.5.2.2.

4.3.10 Identification of Membrane Associated Proteins

Protein identifications obtained from the Mascot search were in the UniProtKB/Swiss-Prot format. An exhaustive list for each separation type was compiled and their associated FASTA files recovered from Uniprot (http://www.uniprot.org/). The FASTA files were submitted to the TMHMM server (http://www.cbs.dtu.dk/services/TMHMM-2.0/) for analysis.

4.4 Protein Modification and Separation Techniques

Due to the extreme complexity of the liver microsome sample robust protein/peptide separation techniques orthogonal to reversed phase chromatography were applied. There exist many well established and robust proteomics techniques for the separation of peptides and proteins Issaq et al., 2002; Giddings, 1987; Mitulovic, 2004). Separations based on physical size, isoelectric point, Coulombic interaction and affinity interaction allowed for an in depth exploration of the liver microsome protein complement. Each of the approaches was coupled to RP-LCMS analysis using both IDA and PI359 based methods to determine MS/MS acquisition

4.4.1 LC-MS Analysis of Modified Protein

All samples were submitted to the same reversed phase liquid chromatography under the same conditions (2.1.2.2).

4.4.1.1 LC-MS Analysis 1d Gel Samples

SDS PAGE has the advantage of solubilising membrane associated and otherwise hydrophobic proteins. Separation takes place at the protein level thus ensuring that all digestion products from any particular protein are present within the fraction. This means that subsequent LC-MS analysis has the potential to provide high levels of protein coverage within single samples when compared to techniques that separate proteolytic digestion products.

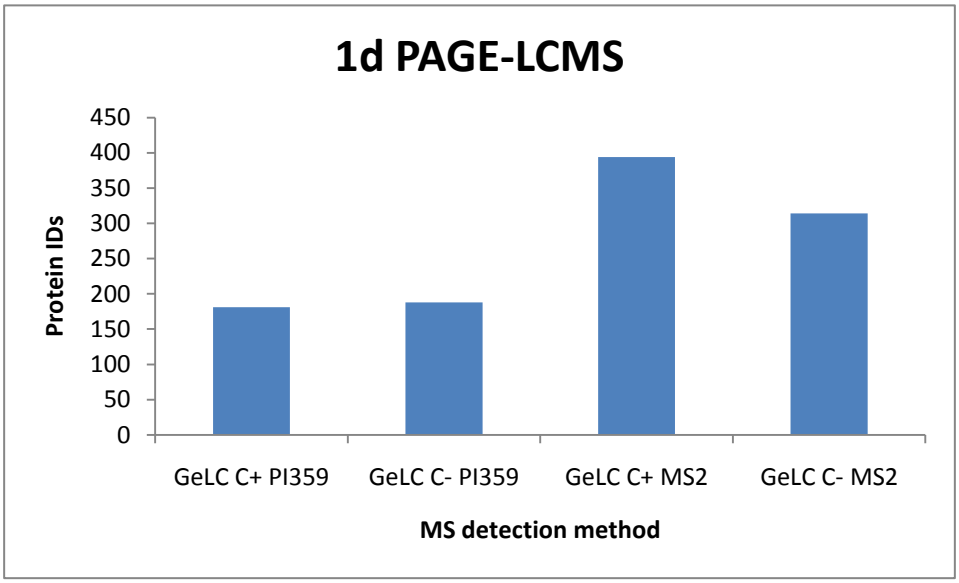


Figure 83. 1d PAGE-LC-MS protein identifications summed across all 12 gel pieces. The false discovery rates (FDR) for all experiments were below 5%.

Detection of proteins by the IDA MS/MS method (2.1.2.3) lead to the identification of 300-400 proteins. The scan is designed to maximise the number of identifications obtained and is selective only to the intensity of any given ion. An exclusion list, written into the program, ensures that once a protein has been subjected to MS/MS it is ignored for 30 seconds. This serves to prevent masking of less abundant ion species, the exclusion time is calculated to ensure that the abundant ion has been completely eluted from the column, the average peak width was measured at about 15-20 seconds.

The PI359 scans have identified between 150-200 proteins, a much higher number than was expected. The high number identified could indicate that the PI359 scan has poor selectivity or simply that the sample complexity is overwhelming. In the second case the low resolution of the quadrupole creates a rather wide window for ion transmission; in order to maximise sensitivity the width was set at 1.2 Da. The downside of this approach is that ions with fragments close enough to the target of 359 are detected and analysed. Additionally, the high complexity of the samples increases the chances that once an ion with a particular nominal mass produces the fragment of interest there will be other ions with similar nominal mass in the ion stream at the same time.

The SDS-PAGE samples should contain more proteins than either of the other two separation methods as even the highly hydrophobic species would be

solubilised and separated. A caveat to this would be that if the separation was poor along the length of the gel then each band may be too heavily loaded with proteins; leading to loss of information caused by ion suppression. After tryptic digestion the peptide fragments of the hydrophobic protein would be more soluble (than the molecule as a whole) and as such would not drop out of solution prior to the RP-LCMS analysis. The gel lanes were each cut into 12 sections, each section having many individual bands, each band many individual proteins. In order to improve separation the gel lanes could be cut into a greater number of pieces. The number of protein species in each of the gel pieces is also unequal. Careful examination of the number of proteins found in each could provide information for optimum cutting of the gel lanes in order to spread the complexity over the different sections.

The cytochrome P450 family of enzymes are of particular interest due to their roles in drug metabolism, proximity to reactive metabolites and accessible cysteine residues (Kyle et al., 2012). CYP450s also represent the presence of membrane associated proteins in the samples, indicating whether or not the sample preparation was effective. The enzymes were detected by both the IDA and PI359 approaches, however the IDA approach managed to identify more than twice as many CYP450s.

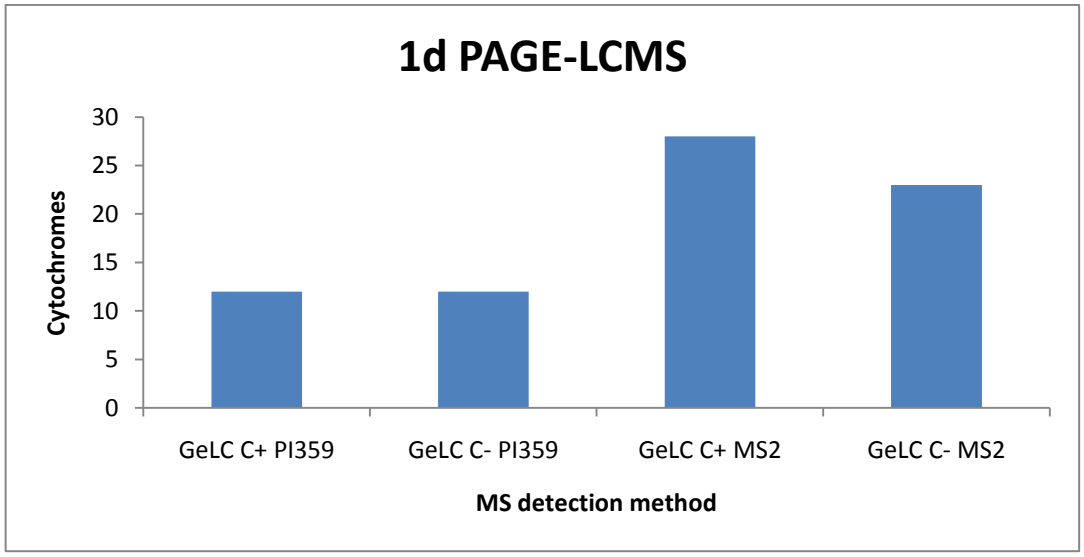


Figure 84. Cytochrome P450 enzymes identified by both the IDA scans and PI359 scans (FDR <5%).

The presence of so many of these membrane integral and associated proteins suggests that the 1d-PAGE approach effectively deals with proteins with hydrophobic domains.

Mascot based searching came back negative for the presence of any metabolite-peptide adducts associated with known proteins. A variety of metabolite-peptide adducts were included at the end of the report amongst the peptides unassigned to proteins; Mascot only reports protein hits that exceed a given MOWSE score threshold, with the total score being a sum of the component peptide scores. Peptides found in this region typically have low ion scores and do not belong to a protein with any other peptide matches. Incomplete digestion of a sample can lead to a greater than expected number of missed cleavages i.e. peptide chains bearing uncut (in this case) tryptic motifs. Physiological or even process-specific post translation modifications of proteins outside of those specified in the Mascot parameters would produce peptides of unpredictable mass. In both cases Mascot would fail to find the true identity of the non-conforming peptide.

Manual inspection of these low scoring unassigned ions with putative clozapine metabolite modifications was carried out. It was found that putative matches were not statistically significant and often included ions (a and z ions) not routinely detected in CID type experiments. As would be expected in the case of statistical artefacts, the IDA MS/MS experiments revealed many more putative matches than did the PI359 experiments. Additionally, a similar range of false positive matches were observed in the negative control samples that did not include any clozapine.

4.4.1.2 LC-MS Analysis of Offgel Samples

The Offgel technique allows for the separation at either the peptide or protein level. Separation is based on isoelectric focusing, as occurs in the 1st dimension of 2D PAGE. Offgel has the advantage of separating samples along the length of an IPG strip whilst having them remain accessible for collection. Fractions are simply pipetted from the tray and can be readily analysed with little further preparation. The technique is not however compatible with less soluble proteins

(Santoni et al., 2000) meaning that in order to analyse the membrane associated species it would first be necessary to digest them to the peptide level.

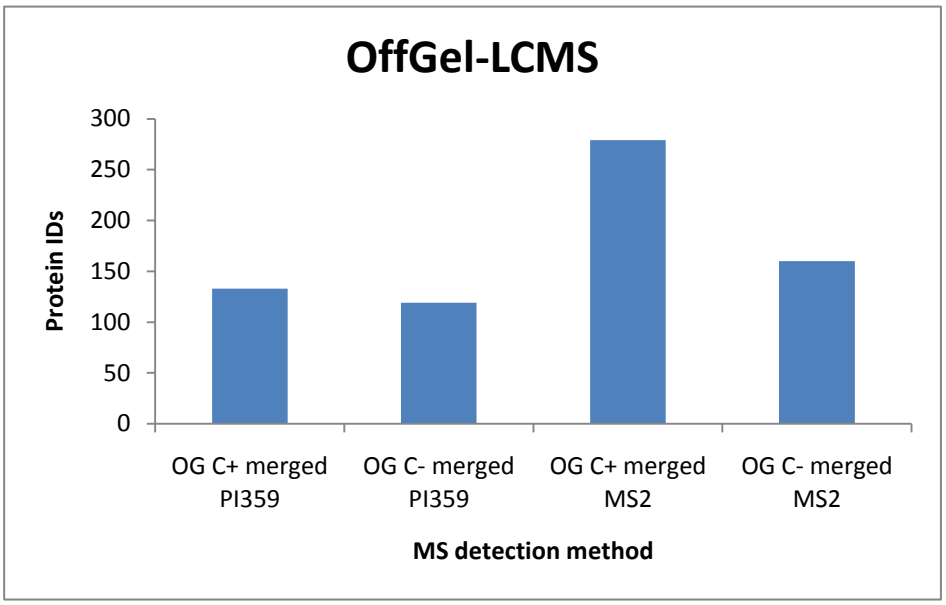


Figure 85. Proteins identified by the Offgel as 1st dimension of separation (FDR <5%).

In order to maximise the presence of lipid associated proteins a delipidation protocol was implemented prior to tryptic digestion (2.1.1.9). This protocol may be responsible for somewhat correcting an oversight in the experimental design, a lack of a denaturing step in the in solution digestion method (further discussed in section 4.5).

Again, the number of proteins identified by the IDA approach is much greater than those identified by the PI359 approach. The Offgel separation afforded more fractions (24) than did the 1d-PAGE method but has yielded considerably fewer protein identifications, particularly by IDA MS/MS. The separation of proteins at the peptide level can considerably decrease the number and certainty of protein identification in individual fractions. This effect is clearly caused by the presence of different tryptic fragments from the same protein being spread across many fractions. As Mascot computes MOWSE scores by summing the scores of peptide assigned to a given protein, the fewer the number of peptides present in a sample the lower the protein scores will be. In order to combat this the mass spectrometric data for each of the fractions in each sample were recombined using the Peak List conversion Tool (Proteomecommons.org IO framework 6.21). The merged data files for each

sample were then submitted to Mascot for searching against the relevant database. This method ensures that proteins receive all of the fragments detected across the IPG strips range and ensures maximum sequence coverage and increases the likelihood of correct identification whilst reducing false matches.

The discrepancy between the number of proteins identified in the clozapine positive and negative samples could easily have been caused by technical variations. Due to the nature of the work and the time required to perform the extensive LCMS analysis it was not possible to continually monitor the performance of the equipment. Variations throughout prolonged runs can occur due to MS related issues, wear on the ESI needle or accumulation of contamination around the ESI orifice or due to some of the less appealing idiosyncrasies related to nanoflow HPLC.

The CYP450s are well represented. Again the IDA scan has identified many more CYP450s than the precursor ion scans. The apparent abundance of the enzymes (around 1:10 proteins) is not surprising considering the sample is a human liver microsomes fraction. The microsomes contain an abundance of endoplasmic reticulum which is the locus of CYP450 activity. The enzymes are there synthesized and cotranslationally inserted into the ER membrane (Negishi et al., 1978).

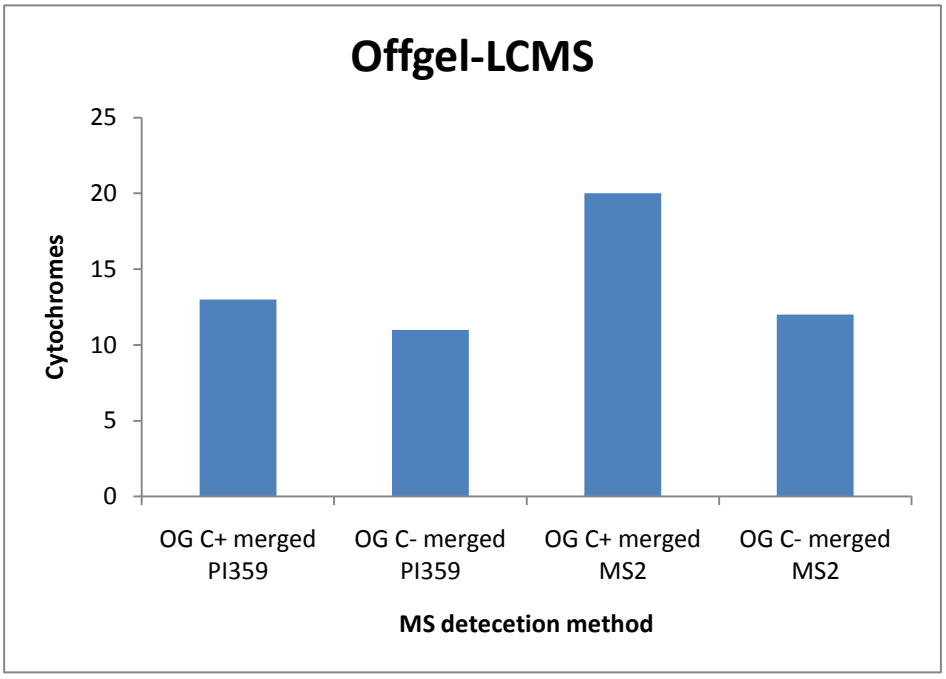


Figure 86. Cytochrome P450 enzymes as detected by Offgel-LC-MS analysis of human liver microsome samples (FDR <5%).

4.4.1.3 LCMS Analysis of IEX Samples

Ion exchange chromatography separates out either proteins or peptides based on Coulombic interactions between a functionalised stationary phase (phosphonic/sulfonic for SCX) acid and charge-bearing regions of the proteins/peptides in a mobile phase (see 1.2.3) (Morris and Morris, 1962; Kopaciewicz et al., 1983). The separation method is readily fine tuned, ideal for automation and interfaces well with reversed phase LC-MS analysis. Another widely used workhouse technique, IEX is well characterised and robust (Masuda et al., 2005).

The complexity of the microsome fraction is evident upon examination of the UV data generated during the SCX separation. Even with a separation gradient of 30 min it is impossible to see sharp individual peaks. The presence of many such peaks is seen as an amorphous 'hill' on the chromatogram; it is likely that the sample contains many tens of thousands of peptides and a wide dynamic range. From the data it was apparent that the majority of the peptides eluted between 15-40 min. The initial spike represents the early elution of dimethyl sulfoxide (DMSO) present in the sample as a means of solubilising the lipophilic clozapine.

Pooling of fractions 3-8 was carried out; based on the evidence of the UV chromatogram (figure 87) there was very little peptide present in these fractions as shown by the lack of any UV response.

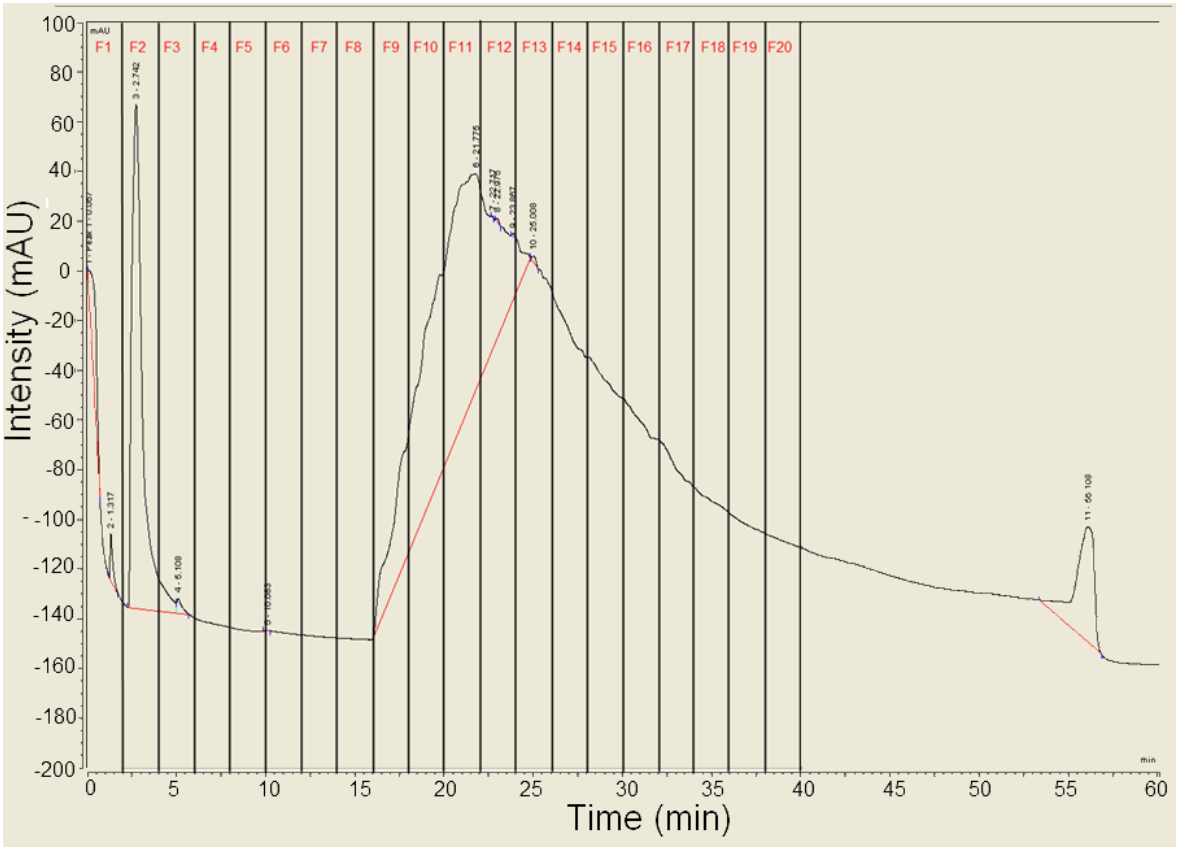


Figure 87. SCX separation of C- HLM sample. UV-VIS 214nm, indicative of peptide bonds. The majority of peptide elution occurs between 15 and 40 min. Individual peaks are not distinguishable; a hallmark of the sample's complexity.

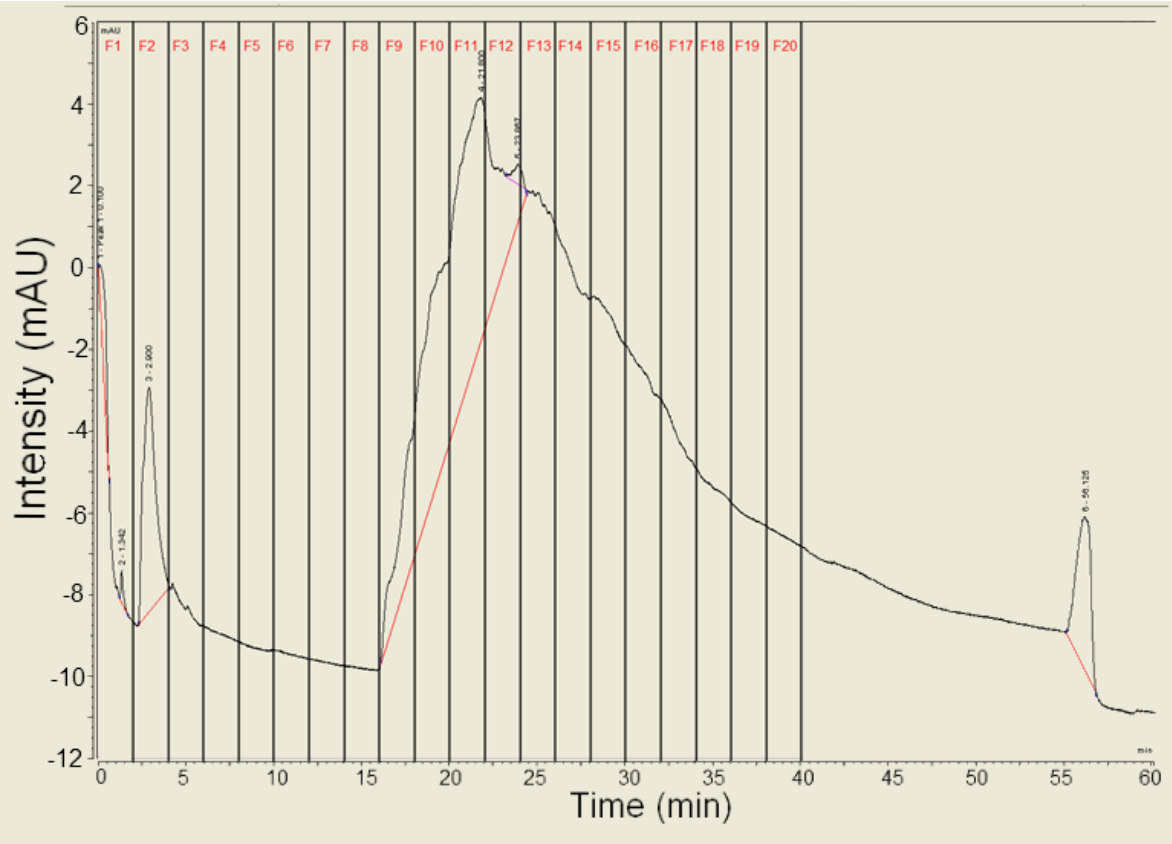


Figure 88. SCX separation of C- HLM UV-VIS 280nm. The absorbance at 280nm is dependent on the presence of the aromatic ring structures associated with particular amino acids. As a consequence the signal is considerably lower than that seen in the 214nm trace which measures the peptide bond associated with all peptides.

SCX provided an effective method for 1st dimension separation of peptides performing almost as well as the GeLC-MS approach. Again the PI359 scans showed up a very high number of protein identifications. There was a lack of protein IDs for the sample SCX C+ merged, examination of the raw data revealed poor acquisition, likely caused by deterioration of ESI quality most likely brought on by a failing needle. It was not possible to correct for this fault due to time limitations. It is more likely that the SCX C- merged IDA sample gives a better representation of the number of protein IDs achievable. The disparity between the PI359 and IDA protein IDs across the other separations approached a roughly 2:1 ratio (figures 83 and 85). SCX C- merged is consistent with this, C+ merged is closer to a 1:1 ratio (figure 86).

Had the acquisition of the clozapine positive sample gone more smoothly then it is probable that the SCX approach would have performed better i.e. provided more protein identifications, than the GeLC approach. As with the Offgel work the liver microsome samples were subjected to delipidation prior to tryptic digestion in order to access the less soluble membrane integral and associated proteins. As a result it was expected that a comparable number of membranes associated proteins be identified.

The number of CYP450s identified in the SCX samples by IDA was considerably lower than the numbers identified using either GeLC or Offgel separations. It is possible that the hydrophobic peptides were retained on the sorbent of the SCX column by hydrophobic interactions (Liu et al., 2006). In order to overcome this problem it would be necessary to increase the acetonitrile content of the SCX buffers from 5% v/v as used in these experiments to around 30% v/v (Liu et al., 2006). Unfortunately it was not possible to repeat the experiment with the improved buffer due to time limitations.

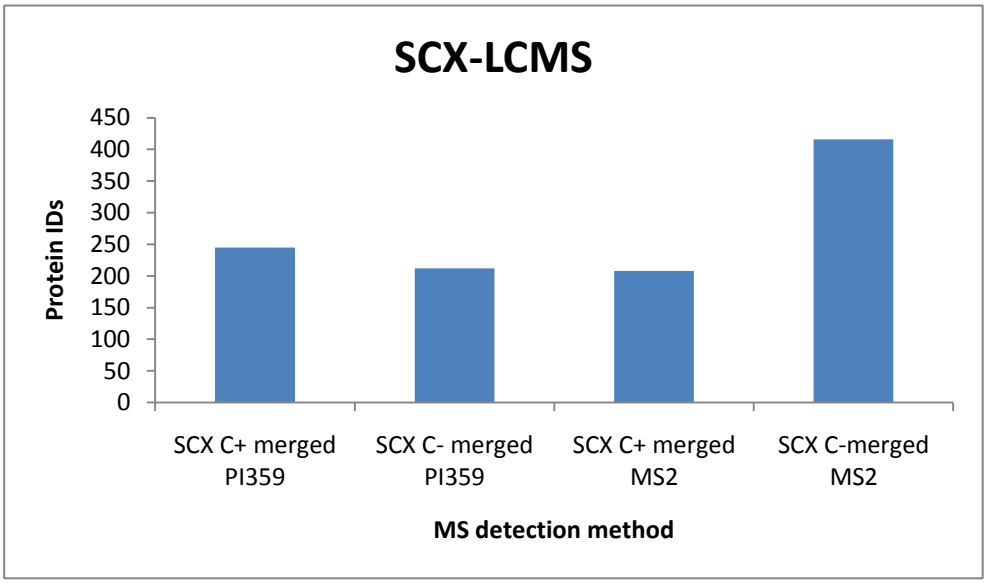


Figure 89. The total number of proteins identified in the IDA experiments (FDR <5%) for SCX C+ is considerably lower than the number identified in the SCX C- IDA runs. Examination of the data suggested that the mass spectrometers ESI interface was not performing well .

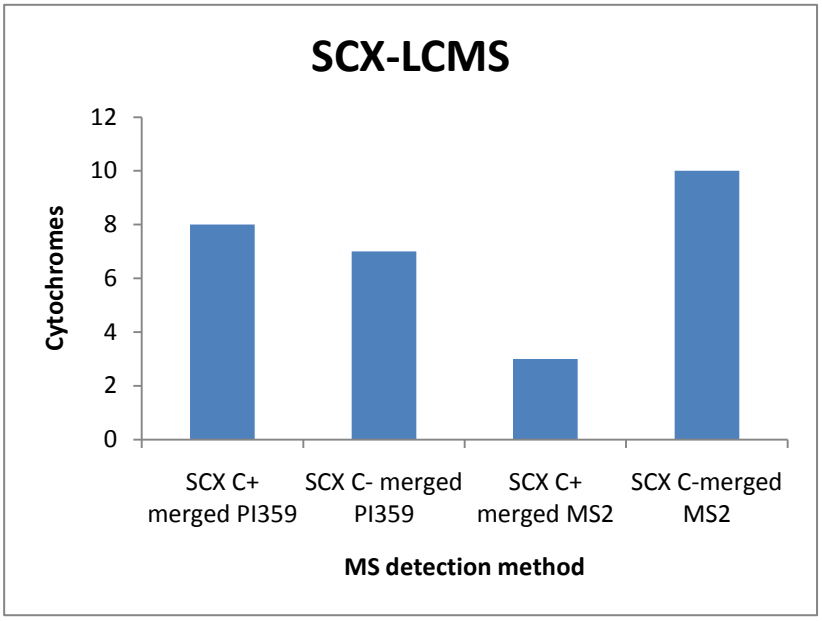


Figure 90. The poor showing of CYP450s is likely caused by retention of hydrophobic peptides on the sorbent of the column. The increasing salt gradient does not effectively disrupt the hydrophobic interaction. (FDR <5%).

4.4.2 Comparisons

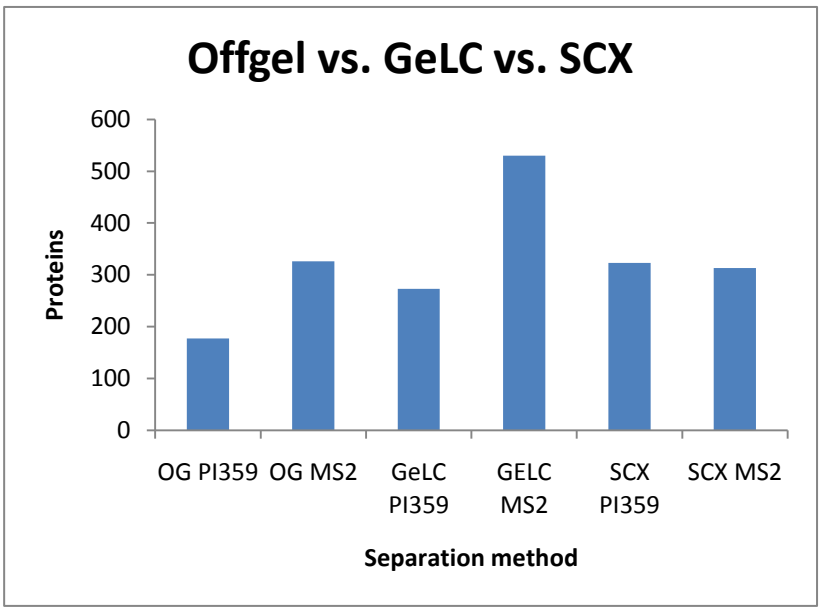


Figure 91. Total unique protein identification for each separation method and scanning method (FDR <5%).

The GeLC approach identified more proteins than any other approach using the IDA method. The detergent SDS used in the 1d-PAGE approach acts to solubilise hydrophobic proteins through the formation of protein-SDS complexes. The interaction between the SDS and the proteins is similar in nature to the interaction of proteins with lipid membranes and other amphiphilic substances (Reynolds and Tanford, 1970; Mascher and Lundahl, 1989). Despite the ease with which the 1st dimensional separation by SDS PAGE handles these otherwise insoluble proteins it would appear that appropriate delipidation steps before tryptic digestion can bring about similar results when used with the other methods (figures 92 and 93).

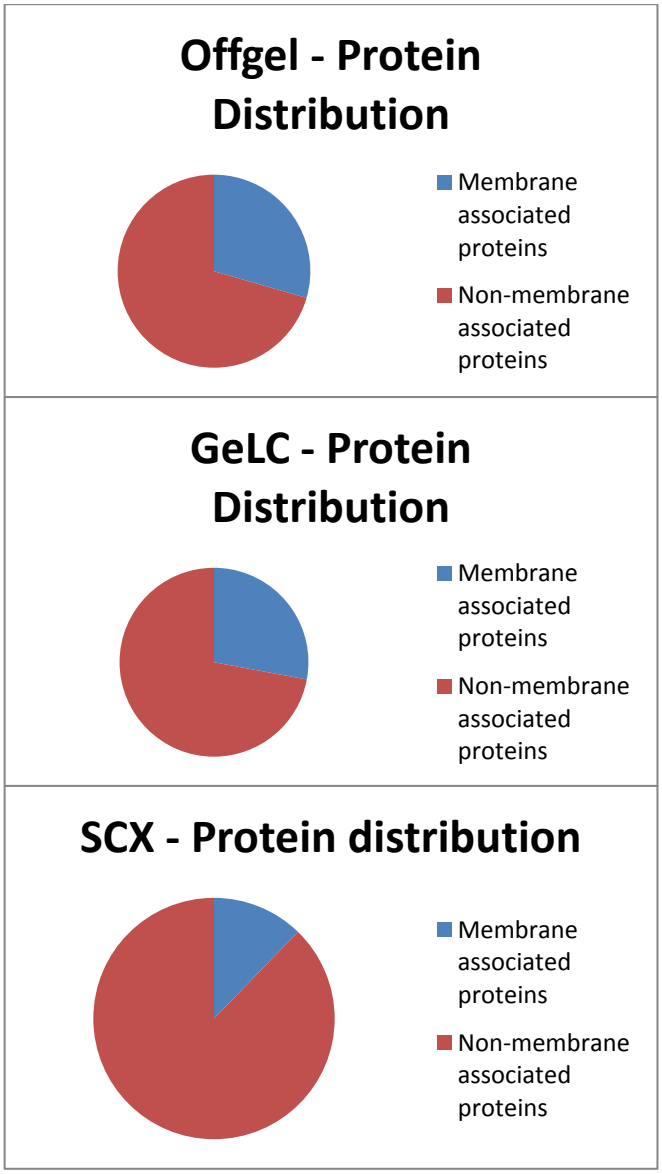


Figure 92. Offgel, GeLC and SCX protein distribution - membrane associated vs. non-membrane associated.

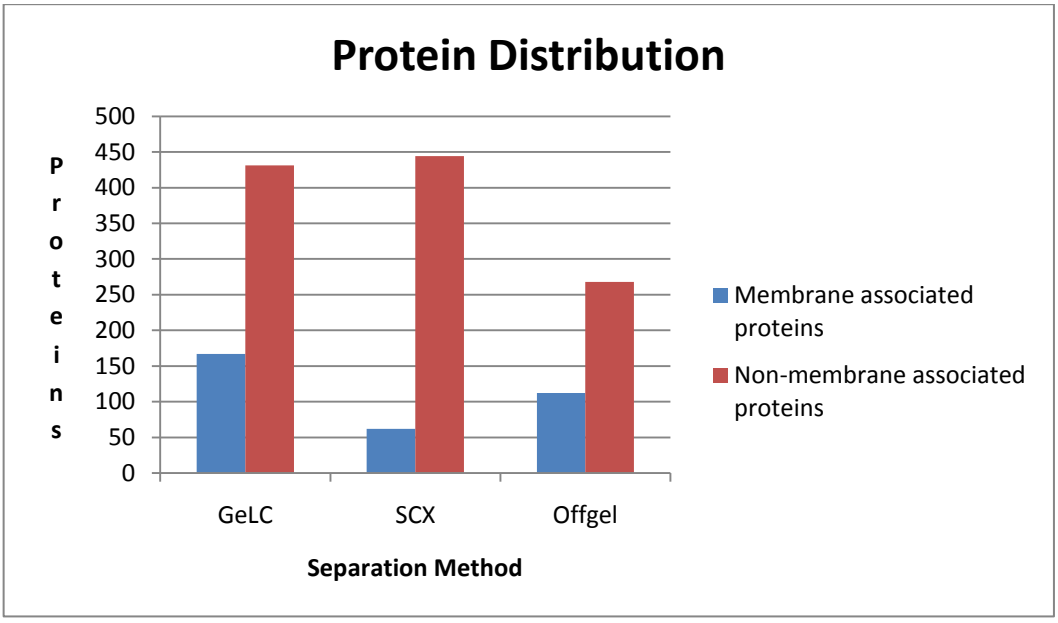


Figure 93. Despite having the greatest number of non-membrane associated protein IDs. The SCX method has identified the fewest membrane associated proteins.

The Offgel approach seems to have the greatest disparity between protein identifications from the PI359 scanning mode to the IDA MS/MS method. It is likely that the sample complexity is lower indicating the possibility that some peptides may have dropped out of solution, likely upon reaching their isoelectric point, or been retained by the IPG strip. The focussing stage involves a long time at room temperature and perhaps there was partial degradation of the sample during that time.

The GeLC approach identified the greatest number of unique CYP450 enzymes. This is likely due to the efficiency with which the method handles proteins with hydrophobic domains. The poor performance of the SCX separation with respect to CYP450 identifications would likely be overcome with a more optimal organic solvent concentration.

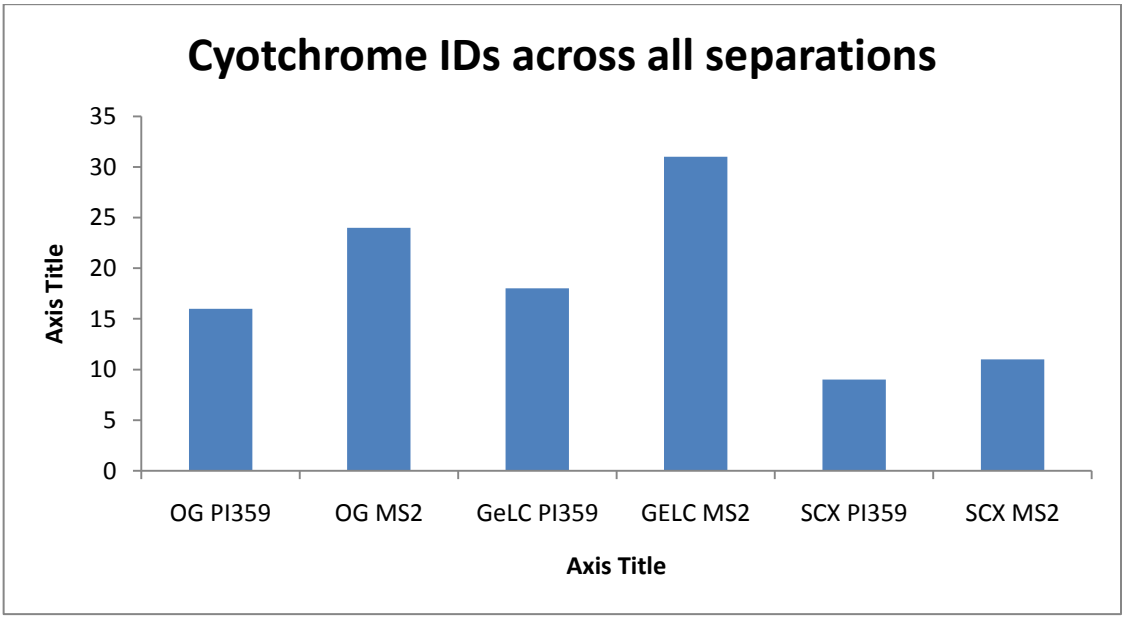


Figure 94. Offgel CYP450 identifications were similar to those identified by GeLC. The SCX approach is markedly less effective.

4.4.3 Overlapping of Protein Identifications

The following figures (95 and 96) serve to illustrate how the different approaches complement each other in the coverage of the proteome being studied. The various means of separation should have resulted in a range of fractions each with particular characteristics. The composition of each fraction should be completely unique and allow for the detection of different peptides. This can be understood best by considering the limitations of mass spectrometric detection. A major obstacle to detection is the limited dynamic range of mass spectrometers, the most abundant ions are detected most commonly and can effectively suppress the detection of lower abundance ions. Each fraction represents a different combination of high and low abundance ions meaning that theoretically the proteins detected in each should vary.

The comparison between the IDA method and PI359 method was quite revealing. It was initially expected that the PI359 method bring about a marked increase in selectivity when searching against a complex background. The data obtained in this project indicates that the PI359 scan detects roughly one protein for every three seen in the non-selective IDA approach, less than an order of magnitude more selective. It is unlikely that such a small improvement would be useful

when dealing with highly complex liver microsome fractions. Further investigation of this revealed that a high number of false negative identifications were made based on the fact that the target ion at m/z 359.1 is not very selective (4.4.4.4).

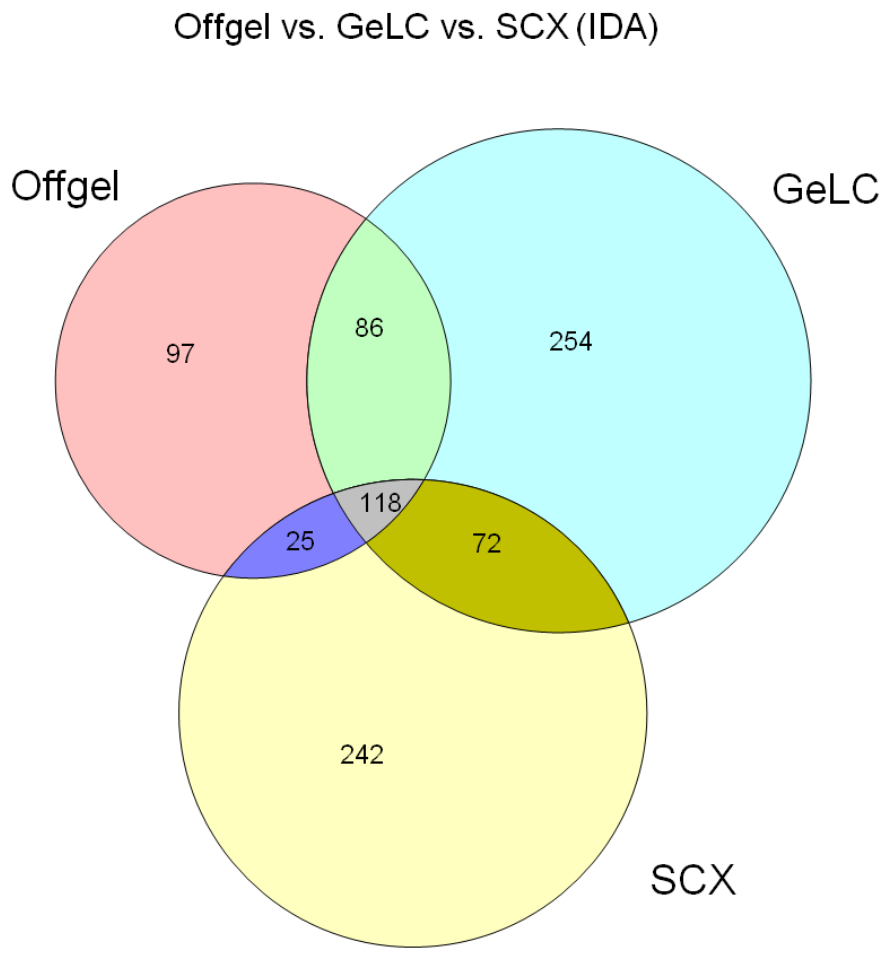


Figure 95. Offgel vs. SCX vs. GeLC protein identifications based on information dependant acquisition. It is immediately apparent how each of the separation techniques contributed to the overall proteome coverage. A large number of proteins were not detected by more than one of the approaches.

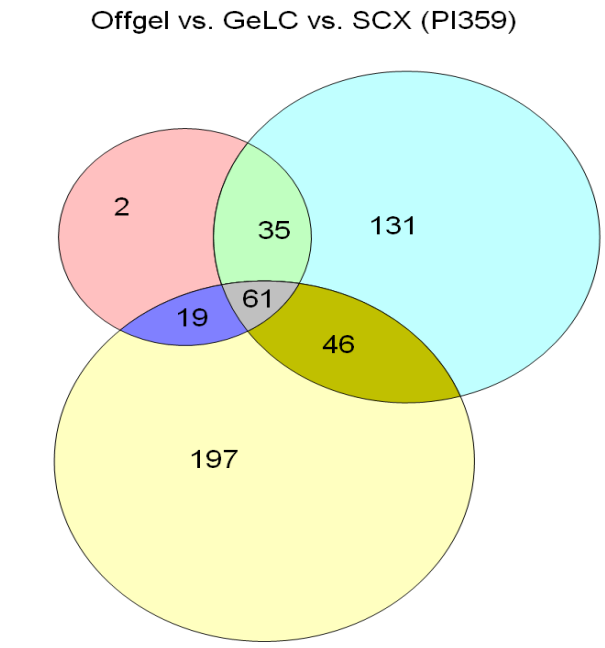


Figure 96. Offgel vs. SCX vs. GeLC protein identifications based on PI359 scanning. As with the IDA experiments a majority of proteins were not detected from more than one of the separation methods.

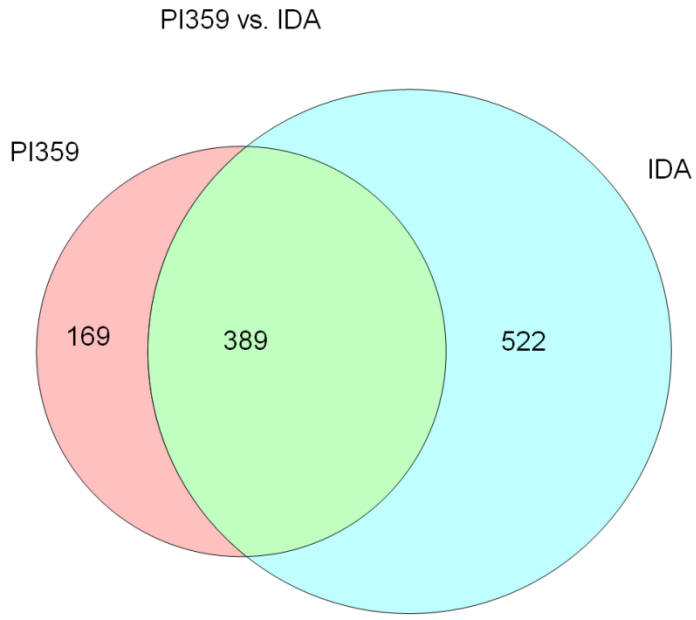


Figure 97. Precursor ion scanning vs. information dependant acquisition protein identifications. The PI359 scan identified approximately 1/3 as many proteins as did the IDA method. This level of selectivity is considerably lower than expected.

4.4.4 Distribution of Protein Identifications Across Multiple Separation Dimensions

In order to better understand how the proteins and peptides were spread across the two dimensional separation spaces in each combination of separation techniques the following heatmaps were created (figures 98-104). The mass spectrometric data in the form of .wiff files (ABI/Sciex) were loaded into the Peak View 1.0 software and the TICs used to generate heatmaps. For the PI359 scans the heatmaps were based on the initial PI359 scan ion chromatograph, for the IDA scans the heatmaps were based on the enhanced MS scan data. In all heatmaps the vertical axis is divided into fractions generated by the 1st dimension of separation, the horizontal axis is based on time. The Intensity of the ions detected is represented by increasing darkness for increasing intensity.

4.4.4.1 GeLC

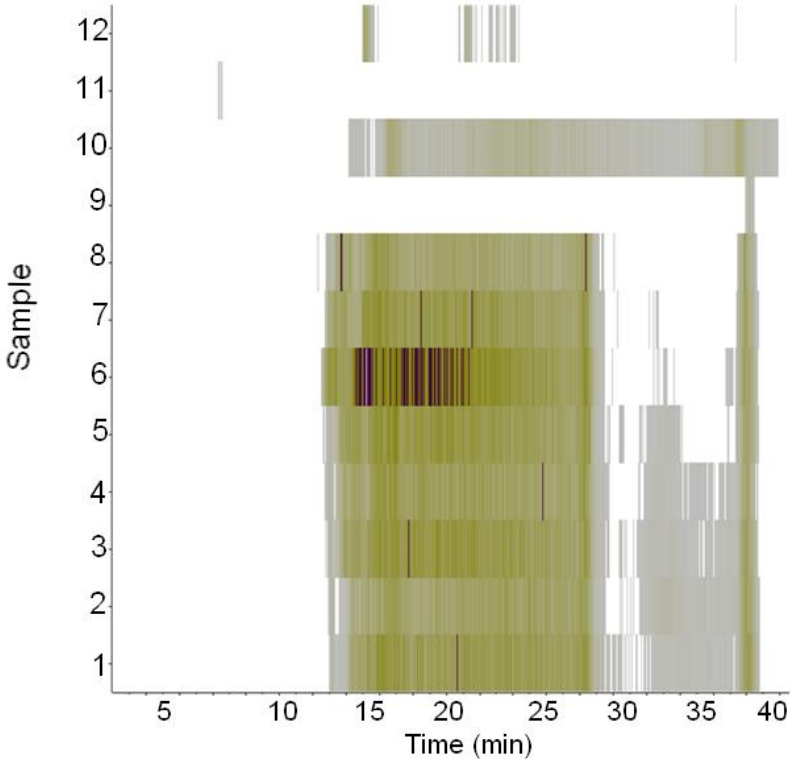


Figure 98. The GeLC separations represent 1st dimension protein separation followed by 2nd dimension peptide separation. From the heatmap the distribution of ion intensity is fairly uniform from around 13-30 min chromatographic time.

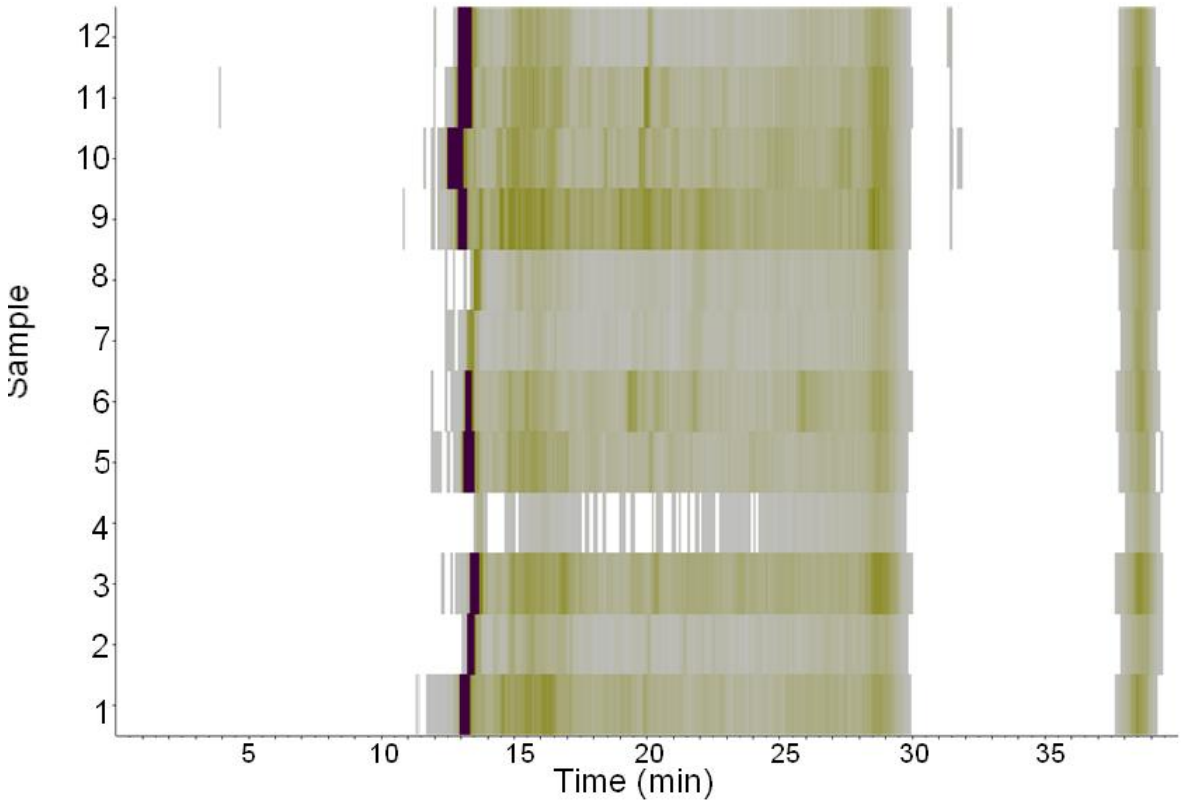


Figure 99. The GeLC PI359 heatmap presents a pronounced intensity of ions at around the 13 minute mark across all fractions with a lower detection from 13-10 min. This ion of m/z 523 was found to be a tryptic peptide fragment hence its presence in all fractions.

4.4.4.2 SCX

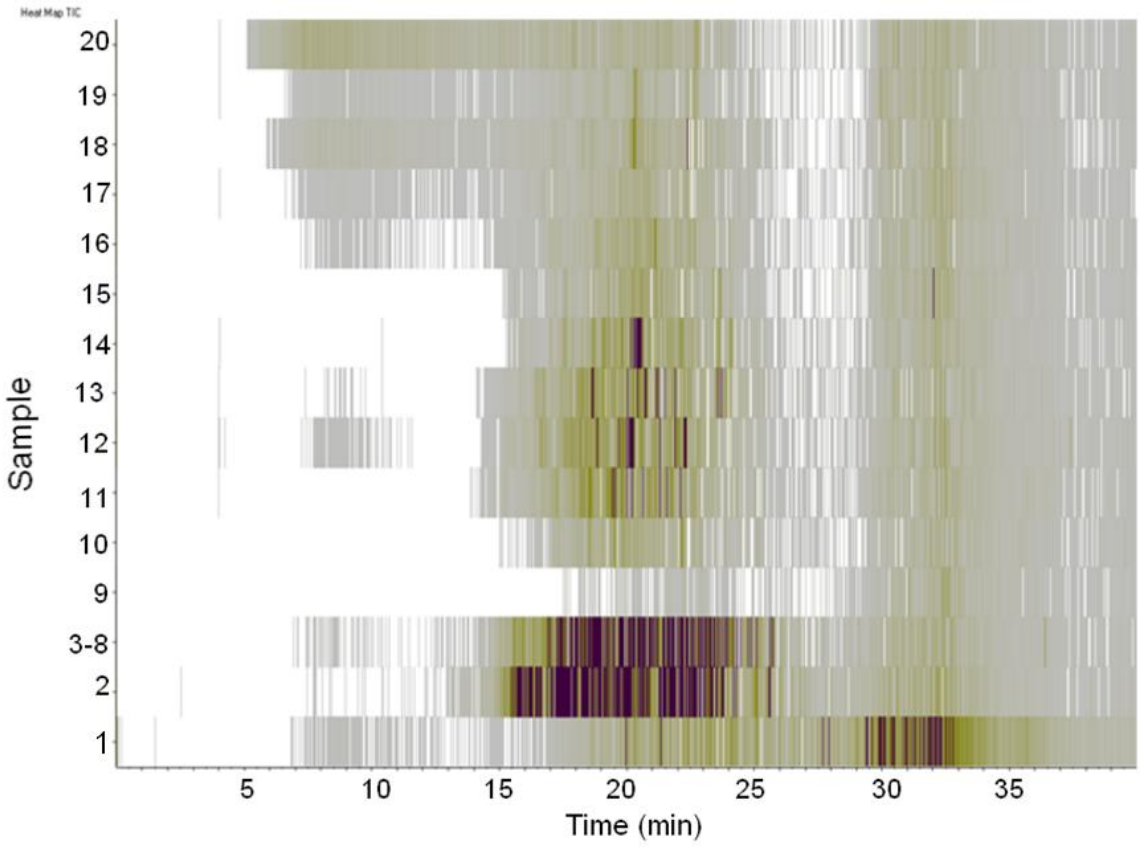


Figure 100. The SCX IDA heatmap represents a non-uniform use of the 2D separation space with areas of particularly high ion intensity.

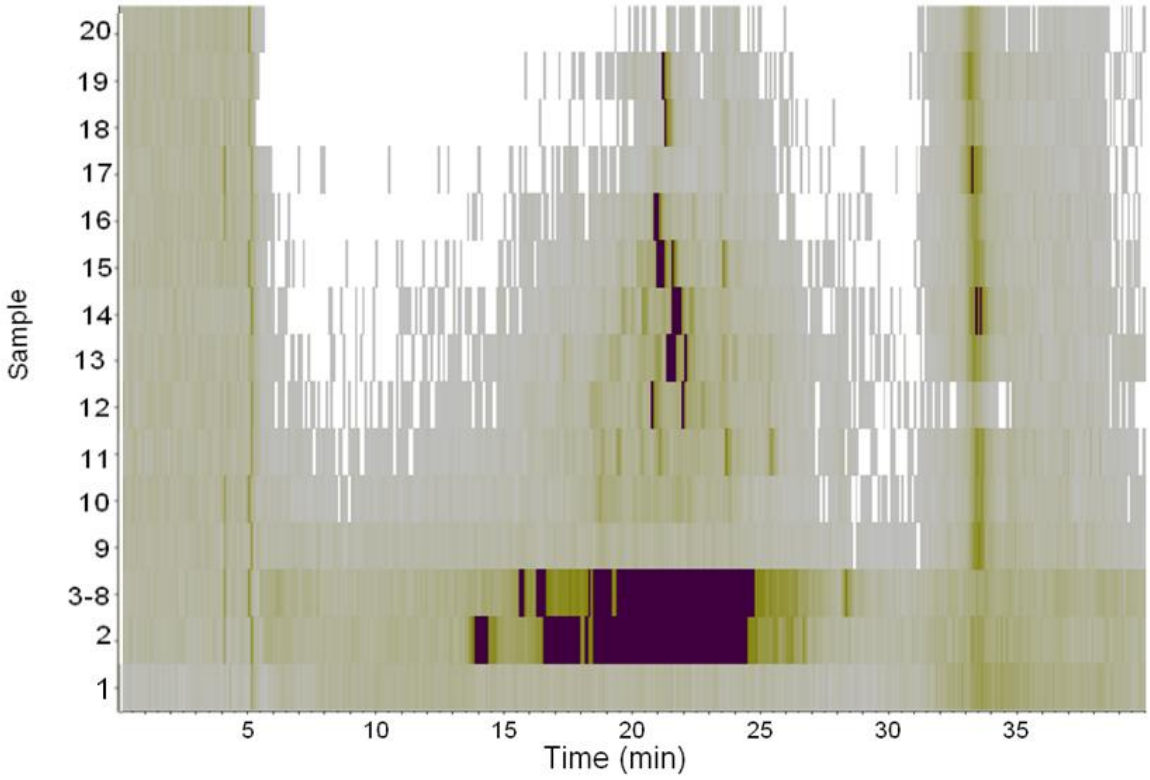


Figure 101. SCX PI359 scan. The pattern of ion intensity is similar to that seen in the IDA scan.

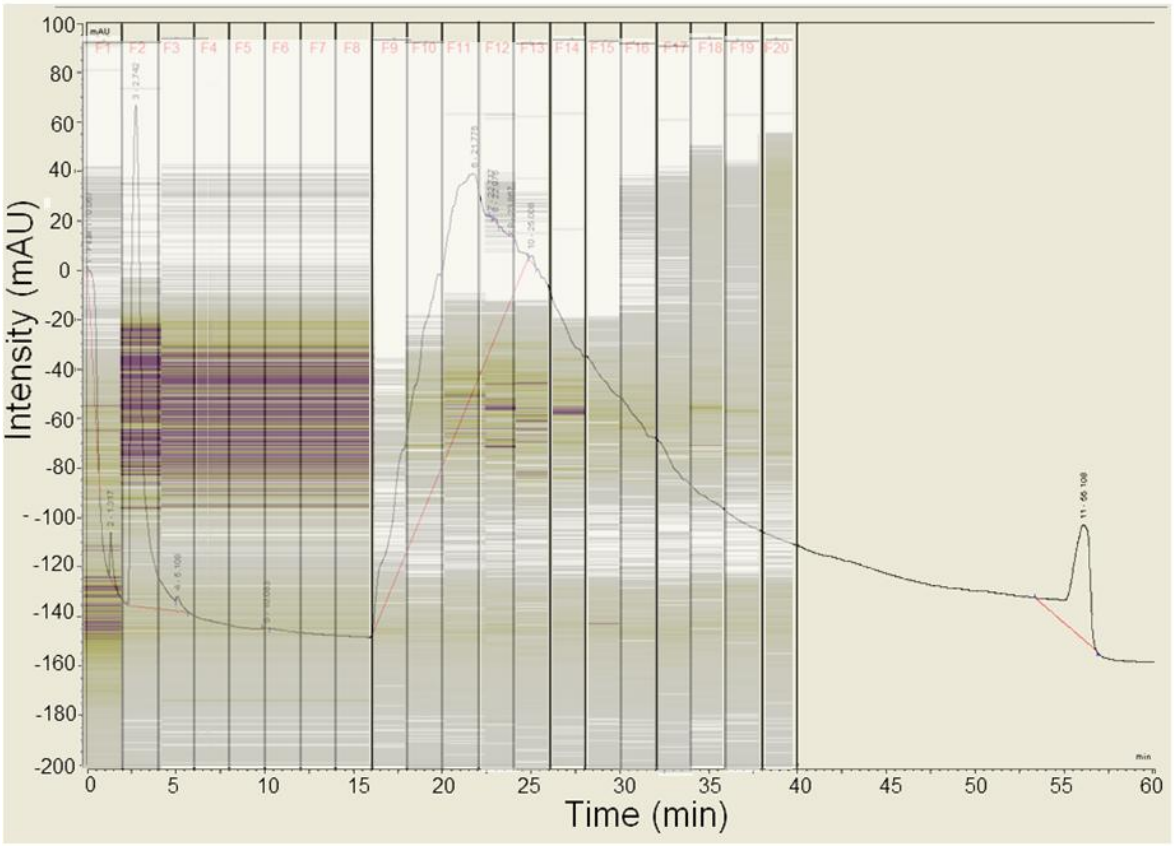


Figure 102. Overlay of the heatmap from figure 100 (SCX IDA) and the UV data from figure 87. The area of high intensity from samples 3-8 represents a single MS run. The lack of UV signal indicated a low level of peptide present and as such these samples were pooled. Samples 9-20 represent the most abundant peptide elution from the SCX run and coincide with the greatest number of peptide identifications.

4.4.4.3 Offgel

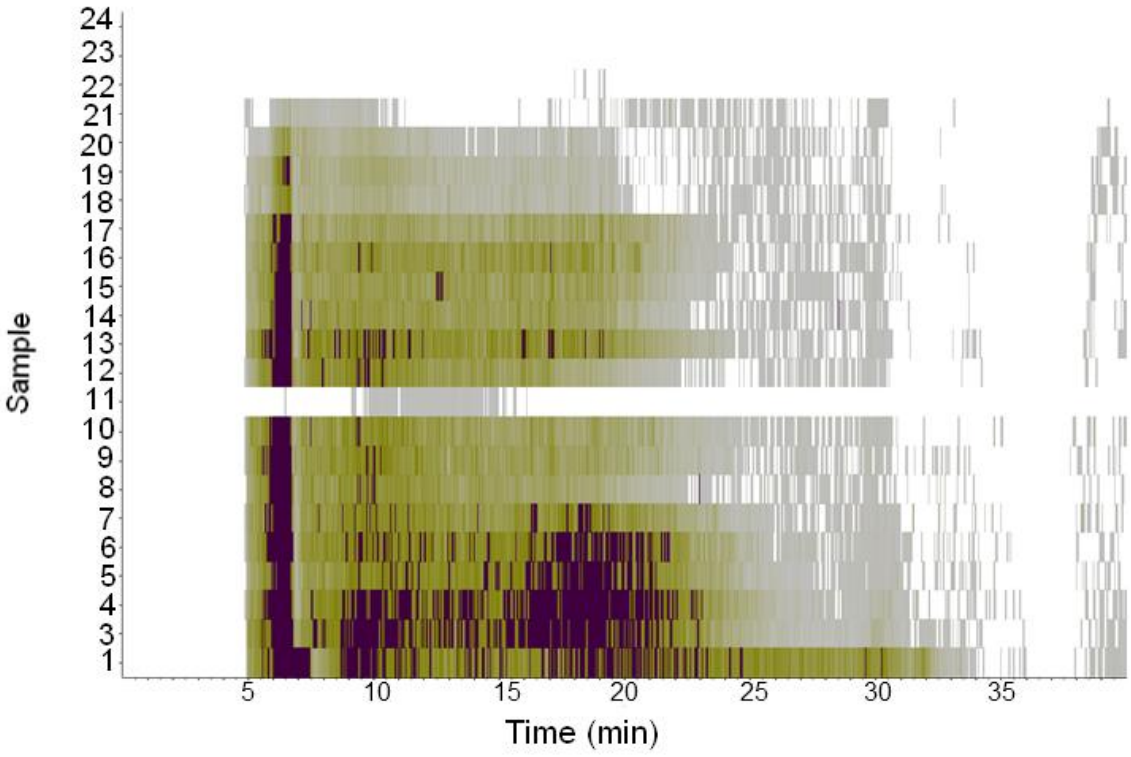


Figure 103. The Offgel IDA heatmap shows a wide distribution of ion intensities.

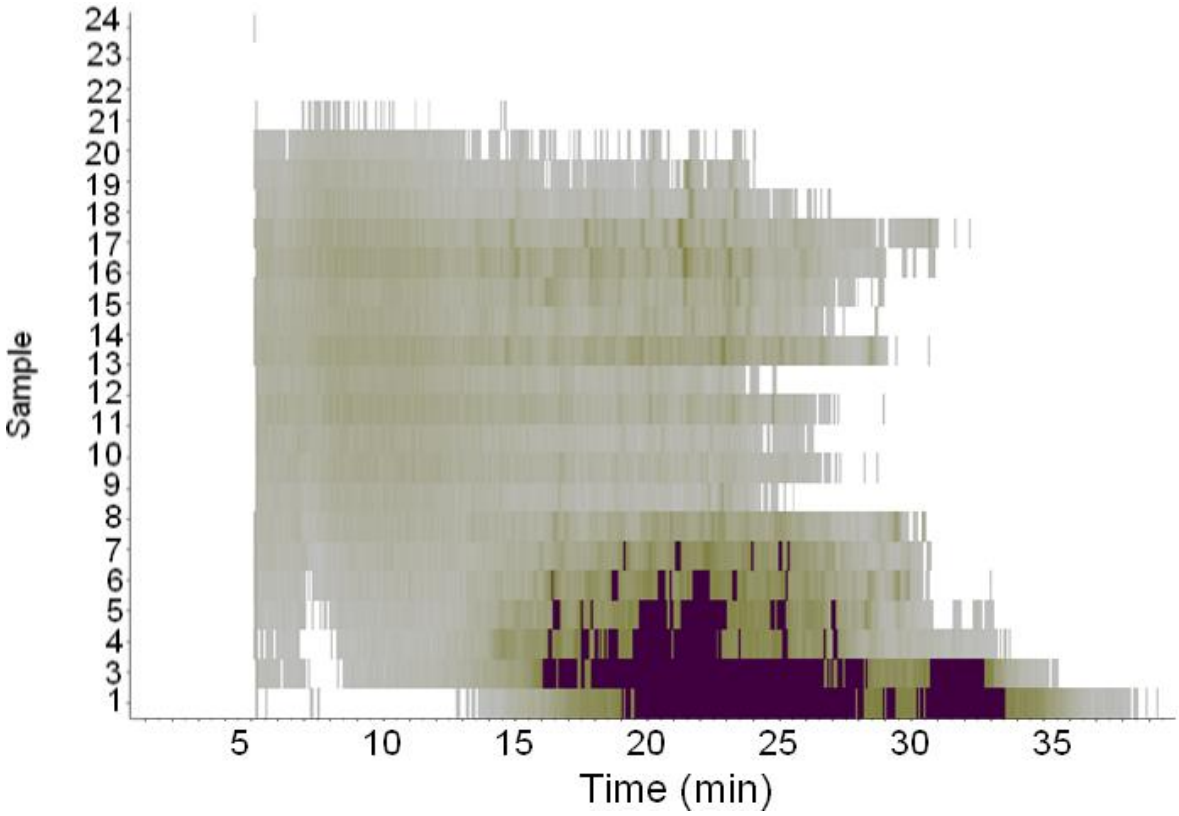


Figure 104. The Offgel PI359 heatmap shows a much narrower distribution of ion intensities.

As would be expected the heatmaps associated with the more selective precursor ion scans tend to be more concentrated with respect to areas of high intensity. The effect is most pronounced in the comparison of the GeLC IDA and PI359 heatmaps. From this data it would appear that the ion intensity peaks at around the 13 minute mark across all of the gel sections in the PI359 sample. This indicates that the PI359 scan mode is reacting to something that is eluted at the same time in all fractions. Due to the nature of the 1st dimension of separation it should not be possible for the same protein to be present in all fractions. An exception to this rule is of course the trypsin used for digestion of all samples, the reason that is does not occur in samples with either Offgel or SCX 1st dimensional separations is that the trypsin added to those samples was subjected to the 1st dimension of separation. In contrast, the GeLC approach is the only one in which tryptic digestion is carried out after the 1st dimension of separation. Trypsin was identified in all samples as was expected and a detailed manual analysis of the data revealed the ion to be a peptide fragment, at m/z 523, belonging to trypsin.

From the heatmaps it appears that the GeLC separation makes the best use of the 2 dimensional separation space followed by the Offgel approach. Interestingly this does not seem to be in agreement with the actual performance of the Offgel method. In this case as only the TIC is taken into account this may indicate the presence of some sort of contaminant in the Offgel fractions. This may also go towards explaining the lack of proteins identifications as a consequence of ion suppression. A known limitation of SCX separations is the tendency for tryptic peptides to bear predominantly either 2 or 3 charges. As separation proceeds based on the charge states of the peptides the elution window is relatively narrow ultimately leading to an inefficient use of separation space (Gilar et al., 2005). Looking at the merged figure of the SCX heatmap and UV data (figure 102) it is clear that peptides are predominantly located within the region of high UV intensity (time 25-40 min). In order to fully utilise the theoretical separation space it would be necessary to increase the length of the SCX elution gradient and to increase in the number of fractions taken. Additionally, peptide elution along the second dimension mainly occurs in the second third of the reverse phase run; lengthening the reverse phase separation gradient should further increase separation space. The obvious downside to this would be the greatly increased analysis time required.

In the all heatmaps the majority of high intensity ions are detected between about 13-30 min. This correlates well with the RP-LCMS gradient used in that maximal detection of peptides occurs during the gradual increase in buffer B. The Offgel IDA heatmap however, has a peak in ion intensity between min 5-7. This is a very early point in the solvent gradient and may be due to some sort of contaminant. The same peak in intensity does not occur in any of the other samples, including the PI359 Offgel runs. This is likely due to the contamination failing to produce the necessary product ion of 359 m/z and as such being unable to trigger the PI359 scan.

The seemingly high number of protein identifications made even when using the precursor ion scan is likely caused by either an almost ubiquitous production of fragments at 359 m/z, overlap of parent ion masses between parent ions that produce the precursor and those that do not; or a combination of both factors. The precursor ion scans selectivity is adversely affected by the lack of mass accuracy inherent to quadrupole instruments (typically >100 ppm). A precursor

ion scan of 359 m/z units translates more accurately to 359.1 +/-0.6, giving a very wide window for non-specific selection of precursor ions. A higher mass accuracy would mean narrowing the window of selection thereby eliminating a high level of false positives. Unfortunately this approach would not serve to close the window on the problem of overlapping parent masses but would effectively reduce the sensitivity of the method. Even with more accurate identification of fragment ion masses the initial parent ion masses would still be limited by the quadrupole's poor resolution. The only truly effective way to eliminate or at least alleviate the problem would be to improve peptide separations and thereby reduce the number of different ion species being introduced to the instrument at any given moment.

4.4.4.4 PI359 candidate ions

A detailed analysis of the data was carried out in order to identify ions that were responsible for triggering the precursor scan. The results from the GeLC experiments were chosen as they are responsible for the greatest number of protein hits. Peptide matches from the mascot results files were manually inspected for any fragment ions with mass 359.1+/- 0.6Da representative of the target mass of the precursor scan coupled with the MS/MS tolerance of the instrument in this scanning mode. The charge state of the parent peptides were not taken into account, only the charge state of the fragment ions were considered as only these are capable of triggering a scan. Ions included in the search criteria were of type b, y and y*; where y* is a y series ion with a loss of 17Da representative of internal fragmentation and loss of ammonia (NH_3). The mass difference was taken as the theoretical exact mass of the detected peptide minus the theoretical exact mass of the adduct fragment (359.1092 Da). The mass difference was then converted into error as expressed in parts per million (ppm). Additionally, the number of possible permutations for a given peptide was calculated using the general formula:

$$Permutations = n!$$

Where n is equal to the number of unique amino acids in the sequence.

And:

$$Permutations = (m!)/2$$

Where m is equal to the total number of amino acids in a sequence with 2 identical amino acids; no sequences were identified that had more than 2 of the same amino acids.

It is important to note that the following table represents only detected peptides and that only one permutation was chosen even in cases where multiple permutations were discovered. The reason being that the purpose of the analysis was to identify which peptides have masses close enough to the target precursor mass to trigger a scan. Every permutation of a given peptide sequence will be isobaric in mass to any other sequence of the same amino acids making the listing of permutations redundant.

Table 9. Peptide fragments detected in the PI359 scans of clozapine positive and negative human liver microsome samples as separated by GeLC. The exact theoretical mass of the precursor ion (359.1092) was compared to the exact theoretical mass of a peptide fragment as matched by the Mascot search engine.

Predicted mass	Peptide	Ion series	Mass difference(Da)	Mass difference (ppm)	Permutations
358.6823	PVTEDR $^{2+}$	y	-0.4269	1189	720
359.1197	NED	b	0.0105	29	6
359.135	GWD	b	0.0258	72	6
359.1361	TEQ	b	0.0269	75	6
359.1561	AADTV	b	0.0469	131	12
359.1561	NDK	y*	0.0469	131	6
359.1561	TGEA	b	0.0469	131	24
359.1561	AADT	b	0.0469	131	12
359.1638	SAGWDAK $^{2+}$	Y*	0.0546	152	2520
359.1674	GGSR	y*	0.0582	162	12
359.1714	NPF	b	0.0622	173	6
359.1748	IMN	b	0.0656	183	6
359.1748	VGMA	b	0.0656	183	24
359.1748	VMQ	b	0.0656	183	6
359.1925	ATSV	b	0.0833	232	6
359.1925	GTLS	b	0.0833	232	24
359.1925	GTVT	b	0.0833	232	12
359.1925	PDASVTK $^{2+}$	y	0.0833	232	5040
359.1925	PDK	y	0.0833	232	6
359.1925	QTK	Y*	0.0833	232	6
359.1925	SAVT	b	0.0833	232	24
359.2037	SPR	y	0.0945	263	6
359.2401	ALR	y	0.1309	365	6
359.2401	IAR	y	0.1309	365	6
359.2653	VIK	y	0.1561	435	6
359.2653	VLK	y	0.1561	435	6
359.6537	SEDDPR $^{2+}$	y	0.5445	1516	360
359.6719	ETESPR $^{2+}$	y	0.5627	1567	360

In total some 149 different motifs were discovered that produced ions close enough to the precursor target mass to trigger the scan. The table describes 28 of these ions but all 149 were permutations of those detailed. It is important to note that these are peptide fragments produced by the CID fragmentation of larger peptides and not full length tryptic fragments produced during digestion of proteins. This means that the limitations imposed on tryptic peptide

fragments (i.e. the position of cleavage sites lysine, and arginine at the start and end of peptides) do not apply and therefore do no limit the number of permutations possible. The total number of permutations possible from the data represented was 9,234. Most fragments were found to be singly charged ions of either three of four residues in length which may seem counterintuitive as fewer permutations are possible when compared to fragments with 5, 6 or 7 amino acids but makes sense when the nature of the dataset being searched is considered. The protein database (Swissprot in this case) contains a finite number of protein sequences of finite length. The probability of randomly matching a given peptide fragment to a protein sequence decreases in proportion to the number of amino acids in the fragment i.e. the sequence EVE is more likely to occur than the sequence EVEKQ.

It should be possible to calculate the total number of peptide sequences that would give rise to the precursor target mass, how often they appear in a given dataset and their variance from the exact mass of the target. With this data it would be possible to ascertain the optimal window at which to perform precursor scan based searches.

In the case of this work none of the matched peptide fragments came within 0.01 Da of the exact mass of the precursor target; a difference of greater than 28ppm. An instrument such as the ABSCIEX TripleTOF © 5600 capable of a resolution of 40,000 FWHM and a mass accuracy of better than 2ppm could theoretically distinguish between the actual target mass and those identified in this analysis. It is extremely unlikely that all of the possible combinations of amino acids with mass similar to that of the target were identified and it is possible that there are masses which come closer than the 28ppm observed; however, the higher resolution and mass accuracy would act to decrease the number of false positives. False negatives can be alleviated with the widening of the mass window to allow for experimental error and allowing for the known precision of any given instrument type.

4.5 Discussion

In the course of these experiments no positive identifications were made with respect to proteins modified with clozapine metabolites. All five of the previously identified metabolites (3.4.1) were included as part of the Mascot search parameters. It is possible that the rarity of modification in the liver microsome system has meant that more abundant ions have completely suppressed the detection of metabolite-protein adducts. Many possible peptide fragments bearing clozapine modifications were identified as unmatched ions i.e. the amino acid sequence did not significantly match any proteins in the databases searched. It is more likely that these matches are statistical artefacts caused by the extremely large amount of data generated by the MS analysis of samples. It is not particularly surprising that the IDA based scans did not reveal any protein-drug adducts. The complexity of the sample and the relative scarcity of the modification (5.4.5) would stack the odds against their discovery. The failure of the more selective precursor ion based scans to identify modifications is more disappointing. The enhanced selectivity should act to effectively reduce the background complexity. However the results of the work carried out in this part of the project demonstrated that the precursor ion scan identified around 1/3 the number of proteins that were identified using the IDA method. This level of selectivity is unlikely to greatly assist in detecting drug modified peptides in such a complex background. This could be partly overcome by using a higher resolution and more accurate instrument in order to increase the selectivity of the precursor scan.

Due to the large number of fractions for analysis it was sometimes necessary to have them stored in the autosampler for up to 24 hours. It was considered possible that sample degradation could occur over this length of time leading to a reduction in the number of peptides detected. However earlier work involving the modification of synthetic peptides with clozapine showed that the modified peptides could be detected in samples after several days at room temperature and multiple freeze/thaw cycles. Additionally, the autosampler device stored the samples at 4 °C further decreasing the likelihood that an unacceptable amount of sample degradation occurred.

Each of the studies involved two dimensions of separation in order to greatly reduce the complexity of any given sample thereby allowing for a more in depth and complete analysis of the liver microsome proteome. Each of these approaches allowed for the identification of several hundred unique proteins. In total, more than 1700 proteins were identified, more than 1000 of which were non-redundant. Despite this fact the low levels of coverage for some proteins means that a large part of the proteome has gone undetected. Additionally, peptides with post translational modifications not accounted for cannot be detected using the Mascot search. Despite the selectivity of the precursor scanning approach limitations on sensitivity are likely to play a part in the detection of low level modifications.

Separation by Offgel fractionation and by ion exchange was carried out at the peptide level in order to maximise the number of hydrophobic proteins solubilised in the samples. In the case of both the SCX and Offgel experiments digestion of proteins was carried out in-solution. With in-solution digestion the primary concern is the prevention of adsorption, protein aggregation and precipitation. In order to prevent these problems, proteins can be treated with chaotropes, detergents (ionic, zwitterionic or non-ionic), organic acids or organic solvents. Work has shown that using these agents as part of in-solution digestion strategies results in a marked improvement in membrane protein identifications, as discussed in the review by Speers and Wu (Speers and Wu., 2007). A good representation of membrane proteins, upon MuDPIT/Offgel analysis, has been obtained simply by carrying out digestion in a high concentration of chaotropic agent (8M urea) (Gonzalez-Begne et al.,2009; Elschenbroich et al., 2009). The chaotrope serves to disrupt protein-protein interaction and prevent precipitation. High concentrations of methanol or acetonitrile also act to denature and solubilise proteins and have been shown effective in dealing with the digestion of membrane proteins (Blonder et al., 2004; Dormeyer et al., 2008).

Detergents are commonly used to solubilise proteins and separate them from membranes. A range of these are used but SDS is typically favoured. SDS forms complexes with the proteins and result in their denaturation (Reynolds and Tanford, 1970). Although it can be used to efficiently and effectively purify proteins from membranes it interferes with mass spectrometric downstream

analysis (Loo et al., 1994). The problem with ionic detergents is that they ionise so efficiently and are in such great quantity that they cause ion suppression and a subsequent loss in sensitivity. In order for mass spectrometric analysis to be carried out the detergent must be stringently removed from the proteins. An approach, known as FASP (filter-aided sample preparation), uses SDS to solubilise proteins before loading them into an ultrafiltration device (Nagaraj, et al., 2008). Once in the device the proteins undergo buffer exchange into an 8M urea solution with the SDS being washed through the filter. Enzymatic digestion is then carried out and the peptide fragments eluted and collected for analysis. Results indicate that the method is highly effective for identification of membrane proteins and also has been shown to improve sequence coverage of identified proteins when compared to standard in gel digestion (Wisniewski et al., 2009).

The delipidation protocol applied in this work with respect to the in-solution digested proteins likely played a part in denaturing the proteins and allowing for more efficient digestion. However it is highly probable that the application of one of the a aforementioned methods for handling membrane proteins would have markedly increase the number of identifications. A review of the literature indicates that in-solution digestion can provide elucidation of membrane proteins at least on par with gel based methods.

Due to the apparent complexity of the samples after the various forms of two dimensional separations applied it may be necessary to either add a further orthogonal separation technique or use a more selective method for identifying modified peptides/proteins or increase the degree of separation in either or both dimensions. The obvious drawbacks to adding further orthogonal separation techniques is the exponential increase in processing time. It could be argued that all of these techniques are limited by the fact that they effectively reduce the resolution of separation by necessitating the collection of fractions. A problem that is only currently overcome when separating proteins by 2D PAGE due to the analogue nature of both dimensions of separation. It would be possible to overcome this limitation by increasing the number of fractions taken to a point at which the fraction length is greatly shorter than the expected chromatographic peak width. However this approach would generate a vast

number of fractions and increase downstream analysis times to an unacceptable extent.

Affinity depletion of the most abundant 4-6 proteins in routinely carried out in order to improve identification of less abundant species(Ramstrom et al., 2009; Echan et al., 2005; Linke et al., 2007), however it has been shown that the 4-6 next most abundant proteins then go on to mask the presence of less abundant species (Stalder et al., 2008). Due to the extremely wide dynamic range (around 10^{10}) the effective application of this technique probably requires several rounds of depletion, during these depletions it is likely that an unacceptable loss of sample will occur through non-specific interactions.

It is worth noting that the failure to detect the presence of modified metabolites is not synonymous with the nonexistence of said molecules. The earlier work carried out involving the formation of both glutathione adducts and synthetic peptide adducts clearly demonstrates the reactivity of the clozapine metabolites with the sulfhydral group of cysteine both with and without the catalytic activity of GST.

The formation of protein adducts would appear to be at least partially specific; with earlier work carried out having demonstrated that the presence of free and accessible cysteine residues does not always lead to adduct formation.

In samples of this complexity a major challenge is that of ion suppression. The phenomenon is poorly understood but is supported by several studies (Tang and Kebarle, 1993; King et al., 2000; Annesley, 2003; Mallet et al., 2004). There are two likely causes of such suppression; firstly, competition between molecules for charge and secondly, a saturation of analyte within the ESI droplets leading to increased viscosity. In the first case molecules with greater basicity will outcompete others for the positive charge (the available charges are limited as droplets evaporate and approach the Rayleigh limit (1.3.1.4.1)) and thereby suppress their transmission into the mass spectrometer. In the second case the analyte concentration can increase the viscosity of the droplets and thereby work against the transition of ions from the liquid to the gas phase. Typically this so called matrix effect is caused by contaminant molecules eluting along with analytes. Sample preparation is an important step in reducing the effects of ion suppression.

Ion suppression occurs with increased sample concentration, previous work has demonstrated that increasing the amount of a tryptically digested BSA sample above 40 fmol lead to a problematic level of ion suppression in a nanoLC-MS setup (Hirabayashi et al., 2007). This is representative of a general suppression of ions and seems more related to the implications it has on the linearity of mass spectrometric response and its applications in quantification (Enke, 1997). However another possible outcome with interesting implications for this work is the suppression of specific ions to such an extent that they are effectively excluded from the MS analysis. The effective exclusion is brought about by their signal being indistinguishable from the background noise.

Metabolite peptide adducts represent only a small fraction of the total peptide content of the highly complex liver microsome digest. Consequently, when competing for free charges in the electrospray ionisation process they will be underrepresented; the number of molecules entering the gas phase as ions would be in proportion to the number present in the liquid phase. However, this is only true when the availability of free charges greatly exceeds the number of analyte molecules in an ESI droplet as the competition between different analytes ends with every molecule entering the gas phase as an ion.

In actuality the dynamic is more complex with the free charges only being accessible at the outer edges/surfaces of the droplets and their number being limited by factors including the electrical conductivity and the flow rate of the solvent (Tang et al., 1989; Cech and Enke, 2001; Tang and Kebarle, 1993). When the number of charges is similar to or less than the number of analytes then molecules that possess properties allowing them to ionise more effectively will acquire more of the available charges. If the disparity between ionisation affinity and the difference in concentration is great enough it is probable that the low abundance species will be effectively entirely suppressed.

As the overall representation of modified peptides is so low it is highly probable that at any time during ESI at which such a species is present it will be accompanied by several non-modified species. The chances of the modified species being the more ionisable will always be lower than that of any of the other group of non-modified peptides simply by the laws of probability.

The ultimate consequence of this is almost paradoxical; it seems to follow that by increasing the concentration of the sample beyond a particular threshold (a number of analyte molecules exceeding the free charges) the sensitivity of the analysis with respect to lower abundance less ionisable species would decrease.

For this reason it has been suggested that decreasing flow rates to low nl/min should decrease charge competition and extend the dynamic range with respect to quantitation (Tang et al., 2004).

Chapter 5: DiGE and Western Blot Analysis

5.1 Aims

The putative stochastic process of electrophilic attack and subsequent protein-metabolite adduct formation suggests a situation in which there exists populations of various protein species with both modified and native configurations. The ability to accurately measure and compare the relative proportions of modified and unmodified proteins of the same species would allow for the creation of a map of proteins at risk from electrophilic attack from any particular xenobiotic. Difference gel electrophoresis (DiGE) allows for the direct comparison of whole proteomes exposed to different physiological conditions (e.g. drug treated vs. control samples) (Alban et al., 2003). Optimisation of the system must always be carried out to ensure proper labelling of samples.

The aims of this project were as follow:

1) Optimisation of a native saturation DiGE protocol applied to the human liver microsome assay products.
2) Analysis of data gathered from a large scale analytical experiment comparing clozapine treated and untreated human liver microsome fractions.

3) Preparative DiGE experiment carried out with proteins of interest identified in 2) excised, digested, analysed by reversed phase LCMS and identified using Mascot.
4) Synthesis of biotinylated desmethyl clozapine for protein binding study by western blot and digestion and LCMS.

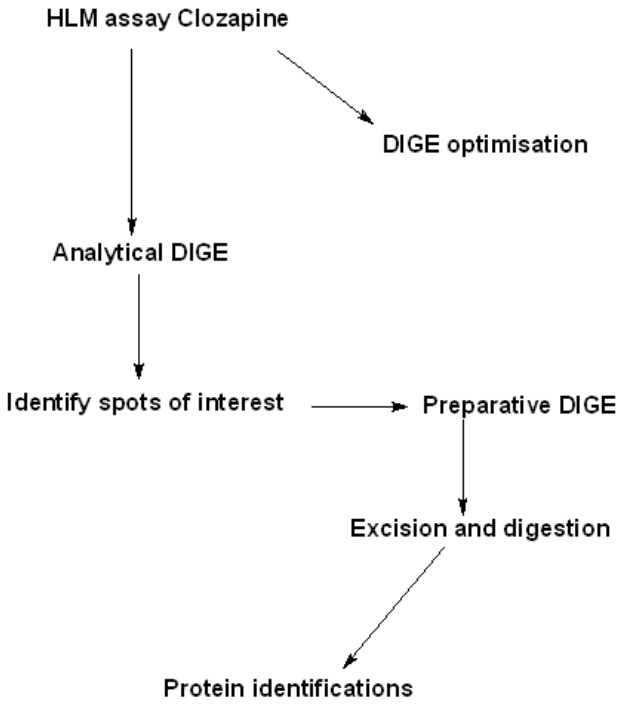


Figure 105. The DiGE workflow used in the following work. After optimisation the changes in apparent protein abundance were analysed by analytical DiGE. Those proteins with markedly altered fluorescence were then subjected to a preparative DiGE experiment for their subsequent identification, excision and tryptic digestion.

5.2 Introduction

5.2.1 DiGE

DiGE allows for the quantification of protein differences between samples of distinct origin whist correcting for problems that arise from gel to gel variation inherent to 2d-gel electrophoresis experiments (Karp et al., 2008; Hrebicek et al., 2007; Viswanathan et al., 2006). The technique is based on

the use of functionalised fluorescent dyes with distinct excitation and emission properties that can be multiplexed on a single 2d gel. The technique has been shown to have dynamic range of at least 4 orders of magnitude (Kolkman et al., 2005).

DiGE minimal labelling sensitivity is about 1ng per spot whereas the saturation dyes detection can be as sensitive as 0.1ng per spot (Shaw et al., 2003). The saturation dyes are spectrally resolvable and have a maleimide reactive group that forms a covalent bond with the sulfhydral group of cysteine. Some protein is lost by precipitation and proteins can be preferentially labelled by either of the dyes (Shaw et al., 2003). This can be compensated for by adding extra replicates and switching the Cy dye used for treated and untreated samples in each ; as was carried out in the following work. A pooled sample, comprising equal parts of all samples analysed, is used to normalise spots across gels (Wheelock et al., 2006). For preparative gels the Cy3 dye is used at much higher concentration for the visualisation of much greater quantities of proteins followed by their subsequent proteolytic digestion and extraction.

A typical DiGE experiment calls for the denaturation and reduction of protein samples, using tri(2-carboxyethyl)phosphine (TCEP), in order to allow for maximal labelling. In the following experiments labelling was carried out on native proteins. This was done in order to selectively label only proteins with surface accessible and reactive cysteines. The labelling is in effect analogous to the protein-drug adduct formation that also occurs at surface accessible and reactive cysteine thiol groups. It is hypothesised that in the drug treated samples those adduct-bearing thiols will not react with the dye. Consequently, those same proteins in the untreated sample will appear to be of a higher intensity due to increased dye abundance.

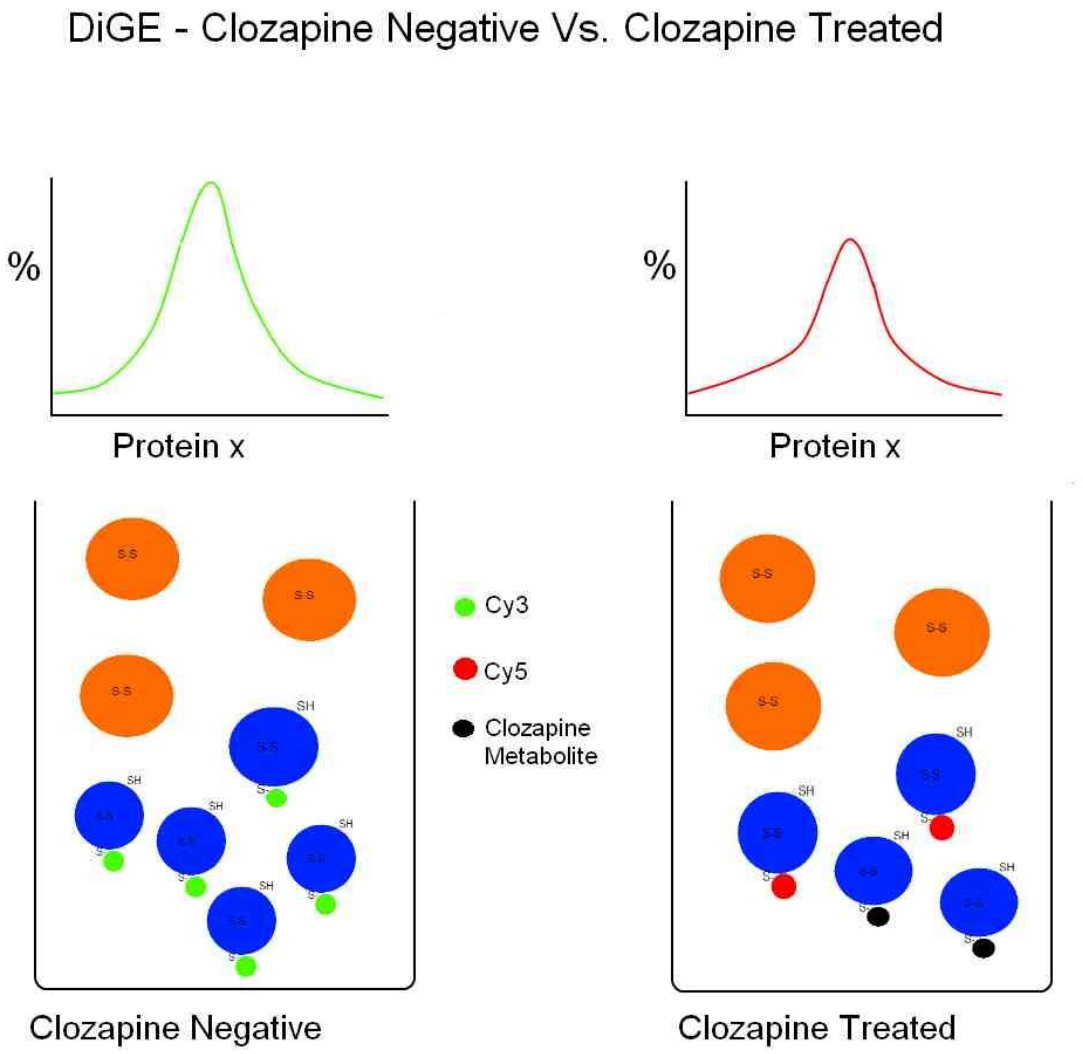


Figure 106. DiGE experiment with only two proteins. In the clozapine negative sample the blue protein (x) has surface accessible cysteine residues and thus can react with the CyDyes. The same protein in the clozapine treated sample has reacted with a clozapine metabolite; consequently fewer protein molecules were available to react with the dye. As a result the fluorescence associated with protein x is decreased in the clozapine treated sample vs. the clozapine negative sample.

Saturation labelling does not affect protein digestion or mass spectrometric analysis (Yan et al., 2002). System variations such as gel to gel variations are corrected for using the internal standards. Spot identification and matching is handled by the DIA (differential in gel analysis) software in order to minimise subjective editing and to ensure that data is consistent. The software

performs background subtraction, quantitation and normalization over the full range of gels analysed. The aim of DiGE is to reduce system variation to a point at which it can be distinguished from biological variation. Spots on the gels are compared to the internal standards to give a ratio. This ratio can be compared directly across all of the gels in the experiment. The internal standard also serves to aid the matching of spot patterns across the gels. Biological replicates are included in order to reduce the effect of biological variation between samples. Biological replicates take into account changes in protein expression between samples with identical treatments but non-identical sources (see section 1.5).

Univariate testing such as Students t-test or analysis of variance (ANOVA) allow for a statistical measurement of changes across gels. These tests provide a so called p (probability) score that describes the likelihood that there has been no change in protein concentration. A low score indicates a low probability that spots are of a similar intensity; consequently the lower the score the higher the probability that proteins are of different concentrations in the samples compared.

5.2.2 Biotinylated Desmethyl Clozapine

The synthesis of a biotinylated form of desmethylclozapine was used to effect a highly sensitive and selective method for the identification of metabolite-drug adducts by western blot analysis followed by protein digestion and LCMS.

Western blotting typically involves the separation of proteins by 1d/2d gel electrophoresis followed by probing with highly specific antibodies coupled to a detection system; usually a secondary antibody-probe molecule (Towbin et al., 1979). The western blot analysis carried out in this work did not make use of antibodies but instead a streptavidin probe conjugated to horse radish peroxidise (Strep-HRP) electro chemiluminescence (ECL) system for imaging. Streptavidin has an extremely high affinity for biotin, with a dissociation constant of 10^{-15} M (Green, 1990) forming a bond with equivalent energy to a covalent bond. The interaction is highly selective and as with antibodies also

used in affinity purifications. The sensitivity of the technique allows for the detection of protein down to about the low picogram-femtogram range.

Duplicate gels stained with Coomassie enabled the identification of proteins highlighted in the western blot analysis. These proteins were digested and recovered from the gel prior to mass spectrometric analysis. Recovery was carried out by making a 1:1 scale hardcopy of the ECL image obtained from the G:BOX. This image was placed under the Coomassie stained gels (figure 122) and used as a template for the excision as indicated in figure 122. The gels were run with the same material and in the same way as was carried out in the analytical and preparatory DiGE experiments (described in the following sections), physical inspection of the Coomassie stained gels (figure 123) reveals many features seen in the DiGE images. The number of features visible however is far fewer than observed in the images obtained from the CyDyes do to their greater sensitivity. The gels seen in figure 123 appear to be almost identical. Glutathione trapping and LC-MS based analysis allowed for the characterisation of reactive species generated via the metabolism of the biotinylated drug. Data from the experiments was used to identify potential characteristic ions that could be applied to selective precursor ion scan MS methods.

5.3 Methods

5.3.1 Optimisation of DiGE Conditions

Titration of the saturation DiGE dyes.

Clozapine treated and untreated samples were obtained from the HLM assay and cleaned up using 3 kDa spin filters. Proteins were recovered from the filter using 25 mM ammonium bicarbonate (AmBic) solution at pH 8. Saturation labelling was carried out using the following concentrations of Dyes: 2nmol, 4nmol and 6nmol. Gel images were examined to determine which dye concentration gave the best results.

5.3.2 Analytical DiGE

Analytical DiGE analysis was carried out as described in (2.1.1.13).

Briefly, three separate batches of HLM were used, each producing both a clozapine treated and untreated sample for a total of six samples. An equal amount of material from each of these was pooled to generate a 7^{th} sample, the pooled internal standard. In total 12 gels were run, gels 1-6 each contained one of the 6 treated or untreated samples labelled with Cy3 and an equal amount of the internal standard labelled with Cy5. Gels 7-12 contained the same but with the internal standard labelled with Cy3 and the samples labelled with Cy 5.

5.3.3 Preparative DiGE

A much greater amount of protein is used in preparative DiGE allowing for its subsequent recovery, digestion and analysis. The large amount of protein protects against the losses inherent to in gel digestion as well as losses by adsorption occurring during manipulation and storage of samples. A full description of the protocols used can be found at (2.1.1.14).

5.3.3.1 Analysis of DiGE Data

The DeCyder Differential in gel analysis (DIA) (GE healthcare) program was used to analyse the data obtained from the gels. The images were loaded into the software and protein spots identified, some manual corrections were made in order to ensure good correlation between the various gel images. The biological variation analysis (BVA) module was used to assign statistical values to changes in protein concentration across the gels. A table of proteins with statistically significant ($p<0.05$) decreases in protein concentration in clozapine treated vs. untreated samples was produced. It should be noted that in the experiment it was necessary to use saturation labelling as the goal was to saturate free cysteine residues. The saturation DiGE system only has two dyes, Cy3 and Cy5, compared to three dyes for minimal labelling (Cy2,

Cy3 and Cy5). This limits experimental design. In a minimal labelling experiment it is possible to have a pooled sample (an internal standard for inter-gel normalisation) as well as two potential test states (e.g. treated and untreated). In this way a pairwise comparison can be carried out. However, the saturation dyes only allow for the inclusion of a pooled sample and a single test state (treated or untreated) per gel. As a consequence, in order to obtain the same number of replicates it is necessary to produce twice as many gels. This is important in order to maintain statistical power as statistical power is directly proportional to the number of replicates being tested (Karp and Lilley, 2009; Hunt et al., 2005).

5.3.4 Biotinylated Desmethylclozapine (b-DMC)

Desmethylclozapine was reacted with pentafluorophenyl biotin to produce b-DMC. The b-DMC was recovered from solution by reversed phase chromatography, fractions were identified by UV (214 nm, 254 nm) and characterised by mass spectrometry (2.1.2.1). The purified b-DMC was metabolised in the HLM assay (2.1.1.17) and the products of the assay collected for analysis.

5.3.5 Trapping and Identification of DMC and b-DMC Metabolites

Biotinylated desmethylclozapine was metabolised by human liver microsomes and trapped with glutathione (2.1.1.15). Samples reconstituted in buffer A were analysed by RP-LCMS in order to characterise b-DMC metabolite-glutathione conjugates. Collision induced dissociation (CID) was carried out on adducts in order to identify prominent fragment ions. These ions were used in order to generate a selective precursor ion scan.

5.3.6 Western Blot Analysis of b-DMC Products

The b-DMC treated sample and the untreated negative control were both separated using 2d-PAGE (2.1.1.5). Appropriately sized PVDF membranes were cut to fit the gels, proteins were transferred from gel to membrane by the application of an electric field. Blocking of the membrane with a 5% w/v solution of powdered milk was followed by labelling with an HRP-streptavidin probe. Enhanced chemiluminescence reagents (Pierce) were added and the gels were visualised in the G:BOX system (Syngene) using its Chemi blot program.

5.3.6.1 Staining, Excision and Digestion of Proteins

A sample treated with b-DMC was run on a 24 cm 2d gel using the protocol specified in (2.1.1.5) with the bind silane treatment. The gel was Coomassie stained, washed in distilled water and placed on top of a full size image taken from the western blot analysis. Spots overlaying those identified by the western blot were excised and tryptically digested (2.1.1.10). The tryptic digests were analysed by LCMS.

5.3.7 Analysis of proteins by Reversed Phase Liquid Chromatography-Mass Spectrometry (RP-LCMS)

The tryptically digested proteins were analysed by RP-LCMS as described in (2.1.2.2) for an IDA scan and (2.1.2.3) for a selective precursor ion scan (2.1.2.7). Data obtained was submitted to Mascot for searching in order to identify proteins and modifications.

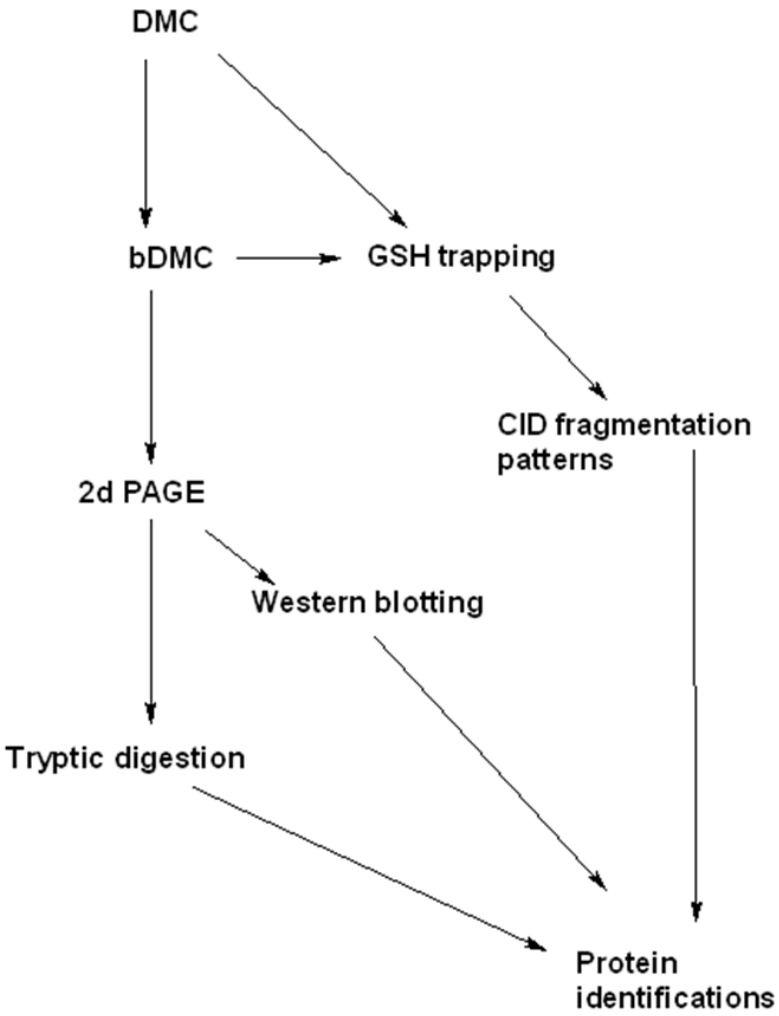


Figure 107. b-DMC workflow. DMC is converted to b-DMC. Glutathione trapping allows for the identification of metabolites. Western blotting is used to locate the metabolite-protein adducts. The corresponding regions of the 2d gel are excised and tryptically digest. The samples are then analysed by LC-MS with data from the glutathione trapping stage used to generate selective precursor ion scan methods.

5.4 Results

5.4.1 Optimisation of DiGE Protocol

Saturation DiGE required careful optimisation for each sample type to be analysed i.e. the fluor: protein ratio must be balanced correctly. Too much

fluor results in side reactions with lysine and to horizontal charge trains; underlabeling results in multiple spots in the vertical dimension.

Cy3 and Cy5 labelled gel images were produced for each of the three dye concentrations. The gels were qualitatively compared for the presence of horizontal or vertical streaking and for the degree of overlay between spots labelled with the different dyes.

The liver microsome fraction is known to have relatively high glutathione levels of between 5-10 mmol/L (Armstrong, 1987; Sies et al 1983); reduced glutathione has a reactive cysteine and as such is capable of reacting with the maleimide functionalised CyDyes. This may lead to depletion of the free dye by means of a reaction between the maleimide functional group and the sulfhydral side group of cysteine. Taking this into account it is likely that a concentration above the recommended 2 mmol would be required in order to produce optimal labelling of a liver microsome sample.

The narrow range, pH 4-7, was chosen to ensure better separation between proteins with similar isoelectric points. Due to the number or proteins in the sample it is likely that any given spot comprises many different proteins species. A consequence of this is that changes in abundance can be masked. An increase in the fluorescence of any given spot may be caused by a change in any one of its several component proteins. A multiple-fold change in the abundance of a very low abundance protein may be lost as background noise when the presence of a much higher abundance protein masks the signal.

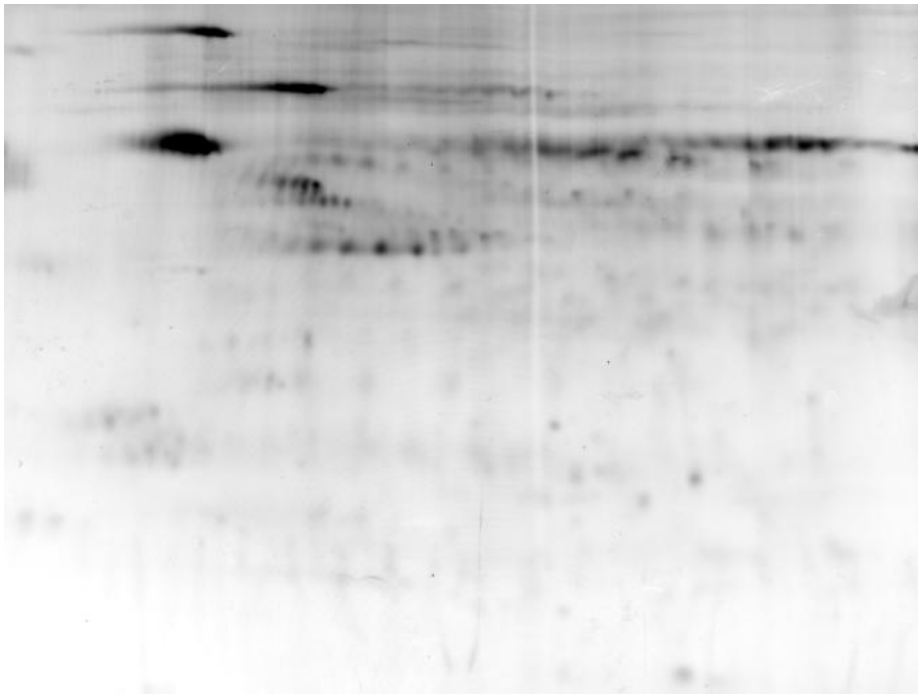


Figure 108. 2 nmol CyDye. There is some evidence of vertical streaking possibly due to underlabeling of proteins. The background noise is relatively high, making identification of some spots difficult.

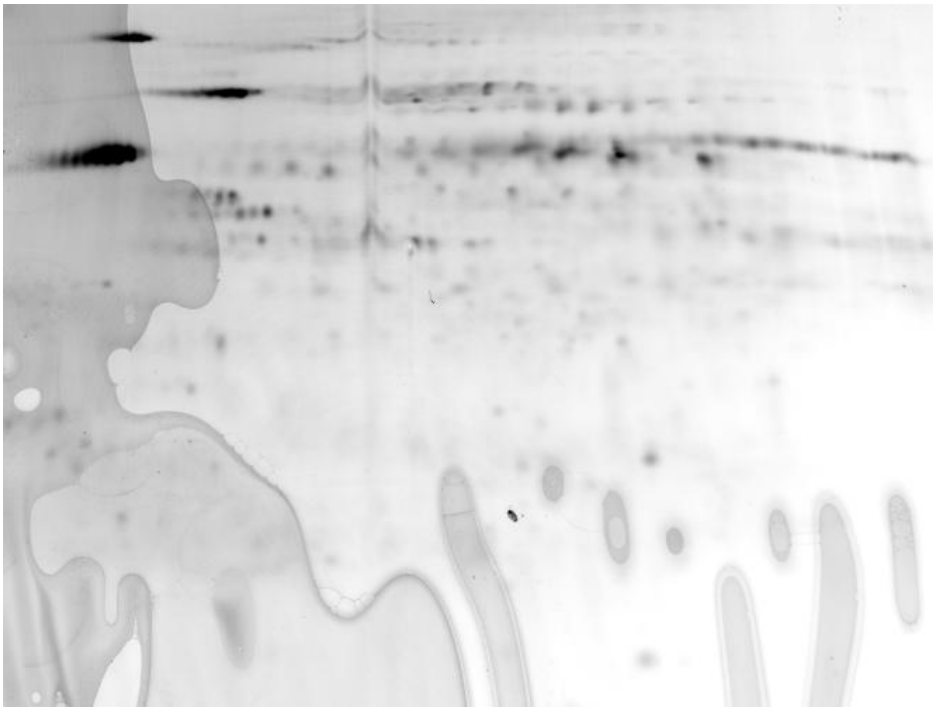


Figure 109. 4 nmol Cy Dye. The gel was partially damaged when removed from the glass plates causing some warping. Spot intensity is better with less background interference. Better resolution of spots is apparent.

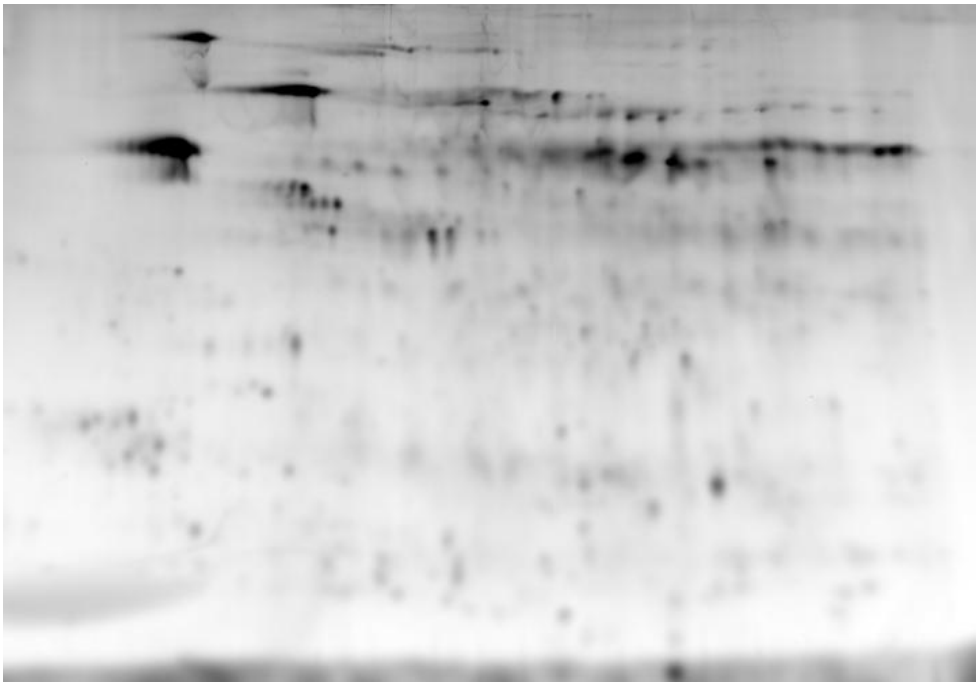


Figure 110. 6 nmol CyDye. Resolution of proteins is further improved and the number of spots visible has increased again with background noise becoming even less apparent.

A general improvement in spot visibility and definition (resolution) is apparent from 2-6 nmol (Figures 94-96). A decrease in background noise (signal: noise) is also obvious and is by 6nmol. From these images it would appear that samples treated with 6 nmol of dye give the best results. This outcome is as predicted and is probably in part accounted for by the high glutathione content typical of liver cells (up to 5mM).

False colours were assigned to the Cy3 (yellow-green) and Cy5 (red) images from each gel in order to compare the overlay of proteins labelled with each of the dyes. Overlaid spots appear as an intense yellow, green and red spots are either poorly overlaid or have one dye at a higher concentration than the other.

Across the three concentrations of dye the overlay is generally good with a trend towards improvement as the concentration increases (Figures 97 and 98). Some failure to overlay is explained by the sensitivity of the electrophoresis to the different dye structures that occurs around the 20-30 kDa mass range and is expected to cause misalignment in roughly 1% of the protein spots (DiGE product booklet, Amersham).

The qualitative analysis clearly indicated that the highest dye concentration (6 nmol) gave the best results and as such was used in the analytical DiGE experiments.

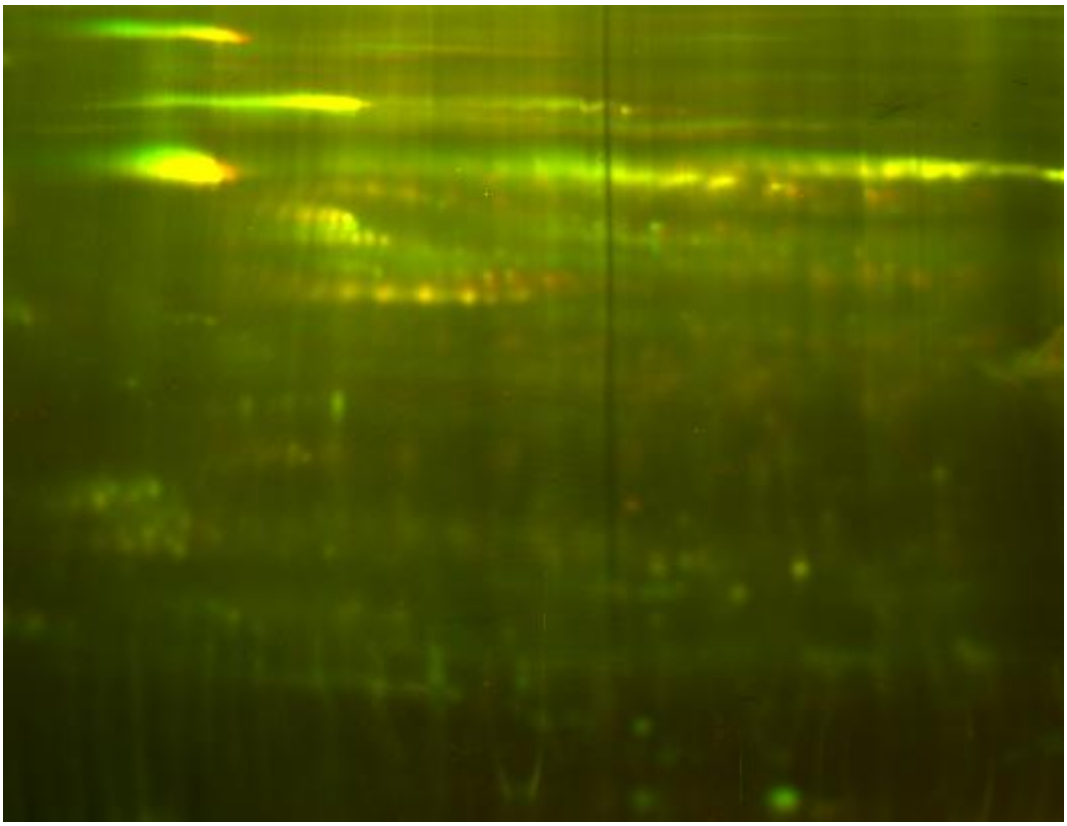


Figure 111. 2 nmol composite. Yellow-green represents Cy3, Red represents Cy5. The intense yellow is the result of the combination of Cy3 and Cy5.

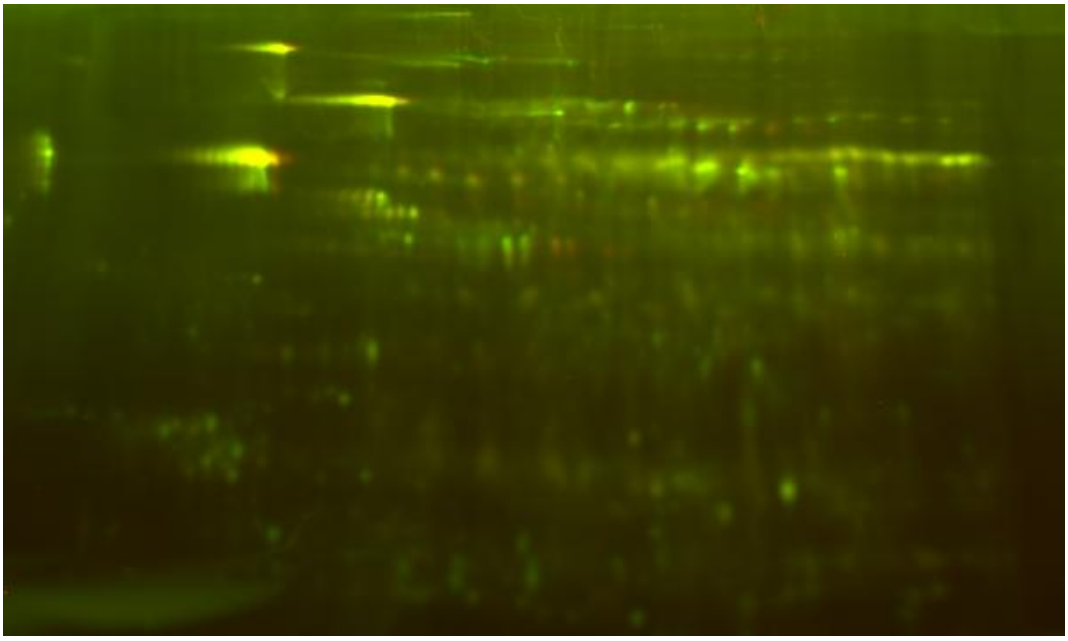


Figure 112. 6 nmol composite. The superior resolution and overlay of proteins is apparent when compared to the 2 nmol composite image.

5.4.2 DiGE of Clozapine Treated Microsomes Vs. Untreated Microsomes

Biological variation analysis (BVA) of the 12 gels yielded the following table of results (table 18). A statistical cut-off p-value of 0.05 was applied to spots showing decreased intensity in the drug treated vs. control samples. Intensity changes range from a factor of 1.16 to 2 were observed within this statistical cross-section. Matching of spots across all 12 gels was a difficult task due to the number of protein spots present and physical differences between the gels. Warping occurred in several of the gels; the bottom portion of the gel became markedly wider than the top, resulting in trapezoid shaped gels. The software has a warping feature that was applied in order to correct for this. There was some variation in the number of spots detected (from 1884-2349 per gel with 1244-1565 matched to the master gel); due to the high number of gels manual spot matching was limited. It should be noted that a principal component analysis for the identification of outliers was not carried out but should have been. Outliers may either represent proteins with strong differential expression or be indicative of mismatched spots.

A consequence of the imperfect matching of spots across the gel series is that some of the protein abundance changes are supported by less than 12 gels thereby reducing their statistical validity. Identification of statistically significant changes is potentially the first step in locating proteins modified by the clozapine metabolites. By running a preparative gel with the same samples, picking the spots identified in this study, digesting and analysing them by LCMS it should be possible to find modified proteins.

A major drawback with this technique is the imperfect separation of proteins across either of the two dimensions. Proteins with similar mass and similar isoelectric points are likely to migrate to approximately the same part of the gel. Consequently each of the spots identified is likely to contain more than a single protein. This can lead to the masking phenomenon described previously (4.4). Additionally it has been demonstrated that DiGE can underestimate the changes in protein concentration as compared to western blot analysis (Hannigan et al., 2007).

Table 10. This table represents all of the spots that showed a decrease in intensity (Clozapine treated vs. Untreated). Only changes with a p-value of <0.05 were accepted (representative of a 95% confidence that there is a real change in intensity and not a false positive).

Master spot	Gels	p-value	Intensity	Excision number
396	12(12)	0.02	-1.32	1
432	11(12)	0.017	-1.39	2
437	11(12)	0.00025	-1.6	3
438	10(12)	0.0082	-1.34	4
491	12(12)	0.034	-1.18	5
498	12(12)	0.035	-1.47	6
504	12(12)	0.034	-1.19	7
575	11(12)	0.028	-1.4	8
601	11(12)	0.062	-1.25	9
612	6(12)	0.0087	-1.16	10

615	10(12)	0.0083	-1.3	11
624	10(12)	0.031	-1.3	12
707	10(12)	0.046	-2	13
748	8(12)	0.0079	-1.8	14
757	11(12)	0.032	-1.25	15

5.4.3 Preparative DiGE

The preparatory gel was imaged as described in (2.1.1.13.5) as the preparatory gel only has a single dye (Cy3) this image was loaded as both channels (Cy3 and Cy5) into the DIA module of DeCyder. After DIA processing the data was loaded into the BVA module with the analytical DiGE data to allow for matching of spots. The spots identified in the previous table were added to the pick list in the DeCyder software. The program records the coordinates of the target spots and communicates the information to the Ettan Spot Handling Workstation (Amersham Biosciences, UK) which then physically removes the gel pieces for further analysis. The gel pieces were then subjected to tryptic digestion (2.1.1.10). Analysis of the 15 excised spots by RP-LCMS using both an IDA approach and the more selective PI359 scanning mode was carried out. Due to material constraints a single preparatory DiGE gel was created.

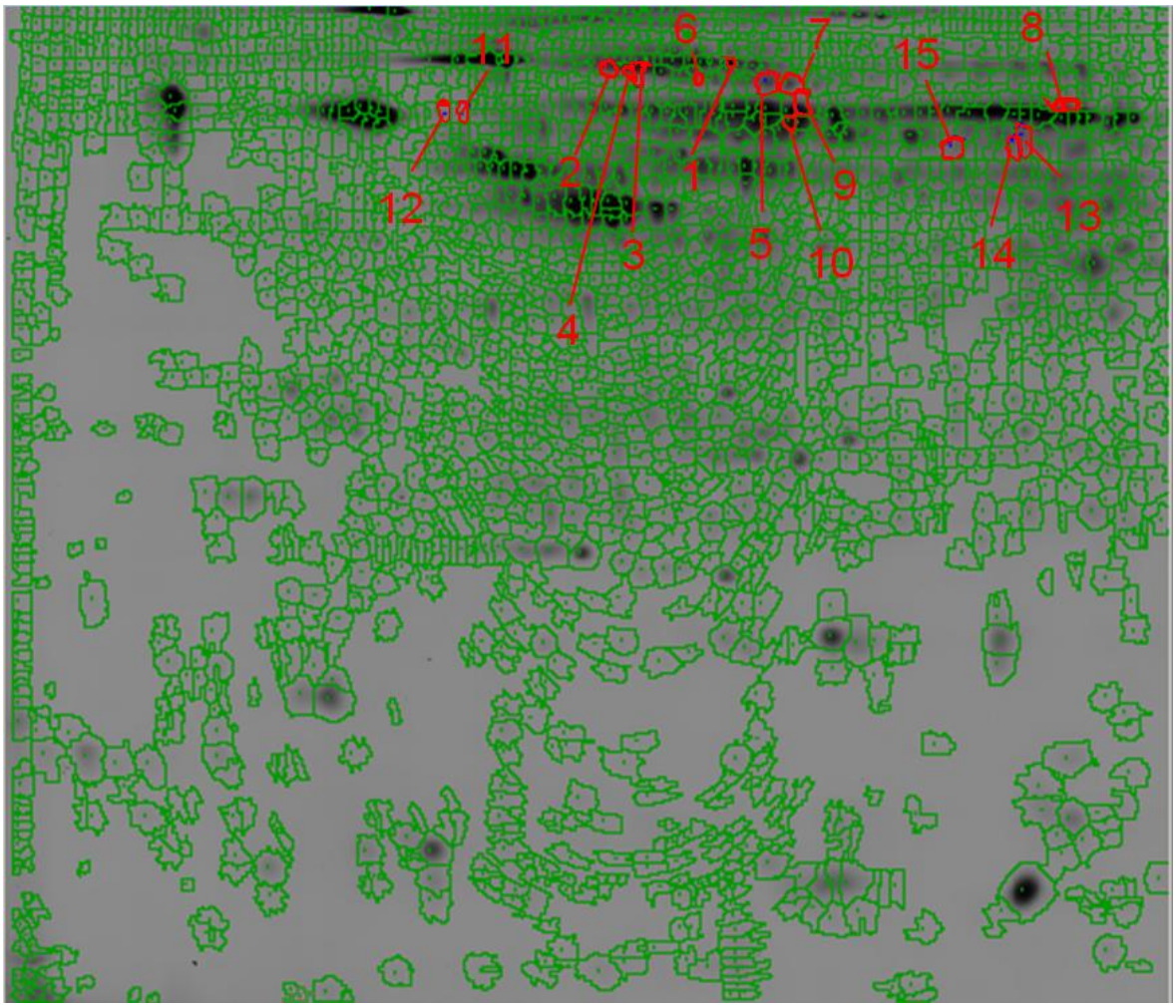


Figure 113. An image of the preparative gel; Spot identification was carried out in the DeCyder DIA and BVA modules. Spots added to the pick list are designated in red.

As previously predicted the number of proteins detected greatly outnumbers the number of spots excised. The excision of spots was slightly altered in this approach in order to compensate for the drift in mass imparted by the clozapine metabolite. This adaptation lead to the use of larger than normal spots (2mm picker head) and the increased number of protein identifications. A likely consequence of the larger spots is the inclusion of more proteins in the gel section taken and analysed.

The mass spectrometric data was searched using Mascot; the parameters were set as previously described in (2.1.4.3). The PI359 scan when used on such small populations of proteins should have enhanced effectiveness due to a reduced likelihood of false positive precursor ions. From the data 18 proteins were

identified compared with the 147 proteins identified in the IDA type experiments . These numbers exclude trypsin and the various keratins that were likely contaminants picked up during sample handling and preparation. From these results one of two things is happening, either selectivity is being increased or sensitivity is being reduced. The lack of adduct identification seems to indicate that the latter is more likely.

It was not possible to detect the presence of any metabolite-peptide adducts in the mass spectrometric data; including both the PI359 and IDA scanning methods. There could be several reasons for this i) The data gathered from the analytical gels may not indicate the presence of protein adducts; ii) a relatively low abundance modified protein may be further depleted in the gel digestion and extraction steps iii) the adduct may not be of the same character as those searched for.

It is possible that 12 gels does not give enough statistical significance to identify spots with changes of such a low degree (1.2-2 fold); in order to determine the statistical significance of the experiment with respect to making false negative errors a power analysis can be carried out. The power analysis requires the effect size to be measured (the minimum difference between two states i.e. control and treated) the significance level (typically 0.05), sample size (replicates) and the standard deviation observed in each group (control vs. treated). It is possible to carry out a post-hoc calculation to determine the statistical power of any given experiment but it is generally believed by statisticians that the results have little meaning or value. Work by Hoenig and Heisey showed that post-hoc calculation of statistical power provides results that are directly proportional to the p-value and so provides no new information (Hoenig and Heisey, 2001). The power analysis only carries real meaning when carried out prospectively and can be used to determine the number of samples required to detect a change of a given size with a particular statistical significance (usually set at 80%). The DeCyder software used in this work did not offer the functionality required to provide this information. Other software however e.g. Progenesis SameSpots (Non-linear Dynamics) can carry out power analysis calculations. An underpowered experiment would produce an high degree of false negatives leading to potentially important proteins going undetected. In this case it would be necessary to either increase the number of

gels examined, which would considerably add to analysis time and costs, or to consider the possibility that the degree of modification is beneath that detectable by the DiGE approach. Relatively little is understood about the mechanisms of drug-protein adduct formation as a generalised concept. It is entirely possible that metabolites could have affinity for particular protein targets (Labenski et al., 2009; Bartolone et al., 1989; Pumford et al., 1990; Nakayama et al., 2010; Fisher et al., 2011). In this case the high concentration of drug used in these studies should ensure a considerable amount of modified protein. If this were the case then this DiGE based approached should easily identify a subsequent depletion of the unmodified protein in the sample exposed to the reactive metabolites.

As to the second possibility, it is well known that losses occur when carrying out in gel digestion and extraction of peptides. The factor of depletion is estimated to be around 15-30% with subsequent handling steps seeing further losses of 10-15% caused by adsorption of peptides to plastic surfaces (Speicher et al., 2000). With saturation DiGE's lower detection limit of around 0.1 ng of material (Shaw et al., 2003) this would mean that after digestion and handling losses (after handling: 0.75 ng) there would be around 15 fmol of protein (for a protein of 50 kDa); well within the detection limits of a Qtrap instrument (Wilm et al., 1996). Additionally, the proteins selected for digestion and extraction were of relatively high intensities indicating that protein abundance was fairly high and as such are unlikely to be at the lower end of the DiGE detection limit.

The formation of a metabolite not detected in the early glutathione trapping experiments is entirely possible and with a mass not added to the Mascot search parameters peptides bearing these modifications would be effectively invisible. However, the metabolites detected and added to the Mascot database comprised the sum of all adducts detected for clozapine based on a thorough search of the literature (Fisher et al., 1991; Jian et al., 2009; Jegouzo et al., 1999; Van Leeuwen et al., 2005; Inoue et al., 2009; MacDonald et al., 2011; Zhu et al., 2007; Zhang and Yang, 2008; Yan et al., 2005). Whilst it is possible that other metabolites exist, the major metabolite is likely amongst those included in the search parameters.

5.4.3.1 Protein Identifications

Proteins identified from both the IDA and PI359 methods were catalogued and compared. The following tables (tables 19 and 20) represent proteins with relatively high MOWSE scores and good protein coverage. The false discovery rates were below 5% for all experiments.

Table 11. High MOWSE scoring proteins identified in the preparative DiGE experiment. These results were taken from the IDA experiments and have an associated false discovery rate of 4.3%.

Spot	Mascot ID	Protein name	Mowse score	Coverage (%)	MW kDa	pI
1	CPSM_HUMAN	Carbamoyl-phosphate synthase [ammonia]	475	31	165	6.3
1	MYH9_HUMAN	Myosin-9	202	11	160	5.5
2	GRP78_HUMAN	78 kDa glucose-regulated protein	438	40	78/72	5.1
2	PDIA4_HUMAN	Protein disulfide-isomerase A4	129	21	72	5
2	HS71L_HUMAN	Heat shock 70 kDa protein 1L	110	7	70	5.7
2	HSP7C_HUMAN	Heat shock cognate 71 kDa protein	102	4	71	5.4
3	GRP78_HUMAN	78 kDa glucose-regulated protein	595	48	72	5.1
3	PDIA4_HUMAN	Protein disulfide-isomerase A4	409	32	73	5
3	HSP7C_HUMAN	Heat shock cognate 71 kDa protein	232	11	71	5.4
4	GRP78_HUMAN	78 kDa glucose-regulated protein	437	33	78	5.1
4	PDIA4_HUMAN	Protein disulfide-isomerase A4	293	32	73	5
4	HSP7C_HUMAN	Heat shock cognate 71 kDa protein	112	13	71	5.4
5	HSP7C_HUMAN	Heat shock cognate 71 kDa protein	356	21	71	5.4
5	HSP71_HUMAN	Heat shock 70 kDa protein 1	166	17	70	5.5
5	GRP75_HUMAN	Stress-70 protein, mitochondrial	149	20	74	5.5
6	GRP75_HUMAN	Stress-70 protein, mitochondrial	237	29	74	5.9
6	ANXA6_HUMAN	Annexin A6	217	24	76	5.4
6	HSP7C_HUMAN	Heat shock cognate 71 kDa protein	134	12	71	5.4
7	ANXA6_HUMAN	Annexin A6	260	25	76	5.4
7	GRP75_HUMAN	Stress-70 protein, mitochondrial	184	27	74	5.9
7	NCPR_HUMAN	NADPH--cytochrome P450 reductase	128	9	77	5.4
7	NDUS1_HUMAN	NADH-ubiquinone oxidoreductase 75 kDa subunit	126	15	79	5.4
7	HSP71_HUMAN	Heat shock 70 kDa protein 1	113	11	70	5.5
8	HSP71_HUMAN	Heat shock 70 kDa protein 1	82	8	70	5.5
9	ALDH2_HUMAN	Aldehyde dehydrogenase	224	24	56/51	6.6
9	PDIA3_HUMAN	Protein disulfide-isomerase A3	221	29	57	6
9	EST1_HUMAN	Liver carboxylesterase 1	120	11	62	6.1
10	ALDH2_HUMAN	Aldehyde dehydrogenase	170	16	56/57	6.6
10	PDIA3_HUMAN	Protein disulfide-isomerase A3	132	20	57	6
10	EST1_HUMAN	Liver carboxylesterase 1	100	8	62	6.1
11	PDIA3_HUMAN	Protein disulfide-isomerase A3	200	28	57	6
11	ALDH2_HUMAN	Aldehyde dehydrogenase	147	21	56	6.6
11	EST1_HUMAN	Liver carboxylesterase 1	82	11	62	6.1

12	EST1_HUMAN	Liver carboxylesterase 1	168	14	62	6.1
14	EST1_HUMAN	Liver carboxylesterase 1	147	12	62	6.1
15	EST1_HUMAN	Liver carboxylesterase 1	170	20	62	6.1

Table 12. High MOWSE scoring proteins from the preparative DiGE experiment. These results were taken from the PI359 experiments and have a false discovery rate of 3.5%.

Spot	Mascot ID	Protein name	MOWSE score	Coverage (%)	MW kDa	pI
1	PDIA1_HUMAN	Protein disulfide-isomerase	101	11	57	4.8
1	NUCB1_HUMAN	Nucleobindin-1	92	8	54	5.2
1	PDIA6_HUMAN	Protein disulfide-isomerase A6	63	7	48	5.0
2	GRP78_HUMAN	78 kDa glucose-regulated protein	528	34	72	5.1
2	HS71L_HUMAN	Heat shock 70 kDa protein 1L	113	6	70	5.8
2	HSP72_HUMAN	Heat shock-related 70 kDa protein 2	85	2	70	5.6
2	HSP7C_HUMAN	Heat shock cognate 71 kDa protein	84	3	71	5.4
3	GRP78_HUMAN	78 kDa glucose-regulated protein	228	23	72	5.1
3	PDIA4_HUMAN	Protein disulfide-isomerase A4	178	8	73	5.0
4	GRP78_HUMAN	78 kDa glucose-regulated protein	153	14	72	5.1
4	PDIA4_HUMAN	Protein disulfide-isomerase A4	150	8	73	5.0
5	GRP75_HUMAN	Stress-70 protein, mitochondrial	283	17	74	5.9
5	HSP7C_HUMAN	Heat shock cognate 71 kDa protein	283	20	71	5.4
5	HSP72_HUMAN	Heat shock-related 70 kDa protein 2	217	13	70	5.6

5	GRP78_HUMAN	78 kDa glucose-regulated protein	153	6	72	5.1
6	GRP75_HUMAN	Stress-70 protein, mitochondrial	105	8	74	5.9
7	GRP75_HUMAN	Stress-70 protein, mitochondrial	86	7	74	5.9
8	NCPR_HUMAN	NADPH--cytochrome P450 reductase	56	1	77	5.4
9	PDIA3_HUMAN	Protein disulfide-isomerase A3	110	8	57	6.0
9	ALDH2_HUMAN	Aldehyde dehydrogenase, mitochondrial	72	13	56	6.6
9	EST1_HUMAN	Liver carboxylesterase 1	44	2	62	6.2
11	ALBU_HUMAN	Serum albumin	83	6	69	5.9
11	ALDH2_HUMAN	Aldehyde dehydrogenase, mitochondrial	62	12	56	6.6
11	EST1_HUMAN	Liver carboxylesterase 1	51	2	62	6.2
12	EST1_HUMAN	Liver carboxylesterase 1	47	3	62	6.2
14	CES1P_HUMAN	Putative inactive carboxylesterase 4	42	5	31	7.8
15	EST1_HUMAN	Liver carboxylesterase 1	56	4	62	6.2

Liver carboxylesterase (EST1), Protein disulfide isomerases (PDIA1, 3 and 4) aldehyde dehydrogenase (ALDH2) and several heatshock proteins were detected in multiple gel spots.

5.4.4 Glutathione Trapping of Desmethyl Clozapine (DMC) and Biotinylated-DMC (b-DMC)

Synthesis of b-DMC was a straightforward reaction followed by purification by RP-LCMS and characterisation by mass spectrometry.

The neutral loss of 129 (NL129) scan was again used in conjunction with the GSH trapping assay in order to identify the major metabolites of DMC and b-DMC. The mechanism of adduct formation is likely through an intermediate nitrenium ion and subsequent electrophilic attack on the nucleophilic sulphur molecule of glutathione (Uetrecht, 1992; Williams et al., 2003). MS/MS data collected from the DMC sample was collected (figure 114) and interpreted (Figure 116). The

spectra collected contained a series of ions consistent with the CID fragmentation of the proposed glutathione-metabolite adduct. It was noted that the ion present at 345.1m/z was the equivalent of the ion used as part of the precursor scanning approach for the identification of clozapine metabolite conjugates (359 m/z).

Analysis and interpretation of the data collected from the b-DMC sample (figures 117 and 119) indicates that the addition of a biotin tag does not considerably alter the route of metabolism and adduct formation when compared to both unmodified DMC and clozapine. A range of ions analogous to those found in both clozapine and DMC (mass shifted for the biotin tag) were identified (figure 119). These included a marker ion, with a m/z value of 571.1, analogous to the previously identified ions representing the drug metabolite with the added mass of sulphur.

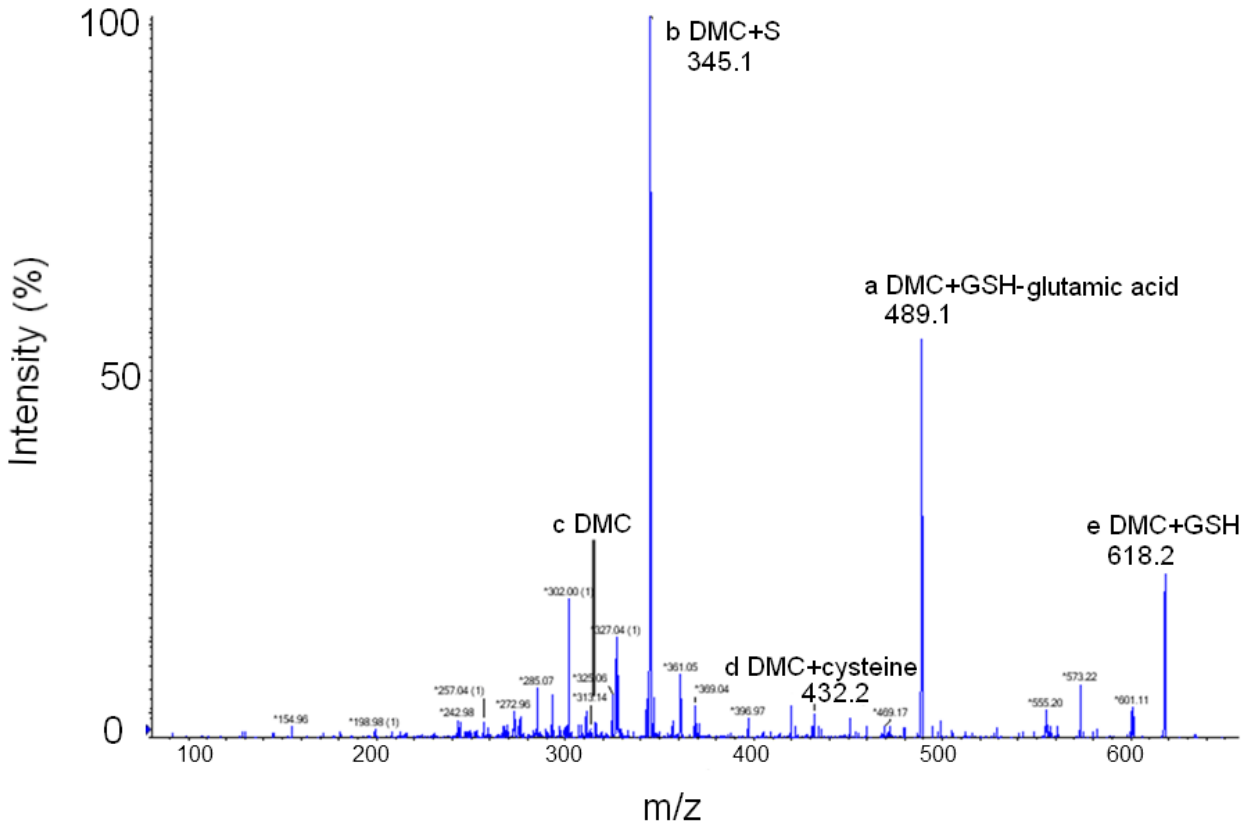


Figure 1134. MS/MS scan of GSH trapped DMC metabolite identified using the NL129 scanning approach. The predominant ion at 345 m/z is equivalent to that used in the previous PI359 scanning approach.

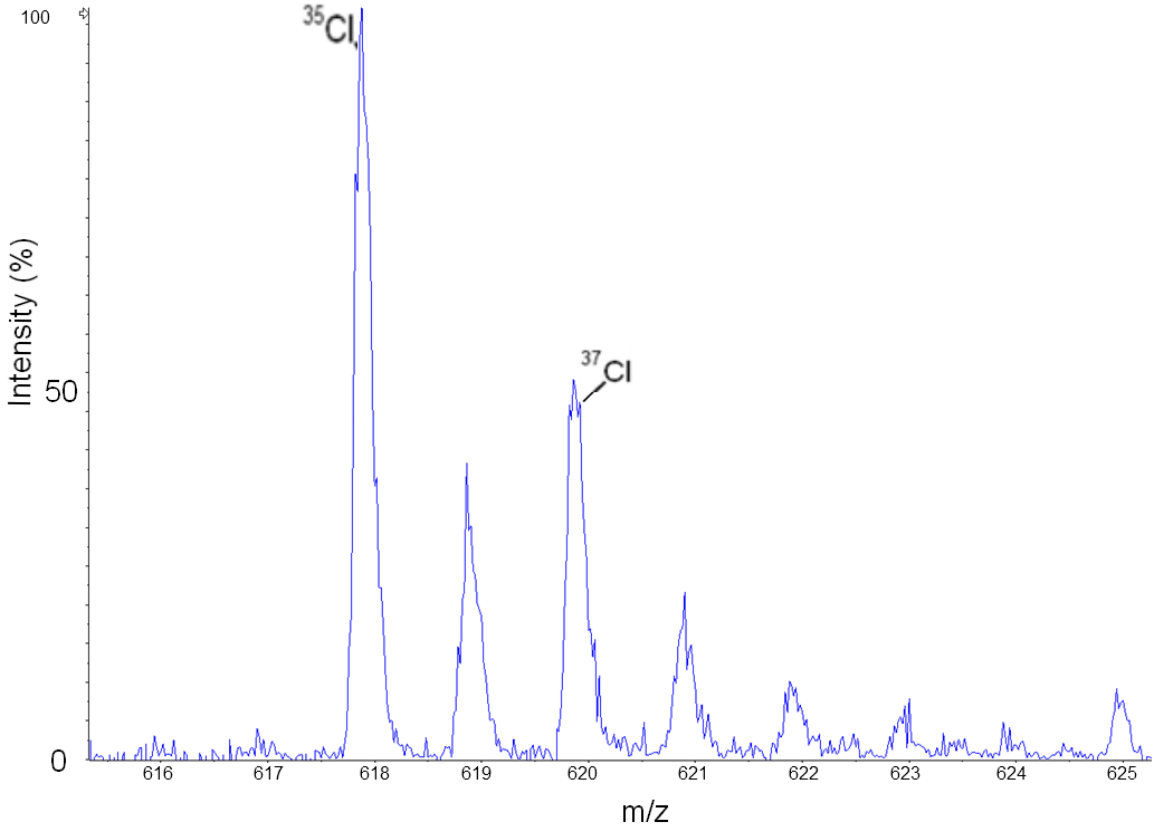
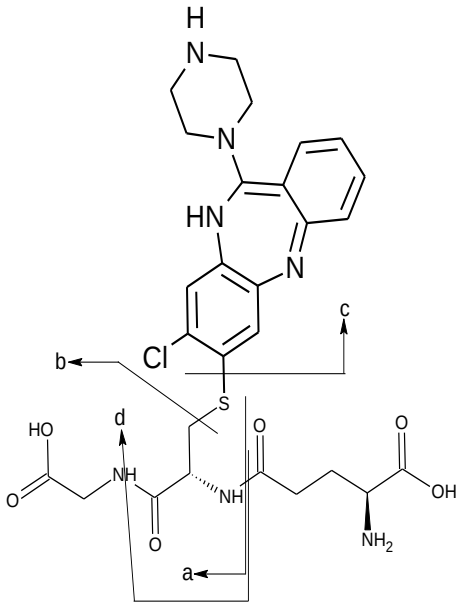


Figure 115. An enhanced resolution scan showing the GSH-DMC conjugate at m/z 618. The unusual chlorine isotope distribution pattern is clearly visible. The peak at m/z 620 represents the presence of ^{37}Cl and is of much higher abundance than the peak at 619 which contains ^{36}Cl.

DMC-GSH



Fragment Ion	m/z	Structure
a	489.1	DMC+GSH-E
b	345.1	DMC+Sulfur
c	313.1	DMC
d	432.1	DMC+Cysteine
e	618.2	DMC+GSH

Figure 116. Proposed CID fragmentation route of DMC-GSH. The ions are putatively described in the table to the right.

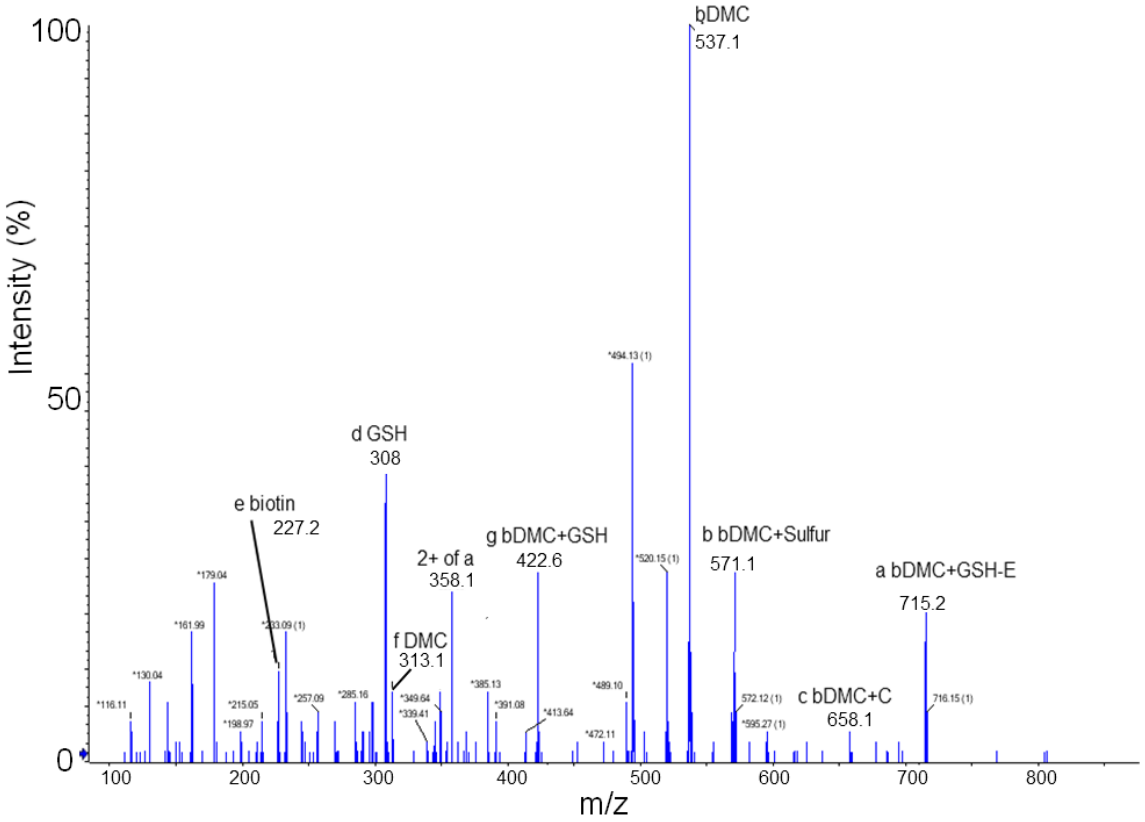


Figure 117. MS/MS scan of GSH trapped b-DMC metabolite. An information dependant acquisition spectrum taken from an NL129 PI scanning method. The ion at m/z 571.1 is analogous to the m/z 359 ion identified in the fragmentation of clozapine glutathione conjugate. The characteristic chlorine isotope pattern is not seen as the spectrum represents the fragmentation of the monoisotopic ion; hence no isotopic data outside of this is visible.

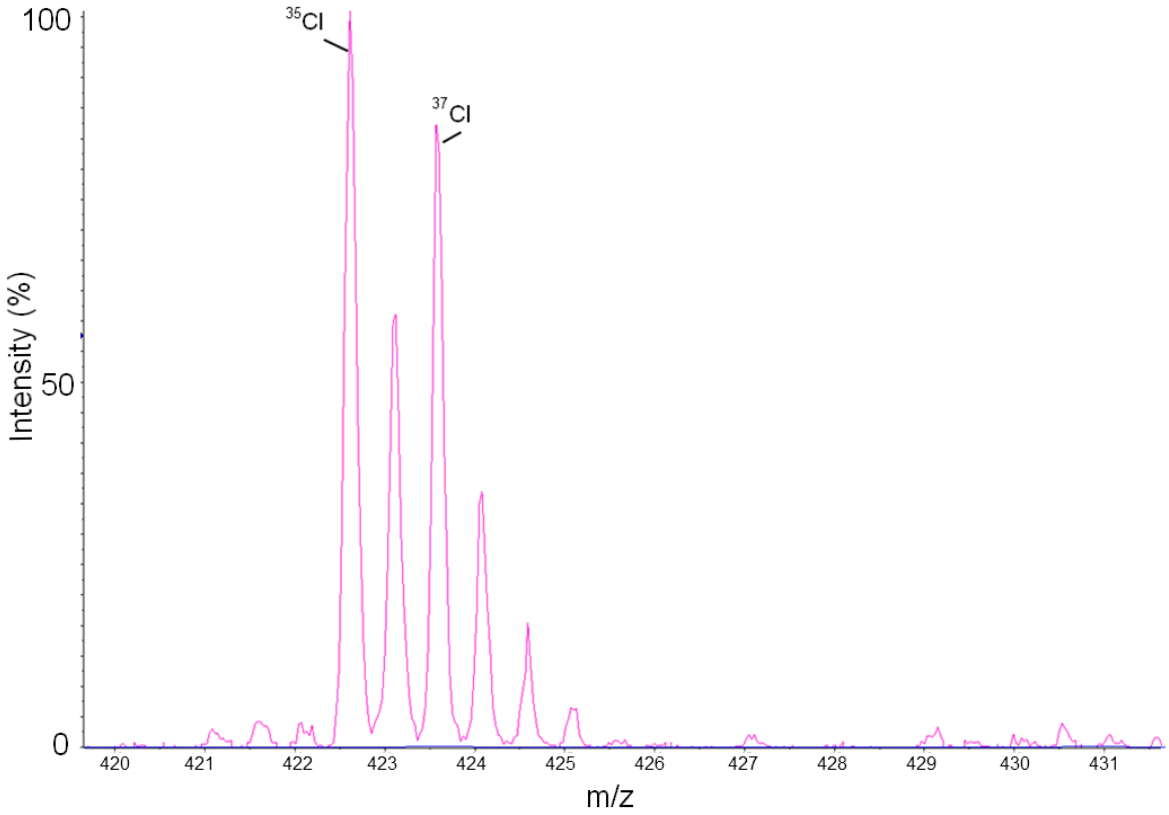
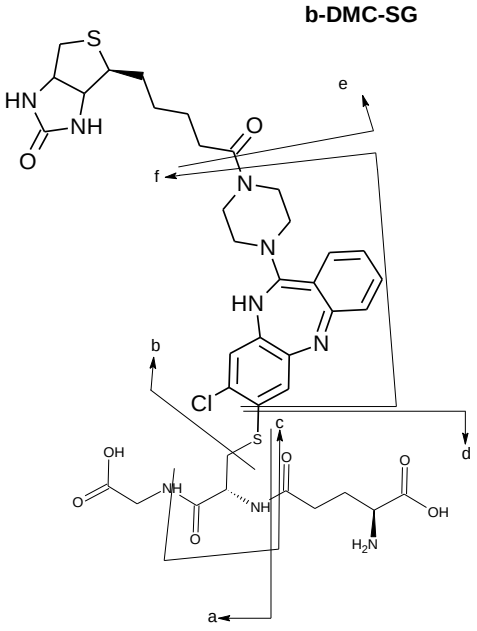


Figure 118. An enhanced resolution scan showing the parent ion from figure 117. The monoisotopic peak at 422.6 contains ^{35}Cl; the peak at 423.6, ^{37}Cl.



Fragment Ion	m/z	Structure
a	715.2	b-DMC+GSH-E
b	571.1	b-DMC+S
c	658.2	b-DMC+C
d	308.1	GSH
e	227.1	biotin
f	313.1	DMC
g	422.6	b-DMC+GSH
h	537.1	b-DMC

Figure 119. Left: proposed CID fragmentation of the b-DMC metabolite-glutathione adduct. Right: compilation of CID fragment ions from the glutathione adduct of the b-DMC metabolite.

These results demonstrate that addition of the biotin tag does not affect the fragmentation pattern of the glutathione adduct other than the addition of mass associated with the tag. It has been demonstrated that the PI359 scan is effective at identifying modified peptide fragments (3.4.5) and by extension the PI571 scan (b-DMC) should have the same capabilities.

5.4.5 2d-PAGE/Western b-DMC

Several spots are clearly visible on the control (clozapine negative) membrane (Figure 120). The same pattern is visible in the clozapine treated sample; they are of a much greater intensity than the background are well focused and appear in roughly the same position. These spots probably represent endogenous biotinylated proteins and the increased intensity is due to the high loading capacity and point focusing characteristics of the 2d gel. Non-specific binding of

the probe to the membrane would not occur in such an orderly and repeatable fashion. A relatively long exposure was used (1 hour) in order to maximise the sensitivity of the technique, this would clearly enhance the signal of both endogenously biotinylated material as well as the synthetically labelled drug protein adducts.

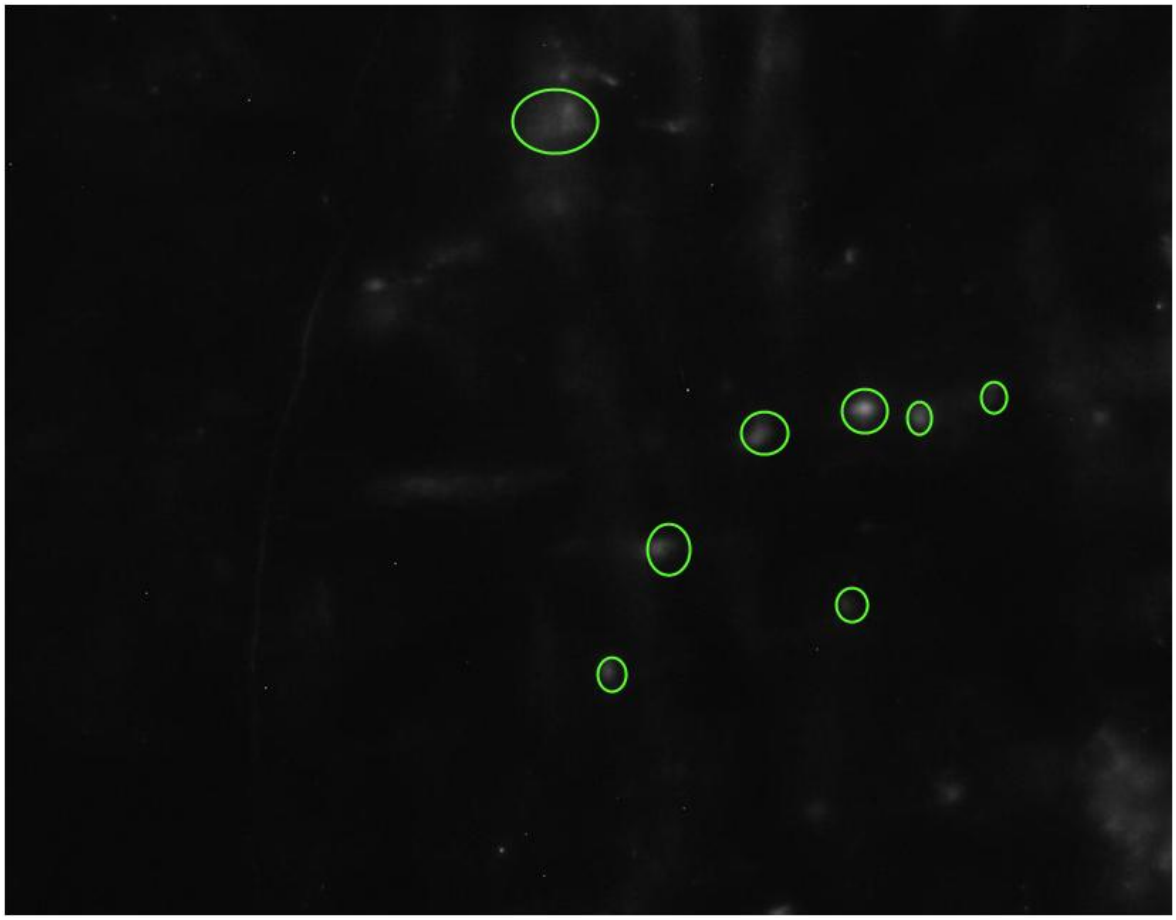


Figure 120. Image of the western blot membrane bearing the untreated sample. No non-specific interactions between strep-HRP probe and the membrane or proteins are evident. The green ellipses mark areas of high biotin concentration (most likely endogenously biotinylated protein).

The membrane representing the b-DMC spiked HLM sample (Figure 121) had an intense band of proteins of apparently the same mass range but with differing isoelectric points, most likely so-called charge trains. The locations of the non-endogenous biotinylated proteins were noted and used to excise these target proteins from coomassie stained gels.

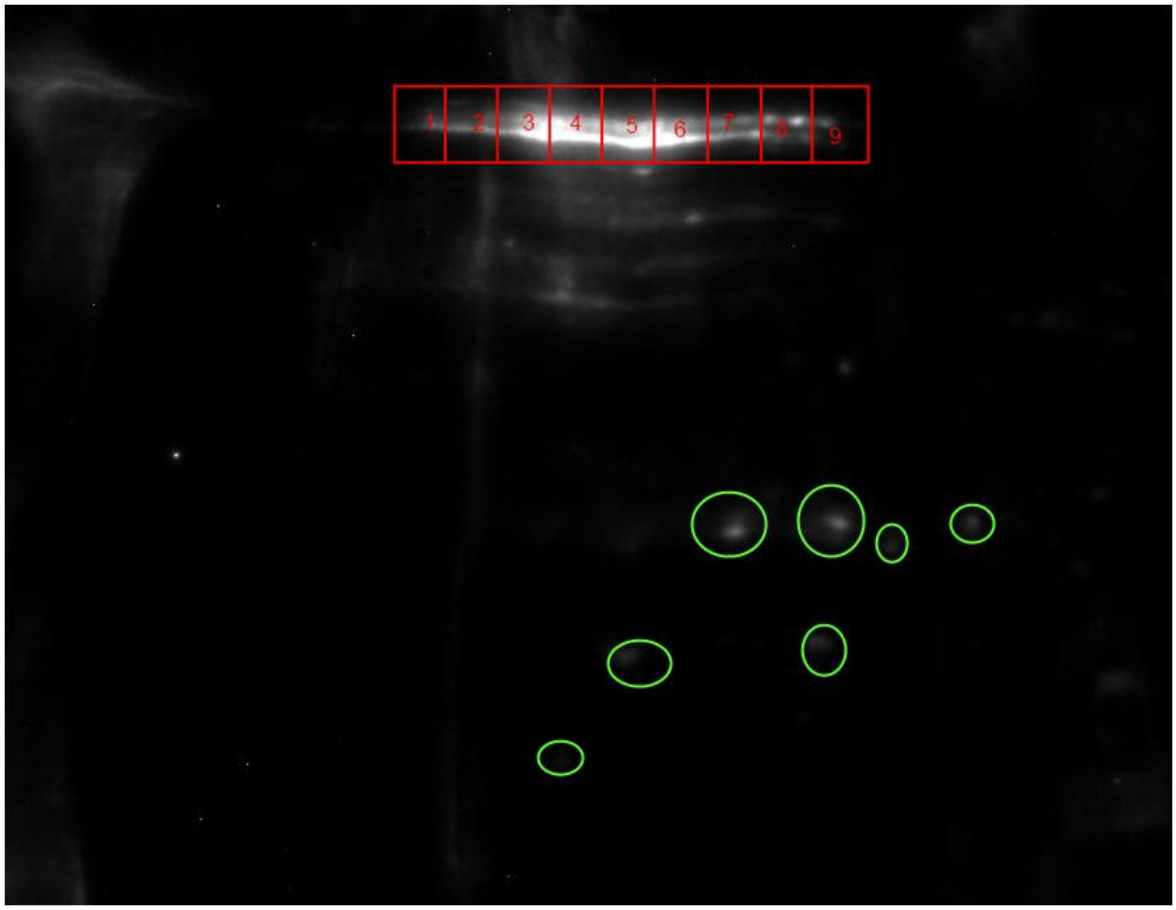


Figure 121. Image of western blot membrane bearing b-DMC treated sample. Overlaid is the region of interest used to define the excision site on the coomassie stained gels. The horizontal "smearing" of the signal may be due to protein charge trains. The green ellipses mark areas with high biotin concentration that coincide with those in figure 119.

5.4.6 2d-PAGE Coomassie Stained

From the Coomassie stained images (Figure 122) it appears that protein separation along both dimensions is optimal. There are no clear signs of horizontal charge trains or vertical mass changes indicative of sample preparation or contamination problems. Spots appear well focussed and distinct. In contrast the Strep-HRP ECL based image shows substantial horizontal streaking. The streaking occurs in the drug treated sample only i.e. the endogenously biotinylated proteins were not affected; this type of streaking can be associated with overloading of the gel. However, the sensitivity of Coomassie blue is limited to around 10-100ng (Fazekas de St. Groth et al., 1963; Neuhoff et

al., 1988) and the sensitivity of the strep-HRP system is approximately three orders of magnitude greater (10-100 pg). If the gel were overloaded it should also appear so in the Coomassie stained images, which it clearly does not.

Another explanation for this is that the protein or proteins bearing the biotinylated metabolite exist with multiple undefined post translational modifications; modifications including phosphorylation, glycosylation, acetylation and many others are common (Packer et al., 1998; Mann and Jensen, 2003; Seo and Lee, 2004) It is even possible that variable amounts of binding between the biotinylated metabolite and protein cause it to spread out along the IPG strip. Artefacts introduced in sample preparation and analysis including oxidation of cysteines, electrolytic reduction of carboxylic acid groups to aldehydes and carbamylation of nucleophilic side groups have also been shown to alter the pI of proteins (Perdivara et al., 2010; Lee and Chang, 2009; Lippincott and Apostol, 1999; Righetti, 2006).

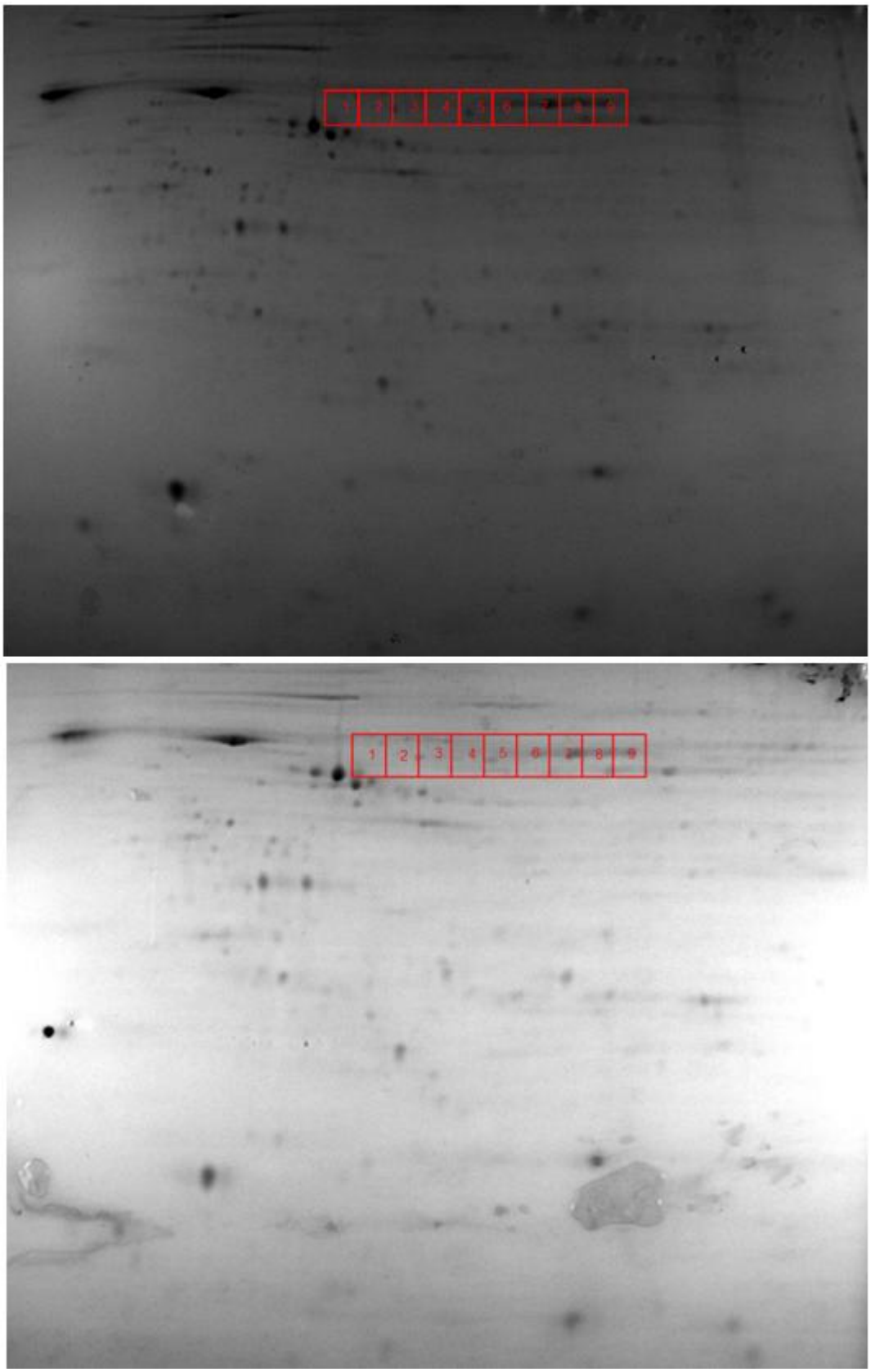


Figure 122. Coomassie stained 2d gel marked for gel excision around the region indicated by the 2d western blot. Each of the 9 boxes were excised, tryptically digested and analysed by RP-LCMS. The top image is that of the negative control, the bottom image is that of the drug treated sample.

Mass spectrometric analysis of the tryptically digested proteins was carried out; the resulting data files were searched against the Swissprot protein database (human) using the Mascot search engine as described previously (2.1.4.3).

In the b-DMC information dependant acquisition (IDA) experiments protein disulfide isomerise A (PDIA3) and a liver carboxyl esterase (EST1) were detected in 8 of the 9 fractions. Mitochondrial aldehyde dehydrogenase (ALDH2) was detected in 7 of the 9. In the precursor ion experiments PDIA3 was detected in 8 of the 9 samples, EST1 in 3 and ALDH2 in 2. MOWSE scores for PDIA3 ranged from 94 to 316 with protein coverage between 11- 34% in the case of IDA experiments and 72 to 224 with protein coverage of 3-10% in precursor ion experiments. The PI scans provide inferior MSMS data when compared to the IDA methods. This is probably due to poor detection of the precursor ions causing triggering of MSMS experiments at non-optimal points i.e. either the leading or trailing edge of their chromatographic elution peaks. No credible identifications of modified peptides have been made so far, the complexity of the samples tested is greatly simplified as a consequence of both dimensions of separation afforded by the 2d-PAGE and the 3rd dimension provided by the RP-LC.

It is possible that although the modified protein is present, as is suggested by the western blot evidence, it is of a negligible quantity and falls outwith the detection range of the mass spectrometer.

5.5 Discussion

5.5.1 DiGE Protein Identifications

The fairly large number of proteins identified in the DiGE experiment was not unexpected. On average around 10 proteins were recovered from each spot excised from the gels. Several proteins were found to be present in multiple gel pieces, these proteins also tended to have high MOWSE (molecular weight search) scores and good protein coverage. Despite this no reliable adduct identifications were made. This list of proteins was compared with the data obtained in from the b-DMC experiments to determine if there were any interesting matches.

5.5.2 b-DMC Experiments Protein Identifications

The following proteins were observed in both the preparative DiGE experiment and in the b-DMC western blot work. These proteins were considered as possible targets for metabolite-adduct formation.

Protein disulfide isomerise A (PDIA3) is a major part of the major histocompatibility complex (MHC) class 1 peptide loading complex. Critical for final antigen conformation and exports from the endoplasmic reticulum to cell surface. PDIA3 acts as a chaperone ensuring the correct folding or isomerisation of nascent proteins through the regulation of disulfide bonding (Laboissiere et al., 1995). Disruption of proper folding is clearly a danger to cell survival and is implicated in disease (Dobson, 2001).PDIA3 is found in close proximity to the cytochrome enzymes implicated in the formation of reactive metabolites meaning that even metabolites with a short half life would have the chance to attack the enzyme.

Liver carboxylesterases (including EST1) are responsible for the metabolism of carboxylic esters into alcohols and carboxylates(Brzezinski et al., 1994; Schindler et al., 1998; Pindel et al., 1997). Involved in the metabolism of drugs with ester or amide bonds, this enzyme is abundant in the liver and fits the profile described in the western blot .

Mitochondrial aldehyde dehydrogenase (ALDH2) functions to detoxify aldehydes (Wang et al., 2009; Jackson et al., 2011) most notably acetaldehyde produced by metabolism of alcohol. Acetaldehyde can also be produced endogenously during lipid peroxidation (Esterbauer et al., 1991), glycation and amino acid oxidation (Anderson et al., 1997). Acetaldehyde has an LD50 about 10 times lower than alcohol, it is a highly active electrophilic molecule and can form adducts with amino, hydroxyl and sulfhydral groups of proteins thereby altering structure, function or elimination.

Irreversible modifications of these proteins leading to a disruption of their functions or to their subsequent proteolytic degradation could result in cellular damage, death or autoimmune reactions (Ohsawa et al., 2003; Smith et al., 1993; Furst et al., 1997; Muller et al., 2011). Clozapine toxicity has no known

connection with these proteins but the detection of adducts would indicate that the proteins are susceptible to electrophilic attack.

The identification of proteins present within each of the gel sections still does not guarantee discovery of the protein-drug adduct. The best represented protein is by no means the most likely candidate. Good sequence coverage and high scoring matches are simply likely to represent abundance. Although some of these proteins mentioned would make for interesting targets no direct evidence was found for their involvement in protein adduct formation in this case.

5.5.3 Selective Protein Adduct Formation

As previously discussed the formation of protein adducts is probably more selective than first thought. The local environment of nucleophilic centres can influence their reactivity (Zhang and Dixon, 1993). Amino acids with basic sidegroups can dramatically reduce the pKa of neighbouring thiol sidegroups of cysteines. The thiolate anions have much greater reactivity than the sulfhydral group and as such are more likely to form adducts.

The primary structure of proteins identified in the b-DMC study were analysed using the program motif_HUNTER (http://proteotools.pharmacy.arizona.edu/proteotools/motif.jsp). The occurrences of the motifs KK, K?K, CH, HC, CR, RC, KC, CK and overall content of each of these residues (Table 13).

KK K?K CH HC CR RC KC CK

Table 13. The presence of basic amino acid residues neighbouring cysteines is a known risk factor for protein adduct formation. The FASTA data for each of these proteins was searched for such sites. All proteins apart from ATPA human were found to posses at least one such domain.

Protein	CK/KC	CR/RC	CH/HC
ALDH2 Human	142 VL**KC**LR 147	-	-
EST1 Human	272 AG**CK**TT 277	-	282 MV**HC**LR 287
PDIA3 Human	58 GH**CK**RL 63	-	57 CG**HC**KR 62
PDIA3 Human	407 GH**CK**NL 412	-	406 CG**HC**KN 411
CH60 Human	234 GQ**KC**EF 239	444 LL**RC**IP 449	-
ATPA Human	-	-	-
FTCD Human	-	468 LA**RC**GN 473	-
FTCD Human	-	252 ET**CR**EA 257	-
FTCD Human	-	474 LA**CR**SD 479	-

For all of the proteins tested there were also multiple instances of either KK or KxK domains. This data was not included in the table because although the reactive nitrenium ion associated with clozapine is a so called intermediate (between hard and soft) electrophile experimental evidence suggests that it does not form adducts with lysine residues (Yan et al., 2007). The presence of these putative binding domains may lend further circumstantial evidence to adduct-formation in the proteins discussed.

In order to check that these domains might be indicative of electrophilic binding targets 50 proteins were selected from those identified in the SCX, OFFGEL and

GeLC experiments. The proteins were selected from the IDA based experiments with only clozapine negative samples being included. The proteins selected has a median length of 347 amino acids. 36 of the 50 (72%) were found to contain at least one of the domains (CK, KC, CH, HC, RC, CR). This would indicate that the presence of the domains is not uncommon and as such may not alone be indicative of electrophilic binding potential.

5.5.4 Western Blot/2d-PAGE Vs. DiGE

The proteins recovered from 2d-PAGE based on the western blot analysis were all around a mass of 60 kDa, this would preclude the presence of the GRP78 protein or any of the other heatshock proteins identified by the DiGE experiment. On the evidence provided by the western blot study it is unlikely that most of these proteins are especially attacked by clozapine metabolites. It is more likely that at least some of the proteins identified in the DiGE analysis were false positives. The sensitivity of the DiGE technique would probably require that a greater number of replicates be used in order to pinpoint changes to protein concentration based on drug modification. With some studies reporting levels of protein modification at less than 5% (DeCaprio and Fowke, 1992; DeCaprio and O'Neill, 1985) it is quite possible that DiGE is incapable of identifying modified proteins regardless of the number of replicates used as the changes would be insignificant in comparison to biological and or system variations.

The DiGE and b-DMC techniques make use of bottom up proteomics as were applied in the case of the protein separation studies (Chapter 4) the major difference however is the presence of entire protein digests within individual samples. This approach enhances the ability to identify proteins and modified peptides as the entire protein sequence can be found in a single sample. The lack of identification of modifications even in the b-DMC samples is interesting. The signal from the ECL images required a long exposure time in order to be visualised indicating that the amount of protein is at the low end of the detection range. After digestion and recovery from the gel the total amount of modified protein is likely to be in the low picogram range (~10 pg) which

translates to ~200 attomoles for a 50 kDa protein. Coomassie staining of the gels revealed proteins in the region containing the metabolite adducts; as this staining is several orders of magnitude less sensitive the concentration of unmodified proteins greatly outnumber the modified ones. This concentration differential decreased the likelihood of detecting any modified peptides . As was seen MS identifications were not a problem. The lack of metabolite modified peptides suggests several possibilities i) the formation of an unknown metabolite ii) limitations in the Mascot search engine for the identification of modified peptides iii) breakdown of the protein adducts before MS analysis iv) poor tryptic digestion v) very low levels of modification.

The glutathione trapping experiments only revealed the presence of a single metabolite and experimentation with clozapine never revealed any metabolites smaller than desmethylclozapine so it is unlikely that any other metabolites exist. The biotin tag must still be in place as the proteins were bound by the streptavidin probe, from the negative control it can be seen that the probe did not have any non-specific binding characteristics. The search engine was capable of identifying clozapine metabolite modified peptides when tested with the synthetic peptides (3.4.6). Experiments were carried out with both glutathione trapped metabolites and modified synthetic peptides. It was found that adducts could still be identified more than a week after modification took place and from samples stored at room temperature in solution. The digestion protocol used has been shown to provide excellent digestion, some of the proteins identified in the study had coverage of greater than 40%, taking into account the relatively complex background this is a good result.

It may be possible that there was not sufficient material for the precursor scan to be effective, the sensitivity of the instrument may be the issue. The machine used in the study, the API 5500™ (AB SCIEX) is at present the most sensitive and advanced instrument of its type. If the problems is one of sensitivity it may be overcome with the development of improved technology. The work involved in designing both precursor ion scans was carried out using a API 4000™ (AB SCIEX) with a sensitivity of at least 1 order of magnitude less than the 5500. The detection of modified peptides using the precursor scan in this case seems to point against an issue of sensitivity.

5.5.5 Mass Spectrometric Detection

It would appear from the data that the degree of adduct formation is below the current threshold of detection. It is likely that the signal is being suppressed by other more intense ions and is being lost as background noise. The precursor ion scan is designed to combat this by filtering out those ions without a characteristic fragment. It was previously shown that the precursor ion (PI359) scan selectivity was poorer than expected; detecting almost 1 third as many proteins as were detected using a basic IDA method (2.1.2.3). In this part of the project the selectivity of the precursor ion scan was increased with only 18 precursor ion based detections to 147 IDA based detections. The precursor ion at 571 m/z may be less common than the ion at 359m/z or the considerably less complex fractions may lead to less overlap in peptide masses eluting from the reversed phase column. However, even with this improvement the selectivity of the precursor scan didn't reduce the number of proteins detected (PI vs. IDA) by a single order of magnitude.

The Q-trap has another scanning method with improved sensitivity and selectivity, the multiple reaction monitoring SRM scan. Using this approach coupled with nanoHPLC it is possible to detect peptides down to about 500 attomoles from a gel purified protein digest (Sinnaeve et al., 2005; Sinnaeve and Bocxlaer, 2004). Running SRM scans however requires detailed fragmentation information of particular protein/peptide targets. In the context of this study in which the mass of any given metabolite modified peptide is unknown SRM scanning is not possible. The selective precursor ion scans are capable of detecting peptides down to low (around 5fmol) femtomole levels (Wilm et al., 1996) typically an improvement of at least one order of magnitude over IDA scanning methods. Calculations have shown that approximately 90% of the total protein content of a cell is made up of around 10% of the known 10,000 – 20,000 protein species (Zuo et al., 2001) and as a consequence many low abundance proteins may be extremely difficult to detect. Furthermore an oversight in this work was the failure to quantitatively determine the performance of the HPLC-mass spectrometer setup. The lower limit of detection was not established, these literature values represent optimum levels. The setup used was monitored

for performance using a 22 fmol injection of a tryptically digested BSA standard that was run daily; in this work this is the lowest confirmed amount of sample detected.

The identification of low level post translation modifications is problematic . The first step usually involves some sort of simplification of the protein background usually in the form of affinity purification (Zhang et al., 2011; Abraham et al., 2000; Engholm-Keller and Larsen, 2011; Sidoli et al., 2012). In this project the background was depleted by means of a 3d separation of proteins (2d of PAGE coupled to RP-LCMS), even with the reduced complexity afforded by analysing the spot removed from the gel the total amount of modified material could still be markedly less than the amount of unmodified material. Assuming the presence of 5 proteins with each protein yielding 20 tryptic fragments and with a single modified protein with 1% adduct formation, modest assumptions, the ratio of modified to unmodified peptides is approximately 1:10000, a dynamic range of 4 orders of magnitude. The added selectivity of the precursor ion scan should increase the odds by eliminating those peptide fragments without the necessary precursor ion. The failure to detect any modified peptides would seem to indicate a problem with the precursor ion scan. It would appear that either the precursor ion is too common or that it is produced at low levels and is indistinguishable from background noise. The collision energy used in the precursor ion scan was used during the synthetic peptides experiments. The CE was set to 47 eV, this value is however not optimal for peptides with masses and compositions decidedly different to those used in the optimisation experiments. This is one possible explanation for a low signal from the precursor ion.

The use of the saturation DiGE approach is only applicable to proteins bearing accessible cysteine groups. The electrophilic nature of the metabolites must be such that they preferentially attack the sulfhydral groups. The technique requires a skilled operator, a lot of time and is a relatively expensive approach. The process is not readily automated. Proteins of low abundance are not readily identified by DiGE and as such critical information may be lost. At best the technique could flag up potential problems but is limited in its ability to handle proteins with extreme isoelectric points and high hydrophobicity.

For these reasons it is unlikely that a DiGE based approach would be suited to a high throughput screening system for the identification of metabolite-protein adducts.

Chapter 6: General Discussion and Conclusions

6.1 Findings

The objective of this project was to develop a methodology for the identification of drug-protein adducts in vitro. Ideally the approach would lend itself to high throughput automation in order to meet the needs of pharmaceutical companies that generate many tens of thousands of new chemical entities (NCEs). The need for such a test is clear; the lengthy timescales and high costs associated with developing new drugs is substantial; eliminating a molecule at an early stage would provide substantial financial savings as well as protect the wellbeing of would-be human test subjects.

The findings of the project were as follows:

- The glutathione trapping assay in combination with the synthetic peptide work was successfully used to identify reactive metabolites of clozapine and design and validate an effective precursor ion scanning method.

- The complexity of the human liver microsome fraction was found to be too great for analysis by the three 2d separation methods applied. This was found to be true even in combination with the precursor ion scan.

- It was not possible to reliably isolate drug-protein adducts using a difference gel electrophoresis (DiGE) approach.

- Drug protein adducts were successfully visualised using a western blot approach to detect biotin modified drug molecules. However it was found that the degree of modification was insufficient for detection by mass spectrometry.

6.2 Trapping of Reactive Metabolites

The glutathione trapping assay is widely used for the identification of reactive metabolites. It performs the task well but is not useful in determining whether or not any given NCE be progressed to the next stage of development. In order to determine this it would be necessary to decide if the metabolites were a threat to cellular function. Currently this means many levels of preclinical and clinical testing.

Being able to identify particular protein targets for an NCE would be a notable step towards characterising its potential toxicity. Looking to the future, it should be possible to correlate drug-adduct formation with adverse reactions. Patterns of protein modification and subsequent toxicity would likely become apparent.

In order to detect drug-protein adducts in a complex background such as that of the human liver microsome fraction it would be necessary to either decrease the complexity of the sample, produce a selective method of scanning or more likely both. The glutathione model is effective at trapping metabolites but the neutral loss scan used to detect the conjugates (drug-GSH) are specific to the glutathione molecule. It was found that the fragmentation of glutathione conjugates produced an ion incorporating both a drug fragment and the sulphur atom from cysteine. This was found to be an effective alternative to the neutral loss scan.

The synthetic peptide work demonstrated both that clozapine metabolites could form adducts with polypeptides other than glutathione and that the precursor ion scan at 359 m/z could be used to detect them. This allowed for the identification of clozapine-protein adducts in the general case i.e. theoretically any clozapine-protein adduct could be detected.

As drug-protein interactions depend on a number of physical factors including electrostatic interactions, physical accessibility and local pKa the synthetic polypeptides, with their limited primary structure, were not ideal analogues but only useful for testing the validity of the precursor ion scanning method.

6.3 Protein/Peptide Separation Methods

The three 1st dimensional separation methods were used to identify some 1700 proteins, more than 1000 of which were non-redundant (4.4.2). The separations appeared to be complimentary in that the degree of overlap was relatively small (4.4.3). The combination of these separation methods and reversed phase LCMS using the selective precursor scanning methods was still insufficient to identify the presence of even a single protein-drug adduct.

The separation techniques chosen are staple in proteomics experiments. It would be possible to further increase the degree of separation attained by increasing the length of the separation gradient used in either the SCX approach (with more fractions being taken) or the reversed phase LC approach or by increasing the number of bands excised from the 1d gel. Unfortunately in so doing the length of time required to carry out the experiments would increase dramatically. Obviously this is not ideal for a high throughout methodology. It is possible that even with increased separation time and/or increased fractions collected that drug-protein adducts still not be detected.

It was shown that the precursor ion scan did not have good selectivity for modified peptides. When compared with a general information dependant acquisition (IDA) the precursor ion scan identified around 1/3 as many proteins, none of which bore drug modifications(4.4.3). This is likely a result of the low resolution inherent to triple quadrupole instruments coupled with the high sample complexity. Examination of the data gathered using this method revealed that there were some 149 fragments not related to adduct formation that could trigger the precursor ion scan and that from these 149 motifs it was possible to predict 9,234 fragments within the same mass range. It was found however that none of these ions came closer than 28ppm of the exact mass of the precursor target (359.1092 Da); meaning that if the scan had been applied to a higher resolution instrument such as an Orbitrap or QqToF none of those false positives would have triggered the precursor scan. Precursor ion scans are used in the discovery of other post translational modifications but usually in conjunction with some form of affinity purification.

6.4 DiGE and Western Blotting

Analysis of the proteins recovered from the DiGE experiment again yielded a distinct lack of any drug-protein adducts. The complexity of each LCMS sample was greatly reduced as only small spots were excised from the gel for subsequent digestion and analysis. The DiGE analytical experiment was used to pinpoint proteins with changes in apparent abundance ranging from 1.2-2 fold. 147 proteins were identified by IDA and 18 by precursor ion scan. The lack of drug-protein adducts in a sample of such relative simplicity was interesting and potentially suggested that the level of modification may be the issue. The evidence from the western blot work suggests that the amount of modification is very low in relation to the total amount of protein (5.5.4). It appears to be at the lower limit of the detection capabilities of the western blot with a 1 hour exposure necessary for visualisation. Consequently it is entirely possible that the total amount of modified protein was too low for reliable mass spectrometric detection. In order to determine the amount of modification occurring it would be necessary to carry out a further experiment. Synthetic peptide could be spiked into the HLM assay and recovered using it's biotin tag. The recovered peptide could then be separated using RP-LC and the modified and unmodified fractions collected. These fractions could then be analysed by mass spectrometry at a range of concentrations. The relative intensity of the parent ions could then be compared and an estimation of modification levels made.

6.5 Conclusions

From the work carried out it would appear that the detection of drug-protein adducts is not trivial. It is apparent that in this case the total amount of modified protein was very low and thus the dynamic range was wide. A single approach for the identification of drug-protein adducts in the general case may be beyond the reach of current technologies and methodologies, at least within reasonable expectations of time and expense. Improvements in liquid chromatography, perhaps ultra performance liquid chromatography (UPLC), and instrument speed and sensitivity are likely to contribute to future developments in the long term. In such a complex system, ion suppression would play an

important role in limiting the dynamic range achieved during mass spectrometric analysis. In the paper by Hirabayashi et al (Hirabayashi et al., 2007) a peptide probe was used to monitor ion suppression. The probe (sequence DSSSSS) was designed to be highly hydrophilic and have a low isoelectric point (pI 3.8); At the pH of a typical RP-LC mobile phase (pH3) it holds a single proton and is not sensitive to gas phase proton transfer reactions. The peptide was spiked into the LC mobile phase and due to its hydrophilicity it is not retained on the column and so is present throughout the mass spectrometric analysis. Analysis of the mass chromatogram associated with the m/z of the probe is analysed with the presence of ion suppression highlighted by marked decreases in the chromatogram. This approach could be applied to the work carried out in this study in order to identify the extent of ion suppression and thereby optimise the amount of sample for analysis. Once an optimal sample load is identified it would then be useful to determine the lower limit of detection for drug-peptide adducts. This could be achieved by spiking an HLM preparation with decreasing (known) amounts of modified peptide. The solution should then be analysed by LCMS, the tandem ms data could then be searched against a protein database. As the amount of modified peptide in the sample decreases the search score should too decrease until it falls below the significance level required for a match. This would determine the lowest amount of modified peptide that can be detected against the complex microsomal background. The experiment could be carried out using both the precursor ion scan and an IDA based approach in order to determine the usefulness of the precursor method.

In addition, an Interesting development was that of the use of peptide aptamer libraries to create affinity purification devices, the so called Proteominer approach (Boschetti and Righetti, 2008). The presence of millions of peptide aptamers allows for the capture of a normalised cross section of a complex protein sample and the detection of low abundance peptides. Although the technology was available at the time of this project, it came to the attention of the author at a late stage when time constraints made further investigations impossible. It has since been demonstrated however, that the proposed mechanism by which the Proteominer approach works is incorrect. Work by a group headed by Friedrich Lottspeich has demonstrated that binding to the beads is not based on specific interactions between the hexapeptides and

proteins but more likely due to hydrophobic interactions (Keidel et al., 2010). Very similar results could be observed when comparing the effects of the Proteominer treatment to treatment using C18 functionalised beads. This is not to say that the approach has no merits. Regardless of the mechanism of action it is clear that the approach can be used to reduce the dynamic range of proteins present in a sample (D'Ambrosio et al., 2008; Farinazzo et al., 2009; Boschetti and Righetti, 2008).

7. References

Abraham, J., Kelly, J., Thibault, P., & Benchimol, S. (2000). Post-translational modification of p53 protein in response to ionizing radiation analyzed by mass spectrometry. *Journal of Molecular Biology, 295*(4), 853-864. doi: 10.1006/jmbi.1999.3415

Adams, C. P., & Brantner, V. V. (2006). Estimating The Cost Of New Drug Development: Is It Really $802 Million? *Health Aff, 25*(2), 420-428. doi: 10.1377/hlthaff.25.2.420

Ahr, H. J., L. J. King, et al. (1982). "The mechanism of reductive dehalogenation of halothane by liver cytochrome P450." Biochemical Pharmacology **31**(3): 383-390.

Alban, A., David, S. O., Bjorkesten, L., Andersson, C., Sloge, E., Lewis, S., & Currie, I. (2003). A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. [Comparative Study]. *Proteomics, 3*(1), 36-44.

Almira Correia, M., Sinclair, P. R., & De Matteis, F. (2011). Cytochrome P450 regulation: the interplay between its heme* and apoprotein moieties in synthesis, assembly, repair, and disposal. *Drug Metabolism Reviews, 43*(1), 1-26. doi: doi:10.3109/03602532.2010.515222

Amunom, I., L. J. Dieter, et al. (2011). "Cytochromes P450 Catalyze the Reduction of Î±,Î²-Unsaturated Aldehydes." Chemical Research in Toxicology **24**(8): 1223-1230.

Anderson, MM., Hazen, SL., Hsu, FF. & Heinecke, JW. (1997). Human neutrophils employ the myeloperoxidase-hydrogen peroxide-chloride system to convert hydroxy-amino acids into glycolaldehyde, 2-hydroxypropanal, and acrolein. A mechanism for the generation of highly reactive alpha-

hydroxy and alpha, beta-unsaturated aldehydes by phagocytes at sites of inflammation, J. Clin. Invest. 99(3):424-432.

Anderson, C.C., and Matzinger, P., (2000). Danger: the view from the bottom of the cliff. seminars in IMMUNOLOGY, Vol. 12, 2000: pp. 231–238

Annan, R. S., Huddleston, M. J., Verma, R., Deshaies, R. J., & Carr, S. A. (2001). A multidimensional electrospray MS-based approach to phosphopeptide mapping. *Anal Chem, 73*(3), 393-404.

Annesley, T. M. (2003). Ion suppression in mass spectrometry. [Review]. *Clin Chem, 49*(7), 1041-1044.

Anusiewicz, I., Berdys-Kochanska, J., & Simons, J. (2005). Electron Attachment Step in Electron Capture Dissociation (ECD) and Electron Transfer Dissociation (ETD). *The Journal of Physical Chemistry A, 109*(26), 5801-5813. doi: 10.1021/jp050218d

Arakawa, T., Kita, Y., & Timasheff, S. N. (2007). Protein precipitation and denaturation by dimethyl sulfoxide. *Biophysical Chemistry, 131*(1–3), 62-70. doi: 10.1016/j.bpc.2007.09.004

Argoti, D., Liang, L., Conteh, A., Chen, L., Bershas, D., Yu, C. P., . . . Yang, E. (2005). Cyanide trapping of iminium ion reactive intermediates followed by detection and structure identification using liquid chromatography-tandem mass spectrometry (LC-MS/MS). [Comparative Study Research Support, N I H , Extramural]. *Chem Res Toxicol, 18*(10), 1537-1544.

Armstrong, R. N. (1987). Enzyme-Catalyzed Detoxication Reactions: Mechanisms and Stereochemistr. *Critical Reviews in Biochemistry and Molecular Biology, 22*(1), 39-88. doi: 10.3109/10409238709082547

Atkins, W. M., Wang, R. W., Bird, A. W., Newton, D. J., & Lu, A. Y. (1993). The catalytic mechanism of glutathione S-transferase (GST). Spectroscopic determination of the pKa of Tyr-9 in rat alpha 1-1 GST. *Journal of Biological Chemistry, 268*(26), 19188-19191.

Attia, S. M. (2010). "Deleterious effects of reactive metabolites." Oxidative Medicine and Cellular Longevity **3**(4): 238-253.

Attwood, T. K., Bradley, P., Flower, D. R., Gaulton, A., Maudling, N., Mitchell, A. L., Moulton, G., Nordle, A., Paine, K., Taylor, P., Uddin, A., & Zygouri, C. (2003). PRINTS and its automatic supplement, prePRINTS. Nucleic Acids Research, 31, 400–402.

Bafna, V., Edwards, N.,(2001) SCOPE: a probabilistic model for scoring tandem mass spectra against a peptide database. Bioinformatics, 17 , S13–S21.

Bailey, A. O., Miller, T. M., Dong, M. Q., Vande Velde, C., Cleveland, D. W., & Yates, J. R. (2007). RCADiA: simple automation platform for comparative

multidimensional protein identification technology. *Analytical Chemistry, 79*(16), 6410-6418.

Baillie, T. A., & Davis, M. R. (1993). Mass spectrometry in the analysis of glutathione conjugates. [Research Support, U S Gov't, P H S Review]. *Biol Mass Spectrom, 22*(6), 319-325.

Barnouin, K. (2004). Two-dimensional gel electrophoresis for analysis of protein complexes. [Research Support, Non-U S Gov't Review]. *Methods Mol Biol, 261*, 479-498.

Bartolone, J. B., Beierschmitt, W. P., Birge, R. B., Hart, S. G., Wyand, S., Cohen, S. D., & Khairallah, E. A. (1989). Selective acetaminophen metabolite binding to hepatic and extrahepatic proteins: an in vivo and in vitro analysis. [In VitroResearch Support, Non-U S Gov'tResearch Support, U S Gov't, P H S]. *Toxicol Appl Pharmacol, 99*(2), 240-249.

Baumann, A.; Karst, U. (2010). Online electrochemistry/mass spectrometry in drug metabolism studies: principles and applications. *Expert Opin Drug Metab Toxicol*, **6(6)** 715-731

Bellec G, D. Y., Lozach P, Ménez JF, Berthou F. (1996). "Cytochrome P450 metabolic dealkylation of nine N-nitrosodialkylamines by human liver microsomes." Carcinogenesis **17**(9): 2029-34.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc Ser B 57:289–300

Benjamini, Y., Yekutelli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Ann Stat 29:1165–1188

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Wheeler, D. L. (2007). GenBank. Nucleic Acids Research, 35, D21–D25.

Bessems, J. G. M., Koppele, J. M. T., Van Dijk, P. A., Van Stee, L. L. P., Commandeur, J. N. M., & Vermeulen, N. P. E. (1996). Rat liver microsomal cytochrome P450-dependent oxidation of 3,5-disubstituted analogues of paracetamol. *Xenobiotica, 26*(6), 647-666.

Bessems, J. G. M., & Vermeulen, N. P. E. (2001). Paracetamol (Acetaminophen)-Induced Toxicity: Molecular and Biochemical Mechanisms, Analogues and Protective Approaches. *Critical Reviews in Toxicology, 31*(1), 55.

Bidlingmeyer, B.; Wang, Q. U.S. Patent 7,125,492 B2, October 24, 2006.

Blades, A. T., Ikonomou, M. G., & Kebarle, P. (1991). Mechanism of electrospray mass spectrometry. Electrospray as an electrolysis cell. *Analytical Chemistry, 63*(19), 2109-2114. doi: 10.1021/ac00019a009

Blonder, J., T. P. Conrads, et al. (2004). "A detergent- and cyanogen bromide-free
method for integral membrane proteomics: Application to Halobacterium purple membranes and the human epidermal membrane proteome." PROTEOMICS 4(1): 31-45.

Bodzon-Kulakowska, A., Bierczynska-Krzysik, A., Dylag, T., Drabik, A., Suder, P., Noga, M., . . . Silberring, J. (2007). Methods for samples preparation in proteomic research. *Journal of Chromatography B, 849*(1–2), 1-31. doi: 10.1016/j.jchromb.2006.10.040

Boor PJ, H. R., Sanduja R. (1990). "A role for a new vascular enzyme in the metabolism of xenobiotic amines." Circulation Research **66**(1): 249-52.

Booth J, B. E. a. S. P. (1961). "An enzyme from rat liver catalysing conjugations with glutathione." Biochemical Journal(79): 516-524.

Boyland, E., & Chasseaud, L. F. (2006). The Role of Glutathione and Glutathione S-Transferases in Mercapturic Acid Biosynthesis *Advances in Enzymology and Related Areas of Molecular Biology* (pp. 173-219): John Wiley & Sons, Inc.

BOSCHETTI E, RIGHETTI P. (2008). *Hexapeptide combinatorial ligand libraries: the march for the detection of the low-abundance proteome continues*. BIOTECHNIQUES. vol. 44, pp. 663-665, ISSN: 0736-6205.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry, 72(1–2), 248-254. doi: http://dx.doi.org/10.1016/0003-2697(76)90527-3

Brizzard, B. L., Chubet, R. G., & Vizard, D. L. (1994). Immunoaffinity purification of FLAG epitope-tagged bacterial alkaline phosphatase using a novel monoclonal antibody and peptide elution. *Biotechniques, 16*(4), 730-735.

Brzezinski, M. R., Abraham, T. L., Stone, C. L., Dean, R. A., & Bosron, W. F. (1994). Purification and characterization of a human liver cocaine carboxylesterase that catalyzes the production of benzoylecgonine and the formation of cocaethylene from alcohol and cocaine. [Research Support, U S Gov't, P H S]. *Biochem Pharmacol, 48*(9), 1747-1755.

Burka, L. T. P., T.M. and Macdonald, T.L. (1983). "Mechanisms of hydroxylation by
cytochrome P-450: metabolism of monohalobenzenes by phenobarbital-induced microsomes." Proceedings of the National Academy of Sciences **80**(21): 6680-6684.

Cairns, D. A., Barrett, J. H., Billingham, L. J., Stanley, A. J., Xinarianos, G., Field,
J. K., Johnson, P. J., Selby, P. J. and Banks, R. E. (2009), Sample size determination in clinical proteomic profiling experiments using mass spectrometry for class comparison. Proteomics, 9: 74–86. doi: 10.1002/pmic.200800417

Cargile B.J., Bundy J.L., Stephenson J.L. Jr. (2004). Potential for false positive identifications from large databases through tandem mass spectrometry. Journal of Proteome Research. 3(5):1082-5.

Castro-Perez, J., Plumb, R., Liang, L., & Yang, E. (2005). A high-throughput liquid chromatography/tandem mass spectrometry method for screening glutathione conjugates using exact mass neutral loss acquisition. [Evaluation Studies]. *Rapid Commun Mass Spectrom, 19*(6), 798-804.

Cech, N. B., & Enke, C. G. (2001). Practical implications of some recent studies in electrospray ionization fundamentals. [Research Support, Non-U S Gov't Research Support, U S Gov't, P H S Review]. *Mass Spectrom Rev, 20*(6), 362-387.

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).

Cohen, S. D., Pumford, N. R., Khairallah, E. A., Boekelheide, K., Pohl, L. R., Amouzadeh, H. R., & Hinson, J. A. (1997). Selective Protein Covalent Binding and Target Organ Toxicity. *Toxicology and Applied Pharmacology, 143*(1), 1-12.

Coles, B. (1984). Effects of modifying structure on electrophilic reactions with biological nucleophiles. [Research Support, Non-U S Gov't Review]. *Drug Metab Rev, 15*(7), 1307-1334.

Colinge, J., Masselot, A., Giron, M., Dessingy, T. & Magnin, J. OLAV: towards high-throughput tandem mass spectrometry data identification. *Proteomics*
3, 1454–1463 (2003).

Craig, R., Beavis, R. C., (2004) TANDEM: matching proteins with tandem mass spectra. Bioinformatics , 20, 1466–1467.

Craig HD, Eklund JD, Seidler NW. (2008). Trifluoroethanol increases albumin's susceptibility to chemical modification. Arch Biochem Biophys. 2008 Dec 1;480(1):11-6. doi: 10.1016/j.abb.2008.09.009. Epub 2008 Sep 22.

Crow, J. A., Bittles, V., Borazjani, A., Potter, P. M., & Ross, M. K. Covalent inhibition of recombinant human carboxylesterase 1 and 2 and monoacylglycerol lipase by the carbamates JZL184 and URB597. *Biochemical Pharmacology*(0). doi: 10.1016/j.bcp.2012.08.017

Cutillas PR, Geering B, Waterfield MD, et al., 2005, Quantification of gel-separated proteins and their phosphorylation sites by LC-MS using unlabeled internal standards: analysis of phosphoprotein dynamics in a B cell lymphoma cell line., *Molecular & Cellular Proteomics*, Vol:4, ISSN:1535-9476, Pages:1038-1051

D C Dahlin, G. T. M., A Y Lu, and S D Nelson (1984). "N-acetyl-p-benzoquinone imine: a cytochrome P-450-mediated oxidation product of acetaminophen." Proceedings of the National Academy of Sciences **81**(5): 1327-1331.

D'Ambrosio, C., Arena, S., Scaloni, A., Guerrier, L. and e. al. (2008). "Exploring the chicken egg white proteome with combinatorial peptide ligand libraries." Journal of Proteome Research(7): 3461-3474.

Damsten, M. C., Commandeur, J. N. M., Fidder, A., Hulst, A. G., Touw, D., Noort, D., & Vermeulen, N. P. E. (2007). Liquid Chromatography/Tandem Mass Spectrometry Detection of Covalent Binding of Acetaminophen to Human Serum Albumin. *Drug Metab Dispos, 35*(8), 1408-1417. doi: 10.1124/dmd.106.014233

Datyner, A., & Finnimore, E. (1973). A new staining method for the assay of proteins on polyacrylamide gels. *Analytical Biochemistry, 52*(1), 45-55. doi: 10.1016/0003-2697(73)90329-1

Davis, K. L., Yang, R.-K., Davidson, M., Mohs, R. C., Ryan, T. M., Schmeidler, J., . . . Gamzu, E. R. (1995). Alzheimer's disease: Tacrine and tacrine metabolite concentrations in plasma and cognitive change. *Drug Development Research, 34*(1), 55-65. doi: 10.1002/ddr.430340109

de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., . . . Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proceedings of the National Academy of Sciences, 100*(13), 7480-7485. doi: 10.1073/pnas.1332608100

Dean, J. A. (1980). Modern Size-Exclusion Liquid Chromatography (Yau, W. W.; Kirkland, J. J.; Bly, D. D.). *Journal of Chemical Education, 57*(11), A324. doi: 10.1021/ed057pA324.4

DeCaprio, A. P., & Fowke, J. H. (1992). Limited and selective adduction of carboxyl-terminal lysines in the high molecular weight neurofilament proteins by 2,5-hexanedione in vitro. *Brain Research, 586*(2), 219-228. doi: 10.1016/0006-8993(92)91630-w

DeCaprio, A. P., & O'Neill, E. A. (1985). Alterations in rat axonal cytoskeletal proteins induced by in vitro and in vivo 2,5-hexanedione exposure. *Toxicology and Applied Pharmacology, 78*(2), 235-247. doi: 10.1016/0041-008x(85)90287-x

Dennehy, M. K., Richards, K. A. M., Wernke, G. R., Shyr, Y., & Liebler, D. C. (2006). Cytosolic and Nuclear Protein Targets of Thiol-Reactive Electrophiles. *Chemical Research in Toxicology, 19*(1), 20-29. doi: doi:10.1021/tx050312l

Di Guan, C., Li, P., Riggs, P. D., & Inouye, H. (1988). Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. *Gene, 67*(1), 21-30.

Di, L., Keefer, C., Scott, D. O., Strelevitz, T. J., Chang, G., Bi, Y. A., . . . Obach, R. S. (2012). Mechanistic insights from comparing intrinsic clearance values between human liver microsomes and hepatocytes to guide drug design. [Journal article]. *Eur J Med Chem, 16*, 16.

Dieckhaus, C. M., Fernandez-Metzler, C. L., King, R., Krolikowski, P. H., & Baillie, T. A. (2005). Negative ion tandem mass spectrometry for the detection of glutathione conjugates. *Chem Res Toxicol, 18*(4), 630-638.

Dobson, C. M. (2001). Protein folding and its links with human disease. [Research Support, Non-U S Gov't Review]. *Biochem Soc Symp, 68*, 1-26.

Dondi, F., Cavazzini, A., Remelli, M., Felinger, A., & Martin, M. (2002). Stochastic theory of size exclusion chromatography by the characteristic function approach. *Journal of Chromatography A, 943*(2), 185-207. doi: 10.1016/s0021-9673(01)01443-1

Dormeyer, W., D. van Hoof, et al. (2008). "Plasma Membrane Proteomics of Human Embryonic Stem Cells and Human Embryonal Carcinoma Cells." Journal of Proteome Research 7(7): 2936-2951.

Dudoit, S., Shaffer, J. P., and Boldrick, ,J. C. (2003). Multiple hypothesis testing in microarray experiments. Statistical Science 18(1):71–103.

Durbin, R., Eddy, S., Krogh, A., & Mitchison, G. (1998). Biological sequence analysis: Probabilistic models of proteins and nucleic acids. Cambridge, UK: Cambridge University Press.

Echan, L. A., Tang, H. Y., Ali-Khan, N., Lee, K., & Speicher, D. W. (2005). Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. [Research Support, N I H ,

Extramural Research Support, Non-U S Gov't Research Support, U S Gov't, P H S]. *Proteomics, 5*(13), 3292-3303.

Elahi, E. N., Wright, Z., Hinselwood, D., Hotchkiss, S. A. M., Basketter, D. A., & Smith Pease, C. K. (2004). Protein Binding and Metabolism Influence the Relative Skin Sensitization Potential of Cinnamic Compounds. *Chemical Research in Toxicology, 17*(3), 301-310. doi: doi:10.1021/tx0341456

Elschenbroich, S., Ignatchenko, V., Sharma, P., Schmitt-Ulms, G., Gramolini, A. O., & Kislinger, T. (2009). Peptide separations by on-line MudPIT compared to isoelectric focusing in an off-gel format: application to a membrane-enriched fraction from C2C12 mouse skeletal muscle cells. [Comparative Study Research Support, Non-U S Gov't]. *J Proteome Res, 8*(10), 4860-4869.

Eng, J. K., McCormack, A. L., & Yates Iii, J. R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. Journal of the American Society for Mass Spectrometry, 5(11), 976-989. doi: http://dx.doi.org/10.1016/1044-0305(94)80016-2

Engholm-Keller, K., & Larsen, M. R. (2011). Titanium dioxide as chemo-affinity chromatographic sorbent of biomolecular compounds – Applications in acidic modification-specific proteomics. *Journal of Proteomics, 75*(2), 317-328. doi: 10.1016/j.jprot.2011.07.024

Enke, C. G. (1997). A predictive model for matrix and analyte effects in electrospray ionization of singly-charged ionic analytes. [Research Support, U S Gov't, P H S]. *Anal Chem, 69*(23), 4885-4893.

Esterbauer, H., Schaur, R. J., & Zollner, H. (1991). Chemistry and biochemistry of 4-hydroxynonenal, malonaldehyde and related aldehydes. *Free Radical Biology and Medicine, 11*(1), 81-128. doi: 10.1016/0891-5849(91)90192-6

Evans, D. C., Watt, A. P., Nicoll-Griffith, D. A., & Baillie, T. A. (2004). Drug−Protein Adducts: An Industry Perspective on Minimizing the Potential for Drug Bioactivation in Drug Discovery and Development. *Chemical Research in Toxicology, 17*(1), 3-16. doi: doi:10.1021/tx034170b

Farinazzo, A., Restuccia, U., Bachi, A., Guerrier, L. and e. al (2009). "Chicken egg yolk cytoplasmic proteome, mined via combinatorial peptide ligand libraries." Journal of Chromatography A 1216(8): 1241-52.

Fazekas de St. Groth, S.; Webster, R. G.; Datyner, A. (1963). "Two new staining

procedures for quantitative estimation of proteins on electrophoretic strips". *Biochimica et Biophysica Acta* **71**: 377–391. doi:10.1016/0006-3002(63)91092-8. PMID 18421828.

Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. (1989). "Electrospray ionization for mass spectrometry of large biomolecules". *Science* **246** (4926): 64–71. Bibcode:1989Sci...246...64F. doi:10.1126/science.2675315. PMID 2675315.

Fink, A. L. (1995). Molten globules. [Review]. *Methods Mol Biol, 40*, 343-360.

Fischer, V., Haar, J. A., Greiner, L., Lloyd, R. V., & Mason, R. P. (1991). Possible role of free radical formation in clozapine (clozaril)-induced agranulocytosis. *Mol Pharmacol, 40*(5), 846-853.

Fisher, A. A., Labenski, M. T., Malladi, S., Gokhale, V., Bowen, M. E., Milleron, R. S., . . . Lau, S. S. (2007). Quinone electrophiles selectively adduct "electrophile binding motifs" within cytochrome c. [Research Support, N I H , Extramural Research Support, Non-U S Gov't]. *Biochemistry, 46*(39), 11090-11100.

Fisher, A. A., Labenski, M. T., Monks, T. J., & Lau, S. S. (2011). Utilization of LC-MS/MS analyses to identify site-specific chemical protein adducts in vitro. [Research Support, N I H , Extramural]. *Methods Mol Biol, 691*, 317-326.

Finn, R. D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L., & Bateman, A. (2006). Pfam: Clans, web tools and services. Nucleic Acids Research, 34, D247–D251.

Flynn, G. C., Pohl, J., Flocco, M. T., & Rothman, J. E. (1991). Peptide-binding specificity of the molecular chaperone BiP. [10.1038/353726a0]. *Nature, 353*(6346), 726-730.

Fodor, I.K., Nelson, D.O., Alegria-Hartman, M., Robbins, K., Langlois, R.G., Turteltaub, K.W., Corzett, T.H. and McCutchen-Maloney, S.L. (2005) Statistical challenges in the analysis of two-dimensional difference gel electrophoresis experiments using DeCyder. Bioinformatics 21(19):3733–40. doi: 10.1093/bioinformatics/bti612

Frank A, Pevzner P, PepNovo: de novo peptide sequencing via probabilistic network modeling, Anal Chem 77 (4):964 973, 2005

Furst, S. M., Luedke, D., & Gandolfi, A. J. (1997). Kupffer cells from halothane-exposed guinea pigs carry trifluoroacetylated protein adducts. *Toxicology, 120*(2), 119-132. doi: 10.1016/s0300-483x(97)03649-4

Galeva, N., & Altermann, M. (2002). 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. [Comparative Study Research Support, U S Gov't, P H S]. *Proteomics, 2*(6), 713-722.

Gallucci S, Lolkema M, Matzinger P. 1999. Natural adjuvants: Endogenous activators of dendritic cells. *Nat. Med.* 5:1249–55

Gan, J., Harper, T. W., Hsueh, M. M., Qu, Q., & Humphreys, W. G. (2005). Dansyl glutathione as a trapping agent for the quantitative estimation and identification of reactive metabolites. [Comparative Study]. Chem Res Toxicol, 18(5), 896-903.

Gan, J., Ruan, Q., He, B., Zhu, M., Shyu, W. C., & Humphreys, W. G. (2009). In Vitro Screening of 50 Highly Prescribed Drugs for Thiol Adduct Formation 顯窩omparison of Potential for Drug-Induced Toxicity and Extent of Adduct Formation. Chemical Research in Toxicology, 22(4), 690-698.

Gardner, I., Popovic, M., Zahid, N., & Uetrecht, J. P. (2005). A Comparison of the Covalent Binding of Clozapine, Procainamide, and Vesnarinone to Human Neutrophils in Vitro and Rat Tissues in Vitro and in Vivo. Chemical Research in Toxicology, 18(9), 1384-1394. doi: doi:10.1021/tx050095o

Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. (2004). Open Mass Spectrometry Search Algorithm. Journal of Proteome Research, 3(5), 958-964. doi: 10.1021/pr0499491

Giddings, J. C. (1967). Maximum number of components resolvable by gel filtration and other elution chromatographic methods. Analytical Chemistry, 39(8), 1027-1028. doi: 10.1021/ac60252a025

Gilar M, Olivova P, Daly AE, Gebler JC. (2005). Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. J Sep Sci. 2005 Sep;28(14):1694-703.

Gion, J. M.; Lalanne, C.; Le, Provost, G.; Ferry-Dumazet, H.; Paiva, J.; Chaumeil, P.; Frigerio, J. M.; Brach, J.; Barre, A.; de Daruvar, A.; Claverol, S.; Bonneu, M.; Sommerer, N.; Negroni, L.; Plomion, C. (2005). The proteome of maritime pine wood forming tissue. PROTEOMICS, 5(14), 3731-3751. doi: 10.1002/pmic.200401197

Gokce, E., Andrews, G. L., Dean, R. A., & Muddiman, D. C. (2011). Increasing proteome coverage with offline RP HPLC coupled to online RP nanoLC–MS. Journal of Chromatography B, 879(9–10), 610-614. doi: 10.1016/j.jchromb.2011.01.032

Gonzalez-Begne, M., B. Lu, et al. (2009). "Proteomic Analysis of Human Parotid

Gland Exosomes by Multidimensional Protein Identification Technology (MudPIT)." Journal of Proteome Research 8(3): 1304-1314.

Gorrod, J. W., Whittlesea, C. M., & Lam, S. P. (1991). Trapping of reactive intermediates by incorporation of 14C-sodium cyanide during microsomal oxidation. [Comparative Study In Vitro Research Support, Non-U S Gov't]. *Adv Exp Med Biol, 283*, 657-664.

Green, N. M. (1990). Avidin and streptavidin. [Comparative Study]. *Methods Enzymol, 184*, 51-67.

Han, X., Jin, M., Breuker, K., & McLafferty, F. W. (2006). Extending Top-Down Mass Spectrometry to Proteins with Masses Greater Than 200 Kilodaltons. *Science, 314*(5796), 109-112. doi: 10.1126/science.1128868

Hannigan, A., Burchmore, R., & Wilson, J. B. (2007). The optimization of protocols for proteome difference gel electrophoresis (DiGE) analysis of preneoplastic skin. [Research Support, Non-U S Gov't]. *J Proteome Res, 6*(9), 3422-3432.

Hansen, B.T., Jones, J.A., Mason, D.E. & Liebler, D.C. SALSA: a pattern recognition algorithm to detect electrophile-adducted peptides by automated evaluation of CID spectra in LC-MS-MS analyses. *Anal. Chem.* **73**, 1676–1683 (2001).

Hanzlik, R. P., Fang, J., & Koen, Y. M. (2008). Filling and mining the reactive metabolite target protein database. *Chemico-Biological Interactions, In Press, Corrected Proof*.

Hargus SJ, A. H., Pumford NR, Myers TG, McCoy SC, Pohl LR. (1994). Metabolic activation and immunochemical localization of liver protein adducts of the nonsteroidal anti-inflammatory drug diclofenac. *Chem. Res. Toxicol., 12*, 387-395.

Hazai, E., Vereczkey, L., & Monostory, K. (2002). Reduction of Toxic Metabolite Formation of Acetaminophen. *Biochemical and Biophysical Research Communications, 291*(4), 1089-1094.

Hazen, S. L., d'Avignon, A., Anderson, M. M., Hsu, F. F., & Heinecke, J. W. (1998). Human Neutrophils Employ the Myeloperoxidase-Hydrogen Peroxide-Chloride System to Oxidize α-Amino Acids to a Family of Reactive Aldehydes. *Journal of Biological Chemistry, 273*(9), 4997-5005. doi: 10.1074/jbc.273.9.4997

Hersh, E. V., Pinto, A., & Moore, P. A. (2007). Adverse drug interactions involving common prescription and over-the-counter analgesic agents.

Clinical Therapeutics, 29(11, Supplement 1), 2477-2497.

Higdon, R., Hogan, J. M.., Belle, G. V., Kolker, E. (2005). Randomized Sequence Databases for Tandem Mass Spectrometry Peptide and Protein Identification. OMICS: A Journal of Integrative Biology. 9(4): 364-379. doi:10.1089/omi.2005.9.364.

Hinson, J. A. (1983). Reactive metabolites of phenacetin and acetaminophen: a review. [Review]. *Environ Health Perspect, 49*, 71-79.

Hirabayashi, A., Ishimaru, M., Manri, N., Yokosuka, T., & Hanzawa, H. (2007). Detection of potential ion suppression for peptide analysis in nanoflow liuid chromatography/mass spectrometry. *Rapid Communications in Mass Spectrometry*, *21*(17) 2860-2866. Doi: 10.1002/rcm.3157

Y. Ho, P. Kebarle. (1997). Studies of the Dissociation Mechanisms of Deprotonated Mono-nucleotides by Energy Resolved Collision Induced Dissociation. Int. J. Mass Spectrom. and Ion Processes 165/166, 433-455

Hoenig and Heisey (2001). The Abuse of Power The American Statistician 55(1):19-24

Holcapek, M., Kolárová, L., & Nobilis, M. (2008). High-performance liquid chromatography-tandem mass spectrometry in the identification and determination of phase I and phase II drug metabolites. *Analytical and bioanalytical chemistry, 391*(1), 59-78.

Hollenberg, P. F., Kent, U. M., & Bumpus, N. N. (2007). Mechanism-Based Inactivation of Human Cytochromes P450s: Experimental Characterization, Reactive Intermediates, and Clinical Implications. *Chemical Research in Toxicology, 21*(1), 189-205.

Hong, F., Sekhar, K. R., Freeman, M. L., & Liebler, D. C. (2005). Specific Patterns of Electrophile Adduction Trigger Keap1 Ubiquitination and Nrf2 Activation. *J. Biol. Chem., 280*(36), 31768-31775. doi: 10.1074/jbc.M503346200

Hoos, J. S., Damsten, M. C., de Vlieger, J. S. B., Commandeur, J. N. M., Vermeulen, N. P. E., Niessen, W. M. A., . . . Irth, H. (2007). Automated detection of covalent adducts to human serum albumin by immunoaffinity chromatography, on-line solution phase digestion and liquid chromatography-mass spectrometry. *Journal of Chromatography B, 859*(2), 147-156.

Hopfgartner, G., Husser, C., & Zell, M. (2003). Rapid screening and characterization of drug metabolites using a new quadrupole-linear ion trap mass spectrometer. *J Mass Spectrom, 38*(2), 138-150.

Hrebicek, T., Durrschmid, K., Auer, N., Bayer, K., & Rizzi, A. (2007). Effect of CyDye minimum labeling in differential gel electrophoresis on the reliability of protein identification. [Research Support, Non-U S Gov't]. *Electrophoresis, 28*(7), 1161-1169.

Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., & Graham Cooks, R. (2005). The Orbitrap: a new mass spectrometer. [Research Support, Non-U S Gov't Research Support, U S Gov't, Non-P H S]. *J Mass Spectrom, 40*(4), 430-443.

Huang, H. J., Tsai, M. L., Chen, Y. W., & Chen, S. H. (2011). Quantitative shot-gun proteomics and MS-based activity assay for revealing gender differences in enzyme contents for rat liver microsome. [Research Support, Non-U S Gov't]. *J Proteomics, 74*(12), 2734-2744.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P. S., Pagni, M., & Sigrist, C. J. A. (2006). The PROSITE database. Nucleic Acids Research, 34, D227–D230.

Hunt, S. M. N., Thomas, M. R., Sebastian, L. T., Pedersen, S. K., Harcourt, R. L., Sloane, A. J., and Wilkins, M. R. (2005). Optimal replication and the importance of experimental design for gel-based quantitative proteomics. J Proteome Res 4:809{819.

Hurtley, S. M., & Helenius, A. (1989). Protein oligomerization in the endoplasmic reticulum. [Research Support, Non-U S Gov't Research Support, U S Gov't, P H S Review]. *Annu Rev Cell Biol, 5*, 277-307.

Huttlin EL, Hegeman AD, Harms AC, Sussman MR. Prediction of error associated with false-positive rate determination for peptide identification in large-scale proteomics experiments using a combined reverse and forward peptide sequence database strategy. J Proteome Res. 2007;6:392–398.

Inoue, K., Shibata, Y., Takahashi, H., Ohe, T., Chiba, M., & Ishii, Y. (2009). A trapping method for semi-quantitative assessment of reactive metabolite formation using [35S]cysteine and [14C]cyanide. *Drug Metab Pharmacokinet, 24*(3), 245-254.

Issaq, H. J., Conrads, T. P., Janini, G. M., & Veenstra, T. D. (2002). Methods for fractionation, separation and profiling of proteins and peptides. [Research Support, U S Gov't, P H S Review]. *Electrophoresis, 23*(17), 3048-3061.

Jackson, B., Brocker, C., Thompson, D. C., Black, W., Vasiliou, K., Nebert, D. W., & Vasiliou, V. (2011). Update on the aldehyde dehydrogenase gene (ALDH) superfamily. *Human genomics, 5*(4), 283-303.

Jakoby WB, Ziegler DM. (1990). The enzymes of detoxication. J Biol Chem. 1990 Dec 5;265(34):20715-8.

Jegouzo, A., Gressier, B., Frimat, B., Brunet, C., Dine, T., Luyckx, M., . . . Cazin, J. C. (1999). Comparative oxidation of loxapine and clozapine by human neutrophils. [Comparative Study]. *Fundam Clin Pharmacol, 13*(1), 113-119.

Jemnitz, K., Veres, Z., Monostory, K., K鼽ori, L., & Vereczkey, L. (2008). Interspecies differences in acetaminophen sensitivity of human, rat, and mouse primary hepatocytes. *Toxicology in Vitro, 22*(4), 961-967.

Jian, W., Yao, M., Zhang, D., & Zhu, M. (2009). Rapid detection and characterization of in vitro and urinary N-acetyl-L-cysteine conjugates using quadrupole-linear ion trap mass spectrometry and polarity switching. *Chem Res Toxicol, 22*(7), 1246-1255.

Jiang, L., He, L., & Fountoulakis, M. (2004). Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. *Journal of Chromatography A, 1023*(2), 317-320. doi: 10.1016/j.chroma.2003.10.029

Jiang, X., Feng, S., Tian, R., Han, G., Ye, M., & Zou, H. (2007). Automation of nanoflow liquid chromatography-tandem mass spectrometry for proteome analysis by using a strong cation exchange trap column. [Research Support, Non-U S Gov't]. *Proteomics, 7*(4), 528-539.

Johnson, R. S., Martin, S. A., Biemann, K., Stults, J. T., & Watson, J. T. (1987). Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. [Research Support, U S Gov't, P H S]. *Anal Chem, 59*(21), 2621-2625.

Ju, C. (2009). The Role of Haptic Macrophages in Regulation of Idiosyncratic Drug Reactions. *Toxicol Pathol*, 0192623308329475. doi: 10.1177/0192623308329475

Jurva, U., Wikstrom, H. V., Weidolf, L., & Bruins, A. P. (2003). Comparison between electrochemistry/mass spectrometry and cytochrome P450 catalyzed oxidation reactions. [Comparative Study]. *Rapid Commun Mass Spectrom, 17*(8), 800-810.

Kalish RS. Antigen processing: the gateway to the immune response. J Am Acad Dermatol. 1995;32:640–652.

Kang, D., Nam, H., Kim, Y.-S., & Moon, M. H. (2005). Dual-purpose sample trap for on-line strong cation-exchange chromatography/reversed-phase liquid chromatography/tandem mass spectrometry for shotgun proteomics:

Application to the human Jurkat T-cell proteome. *Journal of Chromatography A, 1070*(1–2), 193-200. doi: 10.1016/j.chroma.2005.02.058

Kalgutkar AS, G. I., Obach RS, Shaffer CL, Callegari E, Henne KR, Mutlib AE, Dalvie DK, Lee JS, Nakai Y, O'Donnell JP, Boer J, Harriman SP. (2005). "A comprehensive listing of bioactivation pathways of organic functional groups." Current drug metabolism **6**(3): 161-225.

Karger, B. L., Gant, J. R., Martkopf, A., & Weiner, P. H. (1976). Hydrophobic effects in reversed-phase liquid chromatography. *Journal of Chromatography A, 128*(1), 65-78. doi: 10.1016/s0021-9673(00)84032-7

Karp, N. A., Kreil, D. P., Lilley, K. S., Determining a significant change in protein expression with DeCyder during a pair-wise comparison using two-dimensional difference gel electrophoresis. Proteomics 2004,4, 1421–1432.

Karp, N. A. and Lilley, K. S. (2009). Investigating sample pooling strategies for DIGE experiments to address biological variability. Proteomics 9:388{397.

Karp, N. A.; Spencer, M.; Lindsay, H.; O'Dell, K.; Lilley, K. S. Impact of replicate types on proteomic expression analysis. J.Proteome Res.2005,4, 1867-1871

Karp, N. A., Feret, R., Rubtsov, D. V., & Lilley, K. S. (2008). Comparison of DIGE and post-stained gel electrophoresis with both traditional and SameSpots analysis for quantitative proteomics. [Comparative Study Evaluation Studies

Research Support, Non-U S Gov't]. *Proteomics, 8*(5), 948-960.

Kedderis GL, B. R., Koop DR. (1993). "Epoxidation of acrylonitrile by rat and human

cytochromes P450." Chemical Research in Toxicology **6**(6): 866-71.

Keidel, E., Ribitsch, D., and Lottspeich, F., (2010). "Equalizer technology – Equal rights for disparate beads." Proteomics(10): 2089-2098.

Keller A, Nesvizhskii AI, Kolker E, Aebersold R. (2002). Empirical statistical model

to estimate the accuracy of peptide identifications made by MS/MS and database search. *Analytical chemistry*, vol 74(20):5383-92.

Kent, U. M., Juschyshyn, M. I., & Hollenberg, P. F. (2001). Mechanism-based inactivators as probes of cytochrome P450 structure and function. [Research Support, U S Gov't, P H S Review]. *Curr Drug Metab, 2*(3), 215-243.

Kersey, P. J., Duarte, J., Williams, A., Karavidopoulou, Y., Birney, E., & Apweiler,

R. (2004). The international protein index: An integrated database for proteomics experiments. Proteomics, 4, 1985–1988

Kersey, P., Hermjakob, H., Apweiler, R. (2000). VARSPLIC: alternatively-spliced protein sequences derived from SWISS-PROT and TrEMBL. BIOINFORMATICS Vol. 16 no. 11 Pages 1048–1049

Khurana, R., Gillespie, J. R., Talapatra, A., Minert, L. J., Ionescu-Zanetti, C., Millett, I., & Fink, A. L. (2001). Partially folded intermediates as critical precursors of light chain amyloid fibrils and amorphous aggregates. [Research Support, Non-U S Gov't Research Support, U S Gov't, Non-P H S Research Support, U S Gov't, P H S]. *Biochemistry, 40*(12), 3525-3535.

King CD, R. G., Green MD, Tephly TR. (2000). "UDP-glucuronosyltransferases." Current drug metabolism **1**(2): 143-61.

King, R., Bonfiglio, R., Fernandez-Metzler, C., Miller-Stein, C., & Olah, T. (2000). Mechanistic investigation of ionization suppression in electrospray ionization. *J Am Soc Mass Spectrom, 11*(11), 942-950.

Kingdon, K. H. (1923). A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Physical Review, 21*(4), 408-418.

Kolkman, A., Dirksen, E. H., Slijper, M., & Heck, A. J. (2005). Double standards in quantitative proteomics: direct comparative assessment of difference in gel electrophoresis and metabolic stable isotope labeling. [Comparative Study

Research Support, Non-U S Gov't]. *Mol Cell Proteomics, 4*(3), 255-266.

Kopaciewicz, W., Rounds, M. A., Fausnaugh, J., & Regnier, F. E. (1983). Retention model for high-performance ion-exchange chromatography. *Journal of Chromatography A, 266*(0), 3-21. doi: 10.1016/s0021-9673(01)90875-1

Kostiainen R, Kotiaho T, Kuuranne T, Auriola S. (2003). Liquid chromatography/atmospheric pressure ionization-mass spectrometry in drug metabolism studies. J Mass Spectrom. 2003 Apr;38(4):357-72.

Kouranov, A., Xie, L., de la Cruz, J., Chen, L., Westbrook, J., Bourne, P. E., & Berman, H. M. (2006). The RCSB PDB information portal for structural genomics. Nucleic Acids Research, 34, D302–D305.

Koyama, E., Chiba, K., Tani, M., & Ishizaki, T. (1997). Reappraisal of human CYP isoforms involved in imipramine N-demethylation and 2-hydroxylation: a study using microsomes obtained from putative extensive and poor metabolizers of S-mephenytoin and eleven recombinant human CYPs. [Research Support, Non-U S Gov't]. *J Pharmacol Exp Ther, 281*(3), 1199-1210.

Krogh, A., Brown, M., Mian, I. S., Sjolander, K., & Haussler, D. (1994). Hidden

Markov models in computational biology. Applications to protein modeling. Journal of Molecular Biology, 235(5), 1501–1531.

Kuehner, D. E., Blanch, H. W., & Prausnitz, J. M. (1996). Salt-induced protein precipitation: Phase equilibria from an equation of state. *Fluid Phase Equilibria, 116*(1–2), 140-147. doi: 10.1016/0378-3812(95)02882-x

Kulikova, T., Akhtar, R., Aldebert, P., Althorpe, N., Andersson, M., Baldwin, A., Bates, K., Bhattacharyya, S., Bower, L., Browne, P., Castro, M., Cochrane, G., Duggan, K., Eberhardt, R., Faruque, N., Hoad, G., Kanz, C., Lee, C., Leinonen, R., Lin, Q., Lombard, V., Lopez, R., Lorenc, D., McWilliam, H., Mukherjee, G., Nardone, F., Garcia-Pastor, M. P., Plaister, S., Sobhany, S., Stoehr, P., Vaughan, R., Wu, D., Zhu, W., & Apweiler, R. (2007). EMBL Nucleotide sequence database in 2006. Nucleic Acids Research, 35, D16–D20.

Krueger, S. K. and D. E. Williams (2005). "Mammalian flavin-containing monooxygenases: structure/function, genetic polymorphisms and role in drug metabolism." Pharmacology & Therapeutics **106**(3): 357-387.

Kyle PB, Smith SV, Baker RC, Kramer RE. (2012). Mass spectrometric detection of CYP450 adducts following oxidative desulfuration of methyl parathion. J Appl Toxicol. 2013 Jul;33(7):644-51. doi: 10.1002/jat.1792. Epub 2012 Jan 23.

Labenski, M. T., Fisher, A. A., Lo, H. H., Monks, T. J., & Lau, S. S. (2009). Protein electrophile-binding motifs: lysine-rich proteins are preferential targets of quinones. [Research Support, N I H , Extramural]. *Drug Metab Dispos, 37*(6), 1211-1218.

Laboissiere MC, Sturley SL, Raines RT. (1995). The essential function of protein-disulfide isomerase is to unscramble non-native disulfide bonds. J Biol Chem. 1995 Nov 24;270(47):28006-9.

Laitinen, O. H., Nordlund, H. R., Hytönen, V. P., Uotila, S. T. H., Marttila, A. T., Savolainen, J., . . . Kulomaa, M. S. (2003). Rational Design of an Active Avidin Monomer. *Journal of Biological Chemistry, 278*(6), 4010-4014. doi: 10.1074/jbc.M205844200

Laemmli U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227(5259): 680-5.

Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., & Jorgensen, T. J. (2005). Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. [Comparative Study

Research Support, Non-U S Gov't]. *Mol Cell Proteomics, 4*(7), 873-886.

Lazarou, J., Pomeranz, B. H., & Corey, P. N. (1998). Incidence of Adverse Drug Reactions in Hospitalized Patients: A Meta-analysis of Prospective Studies.

JAMA, 279(15), 1200-1205. doi: 10.1001/jama.279.15.1200

Lee, D.-Y., & Chang, G.-D. (2009). Electrolytic Reduction: Modification of Proteins Occurring in Isoelectric Focusing Electrophoresis and in Electrolytic Reactions in the Presence of High Salts. *Analytical Chemistry, 81*(10), 3957-3964. doi: 10.1021/ac900281n

Lemoine, A., Gautier, J. C., Azoulay, D., Kiffel, L., Belloc, C., Guengerich, F. P., . . . Leroux, J. P. (1993). Major pathway of imipramine metabolism is catalyzed by cytochromes P-450 1A2 and P-450 3A4 in human liver. [In Vitro]. *Mol Pharmacol, 43*(5), 827-832.

L. Lennard, J. C. W., and J. S. Lilleyman (1997). "Thiopurine drugs in the treatment of childhood leukaemia: the influence of inherited thiopurine methyltransferase activity on drug metabolism and cytotoxicity." British Journal of Clinical Pharmacology **44**(5): 455-461.

Léonil, J., Mollé, D., Bouhallab, S., & Henry, G. (1994). Precipitation of hydrophobic peptides from tryptic casein hydrolysate by salt and pH. *Enzyme and Microbial Technology, 16*(7), 591-595. doi: 10.1016/0141-0229(94)90124-4

Letunic, I., Copley, R. R., Pils, B., Pinkert, S., Schultz, J., & Bork, P. (2006). SMART 5: Domains in the context of genomes and networks. Nucleic Acids Research, 34, D257–D260.

Liebler, D. C. (2007). Protein Damage by Reactive Electrophiles: Targets and Consequences. *Chemical Research in Toxicology, 21*(1), 117-128.

Lin, D., Saleh, S., & Liebler, D. C. (2008). Reversibility of Covalent Electrophile−Protein Adducts and Chemical Toxicity. *Chemical Research in Toxicology, 21*(12), 2361-2369. doi: doi:10.1021/tx800248x

Linke, T., Doraiswamy, S., & Harrison, E. H. (2007). Rat plasma proteomics: effects of abundant protein depletion on proteomic analysis. [Research Support, N I H , Extramural

Research Support, U S Gov't, Non-P H S]. *J Chromatogr B Analyt Technol Biomed Life Sci, 849*(1-2), 273-281.

Linnet, K., & Olesen, O. V. (1997). Metabolism of clozapine by cDNA-expressed human cytochrome P450 enzymes. [Research Support, Non-U S Gov't]. *Drug Metab Dispos, 25*(12), 1379-1382.

Lipka, E., Guelzim, A., Yous, S., Bonte, J.-P., & Vaccher, C. (2005). Preparative HPLC separation of methoxytetralins, ligands for melatonin receptors, containing two chiral centers with polysaccharide chiral stationary phases. Determination of enantiomeric purity. *Journal of Biochemical and*

Biophysical Methods, 64(1), 46-58. doi: 10.1016/j.jbbm.2005.05.008

Lippincott, J., & Apostol, I. (1999). Carbamylation of cysteine: a potential artifact in peptide mapping of hemoglobins in the presence of urea. *Anal Biochem, 267*(1), 57-64.

Liska, A. J. and Shevchenko, A. (2003) Expanding the organismal scope of proteomics: cross-species protein identification by mass spectrometry and its implications. *Proteomics* **3,** 19–28.

Liu, G., Eggler, A. L., Dietz, B. M., Mesecar, A. D., Bolton, J. L., Pezzuto, J. M., & van Breemen, R. B. (2005). Screening Method for the Discovery of Potential Cancer Chemoprevention Agents Based on Mass Spectrometric Detection of Alkylated Keap1. *Analytical Chemistry, 77*(19), 6407-6414. doi: doi:10.1021/ac050892r

Liu, H., Finch, J. W., Lavallee, M. J., Collamati, R. A., Benevides, C. C., & Gebler, J. C. (2007). Effects of column length, particle size, gradient length and flow rate on peak capacity of nano-scale liquid chromatography for peptide separations. [Comparative Study]. *J Chromatogr A, 13*(1), 30-36.

Liu, H., Finch, J. W., Luongo, J. A., Li, G. Z., & Gebler, J. C. (2006). Development of an online two-dimensional nano-scale liquid chromatography/mass spectrometry method for improved chromatographic performance and hydrophobic peptide recovery. *J Chromatogr A, 24*(1), 43-51.

Liu, H., Finch, J. W., Luongo, J. A., Li, G.-Z., & Gebler, J. C. (2006). Development of an online two-dimensional nano-scale liquid chromatography/mass spectrometry method for improved chromatographic performance and hydrophobic peptide recovery. *Journal of Chromatography A, 1135*(1), 43-51. doi: 10.1016/j.chroma.2006.09.030

Liu, Z. C., & Uetrecht, J. P. (1995). Clozapine is oxidized by activated human neutrophils to a reactive nitrenium ion that irreversibly binds to the cells. *J Pharmacol Exp Ther, 275*(3), 1476-1483.

Loo, R. R. O., N. Dales, et al. (1994). "Surfactant effects on protein structure examined by electrospray ionization mass spectrometry." Protein Science 3(11): 1975-1983.

Lundblad RL, Noyes CM. Chemical reagents for protein modification. Ch 1 CRC Press; Boca Raton Fl: 1984.

Ma B, Zhang K, Liang C, An (2005) e®ective algorithm for peptide de novo sequencing from MS/MS spectra,J Comput System Sci 70:418 430.

Ma, S., Chowdhury, S. K., & Alton, K. B. (2006). Application of mass

spectrometry for metabolite identification. [Review]. *Curr Drug Metab, 7*(5), 503-523.

Ma, S., & Subramanian, R. (2006). Detecting and characterizing reactive metabolites by liquid chromatography/tandem mass spectrometry. [Review]. *J Mass Spectrom, 41*(9), 1121-1139.

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

MacDonald, C., Smith, C., Michopoulos, F., Weaver, R., & Wilson, I. D. (2011). Identification and quantification of glutathione adducts of clozapine using ultra-high-performance liquid chromatography with orthogonal acceleration time-of-flight mass spectrometry and inductively coupled plasma mass spectrometry. *Rapid Commun Mass Spectrom, 25*(13), 1787-1793.

Madden S, Spaldin V, Hayes RN, Woolf TF, Pool WF, Park BK. (1995). Species variation in the bioactivation of tacrine by hepatic microsomes. Xenobiotica. 1995 Jan;25(1):103-16.

Mallet, C. R., Lu, Z., & Mazzeo, J. R. (2004). A study of ion suppression effects in electrospray ionization from mobile phase additives and solid-phase extracts. [Comparative Study Evaluation Studies Validation Studies]. *Rapid Commun Mass Spectrom, 18*(1), 49-58.

Manadas, B., English, J. A., Wynne, K. J., Cotter, D. R., & Dunn, M. J. (2009). Comparative analysis of OFFGel, strong cation exchange with pH gradient, and RP at high pH for first-dimensional separation of peptides from a membrane-enriched protein fraction. [Comparative Study Research Support, Non-U S Gov't]. *Proteomics, 9*(22), 5194-5198.

Manchanda, T., Hess, D., Dale, L., Ferguson, S. G., & Rieder, M. J. (2002). Haptenation of Sulfonamide Reactive Metabolites to Cellular Proteins. *Mol Pharmacol, 62*(5), 1011-1026. doi: 10.1124/mol.62.5.1011

Mann M, Højrup P, Roepstorff P (1993). "Use of mass spectrometric molecular weight information to identify proteins in sequence databases". *Biol. Mass Spectrom.* **22** (6): 338–45.

Mann, M., Hendrickson, R. C., & Pandey, A. (2001). Analysis of proteins and proteomes by mass spectrometry. [Research Support, Non-U S Gov't Review]. *Annu Rev Biochem, 70*, 437-473.

Mann, M., & Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. [Research Support, Non-U S Gov't Review]. *Nat Biotechnol,*

21(3), 255-261.

Mann, M., and Wilm, M. (1994) Error Tolerant identification of Peptides in Sequence Databases by Peptide Sequence Tags. Anal. Chem. 66, 4390-4399

Marchler-Bauer,A., Lu,S., Anderson,J.B., Chitsaz,F., Derbyshire,M.K., DeWeese-Scott,C., Fong,J.H., Geer,L.Y., Geer,R.C., Gonzales,N.R. et al. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins.
Nucleic Acids Res., 39, D225–D229.

Martin S, Weltzien HU. (1994) T cell recognition of haptens, a molecular view. Int Arch Allergy Immunol 104:10–16. [PubMed]

Mascher, E., & Lundahl, P. (1989). Sodium dodecyl sulphate-protein complexes : Changes in size or shape below the critical micelle concentration, as monitored by high-performance agarose gel chromatography. *Journal of Chromatography A, 476*(0), 147-158. doi: 10.1016/s0021-9673(01)93864-6

Massey, V., Komai, H., Palmer, G., & Elion, G. B. (1970). On the Mechanism of Inactivation of Xanthine Oxidase by Allopurinol and Other Pyrazolo[3,4-d]pyrimidines. *Journal of Biological Chemistry, 245*(11), 2837-2844.

Masubuchi Y, Igarashi S, Suzuki T, Horie T, Narimatsu S. (1996). Imipramine-induced inactivation of a cytochrome P450 2D enzyme in rat liver microsomes: in relation to covalent binding of its reactive intermediate. J Pharmacol Exp Ther. 1996 Nov;279(2):724-31.PMID:8930177

Masuda, J., Maynard, D. M., Nishimura, M., Ueda, T., Kowalak, J. A., & Markey, S. P. (2005). Fully automated micro- and nanoscale one- or two-dimensional high-performance liquid chromatography system for liquid chromatography–mass spectrometry compatible with non-volatile salts for ion exchange chromatography. *Journal of Chromatography A, 1063*(1–2), 57-69. doi: 10.1016/j.chroma.2004.11.084

Matsunaga, T., S. Shintani, et al. (2006). "Multiplicity of Mammalian Reductases for
Xenobiotic Carbonyl Compounds." Drug Metabolism and Pharmacokinetics **21**(1): 1-18.

Matzinger, P., (1994). Tolerancem Danger, and the extended family. Annu. Rev.lmmunol. 1994. 12:991-1045

McCracken, N. W., P. G. Blain, et al. (1993). "Nature and role of xenobiotic metabolizing esterases in rat liver, lung, skin and blood." Biochemical Pharmacology **45**(1): 31-36.

Melcher A, Todryk S, Hardwick N, Ford M, Jacobson M, et al. 1998. Tumor immunogenicity is determined by the mechanism of cell death via induction of heat shock protein expression. *Nat. Med.* 4:581–87

Mentlein, R., S. Heiland, et al. (1980). "Simultaneous purification and comparative

characterization of six serine hydrolases from rat liver microsomes." Archives of Biochemistry and Biophysics **200**(2): 547-559.

Merrick, B. A. (2008). The plasma proteome, adductome and idiosyncratic toxicity in toxicoproteomics research. *Brief Funct Genomic Proteomic, 7*(1), 35-49. doi: 10.1093/bfgp/eln004

Meunier, B., Dumas, E., Piec, I., Bechet, D., Hebraud, M. and Hocquette, J.F. (2007) *Assessment of hierarchical clustering methodologies for proteomic data mining*. J Proteome Res 6(1):358–66.

Michel, P. E., Crettaz, D., Morier, P., Heller, M., Gallot, D., Tissot, J. D., . . . Rossier, J. S. (2006). Proteome analysis of human plasma and amniotic fluid by Off-Gel isoelectric focusing followed by nano-LC-MS/MS. [Comparative Study Research Support, Non-U S Gov't]. *Electrophoresis, 27*(5-6), 1169-1181.

Mitulović G, Smoluch M, Chervet JP, Steinmacher I, Kungl A, Mechtler K.(2003). An improved method for tracking and reducing the void volume in nano HPLC-MS with micro trapping columns. Anal Bioanal Chem. 2003 Aug;376(7):946-51. Epub 2003 Jul 11.

Mitulovic, G., Stingl, C., Smoluch, M., Swart, R., Chervet, J. P., Steinmacher, I., . . . Mechtler, K. (2004). Automated, on-line two-dimensional nano liquid chromatography tandem mass spectrometry for rapid analysis of complex protein digests. [Research Support, Non-U S Gov't]. *Proteomics, 4*(9), 2545-2557.

Miyake K (2007) Innate immune sensing of pathogens and danger signals by cell surface Toll-like receptors. Semin Immunol 19:3–10

Mohammed, S., & Heck, A. J. R. (2011). Strong cation exchange (SCX) based analytical methods for the targeted analysis of protein post-translational modifications. *Current Opinion in Biotechnology, 22*(1), 9-16. doi: 10.1016/j.copbio.2010.09.005

Moore R. E., Young M. K., Lee T. D. (2002) Qscore: an algorithm for evaluating SEQUEST database search results. J. Am. Soc. Mass Spectrom. 13,378–386. [PubMed]

Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Buillard, V., Cerutti, L., Copley, R., Courcelle, E., Das, U., Daugherty, L., Dibley, M., Finn, R., Fleischmann, W., Gough, J., Haft, D., Hulo, N., Hunter, S., Kahn, D., Kanapin, A., Kejariwal, A., Labarga, A., Langendijk-Genevaux, P. S., Lonsdale, D., Lopez, R., Letunic, I., Madera, M., Maslen, J., McAnulla, C., McDowall, J., Mistry, J., Mitchell, A., Nikolskaya, A. N., Orchard, S., Orengo, C., Petryszak, R., Selengut, J. D., Sigrist, C. J., Thomas, P. D., Valentin, F., Wilson, D., Wu, C. H., & Yeats,

C. (2007). New developments in the InterPro database. Nucleic Acids Research, 35, D224–D228.

Muller, C., Bandemer, J., Vindis, C., Salvayre, R., & Negre-Salvayre, A. (2011). 259 PROTEIN DISULFIDE ISOMERASE INHIBITION BY 4-HYDROXYNONENAL-ADDUCTS CONTRIBUTES TO OXIDIZED LOW DENSITY LIPOPROTEIN-INDUCED APOPTOSIS. *Atherosclerosis Supplements, 12*(1), 56. doi: 10.1016/s1567-5688(11)70260-4

Mutlib, A., Lam, W., Atherton, J., Chen, H., Galatsis, P., & Stolle, W. (2005). Application of stable isotope labeled glutathione and rapid scanning mass spectrometers in detecting and characterizing reactive metabolites. [In Vitro]. *Rapid Commun Mass Spectrom, 19*(23), 3482-3492.

Nagaraj, N., A. Lu, et al. (2008). "Detergent-Based but Gel-Free Method Allows Identification of Several Hundred Membrane Proteins in Single LC-MS Runs." Journal of Proteome Research 7(11): 5028-5032.

Nakasa, H., Komiya, M., Ohmori, S., Rikihisa, T., Kiuchi, M., & Kitada, M. (1993). Characterization of human liver microsomal cytochrome P450 involved in the reductive metabolism of zonisamide. *Mol Pharmacol, 44*(1), 216-221.

Nakayama Wong, L. S., Lame, M. W., Jones, A. D., & Wilson, D. W. (2010). Differential cellular responses to protein adducts of naphthoquinone and monocrotaline pyrrole. [Research Support, N I H , Extramural Research Support, U S Gov't, Non-P H S]. *Chem Res Toxicol, 23*(9), 1504-1513.

Narayana, C., Suresh, T., Mahender Rao, S., Dubey, P. K., & Moses Babu, J. (2003). A validated chiral HPLC method for the enantiomeric separation of Linezolid on amylose based stationary phase. *Journal of Pharmaceutical and Biomedical Analysis, 32*(1), 21-28. doi: 10.1016/s0731-7085(03)00031-1

Negishi M, Kreibich G. Coordinated polypeptide synthesis and insertion of protoheme in cytochrome P-450 during development of endoplasmic reticulum membranes. J Biol Chem. 1978 Jul 10;253(13):4791–4797.

Nelson, S. D., & Pearson, P. G. (1990). Covalent and Noncovalent Interactions in Acute Lethal Cell Injury Caused by Chemicals. *Annual Review of Pharmacology and Toxicology, 30*(1), 169.

Nesvizhskii AI, Keller A, Kolker E, Aebersold R. (2003). **ProteinProphet:** A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem. 2003 Sep 1;75(17):4646-58.

Neuhoff V et al. (1988). Improved staining of proteins in polyacrylamide gels

including isoelectric focusing gels with clear backgrounds at nanogram sensitivity using G-250 and R–250. Electrophoresis 9, 255–262.

Nguyen, J. K., Fouts, M. M., Kotabe, S. E., & Lo, E. (2006). Polypharmacy as a risk factor for adverse drug reactions in geriatric nursing home residents. *The American Journal of Geriatric Pharmacotherapy, 4*(1), 36-41.

Noort, D., Hulst, A. G., de Jong, L. P. A., & Benschop, H. P. (1999). Alkylation of Human Serum Albumin by Sulfur Mustard in Vitro and in Vivo: Mass Spectrometric Analysis of a Cysteine Adduct as a Sensitive Biomarker of Exposure. *Chemical Research in Toxicology, 12*(8), 715-721. doi: doi:10.1021/tx9900369

Obach, R. S., Kalgutkar, A. S., Soglia, J. R., & Zhao, S. X. (2008). Can In Vitro Metabolism-Dependent Covalent Binding Data in Liver Microsomes Distinguish Hepatotoxic from Nonhepatotoxic Drugs? An Analysis of 18 Drugs with Consideration of Intrinsic Clearance and Daily Dose. *Chemical Research in Toxicology, 21*(9), 1814-1822. doi: doi:10.1021/tx800161s

Obach, R. S., & Reed-Hagen, A. E. (2002). Measurement of Michaelis constants for cytochrome P450-mediated biotransformation reactions using a substrate depletion approach. *Drug Metab Dispos, 30*(7), 831-837.

Ohsawa, I., Nishimaki, K., Yasuda, C., Kamino, K., & Ohta, S. (2003). Deficiency in a mitochondrial aldehyde dehydrogenase increases vulnerability to oxidative stress in PC12 cells. *Journal of Neurochemistry, 84*(5), 1110-1117. doi: 10.1046/j.1471-4159.2003.01619.x

Okubo, K., Sugawara, H., Gojobori, T., & Tateno, Y. (2006). DDBJ in preparation for overview of research activities behind data submissions. Nucleic Acids Research, 34, D6–D9.

Olsen J, Bjornsdottir I, Tjornelund J, and Honore Hansen S (2002) Chemical reactivity of the naproxen acyl glucuronide and the naproxen coenzyme A thioester towards bionucleophiles. J Pharm Biomed Anal 29: 7–152.

Oppenheim JJ, Tewary P, de la Rosa G, Yang D (2007) Alarmins initiate host defense. Adv Exp Med Biol 601:185–194

Packer, N. H., Lawson, M. A., Jardine, D. R., Sanchez, J. C., & Gooley, A. A. (1998). Analyzing glycoproteins separated by two-dimensional gel electrophoresis. [Research Support, Non-U S Gov't]. *Electrophoresis, 19*(6), 981-988.

Pappin DJ, Hojrup P, Bleasby AJ. (1993). Rapid identification of proteins by peptide-mass fingerprinting. Curr Biol. 1993 Jun 1;3(6):327-32.

Park, K. B., Dalton-Brown, E., Hirst, C., & Williams, D. P. (2006). Selection of

new chemical entities with decreased potential for adverse drug reactions. *European Journal of Pharmacology, 549*(1-3), 1-8.

Park BK, Coleman JW, and Kitteringham NR (1987) Drug disposition and drug hypersensitivity. Biochem Pharmacol 36: 581–590.

Parthasarathi, R., Subramanian, V., Roy, D. R., and Chattaraj, P. K. (2004). Electrophilicity index as a possible descriptor of biological activity. Bioorg. Medicinal Chem. 12, 5533-5543.

Patel, D. K., Notarianni, L. J., Bennett, P. N. . (1990). Comparative metabolism of high doses of aspirin in man and rat. *Xenobiotica, 20*, 847-854.

Patel, M., Tang, B. K., Kalow, W. . (1992). Variability of acetaminophen metabolism in Caucasians and Orientals. *Pharmacogenetics, 2*, 38-45.

Paul W., Steinwedel H.; Steinwedel (1953). "Ein neues Massenspektrometer ohne Magnetfeld". *Zeitschrift für Naturforschung A* **8** (7): 448–450. Bibcode:1953ZNatA...8..448P.

Paull, B., & Nesterenko, P. N. (2005). Novel ion chromatographic stationary phases for the analysis of complex matrices. [Review]. *Analyst, 130*(2), 134-146.

Paulo, J. A., Urrutia, R., Banks, P. A., Conwell, D. L., & Steen, H. (2011). Proteomic analysis of a rat pancreatic stellate cell line using liquid chromatography tandem mass spectrometry (LC-MS/MS). [Research Support, N I H , Extramural]. *J Proteomics, 75*(2), 708-717.

Pearson, R. G. (1963). Hard and Soft Acids and Bases. *Journal of the American Chemical Society, 85*(22), 3533-3539. doi: doi:10.1021/ja00905a001

Peng, F., Zhan, X., Li, M.-Y., Fang, F., Li, G., Li, C., . . . Chen, Z. (2012). Proteomic and Bioinformatics Analyses of Mouse Liver Microsomes. *International Journal of Proteomics, 2012*. doi: 10.1155/2012/832569

Perdivara, I., Deterding, L. J., Przybylski, M., & Tomer, K. B. (2010). Mass spectrometric identification of oxidative modifications of tryptophan residues in proteins: chemical artifact or post-translational modification? [Research Support, N I H , Intramural

Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis, 20 , 3551–3567. Research Support, Non-U S Gov't]. *J Am Soc Mass Spectrom, 21*(7), 1114-1117.

Pindel, E. V., Kedishvili, N. Y., Abraham, T. L., Brzezinski, M. R., Zhang, J., Dean, R. A., & Bosron, W. F. (1997). Purification and cloning of a broad substrate specificity human liver carboxylesterase that catalyzes the

hydrolysis of cocaine and heroin. [Comparative Study Research Support, U S Gov't, P H S]. *J Biol Chem, 272*(23), 14769-14775.

Pitman, M. R., & Menz, R. I. (2006). Methods for protein homology modelling. In R. M. B. Dilip K. Arora & B. S. Gautam (Eds.), *Applied Mycology and Biotechnology* (Vol. Volume 6, pp. 37-59): Elsevier.

Polson, C., Sarkar, P., Incledon, B., Raguvaran, V., & Grant, R. (2003). Optimization of protein precipitation based upon effectiveness of protein removal and ionization effect in liquid chromatography–tandem mass spectrometry. *Journal of Chromatography B, 785*(2), 263-275. doi: 10.1016/s1570-0232(02)00914-5

POLVERINO DE LAURETO, P., DE FILIPPIS, V., SCARAMELLA, E., ZAMBONIN, M., and FONTANA, A.(1995). Limited proteolysis of lysozyme in trifluoroethanol Isolation and characterization of a partially active enzyme derivative. Eur. J. Biochem. 230,779-787 (1995) 0 FEBS 1995.

Prakash, C., Shaffer, C. L., & Nedderman, A. (2007). Analytical strategies for identifying drug metabolites. [Review]. *Mass Spectrom Rev, 26*(3), 340-369.

Preissner, S., Kroll, K., Dunkel, M., Senger, C., Goldsobel, G., Kuzman, D., . . . Preissner, R. (2010). SuperCYP: a comprehensive database on Cytochrome P450 enzymes including a tool for analysis of CYP-drug interactions. [Research Support, Non-U S Gov't]. *Nucleic Acids Res, 38*(Database issue), 24.

Pruitt,K.D., Katz,K.S., Sicotte,H. and Maglott,D.R. (2000) Introducing RefSeq and LocusLink: curated human genome resources at the NCBI. Trends Genet., 16, 44–47.

Pruitt,K.D., Tatusova,T., Klimke,W. and Maglott,D.R. (2009) NCBI Reference Sequences: current status, policy and new initiatives. Nucleic Acids Res., 37, D32–D36

Pruitt, P.K., Tatusova, T., Brown, G.R., and Maglott, D.R. (2012). D130–D135 Nucleic Acids Research. Vol. 40, Database issue Published online 24 November 2011 doi:10.1093/nar/gkr1079

Pumford, N. R., Roberts, D. W., Benson, R. W., & Hinson, J. A. (1990). Immunochemical quantitation of 3-(cystein-S-yl)acetaminophen protein adducts in subcellular liver fractions following a hepatotoxic dose of acetaminophen. *Biochem Pharmacol, 40*(3), 573-579.

Qian WJ, Liu T, Monroe ME, Strittmatter EF, Jacobs JM, Kangas LJ, Petritis K,

Camp DG 2nd, Smith RD. (2005). Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. J Proteome Res. 2005 Jan-Feb;4(1):53-62.

Qiu, Y., Benet, L. Z., & Burlingame, A. L. (1998). Identification of the Hepatic Protein Targets of Reactive Metabolites of Acetaminophen in Vivo in Mice Using Two-dimensional Gel Electrophoresis and Mass Spectrometry. *J. Biol. Chem., 273*(28), 17940-17953. doi: 10.1074/jbc.273.28.17940

Qizilbash, N., Whitehead, A., Higgins, J., Wilcock, G., Schneider, L., & Farlow, M. (1998). Cholinesterase inhibition for Alzheimer disease: a meta-analysis of the tacrine trials. Dementia Trialists' Collaboration. [Meta-Analysis Research Support, Non-U S Gov't]. *Jama, 280*(20), 1777-1782.

Rabilloud, T. (2002). Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. [Review]. *Proteomics, 2*(1), 3-10.

Ramstrom, M., Zuberovic, A., Gronwall, C., Hanrieder, J., Bergquist, J., & Hober, S. (2009). Development of affinity columns for the removal of high-abundance proteins in cerebrospinal fluid. [Research Support, Non-U S Gov't]. *Biotechnol Appl Biochem, 52*(Pt 2), 159-166.

Reynolds, J. A., & Tanford, C. (1970). Binding of dodecyl sulfate to proteins at high binding ratios. Possible implications for the state of proteins in biological membranes. *Proceedings of the National Academy of Sciences of the United States of America, 66*(3), 1002-1007.

Righetti, P. G., Castagna, A., Herbert, B., Reymond, F., & Rossier, J. S. (2003). Prefractionation techniques in proteome analysis. [Research Support, Non-U S Gov't Review]. *Proteomics, 3*(8), 1397-1407.

Rinaldi, R., Eliasson, E., Swedmark, S., & Morgenstern, R. (2002). Reactive Intermediates and The Dynamics of Glutathione Transferases. *Drug Metab Dispos, 30*(10), 1053-1058. doi: 10.1124/dmd.30.10.1053

Rock, K.L., Kono, H., (2008).The inflammatory response to cell death. Annual Review of Pathology-Mechanisms of Disease Volume: 3 Pages: 99-126 DOI: 10.1146/annurev.pathol.3.121806.151456 Published: 200

Rodbard, D., & Chrambach, A. (1970). Unified theory for gel electrophoresis and gel filtration. *Proceedings of the National Academy of Sciences of the United States of America, 65*(4), 970-977.

Rosalind E. Jenkins, N. R. K. C. E. P. G. S. M. J. D. J. H. C. S. L. J.-S. B. N. P. E.

V. B. (2008). Glutathione-<B><I>S</I></B>-transferase pi as a model protein for the characterisation of chemically reactive metabolites. *PROTEOMICS, 8*(2), 301-315.

Rosenfeld, J., Capdevielle, J., Guillemot, J. C., & Ferrara, P. (1992). In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. [Research Support, Non-U S Gov't]. *Anal Biochem, 203*(1), 173-179.

Rovere-Querini P, Capobianco A, Scaffidi P, Valentinis B, Catalanotti F, et al. 2004.

HMGB1 is an endogenous immune adjuvant released by necrotic cells. *EMBO Rep*. 5:825–30

Roychowdhury, S., Cram, A. E., Aly, A., & Svensson, C. K. (2007). Detection of Haptenated Proteins in Organotypic Human Skin Explant Cultures Exposed to Dapsone. *Drug Metab Dispos, 35*(9), 1463-1465. doi: 10.1124/dmd.107.015560

Rubino, F.M., Pitton, M., D. Di Fabio, A. Colombi. (2009). Toward an "omic" pathophysiology of reactive chemicals: thirty years of mass spectrometric study of protein adducts with endogenous and xenobiotic compounds. *Mass Spectrometry Reviews*.

Rufer, C. E., Rubino, F. M. Maul, R., Donauer, E., Fabian, E. J., & Kulling, S. E. (2007). In vitro and in vivo metabolism of the soy isoflavone glycitein. [In Vitro Research Support, Non-U S Gov't]. *Mol Nutr Food Res, 51*(7), 813-823.

Ruppen-Cañás, I., P. P. López-Casas, et al. "An improved quantitative mass spectrometry analysis of tumor specific mutant proteins at high sensitivity." Proteomics 12(9): 1319-1327.

Sadygov, R.G., Liu, H. & Yates, J.R. Statistical models for protein validation

using tandem mass spectral data and protein amino acid sequence databases. *Anal. Chem.* **76**, 1664–1671 (2004).

Sadygov, R.G., Cociorva, D., & Yates III, J.R., (2004) Large-scale database searching using tandem mass spectra: Looking up the answer in the back of the book. NATURE METHODS VOL.1 NO.3 195-202

Sadygov, R.G. , and Yates III, J. R. (2003) A Hypergeometric Probability Model for

Protein Identification and Validation using Tandem Mass Spectral Data and Protein Sequence Databases, , *Analytical Chemistry 2003;75:3792-3798.*

Sanderson, J., Naisbitt, D., & Park, B. (2006). Role of bioactivation in drug-induced hypersensitivity reactions. *The AAPS Journal, 8*(1), E55-E64.

Sano, A., & Nakamura, H. (2004). Titania as a chemo-affinity support for the column-switching HPLC analysis of phosphopeptides: application to the characterization of phosphorylation sites in proteins by combination with protease digestion and electrospray ionization mass spectrometry. *Anal Sci, 20*(5), 861-864.

Santoni, V., Molloy, M., & Rabilloud, T. (2000). Membrane proteins and proteomics: un amour impossible? [Review]. *Electrophoresis, 21*(6), 1054-1070.

Santucci, A., Rustici, M., Bracci, L., Lozzi, L., Soldani, P., & Neri, P. (1990). HPLC immunoaffinity purification of rabies virus glycoprotein using immobilized antipeptide antibodies. *Journal of Immunological Methods, 127*(1), 131-138. doi: 10.1016/0022-1759(90)90349-z

Satoh, H., Gillette, J. R., Davies, H. W., Schulick, R. D., & Pohl, L. R. (1985). Immunochemical evidence of trifluoroacetylated cytochrome P-450 in the liver of halothane-treated rats. *Mol Pharmacol, 28*(5), 468-474.

Sayers EW, Barrett T, Benson DA, et al. (2010) Database resources of the National
Center for Biotechnology Information.Nucleic Acids Res;38:D5–16.

Schindler, R., Mentlein, R., & Feldheim, W. (1998). Purification and characterization of retinyl ester hydrolase as a member of the non-specific carboxylesterase supergene family. [Comparative Study]. *Eur J Biochem, 251*(3), 863-873.

Schirle, M., Heurtier, M. A., & Kuster, B. (2003). Profiling core proteomes of human cell lines by one-dimensional PAGE and liquid chromatography-tandem mass spectrometry. *Mol Cell Proteomics, 2*(12), 1297-1305.

Seo, J., & Lee, K. J. (2004). Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. [Research Support, Non-U S Gov't Review]. *J Biochem Mol Biol, 37*(1), 35-44.

Shaw, J., Rowlinson, R., Nickson, J., Stone, T., Sweet, A., Williams, K., & Tonge, R. (2003). Evaluation of saturation labelling two-dimensional difference gel electrophoresis fluorescent dyes. *Proteomics, 3*(7), 1181-1195.

Shen, Y., Tolic, N., Masselon, C., Pasa-Tolic, L., Camp, D. G., 2nd, Hixson, K. K., . . . Smith, R. D. (2004). Ultrasensitive proteomics using high-efficiency on-line micro-SPE-nanoLC-nanoESI MS and MS/MS. [Research Support, U S Gov't, Non-P H S Research Support, U S Gov't, P H S]. *Anal Chem, 76*(1), 144-154.

Shevchenko, A., Sunyaev, S., Loboda, A., et al. (2001) Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. *Anal. Chem.* **73,** 1917–1926.

Shi Y, Zheng W, Rock KL. 2000. Cell injury releases endogenous adjuvants that stimulate cytotoxic T cell responses. *Proc. Natl. Acad. Sci. USA* 97:14590–95

Shi Y, Evans JE, Rock KL. 2003. Molecular identification of a danger signal that alerts the immune system to dying cells. *Nature* 425:516–21

Shin, N.-Y., Liu, Q., Stamer, S. L., & Liebler, D. C. (2007). Protein Targets of Reactive Electrophiles in Human Liver Microsomes. *Chemical Research in Toxicology, 20*(6), 859-867. doi: doi:10.1021/tx700031r

Sidenius, U., Skonberg, C., Olsen, J., & Hansen, S. H. (2004). In vitro reactivity of carboxylic acid-CoA thioesters with glutathione. [Comparative Study Research Support, Non-U S Gov't]. *Chem Res Toxicol, 17*(1), 75-81.

Sidoli, S., Cheng, L., & Jensen, O. N. (2012). Proteomics in chromatin biology and epigenetics: Elucidation of post-translational modifications of histone proteins by mass spectrometry. *Journal of Proteomics, 75*(12), 3419-3433. doi: 10.1016/j.jprot.2011.12.029

Sigrist, C. J. A, Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., Bairoch,
A., & Bucher, P. (2002). PROSITE: A documented database using patterns and profiles as motif descriptors. Briefings in Bioinformatics, 3, 265–274.

Sinnaeve, B. A., & Bocxlaer, J. F. V. (2004). Evaluation of nano-liquid chromatography–tandem mass spectrometry in a column switching setup for the absolute quantification of peptides in the picomolar range. *Journal of Chromatography A, 1058*(1–2), 113-119. doi: 10.1016/j.chroma.2004.07.098

Sinnaeve, B. A., Storme, M. L., & Van Bocxlaer, J. F. (2005). Capillary liquid chromatography and tandem mass spectrometry for the quantification of enkephalins in cerebrospinal fluid. [Research Support, Non-U S Gov't]. *J Sep Sci, 28*(14), 1779-1784.

Smith, D. A., & Obach, R. S. (2005). SEEING THROUGH THE MIST: ABUNDANCE VERSUS PERCENTAGE. COMMENTARY ON METABOLITES IN SAFETY TESTING. *Drug Metab Dispos, 33*(10), 1409-1417. doi: 10.1124/dmd.105.005041

Smith, D. B., & Johnson, K. S. (1988). Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene, 67*(1), 31-40. doi: 10.1016/0378-1119(88)90005-4

Smith, G. C. M., Wolf, C. R., Kenna, J. G., Harrison, D. J., & Tew, D. (1993).

Autoantibodies to hepatic microsomal carboxylesterase in halothane hepatitis. *The Lancet, 342*(8877), 963-964. doi: 10.1016/0140-6736(93)92005-e

Smith, M. C., Furman, T. C., Ingolia, T. D., & Pidgeon, C. (1988). Chelating peptide-immobilized metal ion affinity chromatography. A new concept in affinity chromatography for recombinant proteins. *Journal of Biological Chemistry, 263*(15), 7211-7215.

Souverain, S., Rudaz, S., & Veuthey, J. L. (2004). Protein precipitation for the analysis of a drug cocktail in plasma by LC–ESI–MS. *Journal of Pharmaceutical and Biomedical Analysis, 35*(4), 913-920. doi: 10.1016/j.jpba.2004.03.005

Speers, A. E. and C. C. Wu (2007). "Proteomics of Integral Membrane ProteinsTheory and Application." Chemical Reviews 107(8): 3687-3714.

Speicher, K., Kolbas, O., Harper, S., and Speicher, D. (2000). Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies. J Biomol Tech. 2000 June; 11(2): 74–86.

Stalder, D., Haeberli, A., & Heller, M. (2008). Evaluation of reproducibility of protein identification results after multidimensional human serum protein separation. [Evaluation Studies]. *Proteomics, 8*(3), 414-424.

Sturgill, M. G., & Lambert, G. H. (1997). Xenobiotic-induced hepatotoxicity: mechanisms of liver injury and methods of monitoring hepatic function. *Clin Chem, 43*(8), 1512-1526.

Tabb, D.L., Saraf, A. & Yates, J.R. III. GutenTag: High-throughput sequence tagging via an empirically derived fragmentation model. *Anal. Chem.* **75**, 6415–6421 (2003).

Tanaka-Kawai, H., & Yomoda, S. (1993). Molecular weight and substrate characteristics of human serum arylesterase following purification by immuno-affinity chromatography. *Clinica Chimica Acta, 215*(2), 127-138. doi: 10.1016/0009-8981(93)90120-s

Tang, L., Kebarle, P. (1993). Dependence of Ion Intensity in Electrospray Mass Spectrometry on the Concentration of the Analytes in the Electrospray Solution, Anal. Chem. 65, 3654-3668

Tang, K., Page, J. S., & Smith, R. D. (2004). Charge competition and the linear dynamic range of detection in electrospray ionization mass spectrometry. [Research Support, N I H , Extramural Research Support, U S Gov't, Non-P H S Research Support, U S Gov't, P H S]. *J Am Soc Mass Spectrom, 15*(10), 1416-1423.

Tang, W., & Lu, A. Y. (2010). Metabolic bioactivation and drug-related adverse

effects: current status and future directions from a pharmaceutical research perspective. [Review]. *Drug Metab Rev, 42*(2), 225-249.

Tanner, S., Shu, H., Frank, A., Wang, L. C.et al.,(2005) InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. Anal. Chem. 77 , 4626–4639

Timms JF and Cramer R (2008) Difference gel electrophoresis. Proteomics 8: 4886–4897.

Tirmenstein, M. A., & Nelson, S. D. (1989). Subcellular binding and effects on calcium homeostasis produced by acetaminophen and a nonhepatotoxic regioisomer, 3'-hydroxyacetanilide, in mouse liver. *J. Biol. Chem., 264*(17), 9814-9819.

Tirumalai, R. S., Chan, K. C., Prieto, D. A., Issaq, H. J., Conrads, T. P., & Veenstra, T. D. (2003). Characterization of the Low Molecular Weight Human Serum Proteome. *Mol Cell Proteomics, 2*(10), 1096-1103. doi: 10.1074/mcp.M300031-MCP200

Tornqvist, M., Fred, C., Haglund, J., Helleberg, H., Paulsson, B., & Rydberg, P. (2002). Protein adducts: quantitative and qualitative aspects of their formation, analysis and applications. [Research Support, Non-U S Gov't Review]. *J Chromatogr B Analyt Technol Biomed Life Sci, 778*(1-2), 279-308.

Towbin, H., Staehelin, T., & Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A, 76*(9), 4350-4354.

Tseng, C. F., Huang, H. Y., Yang, Y. T., & Mao, S. J. T. (2004). Purification of human haptoglobin 1-1, 2-1, and 2-2 using monoclonal antibody affinity chromatography. *Protein Expression and Purification, 33*(2), 265-273. doi: 10.1016/j.pep.2003.09.006

Uetrecht, J. (2008). Idiosyncratic Drug Reactions: Past, Present, and Future. *Chemical Research in Toxicology, 21*(1), 84-92. doi: doi:10.1021/tx700186p

Uetrecht, J. P. (1992). Metabolism of clozapine by neutrophils. Possible implications for clozapine-induced agranulocytosis. [Research Support, Non-U S Gov't

Review]. *Drug Saf, 1*, 51-56.

Uetrecht, J. P. (1999). New Concepts in Immunology Relevant to Idiosyncratic Drug Reactions: The 鈥渄anger Hypothesis鈥?and Innate Immune System.

Chemical Research in Toxicology, 12(5), 387-395.

Ulrich, R. G. (2007). Idiosyncratic Toxicity: A Convergence of Risk Factors. *Annual Review of Medicine, 58*(1), 17-34. doi: doi:10.1146/annurev.med.58.072905.160823

Ute M. Kent, M. I. J., Paul F. Hollenberg (2001). Mechanism-based inactivators as probes of cytochrome P450 structure and function. *Drug Metab. Dispos, 30*, 1053-1058

Urfer, W., Grzegorczyk, M., and Jung, K. (2006) Statistics for proteomics: a review of tools for analyzing experimental data. Pract Proteomics 1, 48–55.

van Deemter, J. J., Zuiderweg, F. J. and Klinkenberg, A. (1956) Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. Chemical Engineering Science 5, 271-289

Vailaya, A. (2005). Fundamentals of Reversed Phase Chromatography: Thermodynamic and Exothermodynamic Treatment. *Journal of Liquid Chromatography & Related Technologies, 28*(7-8), 965-1054. doi: 10.1081/jlc-200052969

Vailaya A, Horváth C., (1998). Retention in reversed-phase chromatography: partition or adsorption? J Chromatogr A. 1998 Dec 31;829(1-2):1-27.

van Leeuwen, S. M., Blankert, B., Kauffmann, J. M., & Karst, U. (2005). Prediction of clozapine metabolism by on-line electrochemistry/liquid chromatography/mass spectrometry. [Research Support, Non-U S Gov't]. *Anal Bioanal Chem, 382*(3), 742-750.

Verdoliva, A., Pannone, F., Rossi, M., Catello, S., & Manfredi, V. (2002). Affinity purification of polyclonal antibodies using a new all-D synthetic peptide ligand: comparison with protein A and protein G. *Journal of Immunological Methods, 271*(1–2), 77-88. doi: 10.1016/s0022-1759(02)00341-1

Viswanathan, S., Unlu, M., & Minden, J. S. (2006). Two-dimensional difference gel electrophoresis. *Nat Protoc, 1*(3), 1351-1358.

Wagner, Y., Sickmann, A., Meyer, H. E., & Daum, G. (2003). Multidimensional nano-HPLC for analysis of protein complexes. *J Am Soc Mass Spectrom, 14*(9), 1003-1011.

Wang, M.-F., Han, C.-L., & Yin, S.-J. (2009). Substrate specificity of human and yeast aldehyde dehydrogenases. *Chemico-Biological Interactions, 178*(1–3), 36-39. doi: 10.1016/j.cbi.2008.10.002

Wong, H. L., and *Liebler*, D. C. (*2008*) Mitochondrial protein targets of thiol-

reactive electrophiles. Chem. Res. Toxicol., 21, 796-804.

Wang, X., Stoll, D. R., Schellinger, A. P., & Carr, P. W. (2006). Peak capacity optimization of peptide separations in reversed-phase gradient elution chromatography: fixed column format. [Research Support, N I H , Extramural

Wang, G., et al. (2009), "Decoy Methods for Assessing False Positives and False Discovery Rates in Shotgun Proteomics", Anal Chem. 81(1):146-159 Research Support, Non-U S Gov't]. *Anal Chem, 78*(10), 3406-3416.

Washburn, M. P., Wolters, D., & Yates, J. R., 3rd. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. [Research Support, U S Gov't, P H S]. *Nat Biotechnol, 19*(3), 242-247.

Weltzien HU, Moulon C, Martin S, Padovan E, Hartmann U, Kohler J(1996) : T cellimmune responses to haptens. Structural models for allergic and autoimmune responses.Toxicology 107:141–151,

Wen, B., & Fitch, W. L. (2009). Analytical strategies for the screening and evaluation of chemically reactive drug metabolites. *Expert Opinion on Drug Metabolism & Toxicology, 5*(1), 39.

Wen, B., & Fitch, W. L. (2009). Screening and characterization of reactive metabolites using glutathione ethyl ester in combination with Q-trap mass spectrometry. [Validation Studies]. *J Mass Spectrom, 44*(1), 90-100.

Wen, B., Ma, L., Nelson, S. D., & Zhu, M. (2008). High-throughput screening and characterization of reactive metabolites using polarity switching of hybrid triple quadrupole linear ion trap mass spectrometry. *Anal Chem, 80*(5), 1788-1799.

Wheelock, A. M., Morin, D., Bartosiewicz, M., & Buckpitt, A. R. (2006). Use of a fluorescent internal protein standard to achieve quantitative two-dimensional gel electrophoresis. [Research Support, N I H , Extramural Research Support, Non-U S Gov't]. *Proteomics, 6*(5), 1385-1398.

White, M. Y., & Cordwell, S. J. (2005). 11 Isoelectric focusing and proteomics. In G. David & A. Satinder (Eds.), *Separation Science and Technology* (Vol. Volume 7, pp. 247-264): Academic Press.

Wiesner, J., Premsler, T., & Sickmann, A. (2008). Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. [Research Support, Non-U S Gov't Review]. *Proteomics, 8*(21), 4466-4483.

Williams, D. P., O'Donnell, C. J., Maggs, J. L., Leeder, J. S., Uetrecht, J.,

Pirmohamed, M., & Park, B. K. (2003). Bioactivation of clozapine by murine cardiac tissue in vivo and in vitro. [Research Support, Non-U S Gov't]. *Chem Res Toxicol, 16*(10), 1359-1364.

Williamson, B. L., Marchese, J., & Morrice, N. A. (2006). Automated identification and quantification of protein phosphorylation sites by LC/MS on a hybrid triple quadrupole linear ion trap mass spectrometer. [Research Support, Non-U S Gov't]. *Mol Cell Proteomics, 5*(2), 337-346.

Wilm, M., Neubauer, G., & Mann, M. (1996). Parent ion scans of unseparated peptide mixtures. [Research Support, Non-U S Gov't]. *Anal Chem, 68*(3), 527-533.

Wisniewski, J. R., A. Zougman, et al. (2009). "Universal sample preparation method for proteome analysis." Nat Meth 6(5): 359-362.

Witzmann, F. A., Jarnot, B. M., Parker, D. N., & Clack, J. W. (1994). Modification of Hepatic Immunoglobulin Heavy Chain Binding Protein (BiP/Grp78) Following Exposure to Structurally Diverse Peroxisome Proliferators. *Toxicol. Sci., 23*(1), 1-8. doi: 10.1093/toxsci/23.1.1

Yan, J. X., Devenish, A. T., Wait, R., Stone, T., Lewis, S., & Fowler, S. (2002). Fluorescence two-dimensional difference gel electrophoresis and mass spectrometry based proteomic analysis of Escherichia coli. [Research Support, Non-U S Gov't]. *Proteomics, 2*(12), 1682-1698.

Yan, Z., & Caldwell, G. W. (2004). Stable-isotope trapping and high-throughput screenings of reactive metabolites using the isotope MS signature. *Anal Chem, 76*(23), 6835-6847.

Yan, Z., Maher, N., Torres, R., Caldwell, G. W., & Huebert, N. (2005). Rapid detection and characterization of minor reactive metabolites using stable-isotope trapping in combination with tandem mass spectrometry. *Rapid Commun Mass Spectrom, 19*(22), 3322-3330.

Yan, Z., Maher, N., Torres, R., & Huebert, N. (2007). Use of a trapping agent for simultaneous capturing and high-throughput screening of both "soft" and "hard" reactive metabolites. *Anal Chem, 79*(11), 4206-4214.

Yu,W., Taylor, J.A., Davis, M.T., Bonilla, L.E., Lee, K.A., Auger, P.L., Farnsworth, C.C., Welcher, A.A., and Patterson, S.D., (2010) Maximizing the sensitivity and reliability of peptide identification in large-scale proteomic experiments by harnessing multiple search engines DOI 10.1002/pmic.200900074 Proteomics, 10, 1172–1189

Yu, L. J., Chen, Y., Deninno, M. P., O'Connell, T. N., & Hop, C. E. C. A. (2005). IDENTIFICATION OF A NOVEL GLUTATHIONE ADDUCT OF DICLOFENAC, 4'-

HYDROXY-2'-GLUTATHION-DESCHLORO-DICLOFENAC, UPON INCUBATION WITH HUMAN LIVER MICROSOMES. *Drug Metab Dispos, 33*(4), 484-488. doi: 10.1124/dmd.104.002840

Yukinaga, H., Takami, T., Shioyama, S.-h., Tozuka, Z., Masumoto, H., Okazaki, O., & Sudo, K.-i. (2007). Identification of Cytochrome P450 3A4 Modification Site with Reactive Metabolite Using Linear Ion Trap-Fourier Transform Mass Spectrometry. *Chemical Research in Toxicology, 20*(10), 1373-1378.

Zappacosta, F., Huddleston, M. J., Karcher, R. L., Gelfand, V. I., Carr, S. A., & Annan, R. S. (2002). Improved sensitivity for phosphopeptide mapping using capillary column HPLC and microionspray mass spectrometry: comparative phosphorylation site mapping from gel-derived proteins. [Research Support, U S Gov't, P H S]. *Anal Chem, 74*(13), 3221-3231.

Zhang, D. D., Lo, S.-C., Cross, J. V., Templeton, D. J., & Hannink, M. (2004). Keap1 Is a Redox-Regulated Substrate Adaptor Protein for a Cul3-Dependent Ubiquitin Ligase Complex. *Mol. Cell. Biol., 24*(24), 10941-10953. doi: 10.1128/mcb.24.24.10941-10953.2004

Zhang, H., & Yang, Y. (2008). An algorithm for thorough background subtraction from high-resolution LC/MS data: application for detection of glutathione-trapped reactive metabolites. *J Mass Spectrom, 43*(9), 1181-1190.

Zhang, K., Zhu, Y., He, X., & Zhang, Y. (2011). Systematic screening of protein modifications in four kinases using affinity enrichment and mass spectrometry analysis with unrestrictive sequence alignment. *Analytica Chimica Acta, 691*(1–2), 62-67. doi: 10.1016/j.aca.2011.02.036

Zhang, Z. Y., & Dixon, J. E. (1993). Active site labeling of the Yersinia protein tyrosine phosphatase: The determination of the pKa of the active site cysteine and the function of the conserved histidine 402. *Biochemistry, 32*(36), 9340-9345. doi: 10.1021/bi00087a012

Zheng, J., Ma, L., Xin, B., Olah, T., Humphreys, W. G., & Zhu, M. (2007). Screening and identification of GSH-trapped reactive metabolites using hybrid triple quadruple linear ion trap mass spectrometry. *Chem Res Toxicol, 20*(5), 757-766.

Zheng, N., Zou, P., Wang, S., & Sun, D. (2011). In vitro metabolism of 17-(dimethylaminoethylamino)-17-demethoxygeldanamycin in human liver microsomes. [In VitroResearch Support, N I H , ExtramuralResearch

Support, Non-U S Gov't]. *Drug Metab Dispos, 39*(4), 627-635.

Zhou, S. (2003). Separation and detection methods for covalent drug-protein adducts. *Journal of Chromatography B, 797*(1-2), 63-90.

Zhou, S., Chan, E., Duan, W., Huang, M., & Chen, Y.-Z. (2005). Drug Bioactivation Covalent Binding to Target Proteins and Toxicity Relevance. *Drug Metabolism Reviews, 37*(1), 41 - 213.

Zhu, M., Ma, L., Zhang, H., & Humphreys, W. G. (2007). Detection and Structural Characterization of Glutathione-Trapped Reactive Metabolites Using Liquid Chromatography−High-Resolution Mass Spectrometry and Mass Defect Filtering. *Analytical Chemistry, 79*(21), 8333-8341. doi: doi:10.1021/ac071119u

Zhu, M., Ma, L., Zhang, H., & Humphreys, W. G. (2007). Detection and Structural Characterization of Glutathione-Trapped Reactive Metabolites Using Liquid Chromatography−High-Resolution Mass Spectrometry and Mass Defect Filtering. *Analytical Chemistry, 79*(21), 8333-8341. doi: 10.1021/ac071119u

Zoumaro-Djayoon, A. D., Heck, A. J. R., & Muñoz, J. (2012). Targeted analysis of tyrosine phosphorylation by immuno-affinity enrichment of tyrosine phosphorylated peptides prior to mass spectrometric analysis. *Methods, 56*(2), 268-274. doi: 10.1016/j.ymeth.2011.09.003

Zuo, X., Echan, L., Hembach, P., Tang, H. Y., Speicher, K. D., Santoli, D., & Speicher, D. W. (2001). Towards global analysis of mammalian proteomes using sample prefractionation prior to narrow pH range two-dimensional gels and using one-dimensional gels for insoluble and large proteins. [Research Support, U S Gov't, P H S]. *Electrophoresis, 22*(9), 1603-1615.