Biomarkers for Diagnosis and Prognosis of Heart Failure using label free LC-MS^{E}

Thesis submitted for the degree of

Doctor of Philosophy

University of Leicester

Janica Auluck

Department of Cardiovascular Sciences

Department of Cancer Studies and Molecular Medicine

University of Leicester

December 2013

Abstract: Biomarkers for diagnosis and prognosis of heart failure using label free LC-MS^{E}

In the UK 900,000 people suffer from heart failure, of which 30-40% die within 1 year of diagnosis. Heart failure is a prevalent disease worldwide and is associated with high rates of morbidity and mortality. Current biomarkers suffer from poor levels of accuracy and efficacy. Therefore, accurate, reproducible, and reliable diagnostic and prognostic biomarkers are needed.

In this study, we have chosen mass spectrometry based proteomics to profile patient plasma to discover diagnostic and prognostic biomarkers of heart failure. This experimental method allows simultaneous qualitative and quantitative analysis. Bioinformatic analysis of the protein profiles to detect protein changes has been undertaken to identify potential markers of disease.

Upon method development, plasma protein profiles from one hundred acute heart failure patients were obtained using a Waters Synapt G2 QTOF mass spectrometer post plasma enrichment (Proteominer^{TM}, Bio-Rad) and 2D-RP-RP fractionation. Samples were analysed using a HDLC-MS^{E} experiment and run in triplicate. Statistical comparisons of the protein profiles were made using PLGS v2.5.2 and progenesis LC-MS to identify potential candidates for biomarkers.

Using a label free 2D HDLC-MS^{E} experiment we have found that differences in protein expression of acute heart failure patient profiles exist. Seven candidate proteins have been identified and are shown to be involved in many different physiological processes that play a role in the pathophysiology of heart failure. ELISA analysis of the seven identified markers has identified SAP as a strong predictor of adverse patient outcome in acute heart failure. Using multivariate analysis, SAP has been found to be an independent prognostic marker in acute heart failure patients.

Further studies are needed to verify SAP as a biomarker in a larger patient cohort, and measure SAP alongside current prognostic markers. A mechanistic study to identify the role of SAP in heart failure pathology needs to be undertaken.

Acknowledgments

First and foremost, I would like to thank my supervisors Dr Don Jones and Prof Leong Ng for their continued support throughout my PhD. To Don, thank you for introducing me to mass spectrometry and pushing me to achieve my best. To Prof, thank you for the countless creative ideas along the way and reminding me how much fun science can be. It is with your combined help and guidance that I have reached this point today and for that, I am very grateful.

I would also like to thank the vanGeest foundation for funding this study and much of the equipment used to complete this PhD. I would also like to acknowledge those at Waters Corporation and Non-Linear Dynamics for their help. I have been very lucky to receive the support of so many people throughout the course and especially from those based at the John and Lucille vanGeest biomarker facility.

I would like to take this opportunity to extend a special thanks to Jodie who kept me sane and always knew when it was time to step in and take me for a coffee (even though I don't drink coffee!) Thank you.

Thank you to all my friends and family (old and new) for reminding me of my life outside the lab and providing me with encouragement when I've needed it the most. Thank you to all those who have lived with me and this PhD both in and outside of Leicester.

Special thanks go to my parents and sister, they have been a pillar throughout my time at Leicester, and without their unconditional love and support, this would have been a lot harder to achieve.

Finally yet importantly, I would like to thank Arjun for his continued understanding and patience, especially in my final year when so much of my time has been devoted to my work. You always found a way to restore my faith. Thank you.

Table of contents

List of Figures

Chapter One

Chapter Two

Chapter Three

Chapter Four

Chapter Five

Chapter Six

List of Tables

Chapter One

Chapter Two

Chapter Three

Chapter Four

Chapter Five

Chapter Six

Abbreviations

15-PGDH	15-Hydroxyprostaglandin dehydrogenase
1D	One dimensional
2D	Two dimensional
ACN	Acetonitrile
ADH	Alcohol dehydrogenase
ANP	Atrial natriuretic peptide
AUC	Area under the curve
BMI	Body mass index
BNP	Brain natriuretic peptide
BP	Blood Pressure
BCA	Bicinchonic acid
BSA	Bovine serum albumin
CETP	Cholesteryl ester transfer protein
CHD	Coronary heart disease
CID	Collision induced dissociation
CoV	Coefficient of variance
CRP	C reactive protein
CVD	Cardiovascular disease
CYS D	Cystatin D
DDA	Data dependent acquisition
DIA	Data independent acquisition
DTT	Dithiothreitol
ECD	Electron capture dissociation
ECG	Electrocardiograms
eGFR	Glomerular filtration rate
ECHO	Echocardiogram
ELISA	Enzyme linked immunosorbant assay
ESI	Electrospray ionisation
ETD	Electron transfer dissociation
EWP	Extra wide pore
FA	Formic acid

FDR	False discovery rate
FT-ICR	Fourier transform ion cyclotron
GFP	Glu-fibrinogen peptide
GFR	Glomerular filtration rate
GP	General practioner
HDL	High density lipoprotein
$HDMS^E$	High definition mass spectrometry elevated
HF	Heart failure
HPLC	High pressure liquid chromatography
HUPO	Human Plasma Proteome Project
IAA	Iodoacetamide
ICAT	Isotope coded affinity tag
ICPL	Isotope coded protein labelling
IGF	Insulin growth factor
IGF2	Insulin like growth factor 2
ILMA	Immunoluminometric assay
IM-MS	Ion mobility-mass spectrometry
IT	Ion trap
iTRAQ	Isobaric tags for relative and absolute quantification
JVP	Jugular venous pressure
LC	Liquid chromatography
LDL	Low density lipoproteins
LRG	Leucine rich alpha 2 glycoprotein
m/z	Mass to charge ratio
MACE	Major adverse cardiac event
MAE	Methyl acridinium ester
MALDI	Matrix assisted laser desorption/ionisation
MASP1	Mannan binding lectin serine protease 1
MASP2	Mannan binding lectin serine protease 2
MBL	Mannose-binding lectin
MI	Myocardial infarction
MS	Mass spectrometry
MS^E	Mass spectrometry elevated

nano-ESI	Nano-electrospray
NP	Natriuretic peptide
NPR-A	Natriuretic peptide receptor A
NT-proBNP	N terminal-proBNP
oa-TOF	Orthogonal acceleration time of flight
PCA	Principle component analysis
PLGS	Protein Lynx Global Server
PLTP	Phospholipid transfer protein
Q	Quadrupole
QC	Quality control
QqQ	Triple quadrupole
RAAS	Renin-angiotensin-aldosterone system
RLU	Relative light units
RP	Reverse-phase
RP- SPE	Reverse-phase solid phase extraction
RSD	Relative standard deviation
SAP	Serum amyloid P component
SCX	Strong cation exchange
SDC	Sodium deoxycholate
SDS	Sodium dodecyl sulfate
SILAC	Stable isotope labelling by amino acids in cell culture
SIM	Selected ion monitoring
SPSS	Statistical Package for Social Sciences
SRM	Selected reaction monitoring
TFA	Trifluroacetic acid
TMT	Tandem mass tags
TOF	Time of flight
TWIMS	Travelling wave ion mobility
TWIG-IM	Travelling wave ion guide ion mobility
UPLC	Ultra high pressure liquid chromatography
vLDL	Very low density lipoproteins
WHO	World Health Organisation

Chapter One

Introduction

PART ONE: Heart Failure

1.1 Cardiovascular disease

Cardiovascular disease (CVD) remains one of the leading causes of morbidity and mortality worldwide. The World Health Organisation (WHO) reports that CVD is responsible for more than 17.3 million deaths *per annum*. In 2008, three million of these deaths occurred in those aged below 60 years of age and could have largely been prevented (Mendis *et al.*, 2011). The burden of CVD on health economy and mortality rates is projected to rise, and WHO estimates that there will be approximately 20 million CVD related deaths by 2015 accounting for 30% of death worldwide (Figure 1.1) (Fuster & Kelly, 2010).

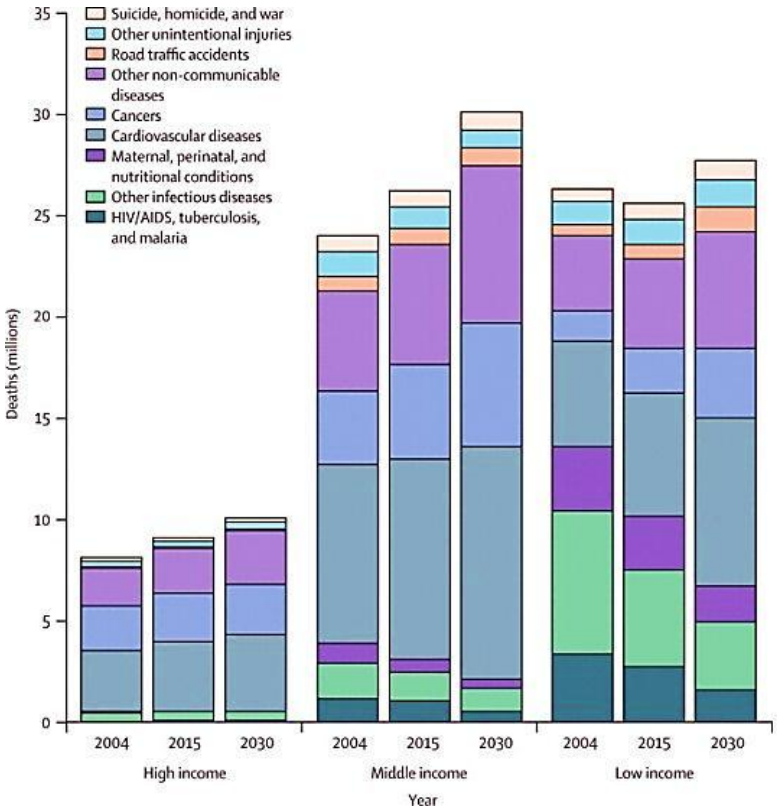


Figure 1.1: Projected global deaths by cause and projected rise in the incidence of CVD disease in the future. Taken from (Beaglehole & Bonita, 2008).

This strikingly high prevalence of CVD worldwide is reflected in the U.K. accounting for almost 180,000 deaths in 2010. Approximately 46,000 of these deaths were premature and 68% of these were in men. CVD places considerable financial burden on the UK health care system, costing £8.7 billion in 2009 (Townsend N *et al.*, 2012).

CVD is part of a class of many diseases that affect the cardiovascular system i.e. disorders of the heart and the blood vessels. Some of the most commonly associated disorders of CVD are coronary heart disease, myocardial infarction (MI), angina pectoris, stroke, heart valve disease, cardiomyopathy, and heart failure.

1.2 Heart Failure

Heart failure (HF) is a syndrome caused by structural or functional abnormalities of the cardiac muscle that leads to the impairment of the heart's circulatory function. The heart is unable to supply sufficient blood flow to meet the oxygen requirements of the major organs. The eventual outcome of heart failure is either left ventricular diastolic or systolic dysfunction (Mosterd & Hoes, 2007). Currently diagnosis of heart failure is determined by using a combination of medical history, clinical examinations, and non-specific laboratory tests. A definitive laboratory test for heart failure diagnosis and prognosis is yet to be discovered (Harris & Thompson, 2012).

1.2.1 Epidemiology

Currently 14 million people in Europe suffer from heart failure with 3.6 million people diagnosed each year (HFM, 2007). In the UK 900,000 people have heart failure with equally as many suffering from myocardial damage that have yet to show signs of eventual heart failure. While the incidence continues to stabilise the prevalence of the disease appears to be increasing. Approximately one in five hundred people are discharged from hospital with heart failure making it one of the most common causes for admission for people aged over 65 years. Of all those hospital admissions, one in four patients are re-admitted within 3 months (National Institute for Health and Clinical Excellence, 2010; McMurray & Pfeffer, 2005).

These statistics coupled with the increasing elderly population may lead to heart failure becoming an increasing economic burden on health care systems worldwide. Biomarkers have the potential to have a huge impact on early detection, disease monitoring, therapeutic decision making and clinical outcome for patients and huge financial savings for health care systems.

1.2.2 Pathophysiology

Heart failure is a multi-organ disorder characterised by abnormalities of the cardiac muscle, skeletal muscle, and renal function; stimulation of the sympathetic nervous system; and a complex pattern of neurohormonal changes (Jackson *et al.*, 2000).

None of the progressive pathological explanations for heart failure is mutually exclusive; they all play a part in progression of the disease. The trigger for progression to heart failure is an initial acute (e.g. MI) or chronic (e.g. longstanding hypertension) deleterious event which causes structural or functional damage to the myocardium. A host of compensatory mechanisms are initiated as the heart attempts to retain the ability to maintain cardiac output, blood pressure, and fluid balance. These compensatory mechanisms begin to alter the muscular and biochemical cardiac environment for damaged and undamaged cells and over time lead to either systolic or diastolic cardiac dysfunction (Krum & Abraham, 2009; Muslin, 2012; Rathi & Deedwania, 2012).

In response to effects on the myocardium, the human body activates a neurohormonal response. This includes the adrenergic nervous system and renin-angiotensin-aldosterone system (RAAS) which is responsible for maintaining cardiac output using peripheral arterial vasoconstriction, increased contractility and retention of salt and water. This also activates inflammatory mediators, which carry out cardiac tissue repair and remodelling. These systems are active in a compensatory manner in acute adaptive phases, the continued activation of the neurohormonal response is when the chronic deleterious cardiac dysfunction takes place and leads to heart failure (Rathi & Deedwania, 2012; MacIver *et al.*, 2013).

The sympathetic nervous system is one of the most important responses and provides the myocardium with inotropic support to maintain cardiac output (Jackson *et al.*, 2000). Inhibitory inputs from high-pressure and low-pressure baroreceptors are the principal inhibitors of the sympathetic outflow. Healthy individuals are able to maintain a variable heart rate and low sympathetic discharge at rest. In heart failure patients, the inhibitory input decreases and the excitatory input increases. The activation of the sympathetic nervous system causes an increase in circulating levels of noradrenalin, which results in increased heart rate, cardiac contractility, and increased

vasoconstriction. Chronic activation has harmful effects causing structural and functional changes. In heart failure patients, the circulating concentrations of noradrenalin are often elevated 2-3 fold in comparison to healthy individuals (Muslin, 2012).

The activation of the sympathetic nervous system also causes stimulation of the renal tissue activating RAAS. This leads to increased concentrations of circulating renin, plasma angiotensin II and aldosterone. Angiotensin is a potent vasoconstrictor, which causes vascular and cardiac remodelling and leads to salt and water retention via aldosterone release. Monocyte released angiotensin II increases cardiac contractility that acts as a compensatory mechanism for initial cardiac injury. Chronic activation leads to the production of growth factors causing myocyte hypertrophy and left ventricular hypertrophy. Hypertrophy is often seen in chronic heart failure (McMurray & Pfeffer 2005; Jackson *et al.,* 2000; Steen & Mann, 2004).

Endothelins are vasoactive peptides that increase as heart failure worsens. Endothelin-1 is the most prevalent in plasma. Endothelins are potent vasoconstrictors on renal vasculature promoting the retention of sodium, cardiomyocyte hypertrophy, cardiac fibrosis, and increased cardiomyocyte contractile function (Muslin, 2012).

Atrial natriuretic peptide (ANP) and brain natriuretic peptide (BNP) are released in response to atrial and ventricular stress and act on the heart, kidneys, and central nervous system (Figure 1.2). Both BNP and ANP have a protective role, as they bind to natriuretic peptide A receptor which results in natriuresis, vasorelaxation, inhibition of renin and aldosterone and inhibition. Circulating levels of natriuretic peptides increase in heart failure. BNP is mostly secreted by the ventricles in healthy and heart failure patients. A precursor of BNP (proBNP) is stored in myocytes and is activated to form the biologically active BNP and N terminal (NT)–proBNP. The specificity of BNP has led to BNP and NT-proBNP being measured in heart failure patients (Lee & Burnett, 2007; Levin *et al.*, 1998).

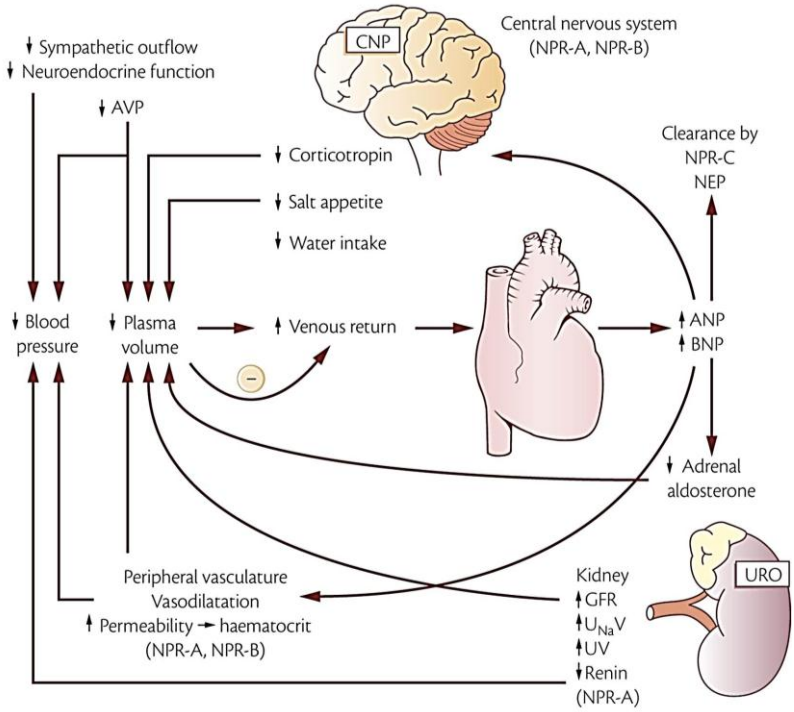


Figure 1.2: Multi organ effect of natriuretic peptides (NP) in HF. Increased NP reduces blood pressure and plasma volume. Urodilatin (URO); neutral endopeptidase (NEP); C-type natriuretic peptide (CNP); natriuretic peptide receptors A, B and C (NPR-A, NPR-B, and NPR-C respectively); arginine vasopressin (AVP); atrial and brain natriuretic peptides (ANP and BNP); glomerular filtration rate (GFR); urinary sodium excretion (UNaV); urinary volume (UV); blood pressure (BP). Taken from (Levin *et al.*, 1998).

The sympathetic response, RAAS, natriuretic peptides are the main neurohormonal responses to heart failure. Other responses include; Prostaglandins (E_2 and I_2) that protect renal function by maintaining the glomerular filtration through dilation of afferent glomerular arterioles. Inflammatory responses to cardiac injury also contribute to heart failure pathology. Oxygen free radicals inactivate nitric oxide, reducing myocardial contractility and may induce apoptosis (Camm AJ, 2009).

Cardiac remodelling, particularly left ventricular remodelling, is the consequence of the continual neurohormonal activation, mechanical and genetic factors that alter ventricular size, shape, and function (Figure 1. 3). Many clinical triggers initiate the remodelling cascade such as cardiomyopathy, hypertension, valvular heart disease, and MI. Consequently, the remodelled cardiac tissue becomes hypertrophied with loss of myocytes and increase interstitial fibrosis (Jessup & Brozena, 2003).

Other mechanisms also contribute to the pathology of hypertrophied ventricles. The extracellular matrix provides scaffolding and differentiates cardiac and non-cardiac muscle. Modification to the matrix causes ventricular remodelling. After a cardiac injury, there is an increased deposition of fibrous proteins such as collagen and fibronectin that alter the mechanical properties of the ventricle. These alterations are propagated by the activation of RAAS and production of aldosterone, which in turn is responsible for extracellular matrix proliferation and increased fibrous tissue deposits within the ventricles, directly affecting the ventricle and altering the coronary microcirculation (Rathi & Deedwania, 2012).

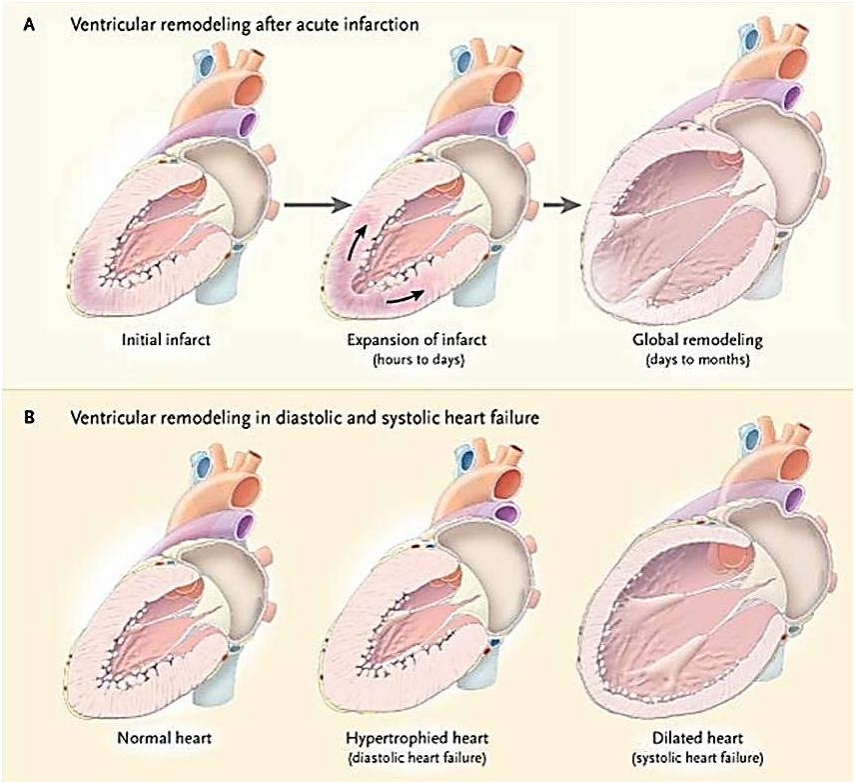


Figure 1. 3: Ventricular remodelling after a MI (A) and in systolic and diastolic heart failure (B). Post-MI there is little change to the ventricular wall; over a few hours, the area affected by the infarct expands and becomes thinner. Over a long period, the entire ventricle can remodel itself to become dilated, causing reduced systolic function, mitral valve dysfunction, and possible aneurysm formation. In diastolic and systolic heart failure the remodelling, results in a normal size left ventricle cavity with thickened ventricle walls and preserved systolic function (diastolic HF). In systolic HF, the ventricle will undergo dilation and become spherical, with decreases systolic function and distortion of the mitral valve (Jessup & Brozena, 2003).

1.2.3 Signs and symptoms

The classic characteristic symptoms in heart failure are shortness of breath, fatigue, and oedema. The most common presenting symptom is shortness of breath at either exertion or rest. Additional symptoms are summarised in Table 1.1 (Harris & Thompson, 2012; Remme & Swedberg, 2001).

Table 1.1: Common symptoms of heart failure

Common symptoms used to make initial diagnosis of Heart failure	**Uncommon symptoms that cannot be used to make diagnosis of heart failure**
Shortness of breath	Nocturia
Exercise intolerance	Anorexia
Fatigue	Abdominal bloating
Oedema	Constipation
	Cerebral symptoms

Determining the severity of cardiac dysfunction from symptoms alone is imprecise as symptoms are extremely non-specific. In contrast, highly specific symptoms also lack sensitivity. For example, upon physical and clinical examination patients may have an elevated jugular venous pressure (JVP) which is highly suggestive of heart failure when it is the suspected diagnosis. While elevated JVP has a specificity of 97%, the sensitivity of the test is only 10% (Harris & Thompson, 2012).

The diagnosis of heart failure based on signs and symptoms requires an accurate patient history to accompany the clinical examination to achieve accurate diagnosis. A biomarker led diagnosis and prognosis could achieve the accuracy, sensitivity, and specificity needed to deal with heart failure.

1.2.4 Risk factors

Many risk factors are associated with, and predispose patients to heart failure as summarised in Figure 1.4. Heart failure does not present in isolation; underlying cardiac defects such as hypertension, coronary heart disease (CHD) and many other risk factors are present. Consequently, heart failure often presents with co-morbidities.

Hypertension is a well-established risk factor of heart failure. Hypertension is thought to initiate cardiac remodelling by promoting myocyte hypertrophy, myocardial fibrosis and loss of myocardial contractile tissue (Kenchaiah *et al.*, 2004). In follow up studies of the original Framingham study and Framingham Offspring study, 91% of patients with hypertension went on to develop heart failure, of which males had a 2-fold increased risk and females a 3-fold risk of developing heart failure in their life time (Mosterd & Hoes, 2007).

Coronary heart disease profoundly increases the risk of developing heart failure due to restricted supply of oxygen to the cardiac muscle. Patients who develop heart failure from CHD are associated with a worse prognosis than those who do not (Cas *et al.*, 2003).

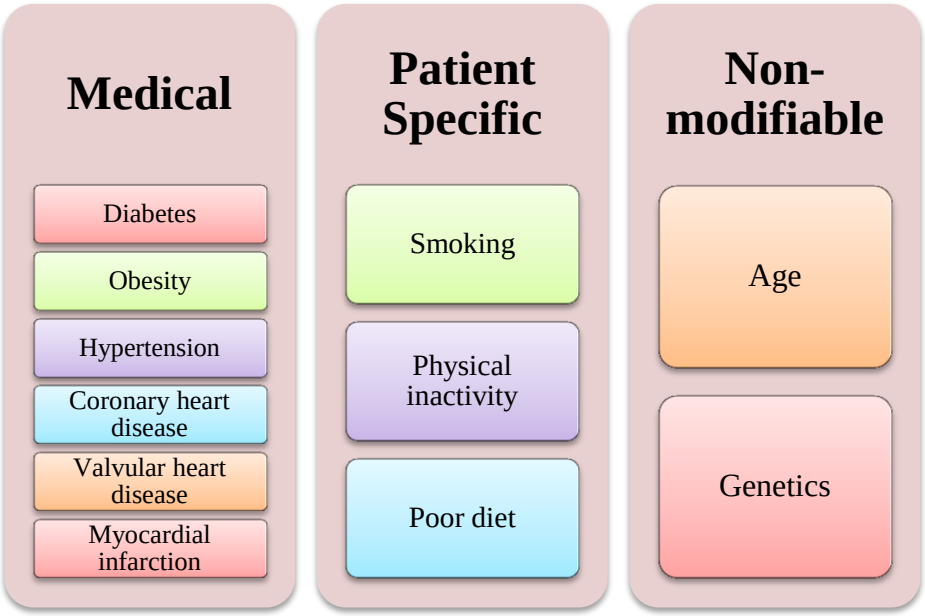


Figure 1.4: Clinical risk factors associated with heart failure

Myocardial infarction influences changes to the hearts muscle leading to heart failure through cardiac remodelling, increasing mechanical wall stress, neurohormonal activation, and cytokine release. Seven to eight years post-MI, 36% of patients go on to develop heart failure and with increased probability in patients with left ventricular systolic dysfunction (Pfeffer & Braunwald, 1990).

Obesity is an increasingly common risk factor of heart failure; obese patients double their risk of developing heart failure (Cas *et al.*, 2003). In addition to these risk factors, valvular heart disease, genetic predisposition and lifestyle risk factors such as smoking, poor diet, and a sedentary lifestyle all confer an increased risk of heart failure.

Diabetes is a well-documented co-morbidity and risk factor of heart failure. Diabetic patients have a grossly increased risk, and are associated with a worse prognosis, once a cardiac event has taken place. The Framingham study found patients who were diagnosed with diabetes had a 4-fold increase in the male population and 8-fold risk in the female population of developing heart failure (Kannel & McGee, 1979).

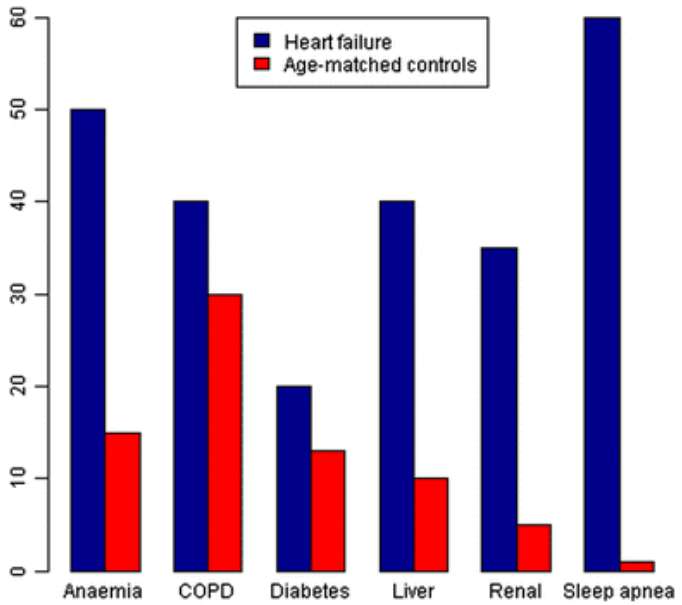


Figure 1.5: Prevalence of co-morbidities in patients with heart failure (BLUE) compared to age matched controls (RED) (Deursen *et al.*, 2012).

Other co-morbidities that are often observed in heart failure include; anaemia, renal and hepatic impairment, obstructive sleep apnoea, and chronic obstructive pulmonary disease (Figure 1.5) (McMurray & Pfeffer, 2005). The co-morbidities along with the pathology of heart failure and increasing age greatly determine the progression and prognosis of disease. Those at risk from heart failure may lead to the progression of asymptomatic or symptomatic heart failure (Figure 1.6).

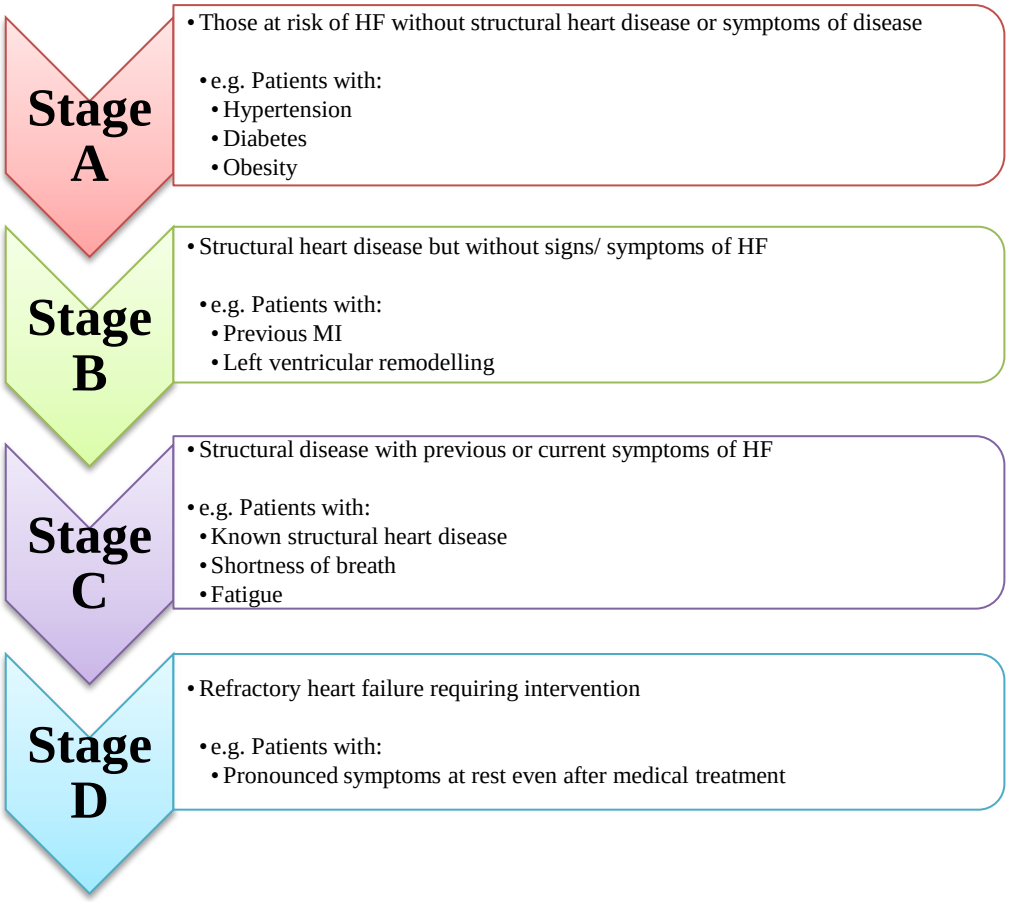


Figure 1.6: Stages of heart failure. The stages show the progression of co-morbidities and risk factors lead to the progression of disease. Adapted from (Krum & Abraham, 2009).

1.2.5 Classification of Heart Failure

1.2.5.1 Chronic and Acute Heart Failure

Heart failure can be chronic and acute, chronic heart failure develops gradually whereas acute heart failure develops rapidly after an event such as MI that leads to death of cardiac tissue or rupture of a cardiac valve. Acute heart failure can clinically present in different forms; acute pulmonary oedema secondary to cardiac dysfunction, cardiogenic shock characterised by hypotension, oliguria and peripheral vasoconstriction and acute worsening of chronic heart failure (Mosterd & Hoes, 2007).

1.2.5.2 Systolic and Diastolic Heart Failure

The heart circulates blood around the circulatory system through a pump-like action by contracting and relaxing the cardiac muscle; these two phases are termed systole and diastole respectively. Traditionally heart failure was thought to result from the

impairment in the heart's ability to pump sufficient volumes of blood into the arteries and circulation (systolic dysfunction). However, heart failure is also found to occur in patients with normal systolic function but where higher pressures are required to refill the heart with blood (diastolic dysfunction) (MacIver *et al.*).

Patients are diagnosed with either systolic of diastolic heart failure after the ejection fraction (stroke volume/end diastolic volume of ventricular chamber) of a patient is determined, usually by echocardiogram (ECHO). Patients with diastolic heart failure have preserved left ventricular ejection fraction greater than 50% and those with systolic heart failure have reduced ejection fraction less than 40-45%. However, this is a very subjective classification as there is still disagreement regarding the cut-off for preserved ejection fractions (Authors/Task Force Members *et al.*, 2012). This thesis will focus on acute systolic heart failure.

1.2.6 Diagnosis

Heart failure can either present in an acute or chronic setting to their general practitioners (GP's) which, leads to both over and under diagnosis. Approximately 25-60% of patients referred to hospital specialists by their GP's due to suspected heart failure symptoms eventually end with the diagnosis of heart failure (Fonseca, 2006).

Diagnosis of heart failure uses a number of techniques and assessments to determine specific types of heart failure (e.g. systolic or diastolic). Patients are diagnosed clinically upon presentation of cardinal signs and symptoms. Diagnosis is confirmed using bio-imaging e.g. electrocardiogram, chest x-ray, and echocardiogram. Laboratory testing of biochemical markers is used to exclude patients that are likely to be free of heart failure and do not need a scan such as BNP and NT-proBNP (Gaggin & Januzzi Jr.).

1.2.6.1 Bio-imaging

Patients who present with reduced myocardial function (i.e. ejection fraction) are easier to diagnose than those with preserved function. Electrocardiograms (ECG) determine the level of myocardial function; display rhythm, rate, and conduction but are best used to diagnose left ventricular systolic dysfunction. Assessment of diastolic function

remains complex, in patients with diastolic dysfunction tissue doppler imaging is used to aid diagnosis but with limited improvement. Chest radiographs identify pulmonary oedema, pleural effusions or cardiomegaly and blood chemistry and haematological tests use circulating biomarkers to identify disease (see section 1.2.6.2). The use of MRI with gadolinium contrast is becoming a popular imaging technique, it provides precise assessment of ventricular structure and function (Harris & Thompson, 2012).

A limitation of bio-imaging techniques is they are user specific, requiring skilled interpretation for the test to be of accurate diagnostic value and leaving little room for misdiagnosis (Remme & Swedberg, 2001; Authors/Task Force Members *et al.*, 2012).

1.2.6.2 Biomarkers

Laboratory tests for circulating biomarkers in patient biofluids are used to confirm diagnosis of heart failure, some of the routine tests include; full blood count, urea and electrolyte, CRP, TNF-α, etc. The use of NP is increasing in the diagnostic process of heart failure. Heart failure biomarkers are broadly divided into groups based on the pathophysiology of heart failure. Figure 1.7 lists all the diagnostic and prognostic biomarkers that are proposed or established in routine laboratory testing.

Myocardial biomarkers such as troponins (T and I) are essential in acute presentation as elimination for MI. Other tests such as serum uric acid, c-reactive protein (CRP) and thyroid stimulating hormone are occasionally used as confirmatory markers. Elevated creatinine and urea is common in treated or severe heart failure as reduced glomerular filtration rate and renal dysfunction can cause these to change. Liver enzymes such as aspartate aminotransferase and alanine aminotransferase are also elevated because of heart failure, are thought to be induced by reduced hepatic blood flow, and may be prognostically significant (Kubo *et al.*, 1987).

BNP and NT-proBNP are currently the most studied and developed for clinical use as heart failure biomarkers. Pro-BNP is cleaved at the N terminus to produce BNP and inactive counterpart NT-proBNP. The levels of these biomarkers are significantly increased above baseline levels upon ventricular stress. This makes them ideal biomarkers to measure, with half lives of 22 min for BNP and 60-120 min for NT-

proBNP (Isaac, 2008). BNP and NT-proBNP are used as elimination tests; a normal plasma concentration of either in untreated patients indicates heart failure is unlikely. Equally, elevated concentrations of either identify patients who require further investigations e.g. ECHO. However, natriuretic peptides are affected by previous patient treatment causing the peptide circulating concentration to fall within the normal range (Camm AJ, 2009).

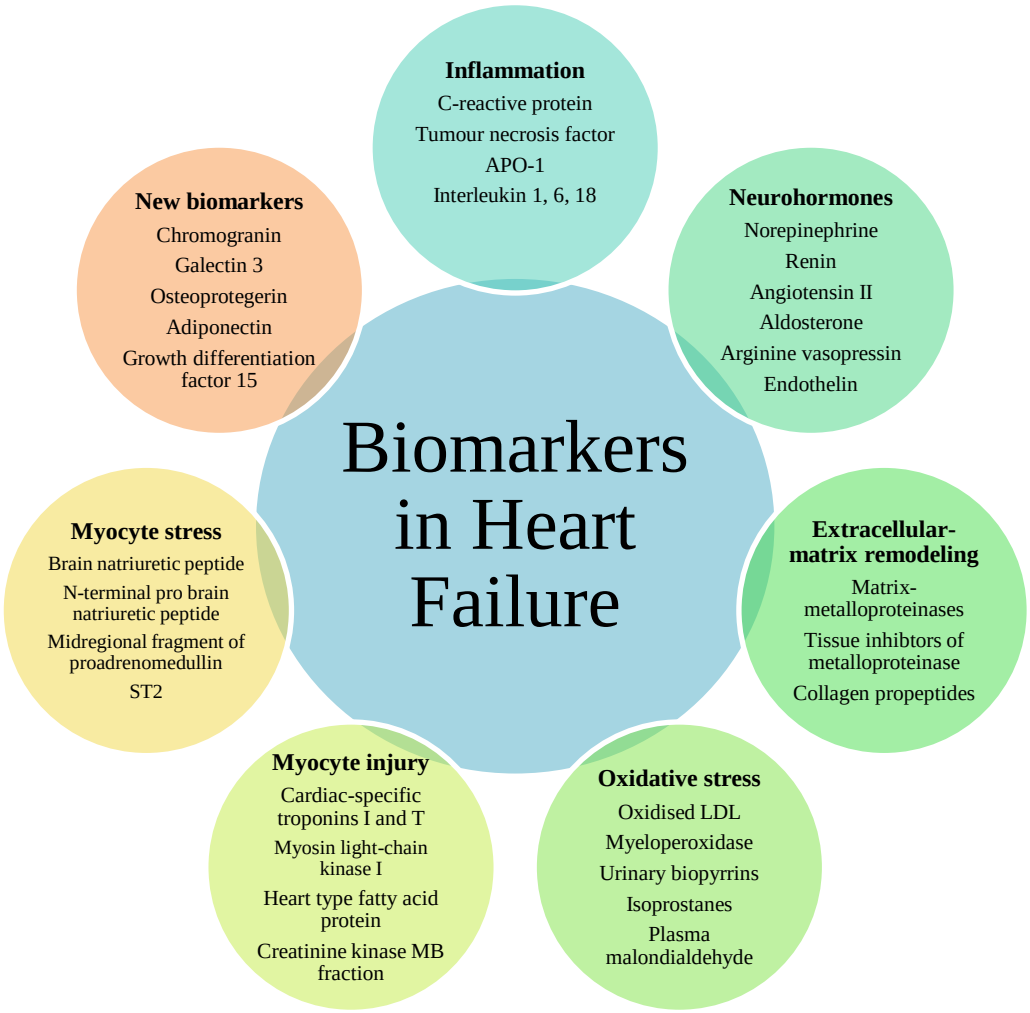


Figure 1.7: Current and proposed biomarkers of heart failure based on disease pathology. Adapted from (Braunwald, 2008)

1.2.7 Prognosis

Poor prognosis is another feature of heart failure, 30-40% of patients diagnosed with acute heart failure die within 1 year of initial diagnosis and 60-70% within 5 years. This five-year survival rate is worse than the survival rate for breast and prostate cancer (Askoxylakis *et al.*, 2010).

Prognosis of heart failure remains difficult. Typically, prognosis is determined using the following criteria: patient characteristics and co-morbidity, laboratory assessments, functional parameters with ventricular function and patient treatment plan. Sudden cardiac death is the major cause of mortality in patients and decompensated heart failure is a frequent occurrence requiring rehospitalisation (Mosterd & Hoes, 2007).

Although assessing different criteria may help clinicians estimate the prognosis of a patient, the method is far from precise. Nor is it a clear indicator of a patient's worsening condition. Prognostic information regarding patients undergoing change in therapy, such as plasma marker variations and long-term prognosis is scarce (McMurray & Stewart, 2000).

Biomarker monitoring for prognosis would allow intervention or change in therapeutic approach to take place before rehospitalisation or cardiac death. Prognostic biomarkers will enable clinicians to monitor patients, make timely and accurate decisions regarding patient therapy, tailored to their condition.

1.2.8 Treatment

There are many pharmacological treatment options for heart failure. Patients can be treated with either one or a combination of the following drugs: angiotensin converting enzyme inhibitors, beta blockers, aldosterone antagonists, angiotensin receptor blockers, hydralazine or isosorbide dinitrate, digoxin, diuretics, anti-coagulants (warfarin), anti-platelet agents, and HMG CoA reductase inhibitor (statins). Surgical or device interventions are typically avoided as a treatment option however, in some cases it is required (National Institute for Health and Clinical Excellence, 2010; Dickstein *et al.*, 2009).

The key to successful treatment and management lies in greater understanding of the underlying pathology and effective use of biomarkers. If greater precision, specificity, and sensitivity are introduced to diagnostic and prognostic biomarkers, individualisation of treatment or personalised targeted medicine may help tailor treatment plans and reduce the mortality and morbidity in heart failure.

1.2.9 Limitations of current Diagnostic and Prognostic methods

ECHO is the gold standard technique for diagnosis and prognosis in heart failure. However, there is some evidence that biomarkers may be more sensitive than ECHOs in some clinical conditions (Emdin *et al.*, 2009). A meta-analysis by Ewald *et al.* discovered that, in symptomatic heart failure, BNP was a better indicator of severity than ECHO assessment (Ewald *et al.*, 2008). Biomarkers are becoming increasingly important in clinical practice as they do not suffer from human error associated with human operated techniques such as bio-imaging.

Current biomarkers also have limitations. Circulating plasma concentrations of BNP and NTproBNP are affected by many variables such as impaired renal function, obesity, age, and gender. The specific use of NP as a diagnostic and prognostic tool is not recommended as it is affected by a multitude of cardiac conditions, making the peptides ideal markers of diagnostic exclusion not inclusion (Chen *et al.*, 2010; McLean *et al.*, 2008).

A study by Latini *et al.* found chronic heart failure could be detected in 92% of patients using a high sensitivity troponin T assay (Schunkert *et al.*, 2011; Latini *et al.*, 2007). However, patients have to wait 12 hours for the troponins to peak in plasma to become detectable. MacLean *et al.* suggested troponins lack sensitivity and specificity as they are elevated in chronic and acute CVD as well as patients with non-CVD diseases such as renal failure, trauma, and sepsis (McLean *et al.*, 2008). Therefore, although the test is sensitive and specific for myocyte injury as a diagnostic and prognostic test, it is not specific for heart failure.

Evidently, due to the high rates of morbidity, mortality, economic burden and the limitations of current biomarkers of heart failure, rapid, reliable and reproducible

diagnostic and prognostic biomarkers are needed which are not subject to variation as seen with the NP and troponin. As discussed, heart failure arises due to the activation of many different systems and multiple organs are involved in the progression of the disease. A multi-marker strategy is the most comprehensive model for obtaining information about a patient's heart failure pathology. The discovery of new biomarkers could open the gateway for more effective diagnosis and prognosis with new and stronger therapeutic targets for heart failure.

PART TWO: Proteomics and biomarker discovery

1.3 Proteomics

Proteomics is the large-scale study of the proteome, particularly in respect of the expression, structure, function, modifications, interactions, and changes in different environments and conditions (Mallick & Kuster, 2010). The term proteome was first introduced by Wilkins *et al.,* in 1994 (Wilkins *et al.,* 1996) as the PROTEin complement of the genOME (Figure 1.8). The number of proteins exceeds the number of genes, as proteins can be expressed in their native form, with post-translational modifications or gene-splicing induced changes.

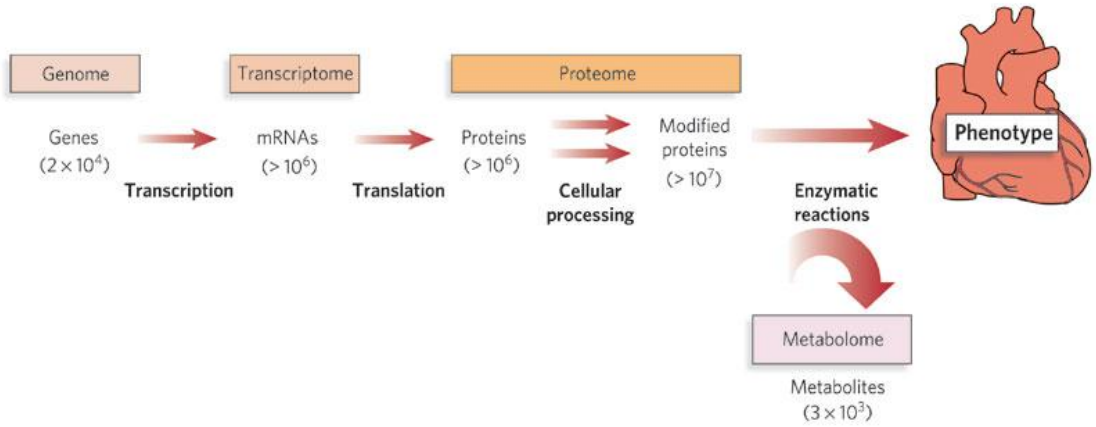


Figure 1.8: Expression of the proteins from genes. Taken from (Gerszten & Wang, 2008)

Proteomics has rapidly gained interest as an approach to help identify novel diagnostic and prognostic protein biomarkers. Bio-fluids are the sample of choice for proteomic biomarkers as they are easily accessible. Hypothetically, tissue that is in contact with biofluids such as plasma, urine and cerebral spinal fluid, release its protein components into the circulation. In a diseased state the proteins released from the tissue or in circulation could increase, decrease or the protein expression profile could change entirely. Proteins that undergo such changes would be ideal candidates for disease biomarkers. Proteomic analysis of the body fluid could build a protein profile of different disease states and reveal new biomarkers of disease. Proteins are currently used in various conditions for diagnosis such as prostate specific antigen for prostate cancer (Wang *et al.*, 1981) or cardiac troponins for acute MI (Mair *et al.*, 1992; Antman *et al.*, 1996). As these proteins exist in low concentrations, profiling or discovery

proteomic studies that search for all tissue leakage proteins could provide new disease biomarker candidates.

1.3.1 Plasma in proteomic analysis

Plasma is the most comprehensive and complex human proteome containing receptor ligands, proteins that have leaked from cells into circulation, secretions from diseased tissue. Despite its complex nature, the plasma proteome is the most comprehensive of the entire human proteome than any other tissue or biofluid and is how most clinical markers are measured. However, its complexity also poses an analytical challenge, which is not trivial (Street & Dear, 2010). Plasma was the biofluid of choice for this study and is discussed with regard to proteomics applications here on in.

1.3.2 Complexity of Plasma

The plasma proteome consists of a variety of different proteins, and as suggested by Anderson and Anderson (Anderson & Anderson, 2002) these proteins can be divided into several functional groups:

- Proteins secreted by solid tissues and act in plasma, largely secreted by the liver and intestines.
- Immunoglobulins, thought to be up to 10 million different sequences of antibodies in circulation.
- Long distance receptor ligands, this includes peptide and protein hormones and come in different sizes.
- Local receptor ligands such as cytokines and other mediators of cellular responses are good indicators of diseases such as sepsis.
- Temporary passengers, non-hormonal proteins such as lysosomes that are secreted and taken up via receptors for sequestration in lysosomes.
- Tissue leakage proteins, these proteins normally function within a cell and release into plasma due to cell damage or death, these make up an important class of plasma biomarkers such as cardiac troponins.
- Aberrant secretions such as tumour proteins and diseased tissue proteins that release into circulation are another functional group and often thought to have little functional requirement.

- Foreign proteins that are secreted by parasites that are released into circulation

From the list, it is clear that the human plasma proteome is incredibly complex. Protein concentrations in plasma range below nanogram per millilitre-milligram per millilitre, with a dynamic range from 10-12 orders of magnitude. As Figure 1.9 demonstrates, twenty two plasma proteins dominate 99% of the proteome, exceeding the dynamic range most analytical techniques available to measure proteins. Therefore, the design and implementation of proteomic strategies needs to be thorough (Surinova *et al.*, 2011).

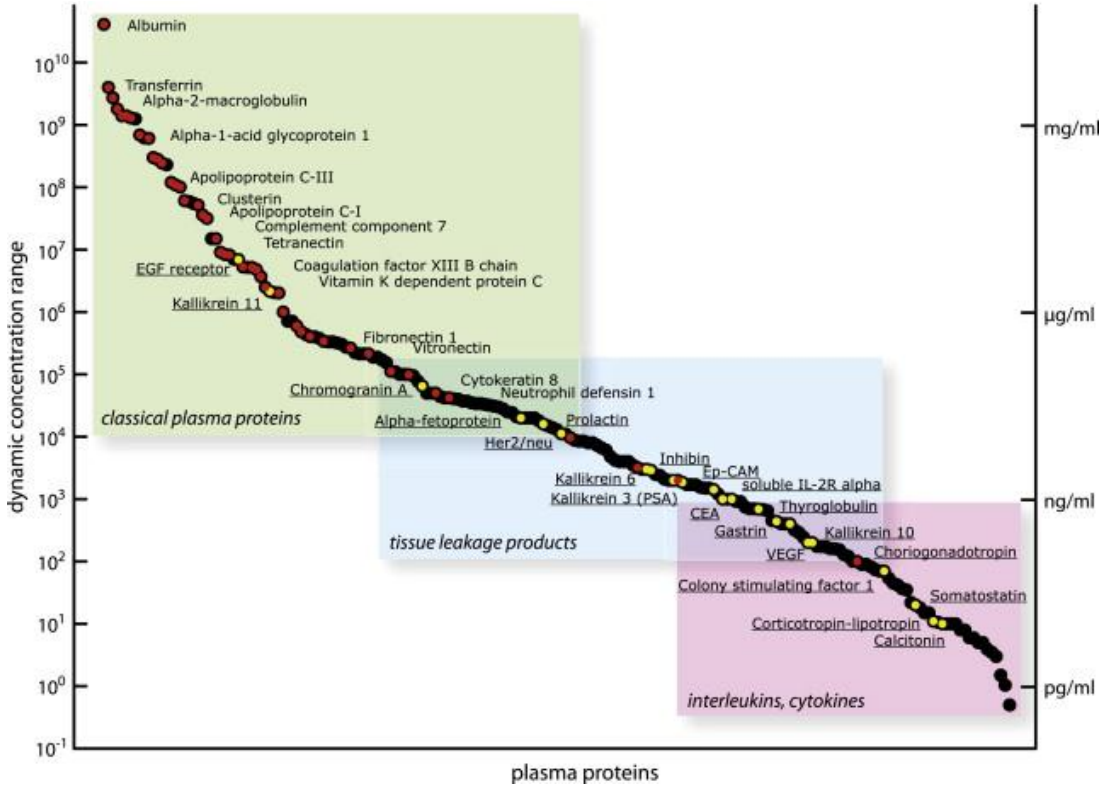


Figure 1.9: Dynamic range of the plasma proteome separated into three main protein groups. The red dots indicate proteins identified by the HUPO plasma proteome project and the yellow dots indicate currently used laboratory biomarker. Taken from (Surinova *et al.*, 2011)

1.3.3 Proteomic strategies in biomarker discovery

When conducting discovery proteomic studies, it is essential to understand the hallmarks of an ideal diagnostic biomarker. A biomarker needs to be specific to the disease and a low false positive rate is essential. The biomarker must be robust and must not be sensitive to rapid or drastic change. Finally, the biomarker must be reliable, reproducible and of clinical utility. Clinicians must be able to use the biomarker(s) to make accurate and informed medical decisions (Boja *et al.*, 2011).

Mass spectrometry (MS) based proteomics has the ultimate advantage over other techniques for biomarker discovery as it can detect, identify and quantify proteins over a wide range of concentrations and can measure the slightest biochemical changes. A discovery proteomic project, which aims to characterise the plasma proteome of a disease state needs to contain two distinct stages. The initial discovery phase where differentially expressed candidate proteins between disease and control are identified and the second verification stage that assesses the validity of the candidates in a larger number of samples.

Briefly, all proteomic studies include:

- Sample preparation to de-convolute the plasma proteome.
- Sample analysis using technologies such as gel electrophoresis or MS.
- Data analysis to produce a list of candidate proteins with protein concentrations (quantification).
- Verification to confirm the concentration of proteins, as well as the change characterised by the discovery phase.

1.4 Pre-analytical strategies

1.4.1 Proteomic biomarker discovery

Extensive pre-analytical treatment is required to overcome the dynamic range of plasma and make samples amenable to MS analysis. This can be achieved through a variety of methods such as: 1D/2D gel electrophoresis, isoelectric focusing, liquid chromatography, immunoprecipitation, immunoenrichment and immunodepletion. The choice of pre-analytical method needs to ensure that in-depth proteomic mass spectrometry analysis can take place as well as be robust, specific, and reproducible.

1.4.2 Reverse-Phase Solid Phase Extraction

Reverse-phase solid phase extraction (RP-SPE) is a method that allows purification of peptides prior to MS and separates analytes according to their relative polarities. Typically, RP-SPE consists of a polar mobile phase (liquid matrix), a non-polar solid stationary phase (C18) and a mid-nonpolar analyte (plasma). Before preparing a sample, the column solid phase needs priming with a solvent such as methanol. Priming activates the surface and enhances the bonding abilities of the column. Methanol is replaced with another solvent with a similar composition to the sample that is to be analysed, conditioning the column in preparation for sample. The sample passes through the SPE column, the plasma proteins adsorb onto the stationary phase, and the column is washed with a solvent. The solvent used does not remove the bounded proteins on the stationary phase but removes any non-proteinaceous material. The bound proteins are eluted from the stationary phase with a nonpolar organic solvent (such as acetonitrile) and collected for MS analysis (Berrueta *et al.*, 1995).

1.4.3 Organic solvent precipitation

Enrichment of low molecular weight species without depletion is possible by protein precipitation using acetonitrile (ACN). Protein solubility is poorest at its pI where it carries no net charge. Insolubility is a result of its polar interactions with aqueous solvents, ionic interactions with salts and repulsive electrostatic forces between similar charge molecules. Above the pI, proteins carry a net negative charge and below the pI, a net positive charge. There are many different types of protein precipitation methods including organic solvent, salt, acid and metal ion precipitation. An organic solvent causes the water molecules within the hydrophobic regions of proteins to displace, which reduces the hydrophobic interactions between proteins and increases the electrostatic interactions causing aggregation of proteins (Polson *et al.*, 2003). Thus, the addition of ACN to plasma or serum causes high molecular weight abundant proteins to form a pellet, precipitating out of solution, leaving low molecular weight proteins in solution, ready for MS analysis and suitable for biomarker discovery. Kay *et al.* found ACN to be the best organic precipitant (Kay *et al.*, 2008).

1.4.4 Immunodepletion

Immunodepletion is the removal of proteins through affinity interactions using pre-selected antibodies. The purpose of immunodepletion is to remove the high abundance proteins that dominate the plasma proteome and achieve greater proteome coverage (Figure 1.10). The cibacron blue dye method which solely depletes albumin and was one of the earliest depletion methods. However, the method lacks specificity and sensitivity and causes mass depletion of multiple proteins. The development of monoclonal antibody methods on resins and chromatography columns has improved the use of targeted removal of proteins (Hu *et al.*, 2006; Fang & Zhang, 2008). Many companies have developed the monoclonal antibody technology to produce commercially available kits or columns for multiple affinity protein removal, allowing the depletion of several high abundance proteins simultaneously. Some examples of the multiple protein removal kits available are the MARS 14 IgG protein depletion HPLC column, Seppro IgY 14 protein HPLC depletion column and the Proteoprep20 protein depletion resin.

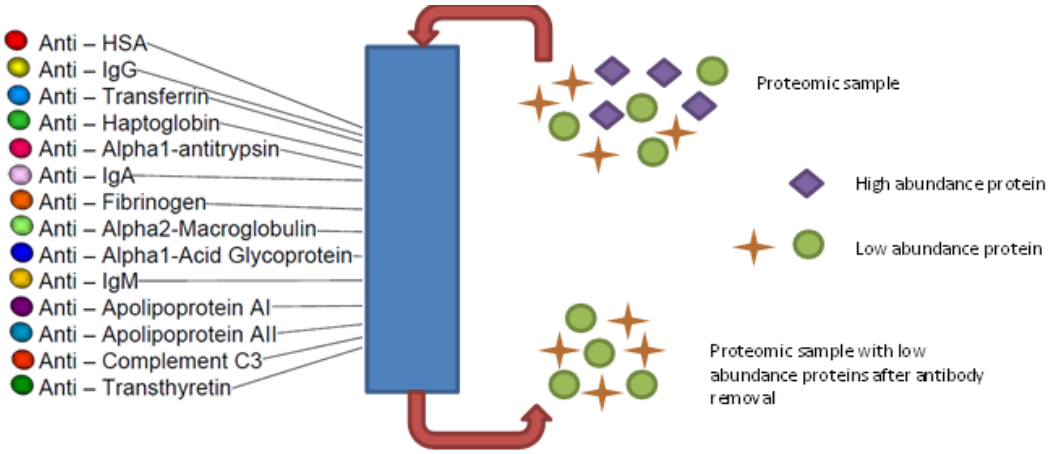


Figure 1.10: Principles of immunodepletion. A sample enters an antibody coated column with pre-selected antibodies raised to high abundance proteins. Through immunoaffinity interactions, high abundance proteins are removed and low abundance proteins remain in solution and flow through the column.

1.4.5 ProteoMiner™ Beads

An alternative to depletion of high abundance proteins is enrichment of the low abundance proteins. The conformational arrangement of amino acids within a protein establishes the protein's properties such as charge, density, hydrophobicity and conformation. The conformation also determines the binding and interaction of proteins with other molecules such as peptide ligands. This forms the basis of ProteoMiner™ or equalizer bead technology, which is the treatment of complex protein samples with a large, highly diverse, library of hexapeptides bound to chromatographic support that perform protein separation by affinity chromatography. The solid-phase combinatorial libraries of hexapeptides are synthesised *via* a short spacer on poly(hydroxymethacrylate) beads, using a modified Merrifield approach using the split, couple, recombine method (Merrifield, 1965). The peptide ligands are represented throughout the beads structure and each bead has millions of copies of a single, unique ligand and each bead could also possibly have different ligands (Thulasiraman *et al.*, 2005). The beads are synthesised from 20 natural amino acids, the library contains a population of linear hexapeptides amounting to $20e^6$, i.e. 64 million different ligands.

The binding capacity of the beads is limited to the number of ligands available, once the binding capacity of a ligand for a specific protein is reached, that specific protein ligand is saturated. This means that high abundance proteins saturate their ligands quickly, leaving excess protein to be washed away. Conversely, low abundance proteins are concentrated on their specific ligands, thereby compressing or decreasing the dynamic range of proteins by affinity interactions. The net result is the enrichment of low abundance proteins and reduction in the concentration of high abundance proteins (Righetti *et al.*, 2005; Boschetti & Righetti, 2009; Righetti *et al.*, 2010). When the bound proteins are eventually eluted, due to equal numbers of ligands on the beads for all proteins, the dynamic range should be reduced and the protein concentrations 'equalized' (Figure 1.11).

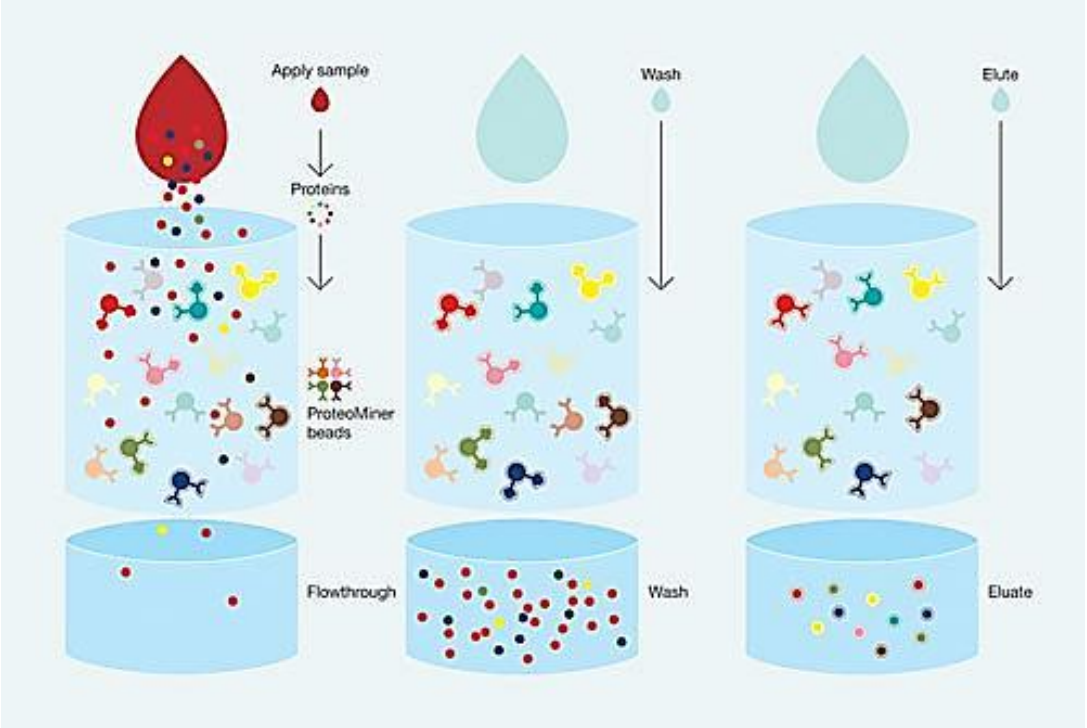


Figure 1.11: Mechanism of equalizer bead interaction with sample. Taken from (Biorad)

1.4.6 Liquid chromatography

Mass spectrometry based proteomics is highly dependent upon separation technologies to reduce the complexity of analytes (peptides) prior to MS analysis and reduce unambiguous identifications. The two main separation technologies employed in protein analysis are either gel-based or gel-free. Liquid chromatography (LC) is a gel-free separation. The development and the direct coupling of LC systems to MS instruments through an electrospray source has overcome some of the limitations faced with gel-based techniques. Due to the development of the electrospray, High Performance Liquid Chromatography (HPLC) or Ultra High Performance Liquid Chromatography (UPLC) has become a standard and essential front-end separation technique for proteomics (Yates *et al.*, 2009). Despite the success of electrospray and LC, a few limitations to the method remain with respect to poor analyte ionisation and subsequent detection. For example, hydrophilic compounds cannot be analysed with good sensitivity using high flow rate techniques. The high flow rate also introduces a large number of analytes to the mass spectrometer at once causing analyte signal suppression (Schmidt *et al.*, 2003).

In 1994, further development of the electrospray source by Wilm and Mann (Wilm & Mann, 1996) lead to the development of nano-electrospray (nano-ESI). The introduction of the nano spray lead to the reduction in the flow rates from µl/min to nl/min, requiring only µl of sample for analysis without loss of analyte ions and most importantly high-throughput quantitative proteomic analysis was now possible. A nano-UPLC system was used in this study.

Reverse-phase (RP) chromatography is a popular choice for the separation of peptides in bottom up proteomics and uses a hydrophobic stationary phase and polar hydrophilic mobile phase. The peptides are the analytes of interest and are injected onto a RP column in low organic solvent, which contains a hydrophobic stationary phase such as C_{18} that bind the peptides. The low organic solvent retains the salts and artefacts from digestion, effectively desalting and concentrating the sample. Separation of the peptides occurs when the organic solvent in the mobile phase is 'ramped' or increased in a gradient fashion. The polar hydrophilic mobile phase (e.g. organic solvent) disrupts the interaction between the stationary phase and the peptides thus, peptides elute according to their hydrophobicity i.e. the hydrophilic peptides elute first (Di Palma *et al.*, 2012).

In proteomic studies, use of buffers and different enzymatic methods may introduce artefacts that can interfere with the nanoUPLC-ESI source causing signal suppression. Although the RP column is equipped to clean the sample, sometimes this step alone is insufficient. Thus, a de-salting step prior to LC analysis can be used and in this study, a trapping column prior to the analytical RP-LC column was used to de-salt peptide samples. The trapping column is similar to the LC column however is of a larger diameter and particle size.

The resolving power or the number of peaks a chromatographic system is able to separate within a given gradient time period is measured as the peak capacity. The peak capacity of a single chromatography system for complex peptides still lacks the resolving power with many peptides co-eluting. Multi-dimensional chromatography using orthogonal separation maximises the peak capacity by combining two LC systems, with two separate peak capacities, prior to MS analysis. Theoretically, the overall separation power is the product of each dimension or LC platform, allowing co-

eluting analytes to be resolved in the orthogonal second dimension (Di Palma *et al.*, 2012). Different multidimensional systems have been reviewed elsewhere (Zhang *et al.*, 2010) in this study two-dimensional (2-D) RP-RP was the multidimensional method of choice (See Chapter Four).

1.5 Analytical strategies

1.5.1 Mass spectrometry

A number of advancements have taken place since Nobel laureate J. J. Thomson's experiment in 1897, where he discovered the electron and determined its mass-to-charge ratio and was later awarded the Nobel Prize in 1906 and knighted in 1908 (Thomson, 1899). His work on the electrons lead to the construction of the first parabola spectrograph (mass spectrometer) where he observed positively and negatively charged ions. The very first mass spectra of oxygen, nitrogen, carbon monoxide and dioxide were obtained alongside isotopes of neon (Thomson, 1899). F. W. Aston continued to develop the mass spectrometer, developing velocity focusing with the addition of electromagnetic focusing, increasing the sensitivity and resolving power. Leading to further naturally occurring isotopes to be identified (Aston, 1919). Aston was awarded the Nobel Prize for his work in 1922 and went on to measure mass defects in 1923. It is from this point that mass spectrometry rapidly gained speed, ever increasing sensitivity and resolving power through the developments of different mass analysers to ion sources. Since 1897, the mass spectrometer has become a universal tool amongst scientists for various applications in physics, chemistry and biology. Since 1984 when J. B. Fenn (Fenn *et al.*, 1989) used electrospray ionisation with mass spectrometry for biomolecules, MS become a vital tool in the characterisation of the human proteome. Further development of the electrospray came in 1994 when Wilm and Mann (Wilm & Mann, 1996) introduced the nano-electrospray allowing hydrophilic compounds (e.g. carbohydrate) to be characterised with greater sensitivity and improved ESI tolerance of salt contaminants.

Mass spectrometry is an analytical technique that measures the mass to charge ratio (*m/z*) of a charged ion in the gas phase under a vacuum. Mass spectrometers are essentially made up of three main sections; an ion source where ionisation occurs, a mass analyser that measures *m/z* ratio and a detector which detects ions at each *m/z*. The

ions are separated according to their *m/z* ratio and are detected by eliciting a change in either charge or current. The more ions the greater the change detected, and the response is recorded as a mass spectrum. The mass spectrum contains all information on ionised components, in this case peptides, in the samples.

1.5.2 Ionisation

Proteins and peptides are polar, non-volatile and thermally unstable structures that readily degrade during ionisation but since the 1980's and the development of soft ionisation techniques, proteins and peptides have been ionised without extensive degradation (Yates *et al.,* 2009). Barber *et al.* described the advent of fast atom bombardment as the first soft ionisation technique developed for biomolecules without the need for chemical derivatisation (Barber *et al.,* 1981). Fast atom bombardment allowed the extension of the mass range and resolution of mass spectrometers. In fast atom bombardment, the sample is mixed with a matrix and the matrix is bombarded with a beam of inert gas ions such as Argon. This action produces either a protonated $[M+H]^+$ or deprotonated $[M-H]^-$ molecule for mass analysis. Although fast atom bombardment was used initially, this technique was quickly replaced with other softer ionisation techniques.

The 1980's saw the development of the two most widely used ionisation techniques in proteomics, matrix assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI). In 2002, J. Fenn and K. Tanaka were awarded the Nobel Prize for their work in soft ionisation methods in mass spectrometry for biomolecules. Michael Karas and Franz Hillenkamp (Karas & Hillenkamp, 1988) improved Tanaka's soft ionisation method to include a matrix, this was known as matrix assisted laser desorption Ionization (MALDI).

In 1988, J. Fenn realised the potential of the ESI technology when the first spectra of protein above 20,000 Da was obtained (Fenn *et al.*, 1989; Yamashita & Fenn, 1984). In ESI, a high voltage is applied creating a stream of ions. The analyte is dissolved in a polar solvent and flows through an emitter set at 2-4 kV. The charged analyte solution forms a Taylor cone and the liquid disperses into droplets. This forms an aerosol when it exits the emitter under atmospheric pressure, producing a stream of ions induced by

the combined effects of electrostatic attraction and a vacuum. These ions are analysed by the mass spectrometer (Figure 1.12) (Aebersold & Mann, 2003).

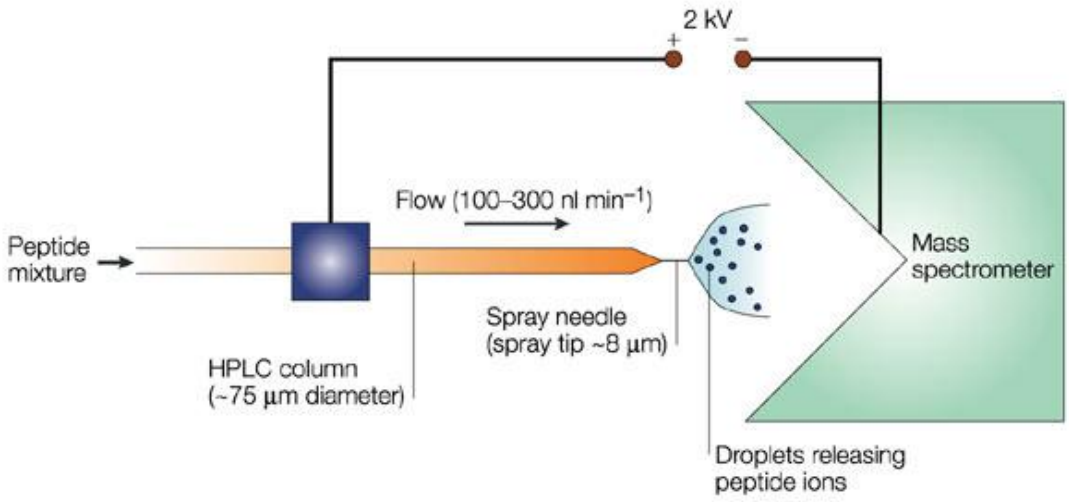


Figure 1.12: Electrospray ionisation. Taken from (Steen & Mann, 2004)

Furthermore, ESI can form highly charged ions without causing fragmentation, where more than one charge state of an ion can be observed. Peptides produced from tryptic digests have two protonation sites, these are the N-terminal and the C-terminal residue (lysine or arginine). These sites can be protonated with a charge remaining on the C-terminus, N-terminus, both C and N-terminus or remain neutral with no charge (Figure 1.13). Tryptically digested proteins produce primarily doubly charged ions in ESI that are readily fragmented into *b* and *y* ion series, making calculation of peptides molecular masses possible. For example, an *m/z* separation of +2 ion are separated by *m/z* value of 0.5, enabling true molecular mass calculations to be determined from the original mass spectrum (Yates, 1998).

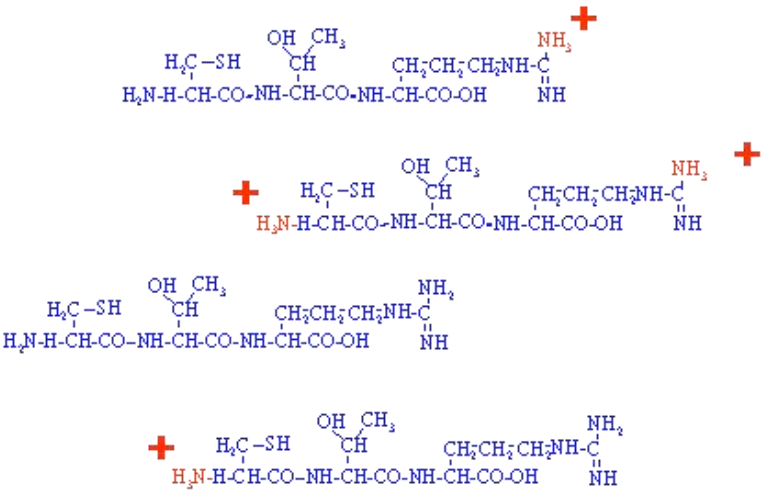


Figure 1.13: Protonation sites for ionisation in ESI. (Taken from ionsource.com)

As mentioned previously mentioned, development of nano-ESI allows flow rates of solutions to be as low as nL/min for ionisation with little sample volume requirement. The lower flow rate in nano-ESI produces smaller droplets, this undergoes the same desolvation process, and less solvent is present in the ion source chamber. Consequently, a greater overall analyte (peptide) volume is present in the droplet thus; a greater number of peptide ions are produced for MS analysis. This was best demonstrated by Wilm and Mann (Wilm & Mann, 1996) where they reported a 510 times higher efficiency of ions transferring from solution to detector with the nano-ESI source in comparison to an ESI. Due to the low volume of solvent present in the droplets produced, the nano-ESI source lends itself to higher solvent salt and contamination tolerance. Advantages offered by the nano flow greatly improve sensitivity of analysis, and its compatibility with RP chromatography makes nano-ESI a more desirable ionisation technique over its normal flow counterpart for bottom-up proteomics.

1.5.3 Mass analyser

Pivotal to MS analysis is the mass analyser, which separates ions according to their *m/z*. The resolving power of a mass analyser i.e. the ability to separate ions with similar *m/z*, varies. For proteomic research the analyser needs to provide high resolution, high mass accuracy, high sensitivity and provide data rich mass spectra of peptide fragmentation.

The most common analysers used for proteomics are ion traps (IT), Fourier transform ion cyclotron (FT-ICR), Orbitrap, time-of-flight (TOF), quadrupole (Q).

1.5.4 Ion trap, Fourier Transform Ion Cyclotron Resonance and Orbitrap

The trapping mass spectrometers Ion Traps, Fourier Transform Ion Cyclotron Resonance and Orbitrap all separate ions by *m/z* according to the ions resonance frequency. Ion trap instruments traps or stores ions in a two or three-dimensional field with an oscillating quadrupolar electric field, by varying the electric field, ions of different *m/z* values are ejected from the IT and detected. In FT-ICR ions are excited with an oscillating electric field, ions with different *m/z* ratios display different trajectories, they cluster or fall in phase and the ions induce a charge detected as a time and frequency function. This information is then fourier transformed into a mass spectrum. The resolution of FT-ICR analysers is inversely proportional to *m/z* and is dependent on observation time and the strength of the vacuum field, which is a notable limitation to this technique (Comisarow & Marshall, 1974).

Proposed by Makarov in a patent in 1999 and an article in 2000 (Makarov, 2000) the Orbitrap is an electrostatic ion trap instrument that uses Fourier transformation to produce a mass spectra. The Orbitrap (Figure 1.14) separates ions by *m/z* based on oscillation frequencies along a central spindle within an ion trap chamber. Every ion oscillates around a central spindle in spirals and induces an oscillating image current which is Fourier transformed to produce a mass spectrum. The resolution of the Orbitrap instrument is inversely proportional to $(m/z)^{½}$ thus, as the mass increases the resolution decreases. For example, when acquiring at a scan rate of 1 second the resolution of the Orbitrap is 60,000 at *m/z* 400. Other mass analysers such as time of flight (TOF) maintain their resolution across a mass range.

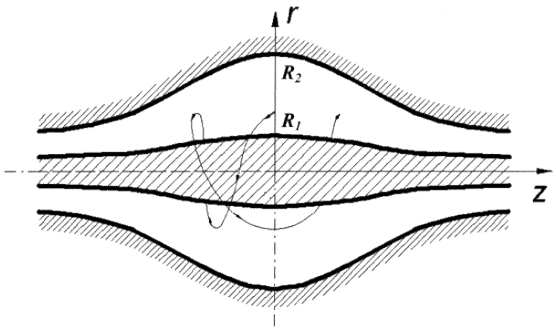


Figure 1.14: The Orbitrap showing a stable ion trajectory. The outer electrode is separated by a small gap, which contains the inner electrode which is spindle like to which a DC voltage is applied. The voltage causes the ions to move around the central electrode while oscillating along the z-axis, giving them a spiral like trajectory. Taken from (Makarov, 2000)

1.5.5 Quadrupole

The quadrupole consists of four parallel rods with opposing electric fields. The quadrupole functions as a mass filter, principally ions of different *m/z* value have different stable oscillations in a quadrupole field. The electric field is created by applying an oscillating electrostatic field to the two pairs of parallel rods. This means at pre-set field frequencies and amplitudes, ions of a single *m/z* value or range can travel through the analyser in spiralling stable trajectories and reach the detector. All ions with unstable trajectories, spiral into the quadrupole rods.

Quadrupoles can scan and detect for specific *m/z* by altering the amplitudes of the applied electric fields, producing a full and detailed mass spectrum, lending themselves as ion guides (see section 1.5.8). In RF-mode, a wide mass range of ions can pass through the quadrupoles with stable trajectories, in a static field a narrow mass range of ions pass through. Quadrupoles are extremely sensitive and thus, lends itself to quantitative mass spectrometry (Paul & Steinwedel, 1953).

1.5.6 Time-of-Flight

Stephens first described the TOF mass analyser in 1946 but it was Wiley and McLaren's linear TOF described in 1955 that commercialised the instrument (Wolff & Stephens, 1953; Wiley & McLaren, 1955). The TOF mass analyser measures the *m/z* value of ions by measuring the time it takes for an ion to travel over a fixed distance in the MS. Time is a function of the square root of the *m/z* value. Therefore, peptides with low *m/z* values travel faster than peptides with high *m/z* values. The flight time of an

unknown ion is converted to *m/z* value by using a calibration standard with a known *m/z* value (Yates, 1998).

The upper mass range for a TOF analyser in theory is non-existent as it relies on flight time and an ions kinetic energy, making it suitable for soft ionisation techniques and the study of biomolecules. TOF analysers obtain a spectrum over a broad mass range in a matter of micro-seconds, this high transmission frequency of ions leads to high sensitivity. Although this provides speedy analysis, it comes at the expense of detector saturation and kinetic energy spread leading to insufficient precision of mass or abundance measurements.

The loss of spectral information and peak broadening due to ions of the same *m/z* having varying velocities and therefore kinetic energies, was addressed by the advent of delayed pulse extraction (primarily applies to MALDI). This applies a time lag between the formation and extraction of ions. The ions first expand into a field free region in the source and after a time delay; a voltage is applied to extract these ions from the source. The ions in the field free region gather and separate according to their kinetic energies. The extraction pulse transmits more energy to the ions that gather in the source for a longer period. Thus, the ions that initially had a less kinetic energy have been supplied enough energy to correct for the energy dispersion of the ions with the same *m/z* ratio correcting for the kinetic energy spread and improving the resolution of the TOF analyser.

Additionally the incorporation of the electrostatic reflectron has greatly improved the resolution and peak broadening of TOF analysers. The reflectron is situated in the flight tube behind the field free region as opposed to the ion source. The reflectron creates a retarding field that acts as a mirror by reducing the ion velocity to zero, deflecting the ions, sending them through the flight tube to the field free region and the ions are detected. This corrects for the kinetic energy dispersion, as ions with greater kinetic energy and greater velocity penetrate the reflectron deeper than those with lower kinetic energy and velocity. Therefore, the ions with greater velocity spend more time in the relfectron and reach the detector as the same time as the slower ions with the same *m/z* (Figure 1.15). This has been further developed to incorporate two reflectron regions

(first region is high electric field and second low), increasing the range of kinetic energies the TOF analyser is able to analyse (Guilhaus *et al.*, 2000).

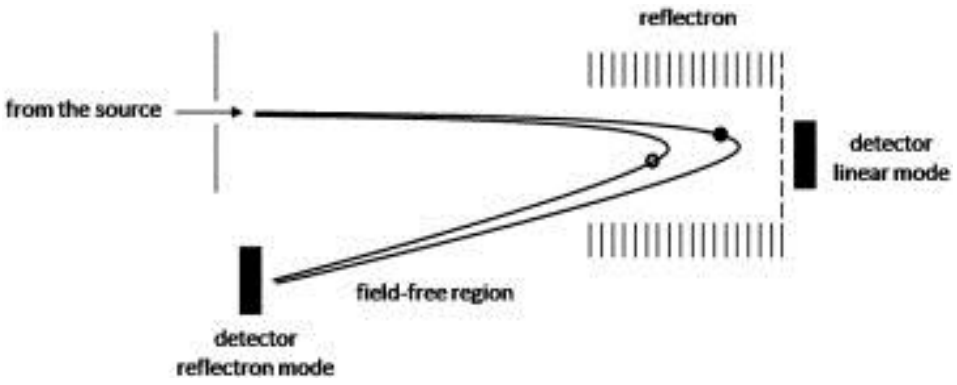


Figure 1.15: Schematic representation of a TOF analyser operating in the reflectron mode. ○ and ● represents bundles of ions having same *m/z* value but different velocities. ● has higher velocity than ○. With appropriate settings, both kinds of ions reach the detector simultaneously. Taken from (Jiwan *et al.*, 2011)

For years TOF analysers were best suited to pulsed ionisation techniques (MALDI) due to TOF systems working best with pulsed ion beams rather than continuous ion beams (ESI). Orthogonal acceleration time of flight mass spectrometry (oa-TOF) is the best technique for coupling ESI with TOF-MS. In oa-TOF the ion are continuously produced at the source, ion guides focus the ions into a beam and directs the beam to the orthogonal accelerator. The orthogonal accelerator space (pusher) fills with ions, and then an injection pulse voltage is applied which pushes the ions in an orthogonal direction to their original trajectories. The ions then enter the flight tube and TOF mass separation takes place. Whilst these ions are analysed by the TOF analyser, more ions are able to enter the pusher chamber, ready for orthogonal acceleration pulse once the highest *m/z* of the previous ions has been recorded (Figure 1.16) (Guilhaus *et al.*, 2000). As the time taken for the ions flight path in the flight tube is approximately equal to the time taken to re-fill the pusher chamber, a high duty cycle of analysis is possible. The duty cycle is further improved with the incorporation of an RF-only ion guide prior to the pusher region, where the ions released in packets by the ion guide are synchronised with a delayed pusher pulse. As the ions travel in *m/z* packets it is possible to

synchronise the oa-TOF pusher with the ion packet to enhance the duty cycle over a defined *m/z* range (Giles *et al.*, 2004). The ability to interface with ESI, which is well suited to peptides and the high duty cycle, makes oa-TOF suitable for proteomic studies and is used in this study.

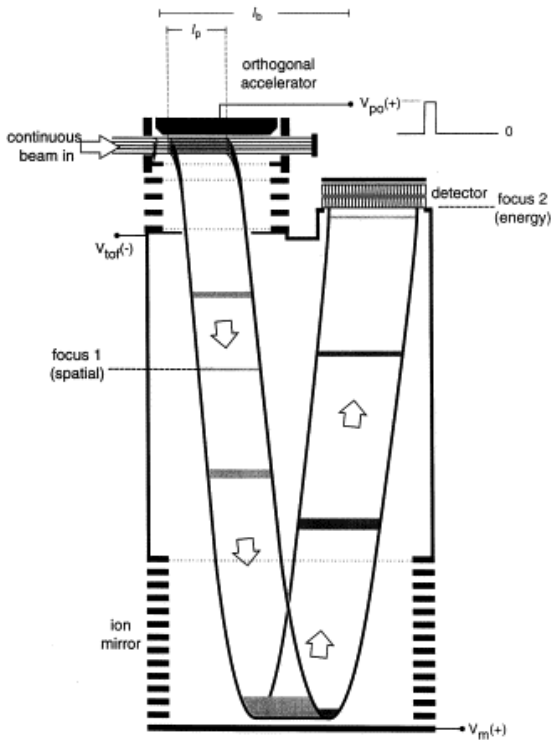


Figure 1.16: Schematic for an oa-TOF taken from (Guilhaus, 2000)

1.5.7 Ion mobility

Ion mobility mass spectrometry (IM-MS) is a gas phase electrophoretic technique that allows analytes to be identified according to their mass, charge and collision cross section area (i.e. size and shape) (Creaser *et al.*, 2004). Ion mobility measures ion velocity through a drift tube filled with a buffer gas, such as helium, in the presence of an electric field gradient. The ions mass, charge, size and shape determine the time (drift time) it takes an ion to migrate through the drift tube and forms the basis of separation. As the ion travels through the drift tube it collides with the buffer gas, the larger the ion (or the greater the collision cross-section) the greater the surface area for the buffer gas collision. This collision action impedes the ion's ability to pass or drift

through the tube. This method allows separation of isomers, isobars and conformers; reduction of chemical noise; and measurement of ion size (i.e. cross sectional area), which is impossible with mass spectrometers alone. Ion mobility has been interfaced with many different mass spectrometers including the TOF-MS adding an additional dimension of separation.

A recent development in ion mobility instrumentation is the travelling wave ion guide ion mobility (TWIG-IM) and is found in Waters Synapt HDMS analyser, and is used in this project. In these instruments, the ion mobility takes place in the centre of a stacked ring ion guide (see section 1.5.8) region known as the TRIWAVE. The first ion guide traps the ions prior to IM separation and the third transfers the ions into an oa-TOF. In the TWIG-IM, the ions travel through a buffer gas in a travelling wave fashion due to sequential superimposed DC and RF voltages that are applied to adjacent electrodes, providing radial confinement of the ions and propelling the ions through the buffer gas. The ions travelling wave time-path through the buffer gas with the applied voltage depends on the ions mobility (Harvey *et al.*, 2011). To improve the resolution, a helium cell is placed before the TWIG-IM and consists of four RF-only ring electrodes, the helium slows down the ions travelling to the TWIG-IM and slowly introduces ions to the travelling wave (Giles *et al.*, 2011). The sensitivity and high throughput analysis makes TWIG-IM ideal for proteomics.

1.5.8 Ion Guide

Transportation of ions without separation and loss of sensitivity is often required in mass spectrometry, especially in tandem mass spectrometry. Ion guides help to re-focus ions and transport them efficiently through the MS analyser, as ions that enter the ion guide through the source can be spread due to mutual ion repulsion or collision with solvent molecules. Quadrupoles operating in the RF-only mode are used to stream line the ions to a focussed trajectory and this has been extended to hexapoles and octapoles. (Giles *et al.*, 2004).

Another focusing ion guide uses a series of stacked ring electrodes with a fixed inner diameter and RF of opposite polarities is applied to consecutive electrodes (Figure 1.17) creating an effective potential that radially confines the ions within the ion guide. This

is known as a stacked ring ion guide and is used in this study. An improvement of the ion guide's performance to reduce the loss of ions saw the introduction of a neutral buffer gas. However, this led to collisions of the ions with the buffer gas and a reduction in the kinetic energy of the ions, increasing transit time through the ion guide. This was a problem for tandem mass spectrometry. Giles *et al.* (Giles *et al.*, 2004) solved this with the use of a travelling voltage wave that is superimposed on the radially confining RF-voltage, causing ion's to 'surf' through the ion guide and reducing the transit time of the ions. This new method has allowed the new travelling wave ion guide to act as a collision cell for tandem mass spectrometry, ion mobility cell and a delivery ion guide for oa-TOF mass spectrometer.

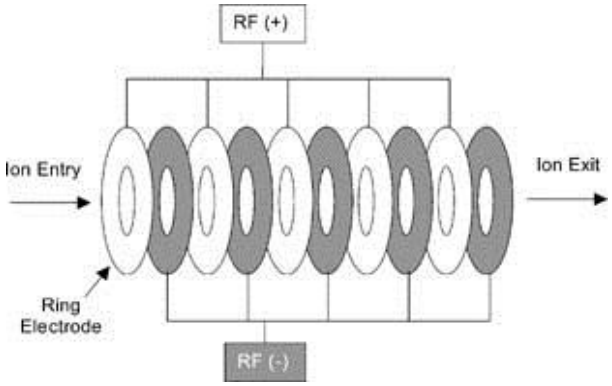


Figure 1.17: Schematic of RF-only stacked ring ion guide taken from (Giles *et al.*, 2004).

1.5.9 Detector

The detector detects the ions at different *m/z* values and channels this information into electrical currents. This electrical current is visualised in the form of a peak on the mass spectrum. The intensity of the peak represents the number of times the ion is detected. The mass spectrum is displayed with the ion intensity on the y-axis and the *m/z* ratio on the x-axis.

1.6 Tandem mass spectrometry

Tandem mass spectrometry is where precursor ions are subjected to sequential separation in different analysers often with the production of product ions because of fragmentation/dissociation. Product ions can be used to elucidate structure or, in the case of peptides, the amino acid sequence. In this study a Q-TOF was used to perform all tandem MS.

In tandem mass spectrometry, there are two different ways of performing experiments, in space and in time. In space experiments use two distinct mass analysers with a collision cell (for product ions to form) separating them, such as the triple quadrupole (QqQ), Q-TOF and TOF/TOF. In time experiments use ion storage devices such as the ion traps and FT-ICR and perform the tandem experiment using a single instrument but in successive steps. In time analysers can repeatedly undergo rounds of fragmentation (MS^n) but this comes at the expense of sensitivity.

There are a number of ways to measure or scan for ions of interest in tandem MS, each with their own strengths. As this project focuses on product ions produced with inert collision gas, there are three main scanning modes:

1. Product ion scan – where the first analyser is static selecting precursor ions of a chosen *m/z* ratio and the second analyser scans all the product ions produced by collision-induced dissociation.
2. Precursor ion scan – where the second mass analyser remains static to a specific *m/z* of a product ion and the first mass analyser scans for all precursor ions that generates the product ion.
3. Selected reaction monitoring (SRM) – where the first and second mass analysers are static to pre-defined *m/z*. The first mass analyser selects the precursor ion *m/z* and the second analyser is static for a corresponding product ion *m/z* characteristic of the analyte in question.

Tandem mass spectrometry requires fragmentation to occur for precursor ions to produce product ions. The different types of fragmentation techniques include collision induced dissociation (CID), electron transfer dissociation (ETD), and electron capture dissociation (ECD) with CID being the most commonly used in Q-TOF instruments. In CID, a precursor ion undergoes multiple collisions with inert gas molecules (e.g. argon) causing the kinetic energy to translate to an internal vibration energy of the ion, breaking chemical bonds at the lowest energy cleavage sites fragmenting the ion. With peptide ions this typically occurs at the peptide bond generating b series (N terminal peptides) and y series (C terminal peptides) product ions (Hoffmann, E.D., Stroobant, V., 2006). In this study, a RF only ion guide was used as the collision cell prior to oa-

TOF analysis. Once the product ions were generated post CID, the product ions were separated according to the *m/z* in the oa-TOF and detected to produce a mass spectrum.

1.6.1 Fragmentation nomenclature

In 1984 Roepstorff and Fohlman (Roepstorff & Fohlman, 1984) introduced the nomenclature for fragmentation observed in peptide ions using tandem MS (

Figure 1.18). For the peptide ion to be detected a charge (which starts typically as a doubly charged ion) must remain on the ion. There are three sites at which peptide fragmentation can occur, these are Cα-C, C-N or N-Cα, producing six types of fragments *a* ,*b*, or *c* when the positive charge remain at the N terminal fragment ion or *x*, *y* or *z* when the positive charge remains at the C terminal fragment ion. The subscript number $_{(n)}$ denotes the number of amino acids in the fragment (Hoffmann, E.D., Stroobant, V., 2006).

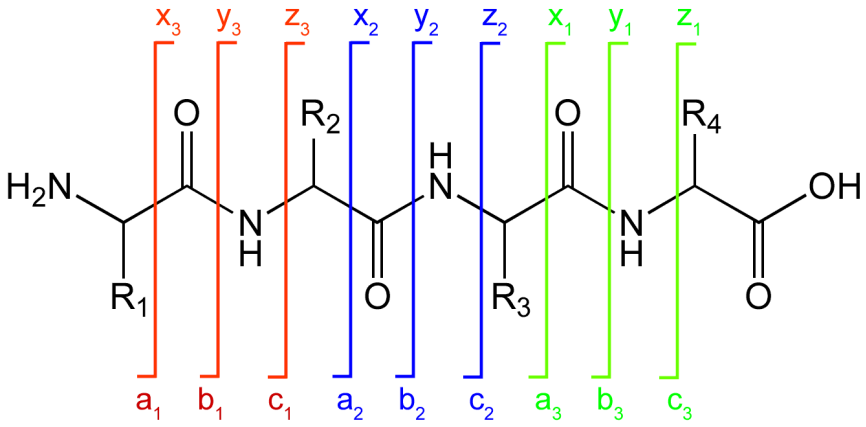


Figure 1.18: Schematic of peptide fragmentation nomenclature.

With this information interpretation of the spectra produced by peptides in tandem MS can be determined manually or *de novo.* However, many software algorithms have been developed to correctly and confidently sequence the peptides and interpret the tandem MS spectra, allowing peptides to be assigned to their proteins.

1.6.2 Proteomic approaches using tandem mass spectrometry

Mass spectrometry based proteomics has three objectives, the first of which is to characterise a proteome, these are typically profiling projects where the objective is to profile an entire proteome of a given sample. The second is comparative projects also known as discovery proteomics, where the objective is to identify and relatively quantify proteins in numerous samples for a given proteome. The relative quantitation identifies proteins that change their expression in different conditions and are reported as fold-changes relative to each other. The third is absolute quantitation or targeted proteomics where the exact amount of protein in a proteome is reported in molarity, mass or copies/cell, with the use of an internal standard. A select panel of proteins are generated from discovery proteomic projects to form the basis of targeted studies. Proteins can be studied intact this is known as top-down proteomics, or studied as digested proteins as a peptide mixture this is known as bottom-up proteomics (Figure 1.19).

1.6.2.1 Bottom up proteomics

Bottom up proteomics is the analysis of digested complete proteomes i.e. peptide mixture and is the method used in this study. The proteins are digested by an enzyme, typically trypsin which hydrolyses protein bonds at C terminal of proteins, specifically lysine and arginine, to form a peptide mixture. The peptide mixture that is produced by trypsin (1-3k Da) span a wide range of hydrophobicities making the peptides ideal from fractionation by liquid chromatography before MS analysis (Zhang & Ge, 2011). This method produces a *y*-ion series of peptide product ions that can be used for protein identification. This approach is particularly attractive for discovering biomarkers as it is more sensitive, capable of greater proteome coverage and is amenable to tandem mass spectrometry thus, accessing the low abundance proteins.

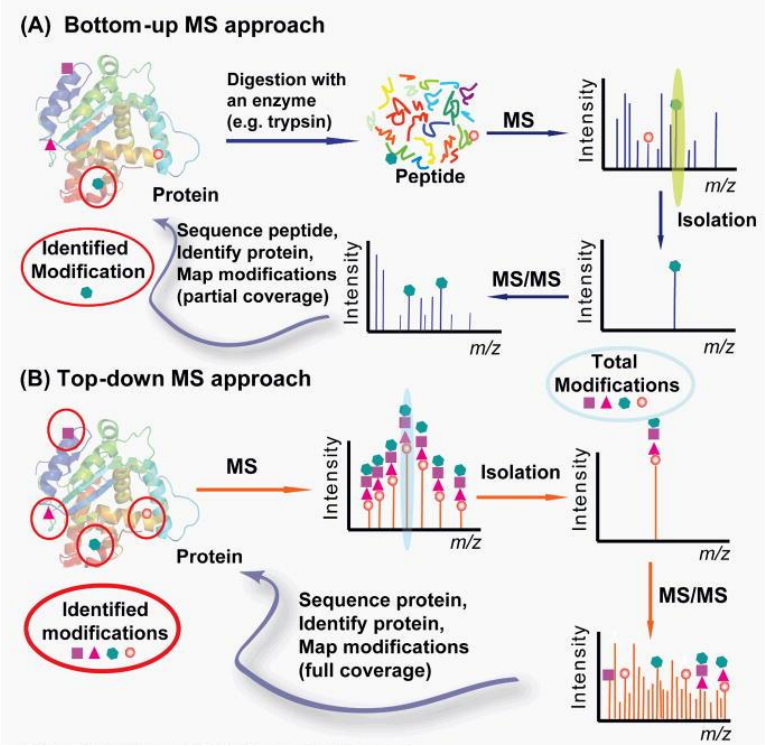


Figure 1.19: Proteomic approaches. (A) Bottom up: a protein is digested with an enzyme (i.e. trypsin) into many small peptides. The recovered peptides are analysed by tandem MS to identify the origin of the peptide i.e. which protein it is derived from. (B) Top-down MS: the whole protein is analysed by MS without digestion so the full information of the modification state is preserved. A specific protein form can be isolated and fragmented by tandem MS to locate the modification sites. All modifications can be identified with full sequence coverage. Taken from (Zhang & Ge, 2011)

1.6.2.2 Data dependant acquisitions

As mentioned previously bottom up proteomics utilises tandem MS, with the capability to apply different scanning modes for data acquisition (see section 1.6). One particular scanning mode is data dependant acquisition (DDA). In DDA the tandem MS performs a survey scan (MS scan) where the most abundant precursor ions are isolated, typically 3-8. The pre-selected precursor ions are fragmented and the associated product ions are scanned. This process is repeated for a defined time-period or number of scans. This scan method is hampered by the pre-selection as there is bias towards higher abundance proteins limiting the dynamic range of MS analysis, co-selection of ions, large number of single peptide identifications and compromised quantification accuracy due to serial MS and tandem MS scanning. Michalski and Mann *et al.* identified more than 100,000 isotope features which were likely to be peptides, eluted in a single LC-tandem MS run but were unidentified by a typical DDA experiment (Michalski *et al.*, 2011).

1.6.2.3 Data independent acquisitions

Data independent acquisition (DIA) attempts to overcome the limitations of DDA analysis by removing the pre-selection of precursor peptides. In place of pre-selected ions, a group of precursor ions within a mass window are co-fragmented as a parallel process. The product ions have the same retention time profiles as the precursors, which allows retention time alignment of all precursor and product ions. In 2003, a pioneering DIA experiment conducted by Purvine *et al.* utilised different nozzle-skimmer voltage differentials on an ion trap. The nozzle-skimmer voltage was set to both high and low to collect product ions and minimise in-source fragmentation respectively. The authors suggested the use of retention times to align precursor and product ions and the experiment was feasible using Q-TOF instruments, providing software could be developed to correctly and confidently analyse the data produced by such experiments (Purvine *et al.*, 2003). These DIA experiments lead to the development of various methods such as: all ion fragmentation using the Orbitrap, SWATH MS, and MS^E. This study utilised MS^E data acquisition, which performs low energy and high energy scans in parallel with a Q-TOF mass spectrometer (Shliaha *et al.*, 2013).

1.6.2.4 Data analysis in mass spectrometry based proteomics

Tandem MS is the preferred platform for bottom-up proteomics and protein profiling. In order to elucidate the identity of proteins, individual peptides undergo tandem MS to reveal partial or full sequence of the peptides. The informatics process first identifies the peptide and then the proteins that are present in the original sample (Vitek, 2009).

The process of data analysis can be described as a simple workflow: peptide sequencing, protein database searching, validation of protein assignment. Data produced by MS is typically transferred in its original format to a data processing machine where the raw data is subjected to data mining methods. This can be de novo sequencing where without any prior knowledge or reference database, all mass spectra are assigned and proteins identified.

This is a long and time-consuming process. The other method is to use a reference database such as UNIPROT (http://www.uniprot.org/) where a FASTA file is

downloaded with all known and theoretical protein sequences. With the latter method, the raw data file contains all the tandem mass spectra information on the ions that are detected by the mass spectrometer is subjected to a database search. The observed raw data is paired to theoretical product ion data that is associated with a known protein. This is an iterative process of running individually observed product ion data against theoretical protein sequences found in FASTA files. Positive identifications are only made when a match between the observed and theoretical data has the same precursor *m/z* within a defined tolerance (Deutsch *et al.*, 2008).

Subsequent to initial identifications, various steps are undertaken to ensure the correct assignments of peptides have been made through database searching. A score is assigned to each candidate peptide match. Then in a two-step process incompatible peptide and precursor ion masses are filtered out and a new score is assigned incorporating new information of features such as peak intensity, post translational modifications (PTM's) etc. To further filter the data and enhance the information mined from the raw data, a false discovery rate (FDR) is used to assign confidence in identifications and quantify error rates. Due to stochastic variation in the spectra or incomplete databases, a statistical tool such as FDR can add a measure of confidence. In this study, a database where the sequences are scrambled, known as a reverse database, is used to determine false positive identifications and the FDR. Once all these measures are taken into consideration protein inferences can be made, meaning all the peptides have now been identified and scored, so proteins can be confidently assigned. It is important to correctly assign peptide sequences to proteins as insufficient discrimination of peptide sequences to proteins can lead to false or incorrect assignments. To ensure minimum mis-assignments occur predefined cut off criteria can be set, an example of controllable criteria is the number of peptide sequences needed for positive protein identification (Vitek, 2009; Domon & Aebersold, 2006).

1.7 Quantitative proteomics

Quantitative analysis in MS proteomics is either relative or absolute with or without the use of peptide labelling. Absolute quantification calculates the exact concentration of every protein present in a sample (e.g. a measure of the protein's concentration in nanomoles, and uses internal standards). Relative quantitation determines the change in

protein concentration by comparing samples (i.e. calculates up or down regulation of a protein relative to different sample groups). The up or down regulation is denoted as a fold change. Relative quantitation between samples can be calculated with either chemical or metabolic labels or be label free where quantitation is based on spectral counting or peak intensity.

1.7.1 Relative quantitation

1.7.1.1 Labelled approaches

Isotopically labelled compounds are chemically and physically identical to their endogenous counterparts except in mass. There are a number of different labelled stable isotope methods for quantitative shotgun proteomics including; isotope coded affinity tag (ICAT), Stable isotope labelling by amino acids in cell culture (SILAC), ^{15}N / ^{14}N metabolic labelling, ^{18}O/ ^{17}O enzymatic labelling, isotope coded protein labelling (ICPL), Tandem mass tags (TMT) and Isobaric tags for relative and absolute quantification (iTRAQ). Even though these methods have been developed extensively they suffer from limitations such as increased sample processing time, compressed dynamic range and adds to sample complexity. Thus, an alternative label free quantitative method has become increasingly popular in shotgun proteomics (Zhu *et al.*, 2010; Elliott *et al.*, 2009).

1.7.1.2 Label free approaches

Relative label free proteomics has two approaches, spectral counting and peak intensity measurements. Relative quantitation using spectral counting is based on the concept that the number of tandem MS spectra identified for a given peptide is correlated to the quantity of that peptide. The spectra are counted and using an algorithm the concentration for all measured proteins and peptides is determined. Lui *et al.* showed that a linear correlation over two orders of magnitude could be observed with spectral counting (Liu *et al.*, 2004). Spectral counting is possible because, as the abundance of proteins and associated peptides increase, so do the number of tandem MS events; resulting in greater number of spectral counts for that given peptide and protein. When spectral counts are compared over multiple data sets, relative quantities of proteins can be determined. However, it was observed that measuring small changes between proteins was less accurate than measuring large changes. To ensure accurate and

reliable detection of proteins, normalisation and statistical analysis of spectral counting methods is necessary to monitor protein changes in biological samples due to bias towards abundant peptides and the masking of lower abundance peptides (Florens *et al.*, 2006; Nilsson *et al.*, 2010).

In relative quantitation using peak intensity ESI LC-MS analysis, an ion's detected intensity directly correlates with the ion's concentration (Voyksner & Lee, 1999). Chelius *et al*. performed a study of myoglobin digests with varying loading concentrations from 10 fmol-100 pmol and analysed these samples using nano-LC ESI tandem MS. The chromatographic MS1 peak areas or area under the curve (AUC) correlates linearly to the concentration of the myoglobin peptide loaded (r^2=0.991). Once all the data for individual peptides has been collated, the correlation between the combined peak area and total protein concentration remains linear. Additionally, myoglobin was also spiked into a human serum sample and the myoglobin peptides were still detected within a complex mixture and the correlation improves with the use of normalised peak areas (r^2=0.9978) (Zhu *et al.*, 2010; Chelius & Bondarenko, 2002).

Li *et al.* (Li *et al.*, 2009) investigated the use of the top three ion approach to provide the absolute quantity of all proteins identified in a sample for their search algorithm. Complex samples were spiked with known concentrations of yeast alcohol dehydrogenase (60, 12 and 15 fmol) to determine absolute concentrations of protein in each sample. The molar amounts show the same relationship as the intensity ratios for the identified peptides thus, peptides are correctly assigned to the correct protein sequence for each sample. Further investigation using product ions show the relative intensities of product ions are consistent and ratios of product ion intensities follow those of the precursors.

Using just accurate mass-retention time pairs, relative abundance of proteins in different samples can be determined through statistical comparisons, where significant changes are highlighted as a relative change in protein abundance. Silva *et al.* have demonstrated that with efficient and reproducible sample preparation methods, the quantitative coefficient of variation (CV) using this label free method lies between 11-14% CV (Silva *et al.*, 2005).

However, there are some limitations when using this quantitation method in complex mixtures. Peak intensities can vary run to run thus; normalisation is required to account for variation. Experimental drift in retention time using a chromatographic column in constant use can also occur. Therefore, high resolution mass measurements and reproducible LC retention time with peak alignment is needed to collect sufficient curve data points and compare sample runs accounting for any drift in retention times.

In MS methods such as DDA, data normalisation is needed to reduce experimental variation between sample injections with the use of internal standards or total ion area. In DDA, AUC data (survey scan) is collected when tandem MS scans are *not* taking place as tandem MS scans in a DDA experiment are for protein identification. In comparison, DIA overcomes this limitation with MS^E. The alternating low energy and high energy MS scan means that quantitative information (low energy scans) and peptide identification (high energy scans) is achieved in a parallel fashion. Removing bias towards higher abundance proteins and reinstating confidence in the quantitative information produced by AUC methods (Wilm, 2009).

1.7.2 Absolute quantitation

It is possible to do absolute label free analysis using emPAI (exponentially modified protein abundance index). Whereby, the number of identified peptides divided by the number of theoretically observed tryptic peptides for each protein, estimates protein abundance. Another method is a modified spectral counting method called absolute protein expression (APEX) where the absolute protein concentration per cell from the proportionality between the protein abundance and the number of peptides is observed (Zhu *et al.*, 2010).

Hi3 absolute quantitation is based on the average MS signal response for the three most intense peptides from a protein per mole being constant or linear. The average MS signal response is determined for the three most intense peptides of an internal standard and this information is used to calculate a universal signal response factor (counts/mol of protein). This is applied to all other identified proteins in the sample and their corresponding absolute concentrations are determined (Silva *et al.*, 2006). Hi3 absolute

and relative quantitation generated by LC-(HD)MS^E data has been used in this study for both one dimensional and two dimensional peptide separation MS analysis.

SRM quantitation is well suited to quantitative proteomics and is ideal for detecting a specific target from a complex background. Typically, in SRM a labelled isotopic internal standard analog (^2H, ^{13}C or ^{15}N) to the target analyte is spiked into the sample and the MS is set to scan for the target analyte and spiked standard.

1.7.3 ELISA

An alternative to MS for quantifying plasma proteins at high sensitivity is enzyme linked immunosorbant assay (ELISA). The ELISA assay was conceptualised and developed by P. Perlmann and E. Engvall *et al.* at Stockholm University in 1971 using the theory of radioimmunoassay's but with an enzyme rather than a radioactive reporter label (Engvall & Perlmann, 1971). Since then ELISA's have become the gold standard technique for biomarker detection and quantitation in clinical laboratories, as the assay can be applied to a large numbers of samples with a high degree sensitivity, robustness and low cost. The limits of detection achieved by the ELISA technique is in the sub-ng/ml range, making it ideal for a clinical setting (Stahl-Zeng *et al.*, 2007).

In principle, ELISA involves an enzyme covalently linked to a specific antibody that recognises the target antigen (target protein). If there is antigen present in the sample, the antibody-enzyme complex binds to the antigen and the enzyme reacts with a substrate, producing a coloured product. The antibodies can be either polyclonal or monoclonal but monoclonal yield more reliable results. There many variations of ELISA methods, direct, indirect, sandwich and competitive. Here we shall consider the indirect ELISA and sandwich ELISA in more detail.

Indirect ELISA detects the presence of an antibody or an antigen (Figure 1.20). In indirect assays, the antigen is absorbed onto the wells, a non-reacting protein (bovine serum albumin) is added to the wells to block any sites in the well that remain uncoated by the antigen. Primary antibodies from a patient a sample is added to the antigen-coated wells and the antigen and antibody are left to bind. Finally, enzyme linked antibodies from another species, such as goat or rabbit, is added and reacts within the

well. All unbound antibodies are washed away and a substrate is applied. If the substrate reacts with the enzyme a colour change is observed, implying that the enzyme link antibody bound to the human antibodies suggesting that the patient sample had antibodies to the antigen. The extent of the colour change is measured using a spectrometer to allow quantitative information to be derived for patient samples against a standard curve. An example of a clinical application of such an assay is HIV testing (Berg *et al.*, 2002).

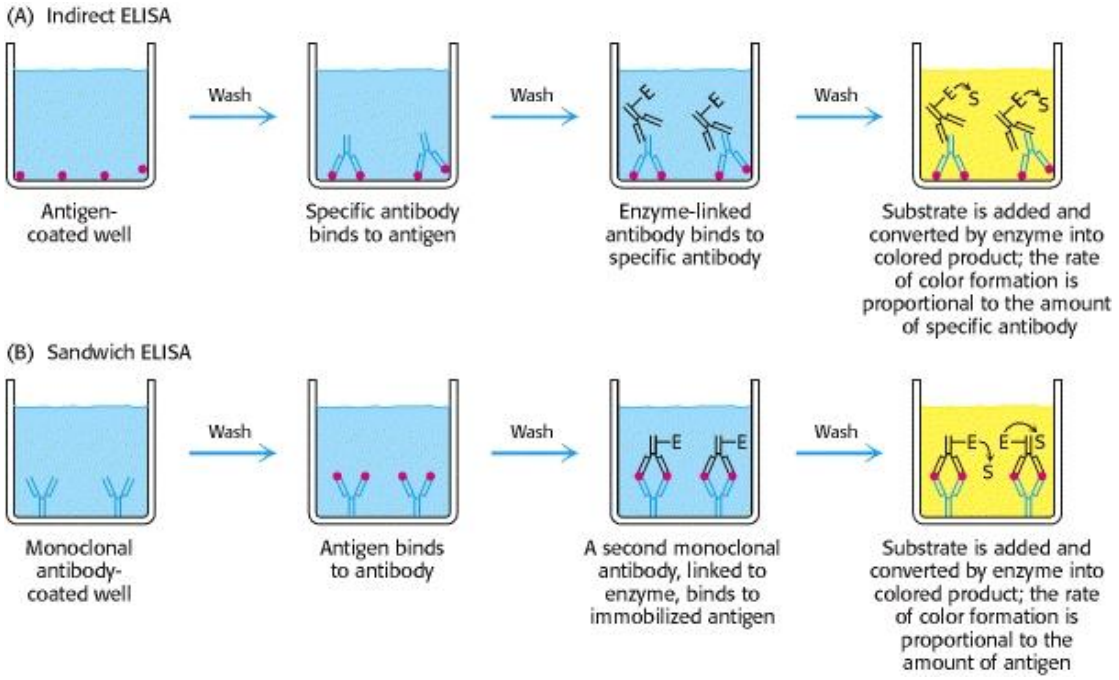


Figure 1.20: Indirect ELISA and Sandwich ELISA (A) In indirect ELISA, the production of colour indicates the amount of an antibody to a specific antigen. (B) In sandwich ELISA, the production of colour indicates the quantity of antigen. Taken from (Berg *et al.*, 2002)

The sandwich ELISA is able to detect and quantify sample antigen (Figure 1.20). A capture antibody to a specific antigen is absorbed to the well, then the antigen (from a patient sample) is added and the antigen binds to the antibody. The specific detecting antibody is added and this binds to the antigen to form an antibody-antigen-antibody sandwich in the wells. In this assay, either the detecting antibody contains the enzyme (direct) or an additional antibody that is enzyme-linked is added to this sandwiched complex (indirect), along with a chemical substrate to produce a signal. In this instance, the signal produced by the reaction is directly proportional to the antigen concentration. With the use of a standard curve, the presence and quantity of antigen can be

determined. Examples of clinical applications for sandwich ELISA is hepatitis B, Giardiasis and food allergy testing (Voller *et al.*, 1978).

The sandwich ELISA was the verification method of choice for this study. The ELISA was used to confirm the results of the discovery mass spectrometry data for a select panel of statistically relevant biomarkers for acute heart failure. Use of ELISA to verify the panel of biomarkers was chosen due to the long establish role and gold standard status in clinical laboratories.

PART THREE: Aims and objectives

Part one and two of this chapter introduced the topics of heart failure and MS led proteomics for biomarker discovery. This project aims to bring these two areas together to identify biomarkers of heart failure. Recent studies in proteomic led discoveries found statistically relevant proteins and peptides for many different forms of CVD as summarised in Table 1.2. Suggesting this approach is able to identify biomarkers in biological samples.

Table 1.2: Biomarkers in plasma and serum for CVD and associated proteomic strategies. Adapted from (Napoli *et al.*, 2013)

Tissue	Protein identified	CVD	Proteomic strategies	References
Plasma	β 2 microglobulin	PAD	SELDI-TOF MS	(Fung *et al.*, 2008)
	α-1-B-glycoprotein, Hakata antigen, tetranectin, tropomyosin	Acute coronary syndrome	2DE-DIGE	(Dardeì • *et al.*, 2010)
	Amyloid A1α, S-sulfate transthyretin	Myocardial infarction	MALDI-TOF	(Kiernan *et al.*, 2006)
	Haptoglobin α, serum amyloid A	Ischemic stroke	2D-DIGE-MS	(Brea *et al.*, 2009)
	Phospholipid-associated proteins	Atherosclerosis vascular disease	MS and density gradient ultracentrifugation	(Gordon *et al.*, 2010) (Davidsson *et al.*, 2010) (Vaisar *et al.*, 2010) (Heinecke, 2009)
	Apolipoprotein A1	Myocardial infarction	1D and 2D SDS–PAGE	(Davidsson *et al.*, 2010)
Serum	Apolipoprotein J	Myocardial infarction	2-DE-MALDI-TOF	(Cubedo *et al.*, 2011)
	cTnI, CK, CK-MB	Myocardial infarction	SELDI-TOF MS	(Silbiger *et al.*, 2011)
	HSP-27	Atherosclerosis vascular disease	2DE	(Martin-Ventura *et al.*, 2004)
	sTWEAK	Atherosclerotic plaques, PAD, Coronary artery disease, heart failure, myocardial infarction	2-DE-MALDI-TOF	(Blanco-Colio *et al.*, 2009) (Urbonaviciene *et al.*, 2011)
	CRP, C3a, C5a	Myocardial infarction, abdominal aortic aneurism	2D-DIGE	(Distelmaier *et al.*, 2009)
	Peroxiredoxin-1	Abdominal aortic aneurism	2D-DIGE	(Martinez-Pinna *et al.*, 2010)
	Leucine-rich α2-glycoprotein	Heart failure	2D-DIGE MS	(Watson *et al.*, 2011)

1.8 Project aims

This study tests the hypothesis that biomarkers of heart failure can be identified in plasma using state of the art label free mass spectrometry technology coupled with sophisticated informatics. The primary aim of this project is to discover blood-based biomarkers of heart failure.

This study will bring together novel pre-analytical techniques with mass spectrometry analysis to characterise the plasma proteome and discern biomarkers of heart failure. The effectiveness and applicability of the pre-analytical methods will be explored and developed to determine which method provides the most comprehensive coverage of the plasma proteome. The optimised method was evaluated using DIA mass spectrometry (MS^E) and bioinformatics for a biomarker discovery study in heart failure for both diagnostic and prognostic markers.

Thus, the main objectives are:

- To assess a number of sample preparation methods that allows the greatest coverage of the proteome and number of proteins to be ascertained. The method will have to generate a rich protein profile for identification and quantification using label free MS^E, the method will have to be reproducible for the human plasma samples.

- Evaluate the use of ion mobility in biomarker discovery for heart failure using a MS^E vs. $HDMS^E$ approach. Compare this approach to the Orbitrap technology using DDA.

- The method will be applied to carefully selected human plasma samples that allow the discrimination of heart failure biomarkers to be achieved for both diagnosis and prognosis using bioinformatic approaches.

- A panel of biomarkers will be verified using ELISA technology to see if the proposed biomarkers can be reliably detected, measured and help move the work into the clinical laboratory.

Chapter Two

Material and Methods

2.1 Materials

2.1.1 General chemicals and kits

All reagents and kits were purchased from Sigma-Aldrich Company Ltd (Poole, UK) at HPLC grade, Fisher Scientific Ltd (Loughborough, UK) at Optima LC-MS grade, Bio-Rad Laboratories Ltd (Hemel Hempstead, UK) and Agilent (Cheadle, UK) unless stated otherwise.

2.1.2 Specimen collection

All samples were collected through written explanation and consent from hospital patients. Whole blood samples were collected in 5 mL vacutainer collection tubes with EDTA as the chosen anti-coagulant (BD Vacutainers, Oxford, UK). The plasma was separated through centrifugation at 3000 rpm for 15 minutes, and stored in 1.5 mL eppendorf tubes (Eppendorf UK Ltd, Cambridge, UK) at -80°C, until required.

2.1.3 Mass spectrometry standards

Lyophilised Waters MassPREP^{TM} alcohol dehydrogenase (ADH) digestion standard (yeast origin, saccharomyces cerevisiae) was used for quantitative analysis. One vial of ADH was reconstituted in 1 mL of 10% acetonitrile (v/v), giving a peptide concentration of ~1 pmol/µl. A 25 µl aliquot of the stock solution was further diluted in 0.1% formic acid (FA) (v/v), giving a final concentration of 50 fmol/µl of the ADH standard in solution.

Lyophilised MassPREP^{TM} Enolase digestion standard (yeast origin, saccharomyces cerevisiae) was used to assess nano-UPLC performance. One vial of enolase was reconstituted in 1 mL 10% acetonitrile (v/v) giving a peptide concentration of ~1 pmol/µl. A 25 µl aliquot of the stock solution was further diluted in 0.1% FA (v/v), giving a final concentration of 50 fmol/µl of the enolase standard in solution.

Lyophilised Glu-Fibrinogen peptide (GFP) was used for mass calibration. One vial was reconstituted in 50% aqueous acetonitrile with 0.1% FA (v/v) to give a final concentration of 500 fmol/µl.

Lyophilised Leucine Enkephalin was used to check the detector performance of the mass spectrometer. One vial was reconstituted in 50% aqueous acetonitrile with 0.1% FA (v/v) to give a final concentration of 5 ng/µl.

2.2 Methods

Many pre–MS methods have been applied to plasma proteomics for biomarker discovery to obtain the maximum proteomic data from MS analysis. For this study three pre-analytical techniques of enrichment with equalizer bead, organic solvent precipitation with acetonitrile and SPE with RP-C18 and C18 extra wide pore (EWP) have been tested. Chromatographic separation was used in one dimensional (1D) and two dimensional (2D) setting. Preceding all MS analysis sample protein concentration was either determined by bicinchoninic acid assay (BCA) or spiking MS samples equally with an internal standard such as ADH. All protein samples were digested using optimum grade trypsin for bottom up proteomic peptide analysis.

2.3 Solid phase extraction

2.3.1 C18

To prepare the plasma sample for the Sep Pak C18 columns (Sep Pak C18 3cc, Waters, Manchester), 1 mL of 0.1% trifluroacetic acid (TFA) was added to 100 µl plasma and precipitated by centrifugation at 16,000 rpm for 2 min. The supernatant was removed and placed in an eppendorf for use later.

The C18 columns were primed with 2 mL methanol and washed with 2 mL 0.1% TFA three times. The supernatant of the precipitated sample was then placed on column, the eluent was collected, this step was repeated. The column was washed again with 2 mL 0.1% TFA three times and the proteins eluted with 2 mL of 60% aqueous ACN in 0.1% TFA. The elutions were collected, evaporated to dryness and lyophilised overnight.

2.3.2 EWP

To prepare the plasma sample for the EWP Bond Elute columns (C18-EWP Bond Elute 3 mL, Varian, USA), 1 mL of 0.1% TFA was added to 100 µl plasma and precipitated by centrifugation at 16000 rpm for 2 min. The supernatant was removed and placed in an eppendorf for use later.

The EWP columns were primed with 2 mL methanol and washed with 2 mL deionised water twice and 2 mL 0.1% TFA twice. The supernatant of the precipitated sample was then placed on column, the eluent was collected and this step was repeated. The column was washed again with 2 mL 0.1% TFA twice and 2 mL of deionised water twice. The proteins were eluted with 1.2 mL of 60% aqueous ACN in 0.1% TFA and 1.2 mL 90% aqueous ACN in 0.1% TFA. The elutions were collected, evaporated to dryness and lyophilised overnight.

2.4 Organic solvent precipitation

Methods were taken and adapted from Kay *et al.* (Kay *et al.*, 2008).

2.4.1 Acetonitrile precipitation

To an eppendorf 20 µl human plasma was transferred and 40 µl deionised water was added and vortexed. To this mix 90 µl ACN was added and the sample was sonicated for 10 min. The sample was vortexed and sonicated again for 10 min. The protein precipitate was pelleted by centrifugation at 12000g for 10 min at room temperature. The supernatant was transferred to a clean eppendorf tube and evaporated to dryness. The sample was lyophilised overnight. The sample was reconstituted in 16 µl 50 mM ammonium bicarbonate at pH 8.2 ready for MS analysis.

2.4.2 Acetonitrile precipitation with urea

To an eppendorf 20 µl human plasma was transferred and 40 µl deionised water was added and vortexed. To this mix 90 µl ACN was added and the sample was sonicated for 10 min. The sample was vortexed and sonicated again for 10 min. The protein precipitate was pelleted by centrifugation at 12000g for 10 min at room temperature. The supernatant was transferred to a clean eppendorf tube and evaporated to dryness. The sample was then lyophilized overnight. The sample was reconstituted in 12 M urea in 50 mM ammonium bicarbonate, reduced with Dithiothreitol (DTT) at final concentration of 15 mM and alkylated with Iodoacetamide (IAA) at a final concentration 20 mM. The urea was dialysed out of the sample with 2000 MW cut off tubes in 50 mM ammonium bicarbonate pH 7.6 for two days, ready for MS analysis.

2.4.3 Acetonitrile precipitation: reconstitution of the pellet

To an eppendorf 20 µl human plasma was transferred and 40 µl deionised water was added and vortexed. To this mix 90 µl ACN was added and the sample was sonicated for 10 min. The sample was vortexed and sonicated again for 10 min. The protein precipitate was pelleted by centrifugation at 12000g for 10 min at room temperature. The supernatant was transferred to a clean eppendorf tube and the protein pellet was re-suspended in 100 µl of the following solvent – 25% ACN: 25% Methanol: 50% deionised water. The sample was evaporated to dryness and lyophilised overnight. The sample was reconstituted in 16 µl 50 mM ammonium bicarbonate at pH 8.2 ready for MS analysis.

2.5 Equalizer beads

The ProteoMiner^{TM} Kit reduces the dynamic range of plasma making it easier to analyse and detect lower abundance proteins. This is achieved by using the pre-packed column which contains beads made up of a library of combinatorial peptide ligands. Once the plasma sample is applied to the beads, the high abundance proteins saturate their ligands quickly depleting the excess in several wash stages whereas, the medium and low abundance proteins take longer to saturate their ligands causing them to concentrate or enrich. This dual action reduces the difference in plasma protein concentration in the proteome without a loss of proteins that were present in the untreated sample.

2.5.1 Equalizer beads protein binding

The pre-packed columns containing the beads (ProteoMiner^{TM} Protein Enrichment Kit, catalog #163-3003, Bio-Rad, Hemel Hempstead, UK) were prepared for plasma incubation according to the manufacturer's instructions till the elution stage, where method development took place. The change in elution buffer was made due to the presence of CHAPS in the kit elution reagent buffer which causes signal suppression in mass spectrometry.

The columns were spun at 1000 x g for 60 seconds to remove storage material. The columns were washed with 600 µl of PBS buffer (150 mM NaCl, 10 mM NaH2PO4, pH 7.4) and 1 mL of plasma was added and incubated for 2 hours at room temperature for protein binding. Unbound plasma proteins were removed by centrifugation. The

column was washed four times with 600 µl PBS buffer at pH 7.4, followed by a wash with 600 µl de-ionised water.

The bead bound proteins were eluted with various elutions reagents through centrifugation at 1000 x g for 2 min. Each of these elution reagents were tested to see which would obtain the maximum protein recovery. The different elution reagents methods are discussed below.

2.5.2 Equalizer beads elution with differential elutions

The bead bound protein was treated with 100 µl of 9 M urea at pH 3.3 for 15 min, with 3 min vortex cycles throughout. The elution was collected, stored and the entire step repeated a second time. Then a differential elution was performed using 100 µl organic solvent (16.6% ACN; 33.3% Isopropanol; 0.5% TFA; 49.5% deionised water). The beads were treated with the organic solvent for 15 min, with 3 min vortex cycles throughout. The organic solvent precipitation was repeated twice. The elutions were pooled and stored at -80°C.

2.5.3 Equalizer beads with peptides and elution with organic solvent

A plasma sample was digested with trypsin estimating 80 mg/ mL protein and 1 mL of digested plasma proteins was used as the sample and the same process of equalizer bead binding took place. The bead bound peptides were treated with 100 µl organic solvent (16.6% ACN; 33.3% Isopropanol; 0.5% TFA; 49.5% deionised water) the beads were treated with the organic solvent for 15 min, with 3 min vortex cycles throughout. The organic solvent precipitation was repeated twice. The elutions were pooled and stored at -80°C.

2.5.4 Equalizer beads elution with trypsin on column

The bead bound protein was treated with 100 µl 50 mM ammonium bicarbonate at pH 7.6 and vortexed. The bead bound proteins were reduced using a final concentration of 15 mM DTT and incubated at 50 °c for 30 min and protein were alkylated using IAA at 20 mM final concentration in the dark for 30 min. The beads were incubated with 10 µg trypsin (proteomics grade, Sigma Aldrich, Poole, UK) at 37 °C overnight. The elution was collected and acidified with formic acid to give an overall 0.1% FA concentration.

2.5.5 Equalizer beads with acetonitrile depleted plasma and trypsin on column

The protein sample (1 mL) was depleted using acetonitrile depletion (see section 2.4.1) and this sample was used for equalizer bead binding, the same process of equalizer bead binding took place. The bead bound protein was treated with 100 µl 50 mM ammonium bicarbonate at pH 7.6 and vortexed. The bead bound proteins were reduced using a final concentration of 15 mM DTT and incubated at 50 °C for 30 min and proteins were alkylated using IAA at 20 mM final concentration in the dark for 30 min. The beads were incubated with 10 µg trypsin (proteomics grade, Sigma Aldrich, Poole, UK) at 37°C overnight. The elution was collected and acidified with FA to give an overall 0.1% concentration.

2.5.6 Equalizer beads with 2% sodium deoxycholate

To the bead bound protein 2% sodium deoxycholate (SDC) in 50 mM ammonium bicarbonate pH 7.6 was added followed by incubation at 60 °C for 60 seconds, then the elution was collected. This step was repeated three times. The elutions were pooled and stored at -80 °C.

2.5.7 Post bead desalting

Empore columns (Sigma-Aldrich, Poole, UK) were used to remove all salts that may interfere with MS analysis. The Empore columns were primed with 1 mL of methanol and washed with 1 mL of 0.1% FA three times. The protein sample was added to the column and recirculated. The column was washed a further three times with 0.1% FA followed by sequential elution of 60 and 80% ACN (both in 0.1% formic acid). The elutions were pooled and the volume was reduced by lyophilisation. The protein pellets were dissolved in 50 mM ammonium bicarbonate pH 7.6 and the protein concentration was determined by BCA.

2.6 Protein assay

To determine the proteins concentration of the samples, a standard known amount of another protein needs to be used to produce a standard curve. The protein amount of an unknown solution can then be ascertained using the known standard curve. In this case bovine serum albumin (BSA) is used as the known protein. BCA was used in this study

to determine all protein concentrations. The assay is formed of two reactions; the first is the reduction of Cu^{2+} to Cu^{1+} by protein in an alkaline medium, followed by chelation of bicinchonic acid with the Cu^{1+} ion which produces a purple coloured product which is absorbed at 562 nm (Smith *et al.*, 1985).

2.6.1 BCA working solution

BCA working solution was made by combining 50 µl solution C (0.4 g Copper sulphate in 10 mL deionised water) to 1.25 mL solution B (0.25 g of Bicinchoninic acid in 6.25 mL of deionised water) and 1.35 mL solution A (0.8 g sodium carbonate, 0.16 g sodium tartrate in 10 mL deionised water and solution made to pH 11.25), in that order.

2.6.2 BCA assay

To a 96 well plate, 100 µl of deionised water was added to an entire row of the plate (i.e. wells A-H) for the standard curve. To this 80 µg BSA was added to well A and B. Then from well B a doubling dilution was carried out until well G, leaving well H with just 100 µl deionised water (BSA protein concentration from 80 µg; 40 µg; 20 µg; 10 µg; 5 µg; 2.5 µg; 1.25 µg; 0 µg). For the samples to be analysed 99 µl deionised water were added to adjacent rows wells and 1 µl of sample was added to the well, each sample was analysed in triplicate. To all the standard curve and sample wells 100 µl BCA working reagent was added and the plate was sealed with film and incubated at 60 °C for 1 hour, the plate was read using a spectrophotometer at 562 nm.

2.7 Tryptic digestion

Digestion of protein samples is usually the penultimate step before MS sample analysis. In order to ensure that the digestion of the protein is thorough, DTT and IAA are used. The DTT is a reducing agent and reduces the disulfide bonds to sulfhydryl groups in the presence of heat to denature and enhance the action which unfolds the proteins. The protein is alkylated using IAA to ensure reformation of disulfide bonds doesn't take place after reduction.

For digestion of all samples they were first reduced with 100 mM dithiothreitol made up in 50 mM ammonium bicarbonate (NH_4HCO_3) and was used at a final concentration of 15 mM. The samples were vortexed and incubated at 50 °c for 30 min. The samples

were cooled to room temperature then alkylated with 100 mM Iodoacetamide in 50 mM NH_4HCO_3 used at final concentration of 20 mM and left in the dark at room temperature for 30 min.

A vial of 20 µg trypsin (Proteomics grade, Sigma Aldrich, Poole, UK) was re-solubilised in 20 µl of 50 mM ammonium bicarbonate and used at a ratio of 1 µg of trypsin for every 50 µg plasma protein. The trypsinised sample was incubated overnight at 37 °c. Following overnight incubation, the samples were acidified using FA at a final sample concentration of 0.1% (v/v), the addition of acid hydrolysed the remaining trypsin and the samples were ready for MS analysis.

2.8 Chromatography

2.8.1 One dimensional configuration

All nano liquid chromatographic separations were performed using a directly coupled NanoAcquity UPLC system (Waters Corporation, Milford, MA, USA). The system was comprised of an auxiliary solvent, binary solvent and sample manager fitted with an analytical column heating and trapping component.

LC separations were performed using a Symmetry C18 trapping column (180 µm x 20 mm 5 µm) and a BEH C18 analytical column (75 µm x 250 mm 1.7 µm). The solvents for the nano separations were made up of solvent A which contained 0.1% v/v aqueous formic acid and solvent B which contained 0.1% v/v formic acid in acetonitrile.

For sample analysis, an aliquot of each sample containing 1:1 internal standard ADH was injected onto the trapping column and flushed with 0.1% solvent B for 3 min at 5 µL/min. The sample was eluted from the analytical column at a flow rate of 300 nL/min, with the organic solvent (solvent B) increasing in concentration from 3-85% over various time periods as discussed in Table 2.1. The sample eluted to a nanoESI sprayer (Waters Corporation, Milford, MA, USA) and a capillary voltage of 2.5 kV was applied. All sample analysis was conducted in triplicates for statistical accuracy.

Table 2.1: Elution gradients used in for 20 min (A), 50 min (B) and 110 min (C) NanoAcquity UPLC analysis

(A) 20 min

Time(min)	Flow Rate(µL/min)	%A	%B	Curve
Initial	0.3	99.0	1.0	
10.00	0.3	50.0	50.0	6
10.33	0.3	15.0	85.0	6
11.60	0.3	15.0	85.0	6
12.00	0.3	99.0	1.0	6

(B) 50 min

Time(min)	Flow Rate(µL/min)	%A	%B	Curve
Initial	0.3	99.0	1.0	
30.00	0.3	60.0	40.0	6
32.00	0.3	15.0	85.0	6
34.00	0.3	15.0	85.0	6
35.00	0.3	97.0	3.0	6

(C) 110 min

Time(min)	Flow Rate(µL/min)	%A	%B	Curve
Initial	0.3	99.0	1.0	
85.00	0.3	60.0	40.0	6
87.00	0.3	15.0	85.0	6
100.00	0.3	15.0	85.0	6
100.10	0.3	97.0	3.0	6

Prior to and after each set of samples, a quality control (QC) injection of 50 fmol Enolase was analysed using LC-MS^E to determine the quality of the chromatographic resolution and the data was processed by PLGS (Figure 2.1). The acceptable peptide sequence coverage was 40%. The starting peptide sequence coverage of enolase was monitored throughout sample acquisition and the peptide sequence coverage was checked to ensure there was not a decrease/increase of coverage by ±10%. Where the peptide sequence coverage fell outside of the pre-determined limits for enolase, no further sample data was collected and the cause of the loss or increase of peptide identification was investigated and resolved.

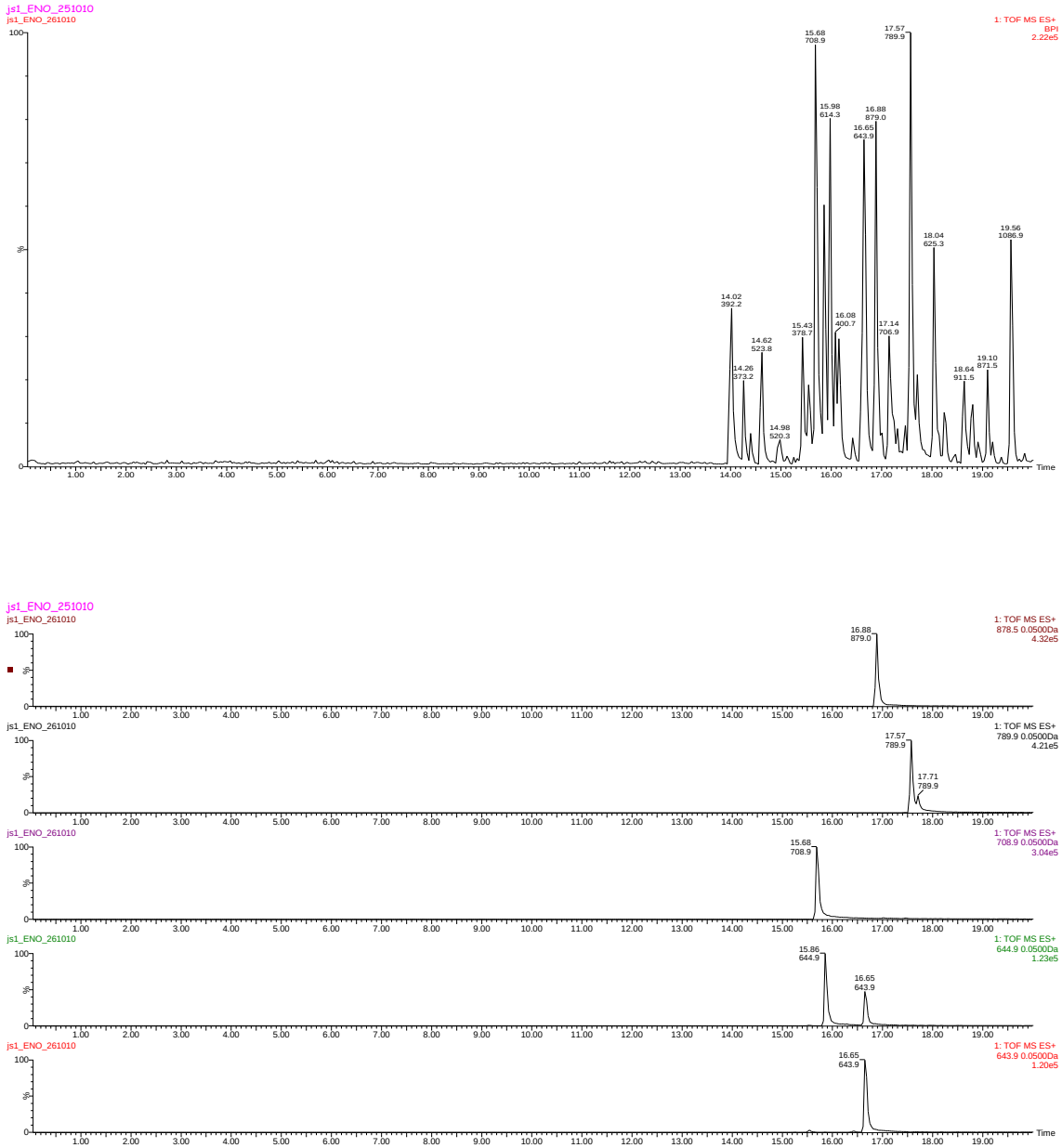


Figure 2.1: Example of an enolase chromatogram used to assess chromatographic resolution (above) and extracted ion chromatogram (below). The nanoUPLC performance is assessed by the chromatogram and intensity of individual peaks at 643.9, 644.9, 708.9, 789.9, 878.6 Da. When judging the chromatogram the retention time, peak width and peak intensity all need to be checked. The raw data is then applied to the PLGS programme and equal to or greater than 40% coverage is expected for the nanoUPLC to be deemed fit for purpose.

2.8.2 Two dimensional configuration

2.8.2.1 High pH RP- Low pH RP concatenation

All high pH RP separations were performed offline using a Waters Alliance 2690 HPLC system (Waters Corporation, Milford, MA, USA). The system consisted of a binary solvent system, an injection valve and fraction collector. The column used was Gemini NX C18 110 Å, 150 x 3 mm analytical column with 3 µm particles (Phenomenex Inc., Cheshire, UK) and was kept at room temperature. Mobile phase A consisted of 10% 200 mM aqueous ammonium formate at pH 10 and mobile phase B consisted of 10% 200 mM ammonium formate in acetonitrile at pH 10. Tryptically digested samples were loaded onto the column and equilibrated with 10% 200 mM aqueous ammonium formate at 0.5 mL/min. 800 µg of sample was injected on column and the following gradient was run over 95 mins for sample separation. In the first 10 mins 0% to 5% B, from 20-70 mins solvent B was ramped from 5 to 35%, from 70-85 mins solvent B was ramped to 70%, held at 70% for an additional 10 mins and ramped back down to 0% at 100 mins. The sample elution was observed at UV 214 nm. The elutions were collected every minute thus, 100 elutions were collected from the first dimension in total. These samples were then concatenated into groups of ten (Table 2.2).

The concatenated samples were evaporated for 2 hours and lyophilised overnight. The samples were reconstituted in 20 µl of 0.1% formic acid in 50 mM ammonium bicarbonate and spiked with 100 fmol ADH at 1:1 ratio.

The sample was applied to the second low pH RP dimension for analytical separation using nanoscale LC separations with a NanoAcquity UPLC system (Waters Corporation, Milford, MA, USA) couple to MS. The system was composed of an auxiliary solvent, binary solvent, and sample manager with a column heating and trapping column section. All chromatographic separations were performed using a BEH C18 column (75 µm x 200 mm 1.7 µm) using 300 nL/ min flow rate.

Table 2.2: Collection order for concatenation of samples, here the minute of collection is stated for each of the pooled samples.

Sample number	Collection time point (min)
0	10,20,30,40,50,60,70,80,90
1	11,21,31,41,51,61,71,81,91
2	12,22,32,42,52,62,72,82,92
3	13,23,33,43,53,63,73,83,93
4	14,24,34,44,54,64,74,84,94
5	15,25,35,45,55,65,75,85,95
6	16,26,36,46,56,66,76,86,96
7	17,27,37,47,57,67,77,87,97
8	18,28,38,48,58,68,78,88,98
9	19,29,39,49,59,69,79,89,99

2.9 Synapt G2 HDMS analysis

2.9.1 The Mass Spectrometer

A high definition nanoUPLC with ESI coupled to a Quadrupole–Travelling wave ion mobility–Time of Flight mass spectrometer (nanoUPLC-ESI-Q-TWIMS-TOF commercially known as the Waters Synapt^{TM} G2), was used to analyse complex plasma samples from various heart failure patients (Figure 2.2). The reason for choosing this platform in place of other available technologies is the Synapt^{TM} G2's ability to extend its analytical dynamic range over five orders of magnitude, operate in data independent mode (MS^{E}), and to incorporate ion mobility separation in a single experiment (HDMS^{E}) without loss of analytical time, ensuring high resolution and high mass accuracy mass spectrometry data.

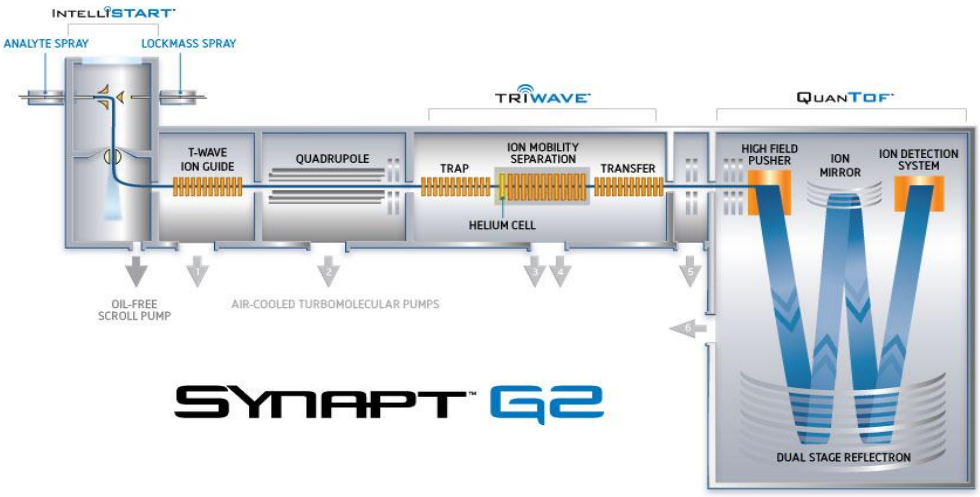


Figure 2.2: Design of the Synapt G2 mass spectrometer

The pre-digested plasma samples are introduced post-chromatographic separation on a C18 RP column from the nanoUPLC to the Q-TWIMS-TOF via a silicon emitter tip. As the peptides exit the emitter tip they undergo electrospray ionisation and enter the Z-spray source. The Z-spray source is designed to reduce background ions and neutral ions from interfering with MS analysis. The ions pass through the source to the quadrupole, the quadrupole is set to allow all ions to filter through to the TRIWAVE region of the mass spectrometer where separation of the ions occur with ($HDMS^E$) or without (MS^E) ion mobility separation.

The TRIWAVE cell is made up of three sections: trap cell; IMS cell; and transfer cell. Both the trap and the transfer cells can act as the collision cell i.e. where CID fragmentation takes place and produce fragments before and/or after separation by ion mobility. In MS^E the collision cell is set to exchange between low (4 eV) and high collision energy (ramped between: 15 – 40 eV) whilst data acquisition takes place. In low collision energy, no fragmentation occurs thus, all ions travel through the mass spectrometer and this is recorded as the precursor ion spectrum. In high collision energy the ions travel from the ion source to the collision cell where the energy ramping occurs producing product ions, these ions are recorded as the product ion spectrum. The parallel scanning between the precursor and product ions produces a peptide fragment pattern which is time aligned (i.e. product ions are paired with their precursor ions) by matching the retention times profiles of each ion, this is known as data independent

acquisition (Figure 2.3). To ensure that accuracy is maintained throughout data acquisition and analysis, a reference spray or LockSpray is used to infuse Glu-Fibrinopeptide B (GFP) and this is scanned every 30 seconds. This detected GFP *m/z* peak can be compared to the theoretical GFP *m/z* peak (785.84265 $[m+2H]^+$), a correction value is calculated with this data and applied to the sample spectra.

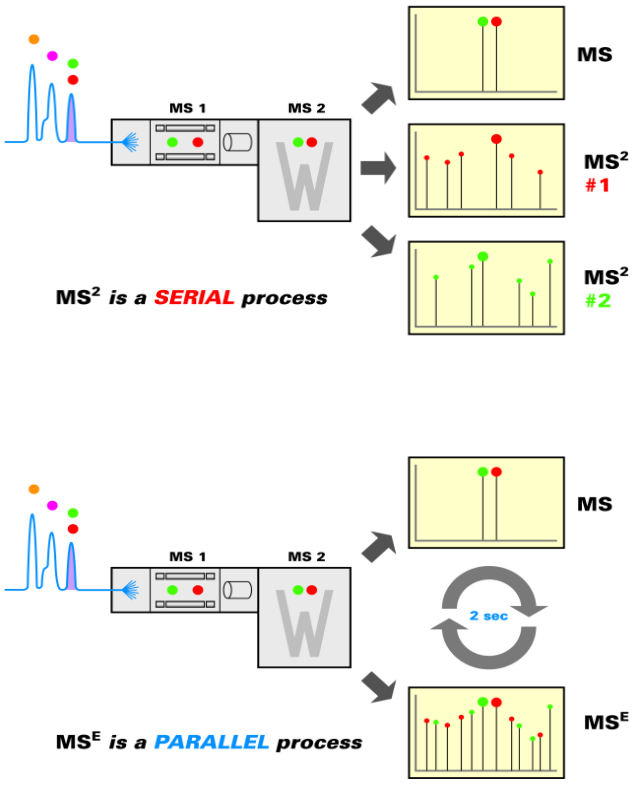


Figure 2.3: Comparison of data dependant acquisition (top) and data independent acquisition (bottom). The Synapt G2 uses data independent acquisition or MS^E.

MS^E experiments can be enhanced using ion mobility (LC-IMS-DIA-MS or $HDMS^E$) There is now the incorporation of the IMS separation cell in the TRIWAVE (Figure 2.4) with further separation of ions using TWIMS (see section 1.4.3.2). The ion mobility cell consists of two chambers, the first chamber is flooded with an inert buffer gas (helium) and acts as passage into the second chamber which is filled with nitrogen. The second nitrogen filled chamber is where the TWIMS takes place. The nitrogen acts as the buffer gas to the travelling ions, which travel in a waveform due to the differential voltage potentials, applied across the cell. The smaller ions are able to overcome the potential gradients with more ease than larger ions therefore, smaller ions

takes less time than the larger ions to travel through the IMS cell. Thus, ions can be separated according to size and shape and the time it takes an ion to *'drift'* through the cell. The drift time is independent of the *m/z* value of an ion adding an extra degree of separation and information than the traditional MS^E analysis. In $HDMS^E$, the fragmentation is set to take place in the transfer cell (low and high collision) prior to TOF analysis and detection, meaning we have both retention time and drift time alignment, improving specificity.

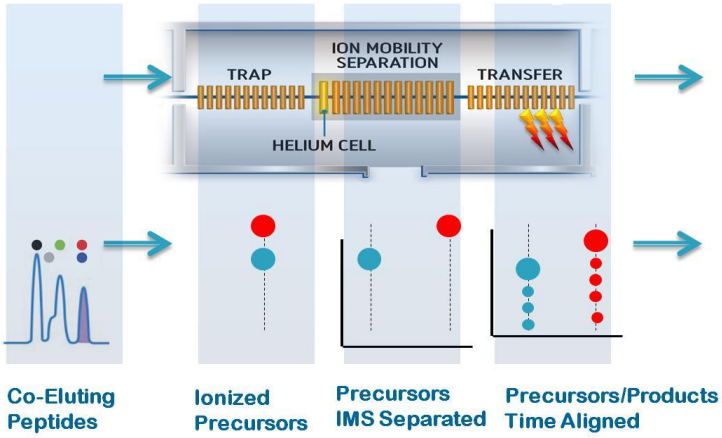


Figure 2.4: Schmeatic of the TRIWAVE cell within the Synapt G2 mass spectrometer

2.9.2 Preliminary checks

To analyse the results produced by the Synapt G2 HDMS, preliminary checks of the Synapt were undertaken to ensure optimal performance. These checks allow the assessment of the nanoUPLC and the MS individually and highlighted any areas of performance that could be a source of error. The preliminary checks which took place prior to sample analysis were; detector check, mass calibration and nanoUPLC check.

The detector check was carried out by infusing 5 ng/µl Leucine Enkephalin at 0.5 µl/min and an *m/z* of 556.2771 was observed and measured. Once this signal was observed the sensitivity of the detector was checked by ramping the detector voltage up and down. The detector checks the voltage that produces the best signal for leucine enkephalin.

GFP was used for lockmass correction and TOF mass calibration. For calibration 500 nL/ min of 500 fmol/ µl GFP was infused via the nanoESI spray source and an MS/MS spectrum was obtained from the doubly charged precursor of GFP peptide (*m/z* 785.84265) from 50 to 1500 Da. The calibration was accepted only when the average ppm error across the mass range was <5 ppm. GFP was also used for lockmass correction and was infused through the NanoLockSpray at 500 nL/ min at 500 fmol/ µl and sampled every 30 seconds during data acquisition.

2.9.3 Data acquisition MS^E

Data independent acquisition was used for all MS analysis in this study. The mass spectrometer was set to exchange between low and high collision energy. In low collision mode data was collected at constant trap collision energy of 4 eV. In high collision energy mode the trap collision energy was set to ramp between 15-40 eV. The transfer collision energy was held at 22 eV and 11 eV for low and high energy respectively. The parallel scanning between the precursor and product ions produced a peptide fragment pattern that was time aligned (i.e. product ions were paired with their precursor ions) by matching the retention time profiles of each ion. The spectral acquisition scan rate was set at 0.9 sec with a 0.1 sec interscan delay. All data was lockmass corrected using the monoisotopic ion of the doubly charged precursor peptide GFP with *m/z* 785.84265.

2.9.4 Data analysis

All MS data generated in this study was processed, identified and quantified using Protein Lynx Global Server (PLGS) version 2.4-2.5.2. PLGS applies a series of algorithms purpose built for MS^E data. These algorithms deconvolute precursor and product ion data by smoothing, removing background noise, centering, deisotoping, reducing the charge state and correcting the mass of data which then lends the data to initial assignments. This iterative process increases the selectivity, specificity and sensitivity of the data analysis and subsequent protein assignments through a series of re-ranking and re-scoring steps until a pre-defined FDR is exceeded.

The raw data (mass spectrums) from all LC-MS^E runs is collected in three different functions: Low energy, High energy and LockSpray. The information is submitted to *ion detection* where data from function 1 and 2 are compiled into a list containing

information on accurate mass-retention time components. The components that are taken into consideration include monoisotopic accurate mass, a calculated mass deviation, summed peak area of all isotopes in all charge states, calculated area deviation, the apex retention time, the chromatographic peak start and end retention times and fractional charge state (Silva *et al.*, 2005).

The precursor and product ions are *time-aligned* from information gathered during the ion detection process and the peak apex. The time-alignment algorithm performs on the basis that the high energy ions calculated apex retention time is equal to the low energy apex retention time; ± 10% of the chromatographic peak width of function 1. In the case of co-eluting peptides, the resultant product ions are assigned to multiple precursors until further along the process where these are refined and depleted using mass accuracy data and protein assignments.

Prior to any search algorithms all time aligned precursor and product ions are *filtered* eliminating all low energy precursors under 750 Da and high energy precursors ion under 350 Da as these are typically tryptic peptides that have high sequence similarity therefore, lacking specificity. These are nonessential until later in the data processing and analysis stage. Further filtering steps include removal of product ions higher in intensity and mass than the precursor ions.

The filtered precursor and product data is prepared for submission to a database search algorithm however, preceding all database searching a database is specified and a randomised database is created from this within PLGS. Once this is established a *pre-assessment survey* where a presearch of time-aligned precursor and product ion tables is conducted. During the presearch the algorithm adjusts the model parameters of peptides in the gas/liquid phase. If the peptide numbers do not meet the criteria or thresholds that are set, a default model of parameters is applied, if the peptide number exceeds approximately 250 the algorithm develops a new model for these set of peptides. The models that are tested and adjusted include: a real-time retention time model, a monoisotopic product ion mass distribution, fragmentation model. Comparisons of sequence length, charge state and precursor intensities to the summed number of identified product ions. The summed product ion intensity is related to the precursor ion

intensity and the total number of continuous and complementary identified b and y ions. Furthermore, comparisons of observed peptide charge states (both sequence order and composition) are made with the summed y/b ion intensity ratios. The data is now ready for the iterative database searching algorithms.

Database search pass 1 is the first iteration where precursor/ product ion tables is queried against the protein sequence database in which only completely cleaved tryptic peptides are considered for identification. For a putative identification to be made the precursor peptides must fit within precursor mass tolerance of < 10 ppm and contain at least 3 product ions with a precursor mass tolerance of < 20 ppm. These putative peptide identifications are scored and then re-scored when compared with a defined model for identifications. Peptide scores are re-defined repeatedly when compared to numerous models generated such as retention time, product ions generated from neutral loss of water or ammonia from peptides containing certain amino acids, presence of certain amino acids near the N-terminus affecting the y/b ion intensities and complementary N- and C- terminal product ions. Additional re-scoring occurs when the peptide identifications are compared with the predicted fragmentation model. The highest scoring tentative peptide identifications are given a peptide rank value of 1 and all others 0.

All peptides now with a rank value of 1 are assigned to their tentative proteins and are re-scored and re-ranked by summing the intensity of product ions. The protein identification with the highest total product ion intensity is given the rank 1 and others are ordered in descending order to this identification. It is at this point that the protein score is derived from matching tryptic peptides with a rank value of 1 and normalised to the peptide length and summed intensity of the three best ionising peptides. These scores are adjusted according to various physiochemical properties and models defined by the presearch. Once the protein scoring process is complete, the highest scoring protein and all precursor and product ions associated with the top rank peptide sequence for that protein are eliminated from subsequent database identifications.

The ranking and scoring process continues until either the FDR exceeds the acceptable value or no protein identification exceeds the minimum score. When the highest scoring reverse/random identification is made, the number of proteins identified prior to this is

determined. This number of identifications is multiplied by the chosen FDR to determine the maximum reverse/random identifications that are allowed to occur. This calculation is repeated with every reverse/random identification in the list until the number of reverse/random identifications is equal to the calculated acceptable number. This marks the end of database pass 1.

Database search pass 2 is set to identify peptide modifications (such as in-source fragmentation ions, loss of water/ammonia) and non-specific cleavage products for proteins assigned in database pass one. A subset database of protein identified in pass 1 is created and using accurate mass-retention time information all non-identified components form pass 1 are re-determined with the addition of programme-modified pass one proteins. The software modifies the subset proteins in various ways such as by adding in-source fragments, in-source loss of water and ammonia, missed cleavages and PTM's.

The first iteration, in-source fragments are assigned to their respective product ions if they can be time aligned using the accurate mass of fragments. The second iteration looks for precursor ion that have a loss of water or ammonia and the third and final iteration identifies missed cleavages, oxidised methionines, deamidations and other modifications. In the event of peptides being assigned to more than one possible variant, the variant containing the largest number of matched product ions and have fragments indicating the point of modification, is selected as the best match. Where the accurate mass-fragment ion information is insufficient to confirm a variant, the modifications are excluded from assignment.

Database search pass 3 once again realigns all data that remains unassigned from the first two databases passes. In this database pass the precursor and product ion tables are searched without product ion intensity restriction. This accounts for potential in-source fragmentation of highly labile peptides. Like pass one all identifications are scored, ranked and assigned to their proteins repeatedly until the minimum protein score and/or FDR criteria is met (Li *et al.*, 2009).

Quantitative information in MS^E is discerned from low and high energy scans using the peak intensity of the mass spectra (see section 1.7.3). The low energy scans are used for quantitative information and the high energy scans for peptide identification. It has been observed that the summed intensity of the top three most intense precursor ions of a given protein is proportional to its molar amount; this relationship can be used to estimate the protein quantity in complex samples (Silva *et al.*, 2006).

2.9.4.1 Data analysis parameters

The processing parameters, workflow and database were set at the following: The database was downloaded from UniProtKB database (http://www.uniprot.org/) in the FASTA format. The database used was for Homo Sapiens, reviewed, canonical sequence. The processing parameters were set to included low energy threshold set to 150 counts, elevated energy threshold set to 40 and the intensity threshold 1000 counts. The workflow was set to identify protein according to the predetermined rules. The fragment ion per peptide had to be greater than 2, fragment ions detected per protein needed to be at least 5 or more and a minimum of 1 peptide per protein was needed for identification using a 4% FDR. ADH was specified as the internal standard and the concentration (fmol) was specified in the workflow for Hi3 estimation of protein concentration. Modifications were also pre-described to include fixed modification of carbamidomethyl cysteine, two missed trypsin cleavage sites with variable modifications including deamidation of asparagine and glutamine, acetyl N-terminus, and oxidation of methionine.

All protein identifications produced in PLGs for a given sample were exported into Microsoft Office Excel™ where protein identifications from technical replicates were filtered for replication to produce a master list of proteins identified from each plasma sample.

2.10 ELISA

Principally ELISA assays use an enzyme to detect the binding of an antigen to an antibody, which is indicated when a colourless substrate is converted to a coloured product through an enzymatic reaction. Initially a specific coating antibody to an antigen is immobilized onto a plastic surface, the antigen is added (plasma sample) and

the coating antibody captures the antigen. Once bound, a detection antibody labelled with a colourless substrate is added, this binds specifically to another antigen site. Excess unbound labelled detection antibody is washed away and an enzymatic reaction of the bound antibody-antigen-antibody complex generates a colorimetric signal. In this study, an immunoluminometric assay (ILMA) was used for verification of proteins, which differs from ELISA only in the final step, where detection utilised flash chemiluminescence of methyl-acridinium ester labelled streptavidin.

2.10.1 Chemiluminescence

In this thesis Methyl Acridinium Ester (MAE) was the chemiluminescent label used for luminescence produced by the biotin-streptavidin reaction. MAE produces light under alkaline conditions with the use of dilute hydrogen peroxide (H_2O_2).

2.10.2 Biotin – Streptavidin Interaction

In this study immunoassays, biotinylated antibodies were used. Streptavidin was labelled with MAE, and biotin – streptavidin interaction linked the antigen and signalling molecule, thereby allowing the detection and quantification of antigen in the sample measured in relative light units (RLU).

2.10.3 Sandwich ILMA

A non-competitive ILMA also known as a sandwich ELISA was designed specifically for the detection of Cystatin D (CYS D), Insulin like growth factor 2 (IGF2) and Serum amyloid P component (SAP). The general procedure for the sandwich assays is detailed below, with specific antibody and concentration is specified in Table 2.3.

96-well ELISA plate were coated with a coating antibody at 100 µl made up in PBS (8 mmol/L NaH_2PO_4, 1.2 mmol/L KH_2PO_4, 2.7 mmol/L KCl and 137 mmol/L NaCl) and were left to coat on a shaker at room temperature overnight. The plates were washed with PBS and blocked with 200 µl of 0.5% BSA for 2 hours at room temperature. The plate was washed in wash buffer (1.5 mmol/L Na_2HPO_4, 8 mmol/L NaH_2PO_4, 340 mmol/L NaCl, 0.5 g/L Tween and 0.1 g/L sodium azide) and a standard curve was constructed using a specific antigen (recombinant human protein) of varying concentrations (Table 2.4) made up in 100 µl ILMA (1.5 mmol/L Na_2HPO_4, 8 mmol/L

NaH_2PO_4, 140 mmol/L NaCl, 1.0 g/L BSA, 1.0 mmol/L EDTA, Triton X100 0.1% and 0.1 g/L sodium azide). To sample wells 100 µl of ILMA buffer was added with chosen plasma volume (Table 2.4) the plate was left on the shaker at room temperature overnight. The plate was washed in wash buffer and the detection antibody was added at 10 ng µl in 100 µl ILMA. The plate was left at room temperature for 4 hours.

The plate was washed in wash buffer and 100 µl of 2:1000 MAE – labelled streptavidin in ILMA buffer was pipetted in each well and left at room temperature to incubate for 1.5 hours in the dark. After washing in wash buffer the plate was read in the Dynex MLX Luminometer (Virginia, USA). The plate was read using 100 µl of 0.1 M nitric acid (HNO_3) with 0.005% hydrogen peroxide (H_2O_2) then 100 µl of 250 mmol/L sodium hydroxide (NaOH) followed by 0.25% cetyl trimethylammonium bromide (CTAB) 4 seconds later. The chemiluminescence was measured in the subsequent 2 seconds and the measurement units were relative light units (RLU).

2.10.3.1 Biotinylation of serum amyloid P component antibody

The detection antibody requires biotin to be attached for the streptavidin to react, for the serum amyloid P component detection antibody the biotinylation was performed separately. The monoclonal mouse IgG anti-serum amyloid P antibody (Chemicon, Millipore, UK) was biotinylated using NHS-biotin (Pierce antibodies, Thermo Scientific, UK). The NHS biotin was added to the antibody in 20 molar excess and mixed for 15 mins. The reaction was quenched with the addition of 100 µl lysine for 1 hour and kept at 4 °C. The biotinylated antibody was filtered using a 30 mwco spin filter (Millipore, UK) at 14000g for 30 min, the filtration step was repeated three times. The supernatant was collected and this was the in house biotinylated monoclonal mouse IgG anti-serum amyloid P antibody, as used in the above serum amyloid P component method.

Table 2.3: Coating and detection antibody source and concentration for sandwich ELISA assays

Protein name	Coating antibody	Coating antibody concentration (per well in PBS)	Detecting antibody	Detecting antibody concentration (per well in ILMA)
Cystatin D	Human cystatin D anti-IgG mouse monoclonal antibody (R&D systems, UK)	100 ng	Polyclonal goat IgG human cystatin D biotinylated antibody (R&D systems, UK)	10 ng
Insulin like growth factor 2	Human IGF - II mouse monoclonal IgG antibody (R&D systems, UK)	100 ng	Polyclonal goat IgG human IGF - II biotinylated antibody (R&D systems, UK)	10 ng
Serum amyloid P component	Anti-serum amyloid P IgG mouse monoclonal antibody (Millipore, UK)	100 ng	In house biotinylated monoclonal mouse IgG anti-serum amyloid P antibody (Chemicon, Millipore, UK)	10 ng

Table 2.4: Antigen, standard curve dilution concentration and plasma sample dilution concentrations for sandwich ILMA assays

Protein name	Recombinant protein name	Standard curve	Plasma sample dilution (in 100 µl ILMA)
Cystatin D	Recombinant human cystatin D (R&D systems, UK)	100 pg – 0 pg	10 µl plasma in 100 µl ILMA
Insulin like growth factor 2	Recombinant human IGF - II (R&D systems, UK)	500 pg – 0 pg	50 µl plasma to 100 µl ILMA
Serum amyloid P component	serum amyloid P component antigen (Calbiochem, EMD millipore, UK)	100 pg – 0 pg	1 in 100,000,000 dilution of plasma in ILMA

2.10.4 Indirect ILMA

An indirect ILMA was designed specifically for the detection of Cholesteryl ester transfer protein (CETP), Mannan binding lectin serine protease 2 (MASP2), Phospholipid transfer protein (PLTP), 15-Hydroxyprostaglandin dehydrogenase (15 PGDH).

96-well ELISA plates were coated with the antigen using plasma samples, plasma samples was diluted 1 in 20,000 in PBS and coated onto the plate. The standard curve was constructed using an antigen containing pooled heart failure plasma sample at various plasma dilutions (1 in 20,000; 1 in 10,000; 1 in 5000; 0) made up in PBS, 100 µl of each dilution was added to the standard curve wells. The plate was left on the shaker at room temperature overnight. The plates were washed with PBS and blocked with 200 µl of 0.5% BSA for 2 hours at room temperature. The plates were washed in wash buffer and the detection antibody (Table 2.5) was added at 10 ng/µl in 100 µl ILMA. The plate was left on the shaker at room temperature overnight.

The plates were washed in wash buffer to remove any unbound detection antibody. As the detection antibodies for MASP2, PLTP and CETP were not biotinylated, a biotinylated anti-rabbit IgG was necessary to detect antibody bound specifically onto the ELISA plate. Biotinylated anti-rabbit (Rockland, Inc., USA) was used at 1:100000 made in ILMA, 100 µl was added to each well, and plate was incubated for 1 hour at room temperature. This step was not needed for 15-PGDH, since that antibody was biotinylated.

The plates were washed in wash buffer and 100 µl of 2:1000 MAE - labelled streptavidin in ILMA buffer was pipetted in each well and left at room temperature to incubate for 1.5 hours in the dark. After washing in wash buffer the plate was read in the Dynex MLX Luminometer (Virginia, USA). The plate was read using 100 µl of 0.1 M nitric acid (HNO_3) with 0.005% hydrogen peroxide (H_2O_2) then 100 µl of 250 mmol/L sodium hydroxide (NaOH) followed by 0.25% cetyl trimethylammonium bromide (CTAB) 4 seconds later. The chemiluminescence was measured in the subsequent 2 seconds and the measurement units were relative light units (RLU).

Table 2.5: Detection antibody source and concentration for the Indirect ELISA assays

Protein name	Detecting antibody	Detecting antibody concentration (per well in ILMA)
Cholesteryl ester transfer protein	Polyclonal rabbit IgG CETP antibody (BIOSS, USA)	10 ng
Mannan binding lectin serine protease 2	Polyclonal rabbit IgG human MASP2 antibody (Abnova, Taiwan)	10 ng
Phospholipid transfer protein	Polyclonal rabbit IgG PLTP antibody (BIOSS, USA)	10 ng
15-Hydroxyprostaglandin dehydrogenase	Polyclonal rabbit IgG 15-PGDH biotinylated antibody (Pierce antibodies, Thermo Scientific, UK)	10 ng

Chapter Three

Method Development

3.1 Introduction

Discovery proteomic studies require a reliable and reproducible workflow, which can confidently identify biomarkers of disease. The three main stages to a proteomic workflow are sample preparation, sample analysis and data analysis (Figure 3.1). This chapter focuses on the development in the first stage, sample preparation. The role of sample preparation and development is three fold, firstly to make samples amenable to mass spectrometry analysis and detection. Secondly, improve the observed dynamic range of a sample and thirdly, ensure samples are reliably prepared and analysed to enable meaningful cross sample comparisons to be made.

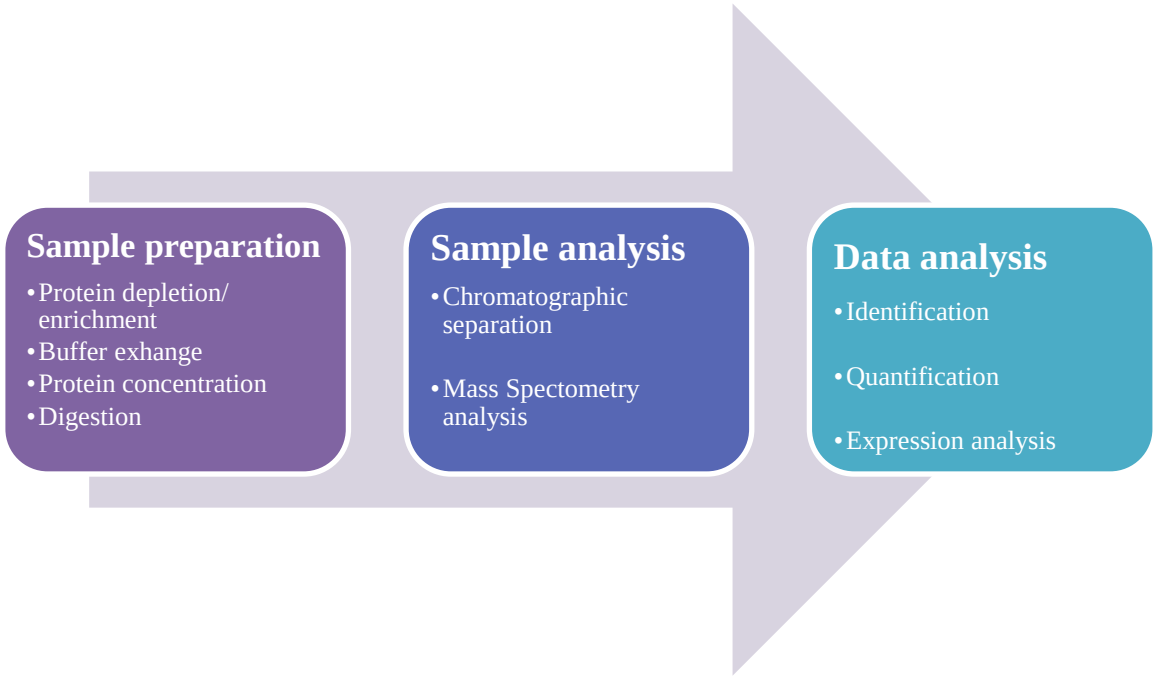


Figure 3.1: Three principle stages in proteomic workflows.

There are many variations to the methods used at each of the stage in a proteomic workflow to identify biomarkers from plasma. In heart failure proteomics, sTWEAK (soluble form of TNF weak inducer of apoptosis) was identified as biomarker using 2D electrophoresis with MALDI-TOF (Blanco-Colio *et al.*, 2009). De Souza *et al.* used dog heart tissue protein extracts and conduced 2D-DIGE-LC-IT-IT experiments and found up-regulation in cardio-protective heat shock protein (De Souza *et al.*, 2010). Watson *et al.* performed an experiment on heart failure serum using 2D-DIGE with nano-LC-Orbitrap and found elevated levels of leucine rich alpha-2-glycoprotein to be indicative

of heart failure (Watson *et al.*, 2011). Most recently, Mebazaa *et al.* identified Quiescin Q6 as a biomarker for acute heart failure in human plasma using metabolically labelled peptides ($^{16}O/^{18}O$) separated by nano-LC and analysed using MALDI-TOF-TOF (Mebazaa *et al.*, 2012).

Many of the techniques used previously involve the use of 2D-gels for protein or peptide separation to identify biomarkers however, limitations of gel-based techniques are widely known. Liquid chromatography is less susceptible to variation, is reproducible, and is able to separate low abundance peptides in complex mixtures (Shi *et al.*, 2004). The most recent development of liquid chromatography systems is nanoLC coupled to mass spectrometers and is commonly used in proteomic analysis (Aebersold & Mann, 2003). Traditionally, nano-LC separated peptides undergo DDA using MS/MS instrumentation. However, further mass spectrometry based development has led to the incorporation of ion mobility within the traditional nano-LC-MS/MS workflow. The incorporation of ion mobility increases the orthogonal separation space for peptides during analysis and provides drift time data for each peptide, increasing confidence in the protein identifications (Valentine *et al.*, 2005). This new ion mobility incorporated MS/MS method is known as $HDMS^E$, this combines DIA acquisition with ion mobility separation. Thus far, there has been no publication on the use of the proposed $HDMS^E$ proteomic workflow with human heart failure plasma. In addition, unlike previous publications, quantification is determined by an absolute label free method termed Hi3 quantitation (Silva *et al.*, 2006).

This chapter introduces a proteomic workflow based on the three stages suggested in figure 3.1. In sample preparation, solid phase extraction, organic solvent precipitation using acetonitrile and equalizer beads were all tested using healthy control human plasma. In the case of organic solvent precipitation and equalizer beads many different variations based on an existing method were made. In sample analysis, the use of different gradient times was evaluated and a comparison of MS^E (DIA) and $HDMS^E$ (DIA with ion mobility) was made.

The aims of this chapter were three-fold:

1. Identify the optimum nano-LC gradient length
2. Compare ion mobility enabled MS analysis with non-ion mobility enabled MS analysis
3. Identify the optimum plasma protein preparation method prior to MS analysis

3.2 Materials and Methods

3.2.1 Materials

Healthy donor plasma was supplied by from the NHS New and Emerging applications of technology funded study *"Screening for heart failure"* (for sample collection protocol see appendix). The samples were centrifuged and plasma was separated from the red blood cells and buffy layer. For all method development work, a superpool was created of ten healthy plasma samples. This superpool sample was used to test all methods and allow direct comparisons to be made between methods.

ProteoMiner^{TM} Protein Enrichment Kit (Bio-Rad, Hemel Hempstead) was used which contained PBS wash buffer and 1 mL equalizer bead columns. Two SPE columns were used, C18 (Sep Pak C18 3cc, Waters, Manchester) and EWP (C18-EWP Bond Elute 3 mL, Varian, USA). Acetonitrile solvent was used for all organic solvent precipitation (HPLC grade, Sigma-Aldrich Company Ltd, UK). Digestion reagents dithiothreitol, iodoacetamide and proteomics grade trypsin were supplied by Sigma Aldrich (Poole, UK). Mass spectrometry standards Glu^{1}-fibrinoginogen peptide, alcohol dehydrogenase and enolase were purchase from Waters (Manchester, UK). Mass spectrometry solvents were all purchased at optima LC-MS grade from Fisher Scientific Ltd (Loughborough, UK).

3.2.2 Sample preparation

3.2.2.1 Solid phase extraction: C18 and C18-EWP

This study exploited the hydrophobicity separation for separating proteins in a plasma sample. Acidified plasma (100 µl of plasma) from control patient samples was separated using SPE, the elutions were collected and protein concentration was estimated by BCA. Once the protein concentration was determined the samples were tryptically digested and analysed using LC-MS^{E} analysis. An internal standard, in this

case ADH, was included at 50 fmol/ µl (at a 1:1 ratio) for estimation of protein concentration for each sample injection. The protein concentration for each sample was determined once the sample was processed using PLGS software.

3.2.2.2 Organic solvent precipitation

A 20 µl healthy control plasma was added to 40 µl deionised water, ACN was added and the mix was vortexted and sonicated twice. The sample was centrifuged and the supernatant extracted, evaporated to dryness and lyophilised. Samples were reconstituted in either 50 mM ammonium bicarbonate or urea, digested and stored at -80 °C prior to MS analysis. In the case of the reconstituted pellet, the pellet was retained rather than the supernatant. The sample was evaporated to dryness and lyophilised overnight. The samples were stored at -80 °C prior to MS analysis.

3.2.2.3 Equalizer beads

Each plasma sample was thawed from -80 °C at room temperature and vortexed to ensure homogeneity, 1 mL plasma sample was incubated with the equalizer beads and eluted with various elutions reagents. These elution reagents were: trypsin, urea, organic solvent and 2% SDC (see chapter 2 for full details). The elutions were digested and lyophilised and stored at -80 °C prior to MS analysis.

Additionally, plasma was pre-treated with acetonitrile precipitation prior to equalizer bead incubation and eluted with trypsin. Plasma was also digested prior to equalizer bead incubation and eluted with organic solvent.

3.2.3 Data Acquisition

All samples were run using the following methods and equipment unless otherwise specified; nanoAcquity UPLC coupled to an ESI source, which is interfaced to a Synapt G2 mass spectrometer (Waters Ltd., Manchester, UK). All samples were run using (HD)MS^E MS acquisition method, all data produced was analysed and proteins assigned and quantified using PLGS version v2.5.2.

3.2.4 Processing of MS^E data

$LC\text{-}MS^E$ data produced was processed using PLGS v2.5.2. Lockspray (GFP) data was collected during the $LC\text{-}MS^E$ runs through a reference spray. The protein identification and assignment process has been described in detail in Chapter two section 2.9.4. Precursor and product ions were time aligned and various data interrogation steps took place to ensure protein identifications were confidently assigned (Li *et al.*, 2009). Data was processed using the following processing parameters; 15, 80, 750 for the low, elevated and intensity threshold respectively.

A Uniprot (http://www.uniprot.org/) human reviewed database was used and amended to include ADH (yeast). ADH was specified as the internal standard with the concentration in fmol, for Hi3 quantification. Protein identification criteria was set to determine detection if the following criteria was met, at least two fragment ions per peptide, five fragment ions per protein and a minimum of one peptide per protein and the false discovery rate was set at 4%.

Fixed modification of carbamidomethyl cysteine and one missed tryptic cleavage site with variable modifications of oxidation of methionine, acetyl N-terminus, deamidation of asparagines and glutamine were all included in the search parameters of the MS^E workflow. All identifications made by PLGS were exported to Microsoft Excel, where further data interrogation and filtering of data to remove duplicates took place.

3.3 Results

3.3.1 Solid phase extraction method

Two SPE methods were tested, these were C18 (Sep Pak C18) and EWP (C18 - EWP Bond Elute). In this study, it was found that both the C18 and EWP method were unable to separate plasma samples enough for in-depth proteomic analysis and protein profiling. For the C18 method technical replicates, 27, 24, and 22 proteins were identified and the protein concentration reported by PLGs analysis was 500 ng ± 20% for each replicate. For the EWP technical replicates, 39, 34 and 35 proteins were identified and protein concentration reported by PLGs analysis was 500 ng ± 34% (Figure 3.2).

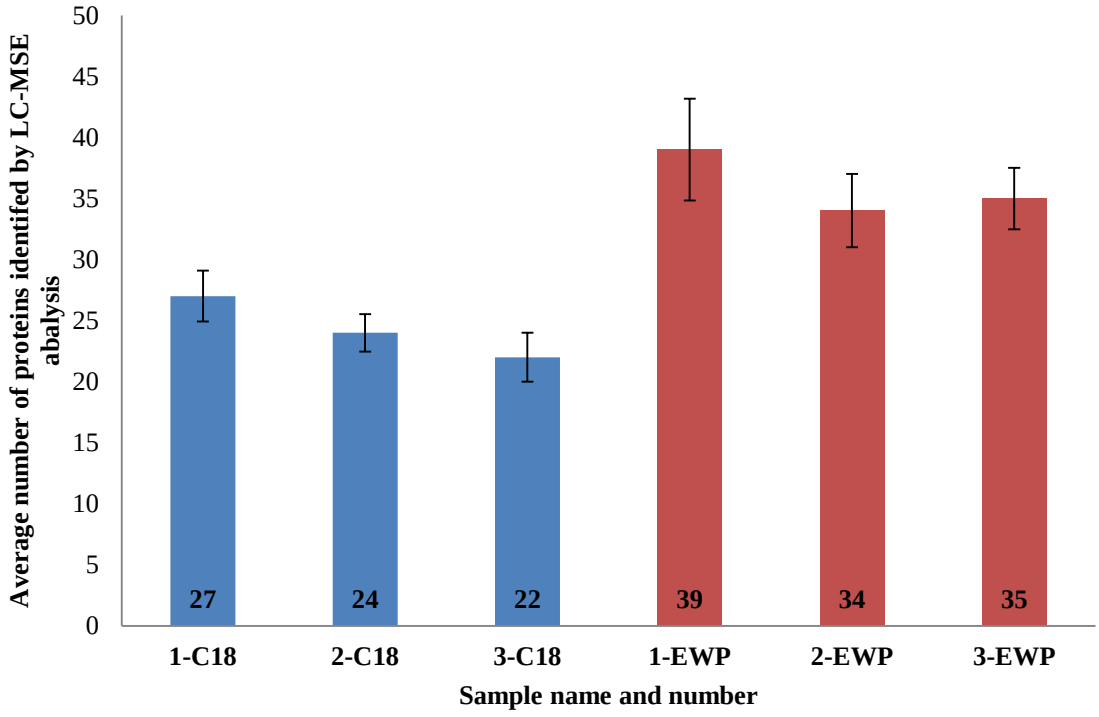


Figure 3.2: This graph displays the average number of proteins identified by both the C18 and EWP method. The **BLUE** columns indicate the C18 solid phase extraction method and the **RED** columns indicate the EWP solid phase extraction method. Three samples of superpool control plasma were used for each sample run and each sample was run in triplicate and error bars indicate the standard deviation of protein identifications indicating the reproducibility of the method.

Further analysis of the protein lists generated by PLGS v2.5.2 for both SPE methods found that the protein lists consisted mostly of the same proteins and was still

dominated by the presence of high abundance proteins such as, serum albumin, apolipoprotein A1, fibrinogen alpha and beta chain and serotransferrin. These five proteins were also the most dominant proteins in the LC-MS^{E} analysis as they ranked amongst the highest proteins with respect to peptide coverage and sequence coverage. For example, serum albumin had a sequence coverage of >70% for both SPE methods. The number of peptides identified for each of these proteins varied between SPE methods (Figure 3.3). The EWP method identified more peptides for each of the high abundance proteins than the C18 method.

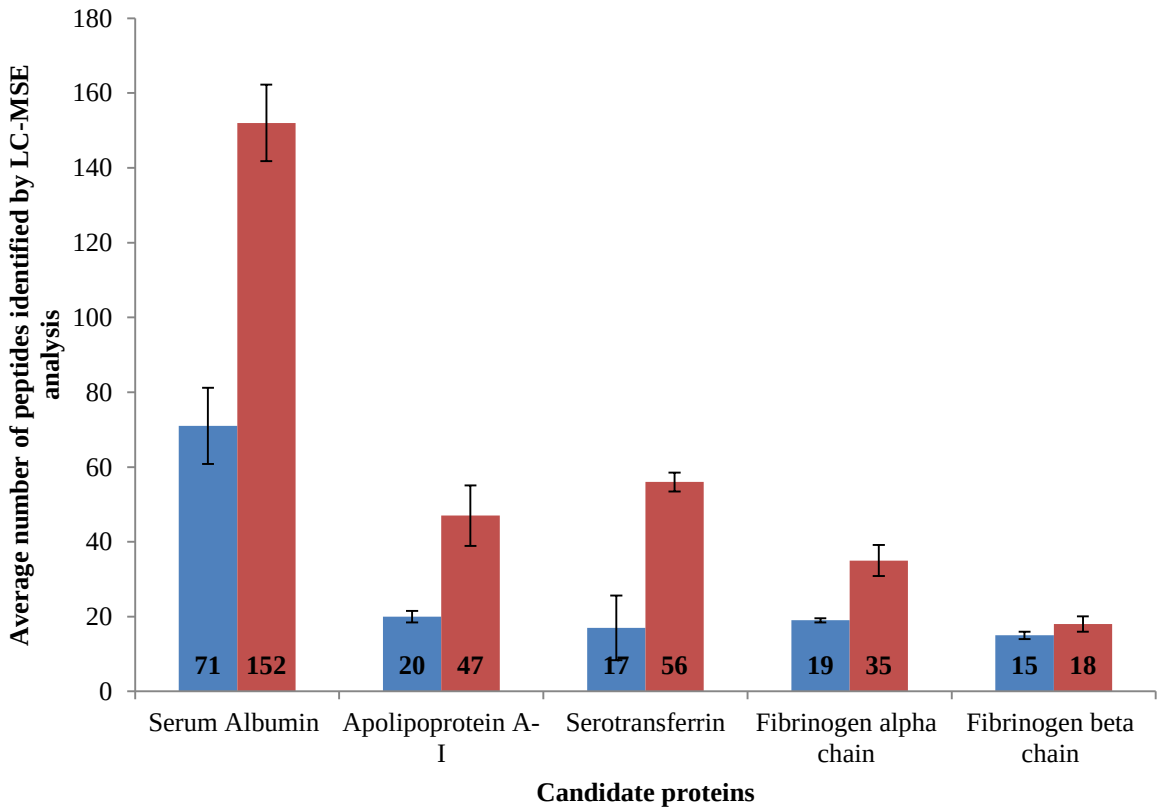


Figure 3.3: This graph displays the average number of peptides identified by both the C18 and EWP method for the five most abundant proteins measured and identified by each method. The **BLUE** columns indicate the C18 solid phase extraction method and the **RED** columns indicate the EWP solid phase extraction method. Three samples of superpool control plasma were used for each sample run and each sample was run in triplicate and error bars indicate the standard deviation of peptide identifications indicating the reproducibility of the method. As the graph shows the number of peptides identified by the EWP method is much greater than the C18 method for the top five most dominant proteins.

The results from both the C18 and EWP methods were much lower than had been expected and this can be attributed to the inability of the method to reduce the dynamic range of plasma which spans over nearly 12 orders of magnitude. As the mass spectrometer typically spans 5 orders of magnitude and the nano-UPLC possibly 2 orders of magnitude, the sample from the SPE methods were too complex for MS and PLGS data analysis to successfully process. Literature detailing the plasma proteome identifies 423 proteins that are known to be present reproducibly in plasma using MS^E analysis (Levin *et al.*, 2010). The ability to identify 423 protein using MS^E analysis highlights the importance of overcoming the dynamic range of plasma for proteomic based biomarker discovery using optimised sample preparation.

3.3.2 Organic solvent precipitation method

3.3.2.1 Acetonitrile precipitation

The acetonitrile precipitation method for serum proteomic analysis was described by Kay *et al.* 2008 (Kay *et al.*, 2008). In this paper, it was demonstrated that acetonitrile precipitation was, reproducible, and successful in extracting low molecular weight proteins using nano-LC-triple quadrupole analysis of SDS-PAGE extracts. In this study, the acetonitrile precipitation method was tested with human plasma and LC-MS^E analysis. Plasma from control samples was tryptically digested and ADH was added at 50 fmol/ µl for estimation of sample protein concentration from each injection. In addition to testing the acetonitrile precipitation method for plasma proteome profiling, the chromatographic run times were also tested. The reported chromatographic method used in Kay *et al.* lasted a total of 95 min, the chromatographic run times in this study were tested with a 50 min and 110 min runs using nanoAcquity UPLC-MS^E. The different runs times were expected to yield different results in protein numbers.

Across the three technical replicates for the 50 min run an average of 15, 16 and 13 proteins were identified for each sample, and for the 110 min runs and average of 15, 13 and 12 proteins were identified. The average sequence coverage for 50 min runs was 40% and for the 110 min runs 38%. As Table 3.1 shows, the number of proteins identified per sample is below a 1.5 standard deviation of the mean number of proteins for both the 50 min and 110 min method. Figure 3.4 shows the average protein numbers obtained from the three samples were approximately the same for both the 50 min and

110 min chromatographic runs for the acetonitrile precipitation method, with the 50 min run showing a slight advantage in protein numbers. A paired two tailed t-test of the average protein numbers produced by both the 50 and 110 min run times calculated a p value of >0.2. Thus, the use of a shorter or longer chromatographic run time for this sample preparation method offers no apparent advantage.

Table 3.1: The table displays the number of proteins identified for each replicate of three different samples using acetonitrile precipitation. The average number of proteins from replicate is shown for both the 50 min and 110 min runs, along with the standard deviation of the average number of proteins for samples 1 -3.

	50 min run			**110 min run**		
Sample name	**Number of proteins**	**Average number of proteins**	**Standard deviation of the average number of proteins for samples 1 – 3**	**Number of proteins**	**Average number of proteins**	**Standard deviation of the average number of proteins for samples 1 – 3**
Sample 1	18	15	1.53	17	15	1.39
	16			15		
	10			13		
Sample 2	19	16		13	13	
	15			14		
	13			12		
Sample 3	14	13		10	12	
	13			14		
	11			13		

The number of proteins identified by the acetonitrile method was lower than those identified by the C18 and EWP SPE methods. Once again this method was unable to profile the plasma proteome using LC-MS^E technology and overcome the dynamic range of plasma. Further analysis of the individual protein profile generated by each sample, found that high abundance proteins were still mostly being observed in each sample run. The most confidently assigned proteins using PLGS software for the acetonitrile precipitation method were apolipoprotein A1, apolipoprotein A2 and serum albumin. The average number of peptides identified for apolipoprotein A1 was 30,

apolipoprotein A2 was 32 and serum albumin 28. In addition, the average percentage coverage for all three proteins was 41% and above. These proteins in theory should remain in the protein pellet post precipitation. The observation of the high abundance proteins lead to two further variations of this method.

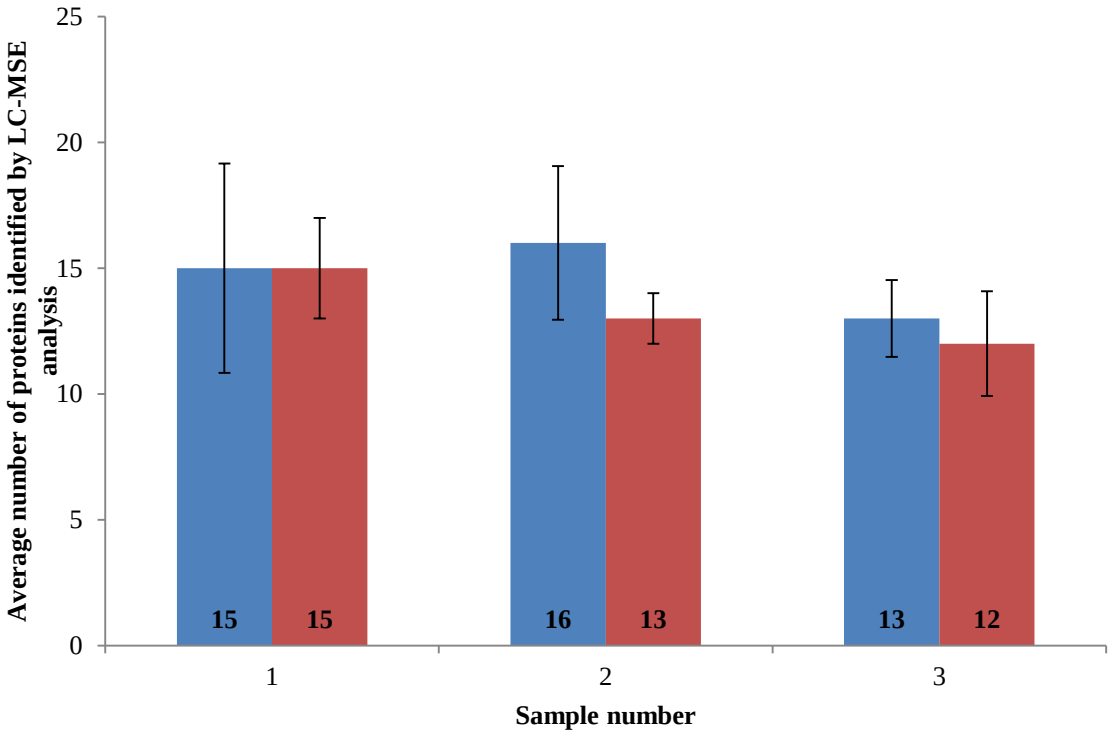


Figure 3.4: This graph displays the average number of proteins identified by the acetonitrile precipitation method. Three samples of superpool control plasma were used for each sample run and each sample was run in triplicate and error bars indicate the standard deviation of protein identifications indicating the reproducibility of the method. The data displayed in **BLUE** is the average number of proteins identified by the 50 min chromatographic method for acetonitrile precipitation. The data displayed in **RED** is the average number of proteins identified by the 110 min chromatographic method for acetonitrile precipitation.

3.3.2.2 Acetonitrile precipitation with urea digestion and with reconstitution of the pellet

Due to the high abundance proteins identified, two approaches for developing the acetonitrile precipitation method were tested using control plasma. The first was the addition of urea to denature proteins to aid tryptic digestion of the sample followed by LC-MS^E analysis. The second method development was the reconstitution of the protein precipitate and LC-MS^E analysis of the precipitated proteins. As the proteins identified by the original acetonitrile method still contained high abundance proteins and so few proteins were identified, it was thought that the protein precipitate may contain proteins of interest.

Plasma samples were prepared using the acetonitrile precipitation method with the addition of urea and acetonitrile precipitation with reconstitution of the pellet. The samples were run using 110 min LC-MS^E runs, 2 µl injection volume was used and all data was processed using PLGS v2.5.2.

Table 3.2: The results for the average number of proteins identified for each sample triplicate for precipitation with urea and reconstitution of the protein pellet.

	Acetonitrile precipitation with urea digestion				**Acetonitrile precipitation with reconstitution of the pellet**			
Sample name	Number of proteins	Average number of proteins	Standard deviation of the triplicate	Standard deviation of the average proteins for samples 1 -3	Number of proteins	Average number of proteins	Standard deviation of the triplicate	Standard deviation of the average proteins for samples 1 -3
Sample 1	13	10	3.06	5.60	5	5	0.58	2.31
	7				4			
	9				5			
Sample 2	14	13	1.15		6	9	3.06	
	12				8			
	14				12			
Sample 3	20	21	1.15		6	5	1.15	
	20				4			
	22				4			

The technical replicates for acetonitrile precipitation with urea digestion produced similar average protein numbers (see Table 3.2) to the original acetonitrile precipitation method. The total average protein numbers profiled by 110 min LC-MS^E analysis for acetonitrile was 13, this was only marginally improved with acetonitrile with urea with an average protein number of 14. However, the average number of proteins identified by the acetonitrile reconstitution of the pellet was a lot lower than the previous methods as shown in Table 3.2. It was also observed that the standard deviation of the two developed acetonitrile methods was higher than that of the original method. Thus, these two developed methods did not succeed in improving the protein profiling of the human plasma proteome.

The best method for the acetonitrile precipitation method was the original as described by Kay *et al.* in 2008. The original method managed to produce a protein profile using LC-MS^E analysis which was reproducible and within the acceptable standard deviation. Although the original acetonitrile precipitation method produced the best overall results for protein profiling in this study, the proteins that dominated were of high abundance and in particular were of apolipoprotein origin which was in keeping with data reported by the original study. Key *et al.* confirmed the identification of 29 of the 57 reported proteins by MRM analysis. In comparison of the average number of protein produce by the original method (57 proteins); in this study, proteins of this number were not achieved. Although similarities of proteins identified can be drawn with the original study, the differences in average number of protein identified cannot be overlooked.

Acetonitrile precipitation, like SPE, failed to provide a method that was successful in producing a comprehensive protein profile of plasma by LC-MS^E analysis. Although the proteins were identified with great confidence with 40% or more coverage for the top proteins, once again the inability to overcome the dominating proteins in plasma is insufficient for biomarker discovery and profiling of low abundance proteins.

3.3.3 Equalizer beads

3.3.3.1 Elution method development

Due to the lack of success in protein profiling using SPE and ACN based methods, an alternative sample preparation method was required. The equalizer beads have been extensively studied for protein enrichment and various method development strategies have been undertaken to produce protein profiles. The best elution methods and type of MS analysis have been studied for a variety of biofluids. Methods using equalizer beads with biofluids such as plasma or serum have been analysed mainly by either 2D-PAGE gel and SELDI-MS (Boschetti & Righetti, 2009; Boschetti *et al.*, 2007; Righetti & Boschetti, 2007; Righetti *et al.*, 2004); Guerrier *et al.*, 2006). Thus, in this study equalizer beads have been used with human plasma samples and analysed by in solution label free nanoUPLC-MS^E analysis in an attempt to create a protein profile characterising the human plasma proteome.

The equaliser bead commercial pack provided an elution reagent however, the elution reagent was for SDS-PAGE analysis and contained CHAPS reagent. CHAPS is highly incompatible with mass spectrometry analysis. Therefore, the initial objective of the equaliser beads method development was to investigate alternative elution reagents that were amenable to label free nanoUPLC-MS^E analysis.

3.3.3.2 Equalizer beads with differential elutions

As an alternative to the CHAPS reagent provided with the equaliser beads, a differential elution method was tested with an acidic urea elution at pH 3.3 followed by a hydro-organic elution using an organic solution. The use of the differential elutions was to ensure maximum protein recovery from the equalizer beads was achieved. This was also the reason the two elution reagents were chosen, as they act on different protein-bead bonds. A study by Righetti *et al.* (Righetti *et al.*, 2006) discussed the use of different elution reagents the have been used with equalizer beads. In this paper, it was noted that acidic urea at pH 3.5 was strong enough to desorb mostly all bound proteins from beads. This was the reason the acidic urea was chosen as the first elution and followed by a hydro-organic elution to recover any protein that may remain on the beads. Additionally, Righetti's paper focused mainly on 2D map analysis of the eluent,

whereas this study used nanoUPLC-MS^E analysis. Thus, elution reagents chosen for this study had to be compatible with MS analysis.

All samples were tryptically digested following elution from the equalizer beads. To ensure all proteins were identified, samples were run as single injections for each elution reagent and protein identifications were combined manually. Each sample contained an internal standard of ADH at 50 fmol/ µl was used at a ratio 1:1 and analysed using 110 min LC-MS^E run, the data was processed using PLGS v2.5.2 and exported to Excel.

The samples were run in triplicate for both the acidic urea elution and the hydro-organic elution. Using Excel, the protein identifications were transferred into a spread sheet and protein information was filtered. The conditional formatting function in Excel highlighted all duplicate protein identifications so these could be removed from the average protein identifications reported. After filtering using Excel the average number of proteins identified by the acidic urea and hydro-organic elutions for samples 1 – 3 are shown in Figure 3.5. In addition to the average number of proteins for each sample per elution method, data for the average number of proteins of combined elutions for samples 1 – 3 is also displayed. The duplicate identifications were removed to ensure no protein identifications were replicated. By combining the entire protein identification data for all samples, a total of 22 protein identifications were made using equalizer beads with the differential elutions with all duplicate identifications removed.

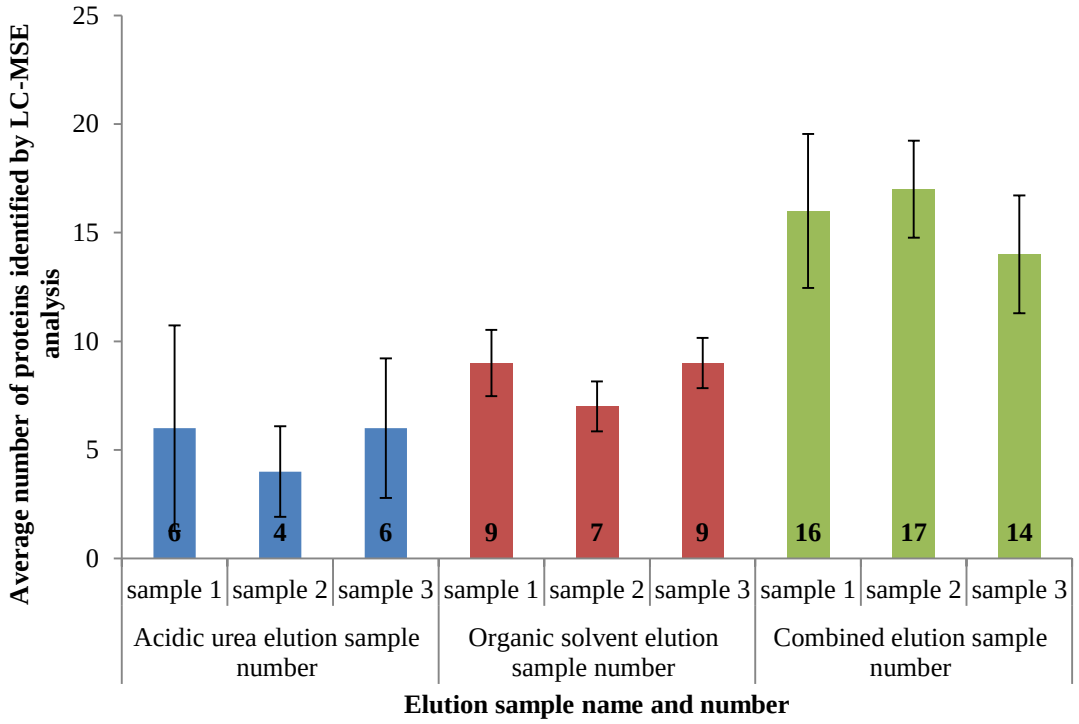


Figure 3.5: This graph displays the average number of proteins identified by the Equalizer beads with differential elutions method. Three samples of superpool control plasma were used for each sample run and each sample was run in triplicate and error bars indicate the standard deviation of protein identifications indicating the reproducibility of the method. The data displayed in **BLUE** is the average number of proteins identified acidic urea elution. The data displayed in **RED** is the average number of proteins identified by hydro-organic elution. The data displayed in **GREEN** is the average number of protein identified by the two combined methods. All duplicate protein identifications were removed.

As the graph of the average number of proteins shows, there is very little protein identification overlap. Sample one identified no overlapping proteins, the acidic urea and hydro-organic elutions produced unique proteins. Sample two showed a single overlapping protein identification, in this case olfactory receptor 5K4. Sample three showed two overlapping protein identifications these were fibrinogen alpha chain and complement factor H-related protein 1. Thus, the two different elution reagents used with the equalizer beads targeted different protein populations.

Although the use of two different elution methods produced different protein identifications and maximised the elution of proteins from the equalizer beads, the total number of proteins identified for the entire method was a lot lower than expected. It has been shown that the equalizer beads are capable of identifying 285 non-redundant

serum proteins using labelled 2D-LC-QTOF analysis from 2D-gels extracts (Dwivedi *et al.*, 2010). Another study by Capriotti *et al.* (Capriotti *et al.*, 2012) found the equalizer beads were able to identify 267 proteins using nano-HPLC-tandem MS with an LTQ Orbitrap using in solution digestion. Although the studies published by Dwivedi *et al.* and Capriotti *et al.* were methodically different, the difference in protein number identified is difficult to overlook therefore, further elution methods needed to be considered.

3.3.3.3 Equalizer beads with peptide elution

The equalizer beads have attached to them a library of combinatorial peptides which bind to the proteins in a sample. The combinatorial peptides consist of amino acid chains; it was hypothesised that so long as there were peptides or amino acids within a sample that were able to bind to the ligands, the mechanism of equalization would remain. Therefore, the objective of the pre-treated plasma study was to test the elution of peptides from the equalizer beads using the best eluent from the differential elution method, in this case organic solvent.

The sample was analysed using 110 min run in triplicate. All samples underwent nanoUPLC-MS^E analysis and the data was processed using PLGS v2.5.2 and exported to Excel.

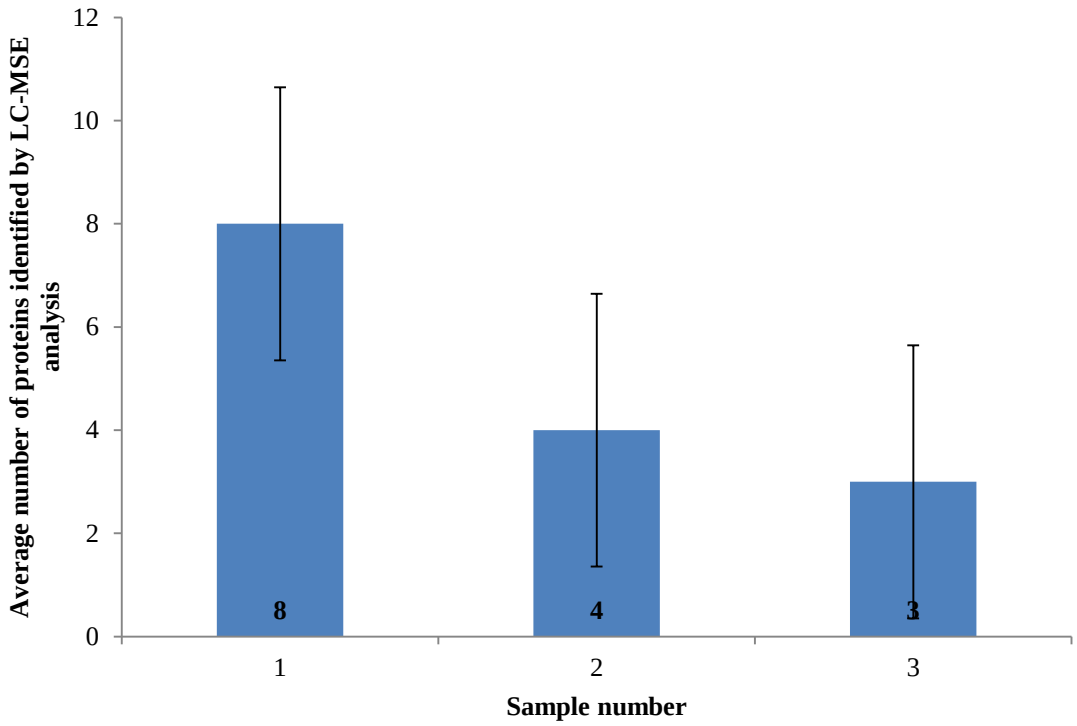


Figure 3.6: This graph displays the average number of proteins identified by the Equalizer beads with peptides method. Three samples of superpool control plasma were used for each sample run and each sample was run in triplicate and error bars indicate the standard deviation of protein identifications indicating the reproducibility of the method.

As shown is Figure 3.6, the highest number of proteins identified for this method was 8. Using Excel, the protein identifications were transferred into a spread sheet and protein information was filtered. The conditional formatting function in Excel highlighted all duplicate protein identifications so that these could be removed from the average protein identifications reported. Using this function, the total number of proteins identified was 10 and the standard deviation for this method was 2.6. The number of proteins identified was too low for proteomic profiling of plasma for biomarker studies upon further analysis it was observed that the most commonly identified protein was serum albumin which had the greatest sequence coverage and approximately 50% more sequence coverage than the second most highly sequenced protein. This method was unsuitable for plasma proteome profiling.

3.3.3.4 Equalizer beads with trypsin

The equalizer beads are attached to combinatorial peptide library ligands and bind protein through amino acid interaction. The addition of trypsin to the bound protein as an elution method has been suggested previously (Righetti *et al.*, 2010). The advantage of using trypsin is twofold, no addition of an elution reagent is required and the supernatant of the centrifuged peptides can be used directly for proteomic analysis with mass spectrometry, post acidification. Due to the advantages of digesting the equalized proteins on the bead, this method was tested with various permutations.

3.3.3.4.1 Equalizer beads with trypsin elution

The first experiment chosen for the equalizer beads with trypsin involved directly reducing, alkylating and digesting proteins bound to the beads. The protein peptides were eluted post-overnight digestion through centrifugation.

All samples were tryptically digested following elution from the equalizer beads, each sample contained an internal standard of ADH at 50 fmol/ µl was used at a ratio of 1:1 and was analysed using 50 min and 110 min runs for all three samples and their replicates. All samples underwent nanoUPLC-MS^E analysis and the data was processed using PLGS v2.5.2 and exported to Excel. Using Excel, the protein identifications were transferred into a spreadsheet and protein information was filtered. The conditional formatting function in Excel highlighted all duplicate protein identifications so that these could be removed from the average protein identifications reported.

The average number of proteins identified by each sample for the 50 min and 110 min chromatographic runs is shown in Figure 3.7. The total number of proteins profiled by nanoLC-MS^E analysis for this method was determined by using Excel to remove all duplicate data. The 50 min run identified 140 proteins, whereas the 110 min run identified 164 proteins. The average sequence coverage for 50 min runs was 38% and for the 110 min runs 39%. As displayed in Figure 3.7, the standard deviation of the mean number of proteins for both the 50 min and 110 min runs was very high. For the 50 min run the standard deviation was calculated at 19.3 and for 110 min run 24.2. As shown in Figure 3.7 the average protein numbers obtained from the three samples showed a significant difference, especially in samples 1 and 3, for both the 50 min and

110 min chromatographic runs. The equalizer beads with trypsin elution, was significantly better with a longer chromatographic run as the 110 min run showed an advantage in protein numbers. A paired two tailed t-test of the average protein numbers produced by both the 50 and 110 min runs calculated a p value of 0.07.

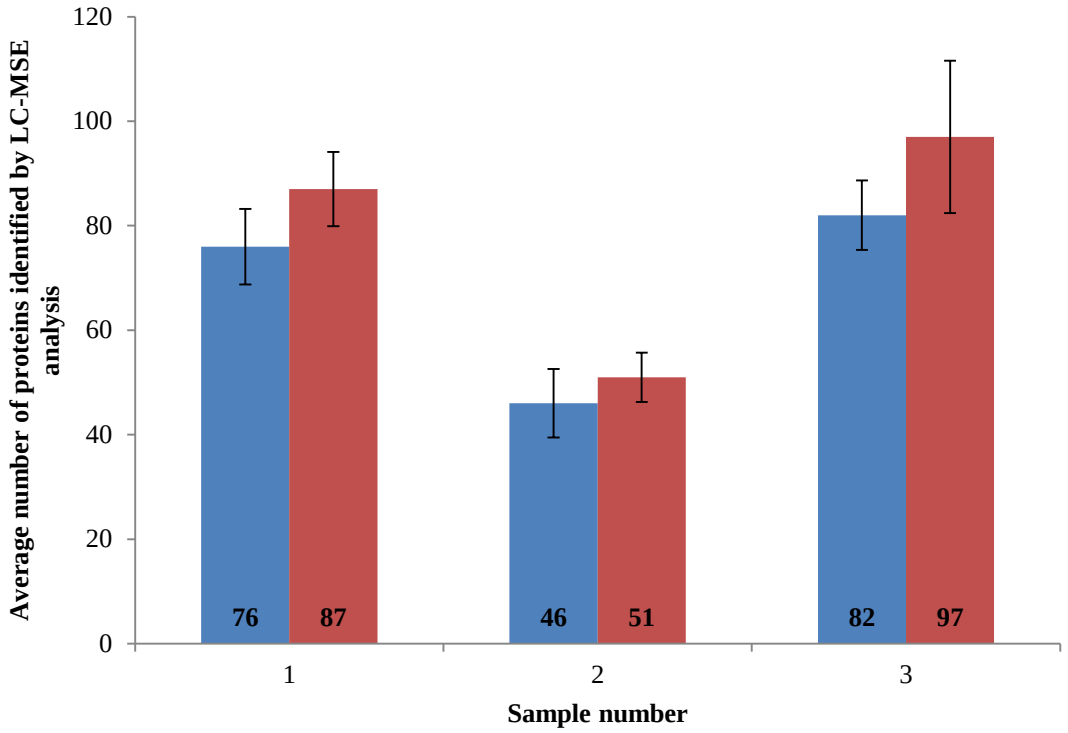


Figure 3.7: This graph displays the average number of proteins identified by the equalizer beads with trypsin elution method. Three samples of superpool control plasma were used for each sample run and each sample was run in triplicate and error bars indicate the standard deviation of protein identifications indicating the reproducibility of the method. The data displayed in **BLUE** is the average number of proteins identified by the 50 min chromatographic run. The data displayed in **RED** is the average number of proteins identified by the 110 min chromatographic run.

Despite the p value falling out of the statistical 0.05 range, it is visually clear from Figure 3.7 that the 110 min run produces a greater number of proteins. Thus, the use of a longer chromatographic run time for this sample preparation method offers advantage in protein identifications, unlike the results seen with the acetonitrile precipitation. Upon further analysis of the 50 min and 110 min runs it was observed that the 110 min run identified an additional 254 peptides than the 50 min run (average peptides identified 50 min run = 1469 and 110 min run = 1723). The average amount of protein

injected per sample for both the 50 min and 110 min run showed a high degree of variation (Figure 3.8). This variation would not allow accurate or reproducible profiling of samples as the amount of peptides (thus, proteins) recovered from the beads could lead to over or under sampling. This method has produced the highest number of proteins thus far and the 110 min is clearly better suited to this method. Thus, all other methods using trypsin elution with the beads shall be run using 110 min chromatographic runs.

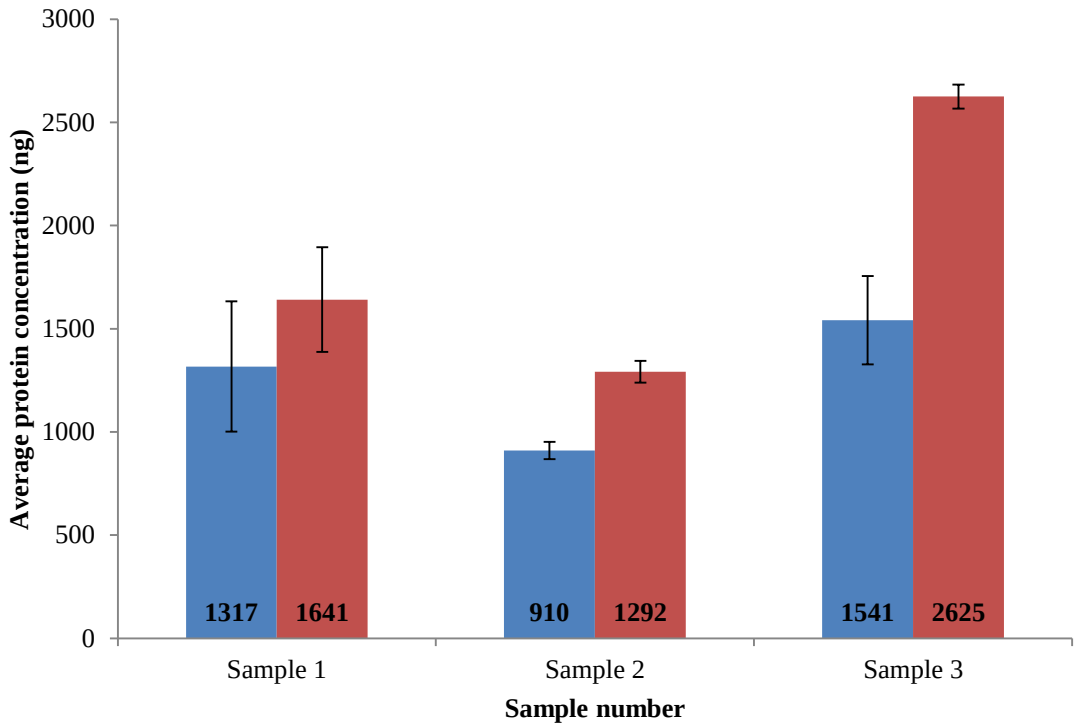


Figure 3.8: This graph displays the average protein concentration (n grams per µl) for sample 1 – 3 using the equalizer beads with trypsin elution method. Three samples of superpool control plasma were used for each sample run and each sample was run in triplicate and error bars indicate the standard deviation of protein concentration indicating the reproducibility of the method. The data displayed in **BLUE** is the average protein concentration for the 50 min chromatographic run. The data displayed in **RED** is the average protein concentration for the 110 min chromatographic run.

3.3.3.4.2 Equalizer beads with trypsin and urea elution

As the results were much improved using trypsin elution and eluting peptides from the column, the addition of urea to aid complete digestion was added to the samples prior to digestion of samples. A control plasma sample was analysed using 110 min nanoUPLC-MS^E run, and all data was processed using PLGS v2.5.2. The number of proteins

identified by the equalizer bead with trypsin and urea elution is summarized (Figure 3.9). The total numbers of proteins profiled by 110 min analysis for this method was determined by using Excel to remove all duplicate data. The 110 min run identified 29 proteins in total. The standard deviation of the mean number of proteins identified was 9.3. This method was more reproducible the original trypsin elution method, however, the standard deviation was still too high for a proteomic method.

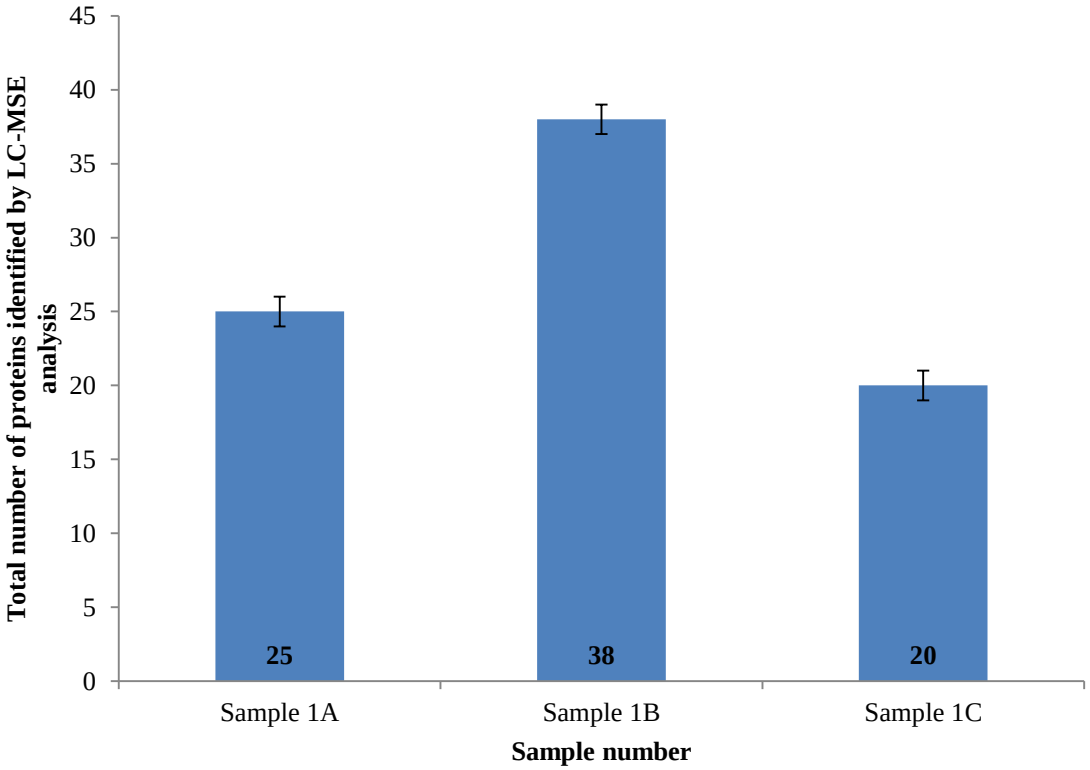


Figure 3.9: This graph displays the total number of proteins identified by the Equalizer beads with trypsin and urea on bead method. N = 1 run in triplicate error bars indicate the standard deviation of protein identifications indicating the reproducibility of the method.

The use of urea with the trypsin on bead method did not improve the number of proteins seen in each plasma sample. Thus, the ability of this method to identify proteins of low abundance where putative biomarkers exist was inadequate for this study. Therefore, the original trypsin elution method remained best for protein profiling but this method produced a more reproducible protein profile with a lower standard deviation.

3.3.3.4.3 Equalizer beads with trypsin elution and overnight agitation

The trypsin elution method produced a more extensive protein profile than the trypsin and urea elution. However, the standard deviation for method utilizing urea was lower than the standard deviation for the trypsin elution. As the protein profiles from the trypsin elution were more comprehensive than the trypsin and urea, this method utilises the trypsin elution method with constant agitation overnight. As the samples were trypsinised with the beads to ensure that the trypsin worked evenly on all bead bound protein, to reduce variation in protein numbers identified, the samples were incubated with trypsin and constantly agitated overnight. Healthy control plasma samples were analysed using 2 µl injections over 110 min nanoLC-MS^E runs, and all data was processed using PLGS v2.5.2.

The number of proteins identified by the equalizer bead with trypsin added to the bead with constant agitation is summarized Figure 3.10. The average total numbers of proteins profiled by 110 min analysis for this method was determined by using Excel to remove all duplicate data. The 110 min run identified 88 proteins in total. The protein amount per injection was 1 µg on column ±79 ng. The standard deviation of the mean number of proteins identified was 7.9. The standard deviation for this method was much improved over the trypsin elution method without agitation therefore, this method produced more reproducible results. In addition, the proteins identified were higher than the trypsin with urea on bead. However, the number of proteins identified with this method was still lower than the trypsin on bead without agitation.

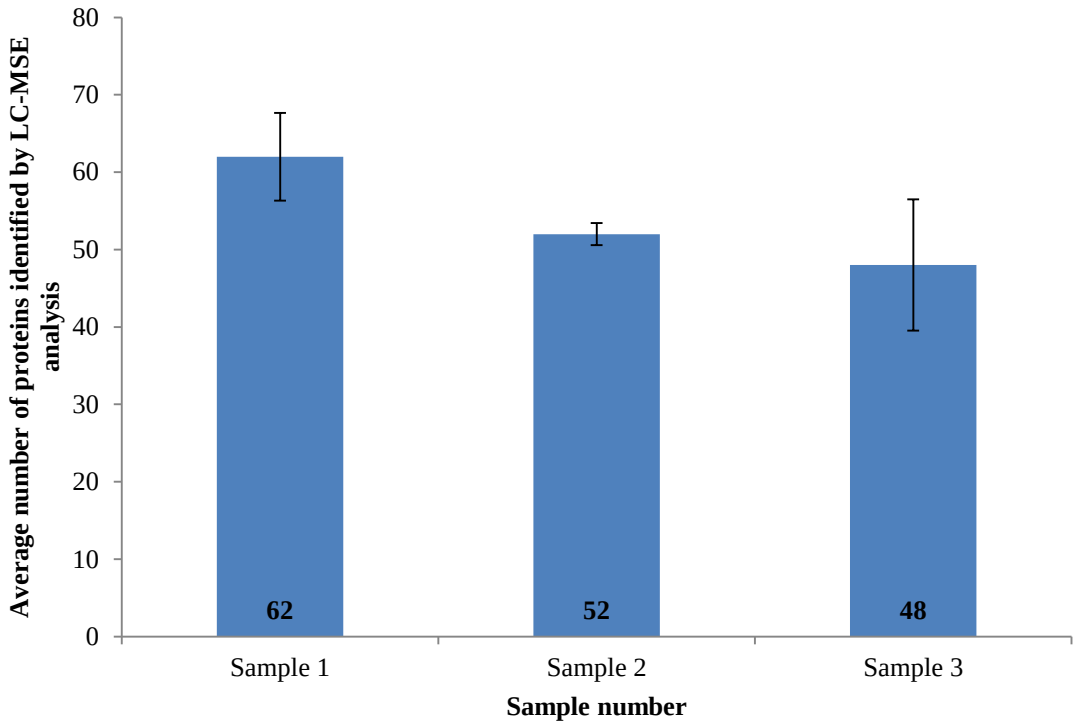


Figure 3.10: This graph displays the average number of proteins identified by the Equalizer beads with trypsin added to the bead and constant agitation method. Three samples of superpool control plasma were used for each sample run and each sample was run in triplicate. Error bars indicate the standard deviation of protein identifications indicating the reproducibility of the method.

3.3.3.4.4 Equalizer beads with acetonitrile depleted plasma and trypsin

Due to the number of proteins identified by the equalizer bead with trypsin on bead method, it was hypothesised that if the plasma proteins were to be pre-depleted and then bound to the beads, mid and lower abundance proteins may be identified more so than the higher abundance proteins. Thus, the objective of this project was to try to implement a change in the protein profile.

Healthy control plasma was taken and depleted with acetonitrile precipitation, the acetonitrile was evaporated off and the remaining pellet was reconstituted in 50 mM ammonium bicarbonate. The depleted plasma was treated with equalizer beads and the sample was digested on the beads, the peptides were eluted through centrifugation. Each sample was analysed using 110 min nanoUPLC-MS^E analysis and the data was processed using PLGS v2.5.2 and exported to Excel. Using Excel the data was filtered

and protein identifications refined to produce a list of proteins which were free from duplicate entries.

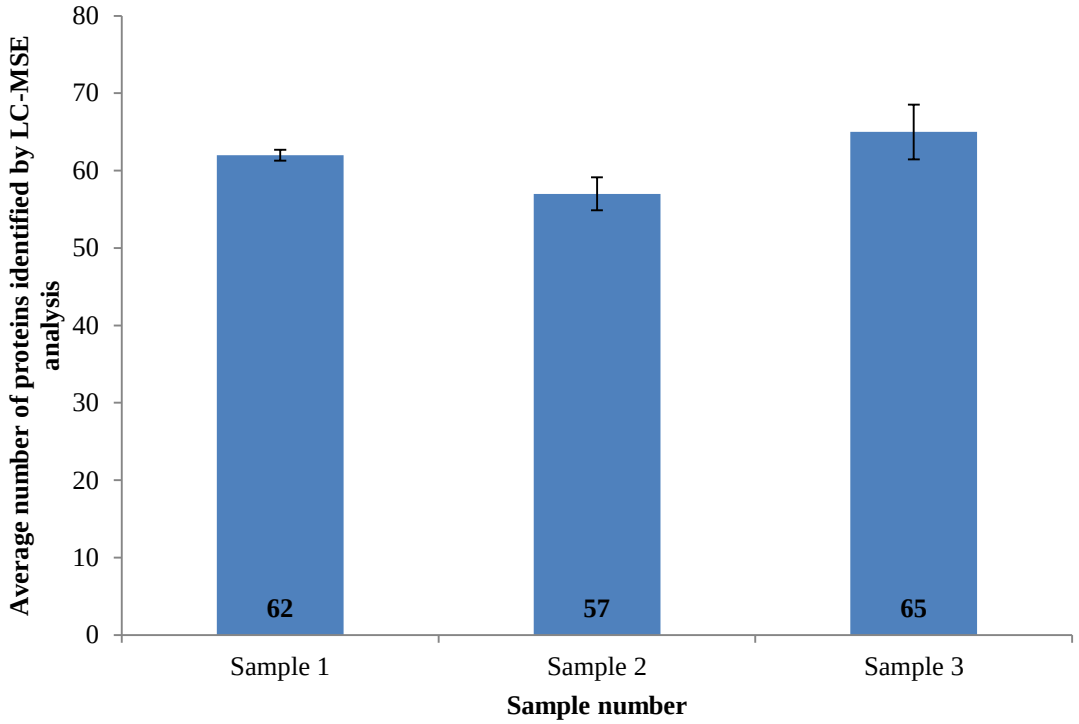


Figure 3.11: This graph displays the average number of proteins identified by the Equalizer beads with acetonitrile depleted plasma. Three samples of superpool control plasma were used for each sample run and each sample was run in triplicate and error bars indicate the standard deviation of protein identifications indicating the reproducibility of the method.

As shown is Figure 3.11, the highest number of proteins identified for this method was 65. Using the conditional formatting function in Excel, all duplicate protein identifications were removed so that these could be removed from the average protein identifications reported. The total number of proteins identified by this method was 91 with a standard deviation of 4.1. This was the best standard deviation identified thus far and serum albumin was now the 8th most commonly identified protein and the sequence coverage for every protein identified was <20%.

Although this method was successful in producing a lower standard deviation, the protein numbers identified were a lot lower than trypsin on bead. Although trypsin has identified the highest number of proteins of all the protein profiling methods tested, the variability in protein profiles was too high. Therefore, an elution method which could

produce reproducible results and improve the number of protein identifications still needed to be identified.

3.3.3.5 Equalizer beads with sodium deoxycholate

The use of sodium dodecyl sulfate (SDS) as an elution reagent for equaliser beads has been well established for 2D-PAGE analysis followed by mass spectrometry (Righetti *et al.*, 2010). However, as this study is an in solution proteomics project, the use of SDS without running a 2D-PAGE gel would require extensive detergent removal strategies to ensure no signal suppression occurred due to the presence of SDS. Signal suppression by SDS has been observed in peptide mass spectrometry, along with reducing trypsin activity (Rundlett & Armstrong, 1996). Therefore, despite effectiveness of SDS as an elution reagent for equalizer bead technology, a detergent that can be removed efficiently and effectively prior to mass spectrometry would be needed to replace SDS.

Sodium deoxycholate is an acid insoluble detergent which has been shown to significantly increase the coverage of tryptic peptides and proteins identified in comparison to SDS (Zhou *et al.*, 2006). The advantage of the acid insolubility is that adding acid prior to mass spectrometry means that the SDC is precipitated out of solution and can be easily removed leaving peptides that are acidified and ready for ionisation and MS analysis. As detergents have been shown to be effective at eluting bead-bound protein from the equalizer beads and SDC is acid insoluble, SDS was replaced with SDC as an elution reagent and evaluated.

All samples contained an internal standard of ADH at 50 fmol/ µl at a ratio 1:1 and analysed using 110 min analysis, the data was processed using PLGS v2.5.2 and exported to Excel. Using Excel, the protein identifications were transferred into a spread sheet and protein information was filtered. The conditional formatting function in Excel highlighted all duplicate protein identifications so these could be removed from the average protein identifications reported. As seen in Figure 3.7, 110 min run produced a greater number of proteins for a richer or more proteinaceous sample. Thus, the use of a longer chromatographic run time for this sample preparation method offered an advantage in protein identifications and was subsequently used.

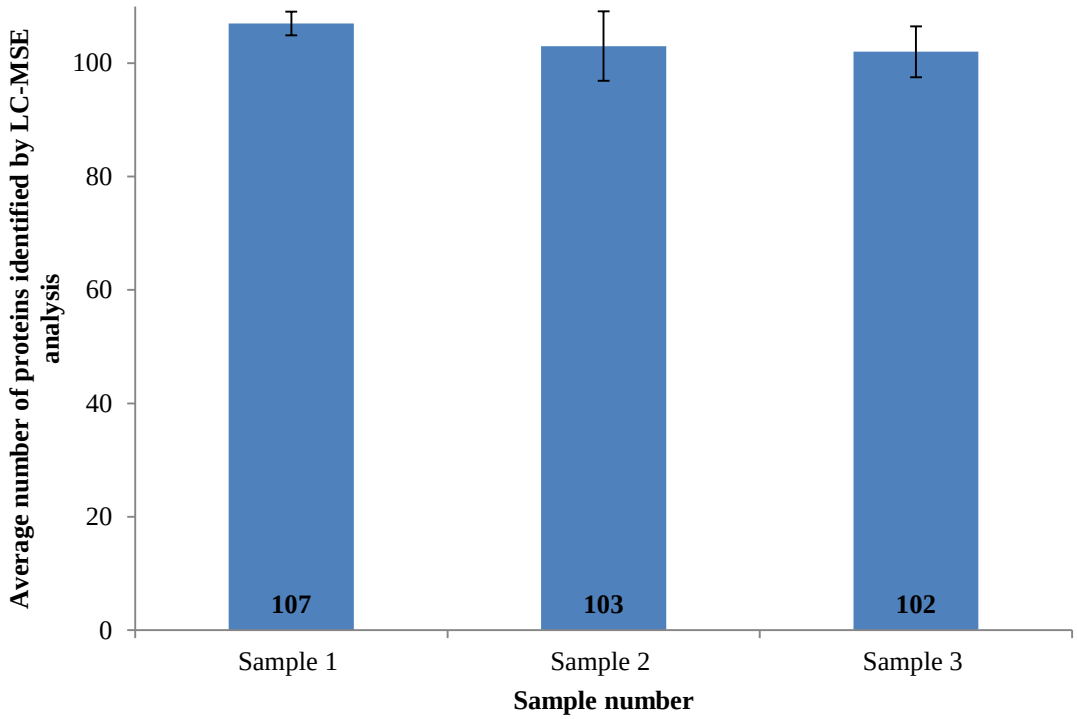


Figure 3.12: This graph displays the average number of proteins identified by the equalizer beads with SDC method. Three samples of superpool control plasma were used for each sample run and each sample was run in triplicate and error bars indicate the standard deviation of protein identifications indicating the reproducibility of the method. The data displayed in **BLUE** is the average number of proteins identified by the 110 min chromatographic run for three samples analysed in triplicate.

The average number of proteins identified by each sample for the 110 min chromatographic runs is shown in Figure 3.12. The total number of proteins profiled by 110 min analysis for this method was determined by using Excel to remove all duplicate data and the total number of proteins identified were 169 (see list in appendix). The average sequence coverage for the 110 min runs was <40%. As displayed in Figure 3.12, the standard deviation of the average number of proteins identified for 110 min runs was 4.7; individually the standard deviation for sample 1 – 3 was 2.1, 6.1 and 4.5 respectively. The average number of peptides identified was 2313, 1975 and 2004 for samples 1 – 3 with a standard deviation of 237.4, 47.4 and 33.6 for samples 1 – 3 respectively. Overall this was the highest peptide identification reported for all methods tested (Figure 3.13).

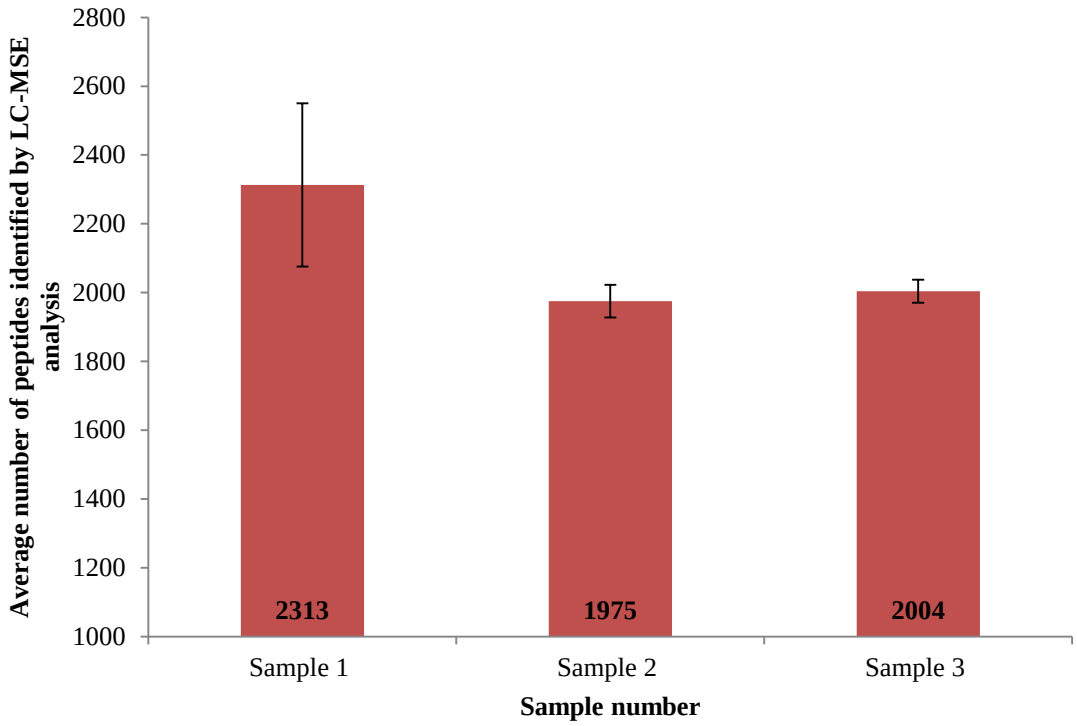


Figure 3.13: This graph displays the average number of peptides identified by the equalizer beads with SDC method. Three samples of superpool control plasma were used for each sample run and each sample was run in triplicate error bars indicate the standard deviation of peptide identifications indicating the reproducibility of the method. The data displayed in **RED** is the average number of peptides identified by the 110 min chromatographic run for three samples analysed in triplicate.

3.3.3.5.1 Reproducibility

Equalizer beads with SDC produced the most extensive protein and peptide profile of all methods tested for proteomic profiling. Therefore, to test SDC elution methods reproducibility a control sample was split into three aliquots and each aliquot was incubated with equalizer beads and eluted with 2% SDC. The samples were analysed using nanoUPLC-MS^E in triplicate over a 110 min LC run and repeated over three days, as summarised in Figure 3.14.

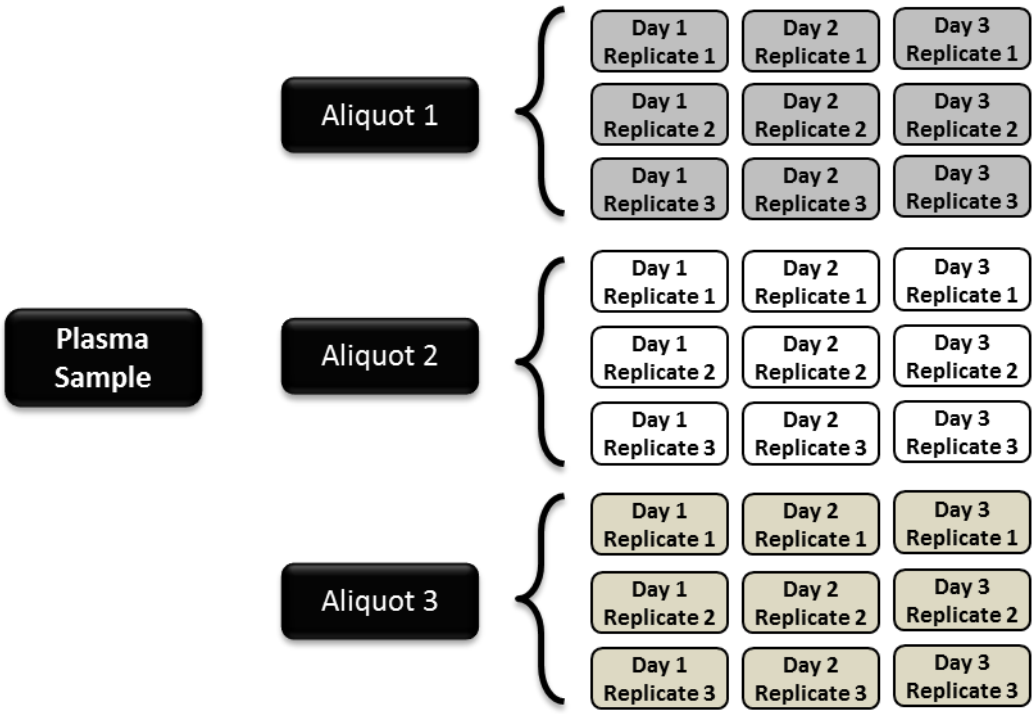


Figure 3.14: A schematic diagram shows the study design to test the reproducibility of equalizer beads with 2% SDC with a single control sample split into three aliquots. All samples were run using a 110 min LC-MS^E run.

Using the study design shown in Figure 3.14, samples 1 – 3 were tested in triplicate and this was repeated over three days. Therefore, each sample was tested a total of 9 times each. The average number of proteins identified by each sample for the 110 min chromatographic runs is shown in Figure 3.15. The total number of proteins profiled by 110 min analysis for this method was determined by using Excel to remove all duplicate data and the total number of proteins identified were 271 (see list in appendix). The average peptides identified for samples 1 – 3 over the three days was 2183 ±178, 1968 ±36 and 1968 ±4.1 respectively. The average on column protein loading per sample was

estimated at 1 µg/µl and using PLGS v2.5.2 the average protein loaded on column was calculated using the ADH top 3 quantification. The average protein on column was 1 µg/ µl ±120 ng.

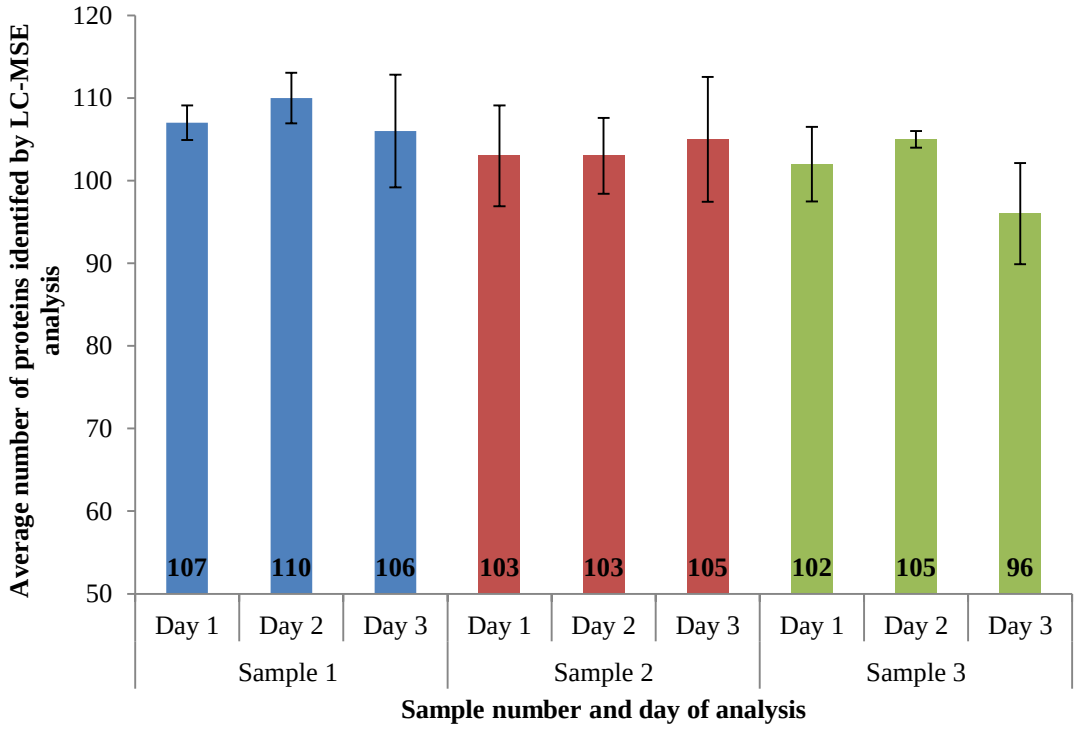


Figure 3.15: This graph displays the average number of proteins identified by the equalizer beads with SDC method to assess the reproducibility of the method. The data displayed in **BLUE** is the average number of proteins identified by the 110 min chromatographic run for three samples analysed in triplicate for sample 1 over day 1 – 3. The data displayed in **RED** is the average number of proteins identified for sample 2 over day 1 – 3. The data displayed in **GREEN** is the average number of proteins identified for sample 3 over day 1 – 3. Error bars indicate the standard deviation of protein identifications indicating the reproducibility of the method.

As mentioned previously the sequence coverage for the 110 min runs remained at < 40%. The standard deviation of the average number of proteins identified for 110 min runs was 4.8 over all samples analysed, individually the standard deviation for samples 1 – 3 over the three days was 4.3, 5.5 and 4.8 respectively. Assessment of the dynamic range of the proteins identified showed that this method was successfully able to analyse proteins over 6 orders of magnitude (Figure 3.16). This is approximately half of the dynamic range of plasma.

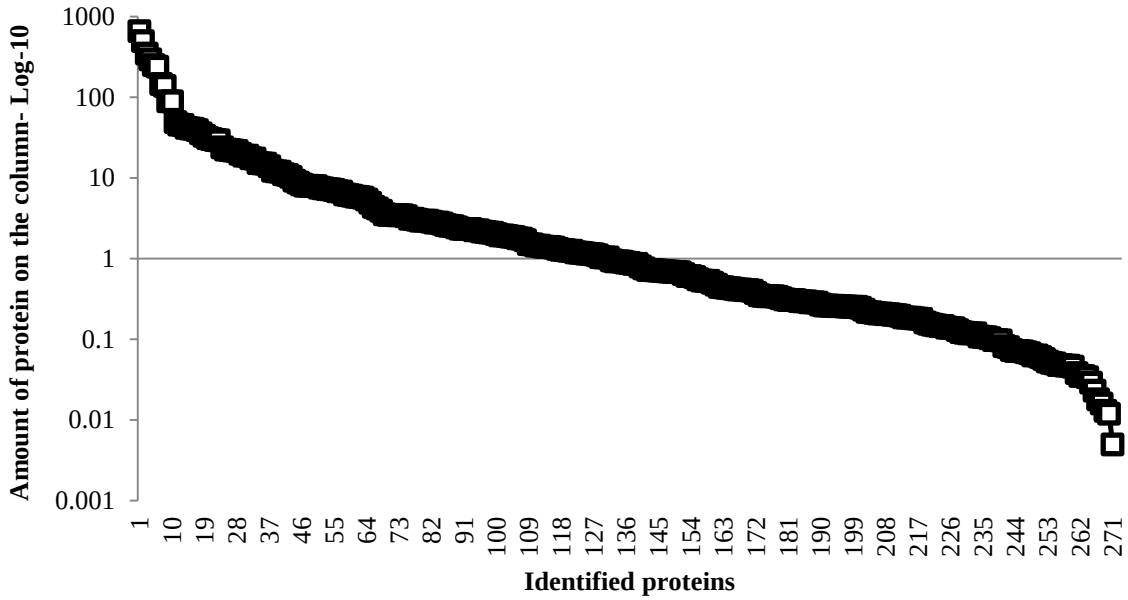


Figure 3.16: This graph shows the protein concentrations on a log scale for the equalizer with SDC method. As the graph shows the protein concentration range spans 6 orders of magnitude.

To assess the use of this method with a LC-MS^E platform for biomarker discovery the reproducibility was measured using various observations. In this case, the three samples were compared. The coefficient of variance for the number or proteins identified over the three days was determined and this on average for samples 1 – 3 over the three days was 4%, 5% and 5% respectively.

For extensive reproducibility evaluation, 18 randomly selected proteins (see appendix) from different plasma concentration ranges were evaluated across the all sample runs. The first form of measurement was the coefficient of variance (CoV) for the protein concentrations calculated from the Hi3 quantitation from PLGS v2.5.2. Eighty nine per cent of protein values measured fall within 10% of the CoV. Over half of the proteins are measured to within 10% CoV and over 80% are within 15% CoV for protein concentration (Hakimi *et al.*, 2013). Figure 3.17 demonstrates for the 18 proteins chosen at random, all 18 fall below the 20% CoV value for a number of samples measured over three days, with a mean average of <3% CoV achieved for this method. The low CoV values calculated for protein concentration highlights the reproducibility of the method.

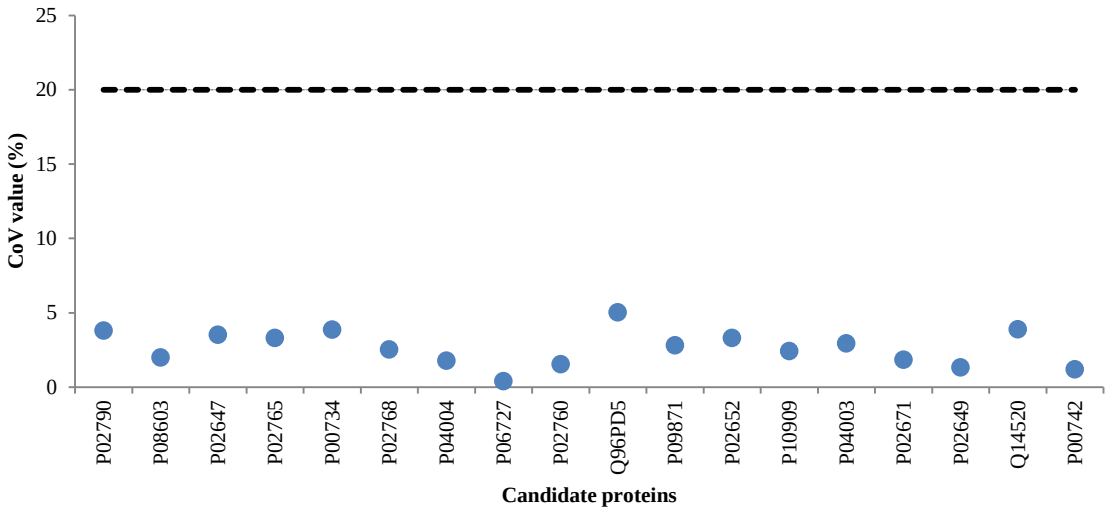


Figure 3.17: The average CoV's of protein concentration calculated by Hi3 quantitation using PLGS v2.5.2 for 18 randomly selected proteins from different plasma concentration ranges of which all of them fall below 20%.

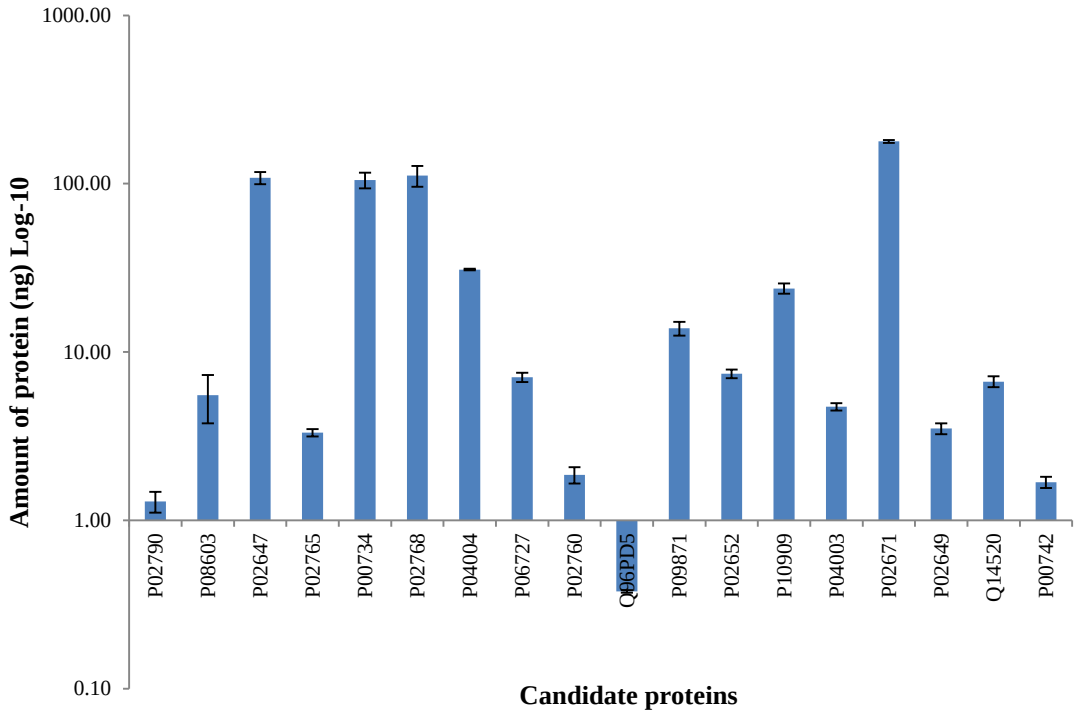


Figure 3.18: The average amount of protein of a sample over 3 days of analysis. The error bars clearly indicates the reproducibility of the analysis.

Figure 3.18 shows the average amount of protein reported from each sample, the error bars indicate the low variance of the measured protein levels. The low variance in the detected and consequently reported levels of protein again confirms along with the findings displayed in Figure 3.18, that the method is highly reproducible and hence suitable for biomarker discovery projects.

3.3.3.5.2 MS^E vs. $HDMS^E$

MS^E allows parallel acquisition of precursor and fragment or product ions. This method as displayed above has been successful in providing comprehensive lists of proteins identified with confidence and reproducibly. The ion mobility cell within the Synapt G2 mass spectrometer (Waters Corp., Manchester) adds the additional drift time dimension, with this additional analysis co-eluting peptides can be separated and proteins can be assigned with more confidence as the additional drift time information is used to assign precursor and product ions along with retention time alignment. In this study, we found $HDMS^E$ is more reproducible as demonstrated by the improved CoVs that is brought about by better assignment of proteins by ion mobility (Figure 3.19).

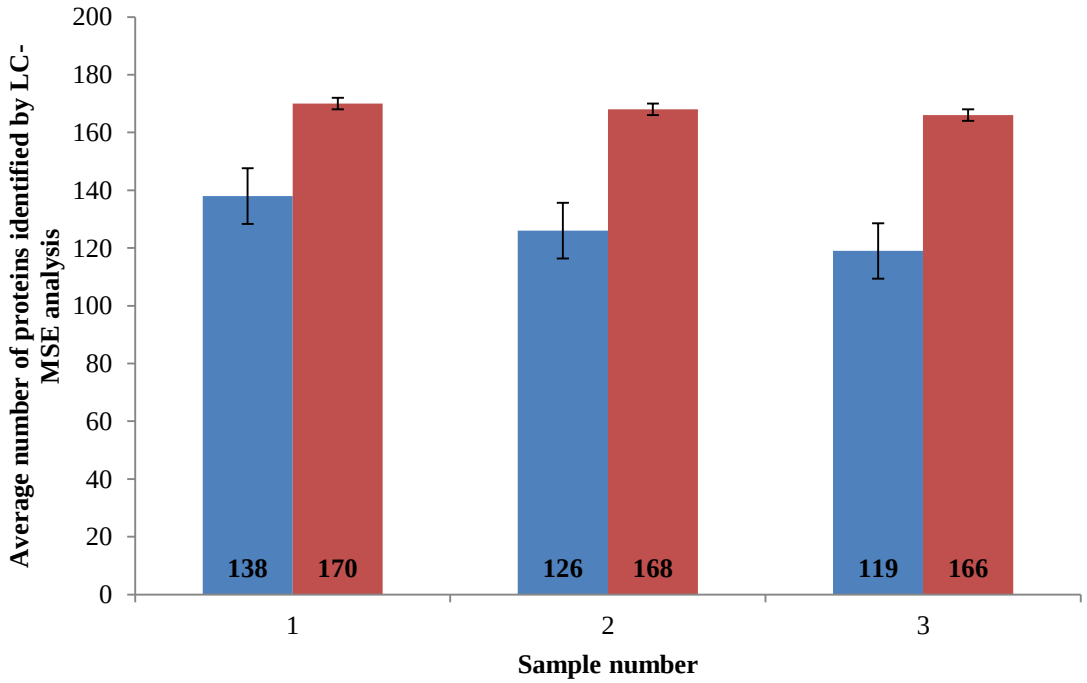


Figure 3.19: This graph displays the number of proteins identified by the equalizer beads with SDC method in both MS^E and $HDMS^E$. Three samples of superpool control plasma were used for each sample run and each sample was run in triplicate and error bars indicate the standard deviation of protein

identifications indicating the reproducibility of the method. The data displayed in **BLUE** is the number of proteins identified by the 110 min chromatographic run in MS^E. The data displayed in **RED** is the number of proteins identified by the 110 min chromatographic run in $HDMS^E$.

The equalizer beads with SDC proved to be the best method for protein profiling for biomarker discovery. Using this method a sample was prepared and run in triplicate using a 110 min chromatographic MS^E and $HDMS^E$ runs. This allowed qualitative comparison of the two protein profiles (see list in appendix). The total number of proteins profiled by 110 min nanoUPLC-MS^E analysis for this method was determined by using Excel to remove all duplicate data and the total number of proteins identified was 165, whereas the total number of protein identified for the 110 min nanoUPLC-$HDMS^E$ run was 228. There was approximately a 30% increase in the number of protein identifications. The standard deviation for the MS^E method was a lot higher than that of the $HDMS^E$ method, 9.6 and 2.0 respectively. However, both of the mass spectrometry methods fall under the 20% CoV with the MS^E producing an CoV of 7% and $HDMS^E$ 1%.

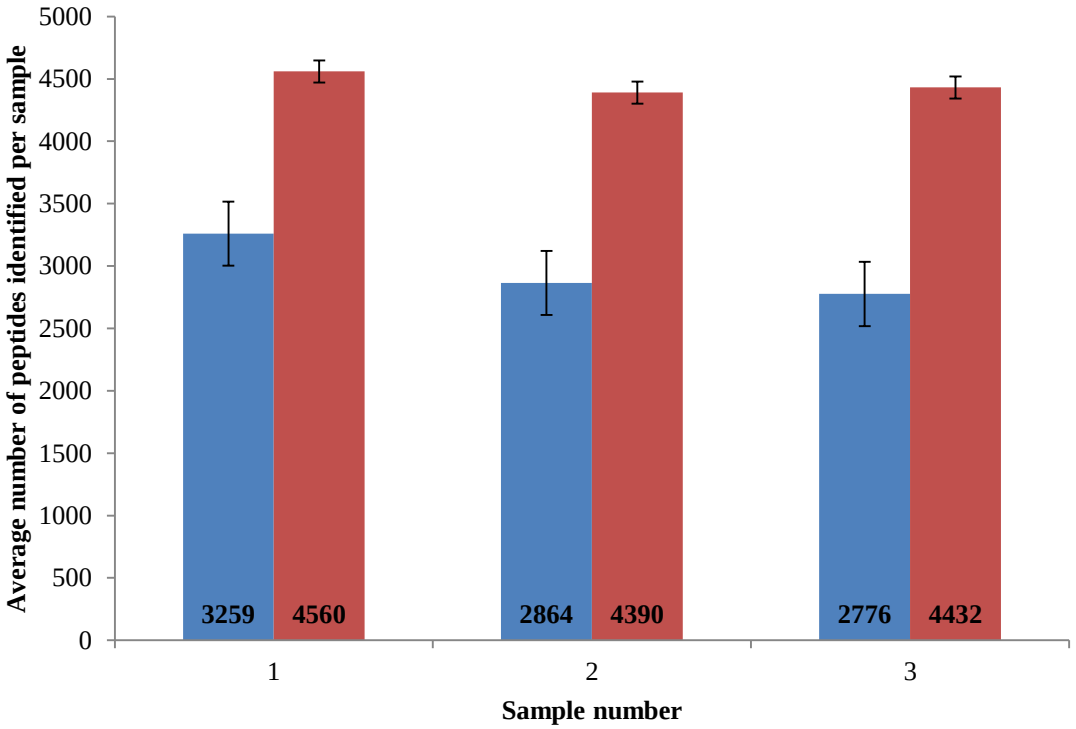


Figure 3.20: This graph displays the average number of peptides identified by the equalizer beads with SDC method in both MS^E and $HDMS^E$. Three samples of superpool control plasma were used for each

sample run and each sample was run in triplicate and error bars indicate the standard deviation of peptide identifications indicating the reproducibility of the method. The data displayed in **BLUE** is the number of peptides identified by the 110 min chromatographic run in MS^E. The data displayed in **RED** is the number of peptides identified by the 110 min chromatographic run in $HDMS^E$.

To understand where the differences in the protein identifications originated the number of peptides identified was studied, as the use of ion mobility should have an effect on the peptides identified and assigned. Figure 3.20 displays the number of peptides identified on average for samples 1 -3. Once again, a 30% increase in the number of peptides identified was observed, with $HDMS^E$ identifying the most peptides, upon comparison the CoV for the average number of peptides using MS^E was 7.9% and $HDMS^E$ 2.0%.

The samples used for the comparison of the mass spectrometric methods were identical. The injection volumes were identical thus, the protein loading on column was expected to be identical. However, it was observed that there was a significant difference in protein quantities reported by the PLGS Hi3 quantitation method. This discrepancy was further explored by comparing two LC runs and the same sample, one run in MS^E and the other in $HDMS^E$ with an injection volume of 1 µl and chromatographic run time of 110 min. The proteins sample was injected to give 1 µg/ µl.

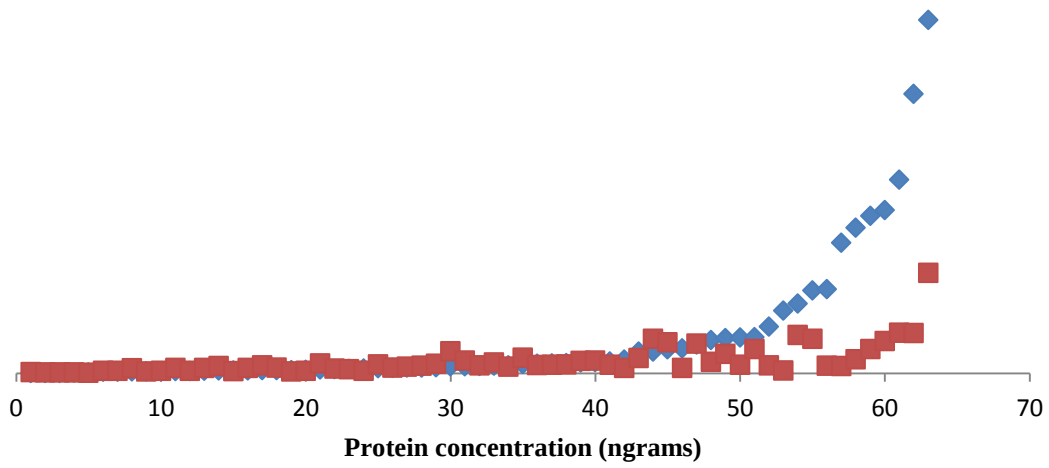


Figure 3.21: This graph displays the protein concentrations of 63 proteins common to both the MS^E and $HDMS^E$ of runs for the equalizer beads with SDC method. The data displayed in **BLUE** is the protein concentrations reported for MS^E. The data displayed in **RED** is the protein concentrations reported for $HDMS^E$.

Using the data analysis, protein assignments, and concentrations reported by PLGS v2.5.2 the total protein on column reported for MS^E was 1.1 µg and for $HDMS^E$ 0.5 µg. Identical sample and sample injection volume was used for both runs, to compare the proteins common to both the MS^E and $HDMS^E$, the protein profiles were manually compared using Excel. The difference in the reported protein concentrations is displayed in Figure 3.21. As shown it appears that the protein concentrations for the lower abundance proteins are comparable for both MS^E and $HDMS^E$ however, as the protein concentration increases and thus the abundance of the proteins, the MS^E and $HDMS^E$ protein concentrations reported differs dramatically. This could be due to detector saturation as there are an increasing number of peptides identified over the same run time in $HDMS^E$ in comparison to MS^E (Shliaha *et al.*, 2013).

$HDMS^E$ offers the qualitative advantage of identifying more proteins than MS^E, as had been reported. Unfortunately, the discrepancy in protein concentration between MS^E and $HDMS^E$ cannot allow accurate protein quantification of samples. Underestimating protein concentrations at the high abundance concentration range means estimating total protein amount or concentration of a sample is challenging. However, as figure 3.20 displays the trend of protein concentrations is mirrored between MS^E and $HDMS^E$ even though the amount protein differs. Therefore, for further studies using $HDMS^E$ for mass spectrometric analysis a *'scouting run'* should be used prior to all $HDMS^E$ analysis. The scouting run will be a single injection of a sample in MS^E to ensure the protein concentration on column does not exceed 1 µg/ µl and is consistent with all other sample concentrations for comparisons. Once the scouting run has been performed, the sample can then be analysed in triplicate using $HDMS^E$ analysis.

3.3.3.5.3 Data dependant vs. Data independent acquisition

Data for mass spectrometric analysis can be acquired using two different workflows these are data independent and data dependent acquisition. In data independent acquisition there is no pre-selection of peptides that are to be fragmented to produce product ion information, all ions are fragmented. In data dependent acquisition, a specified number of pre-selected precursor ions with the highest signal intensities are fragmented for product ion information.

In this study, data dependent and data independent acquisition were tested using the same sample. The sample was prepared from control human plasma and treated with the equalizer beads and eluted with SDC. The sample was split into two different aliquots for the different machines. For the data independent acquisition, the Waters Synapt G2 mass spectrometer was used in MS^E mode. For data dependent acquisition ThermoVelos Orbitrap and DDA function for the Synapt™ G2 was used. The samples were run on a nanoUPLC 110 min run and the data was collected and analysed.

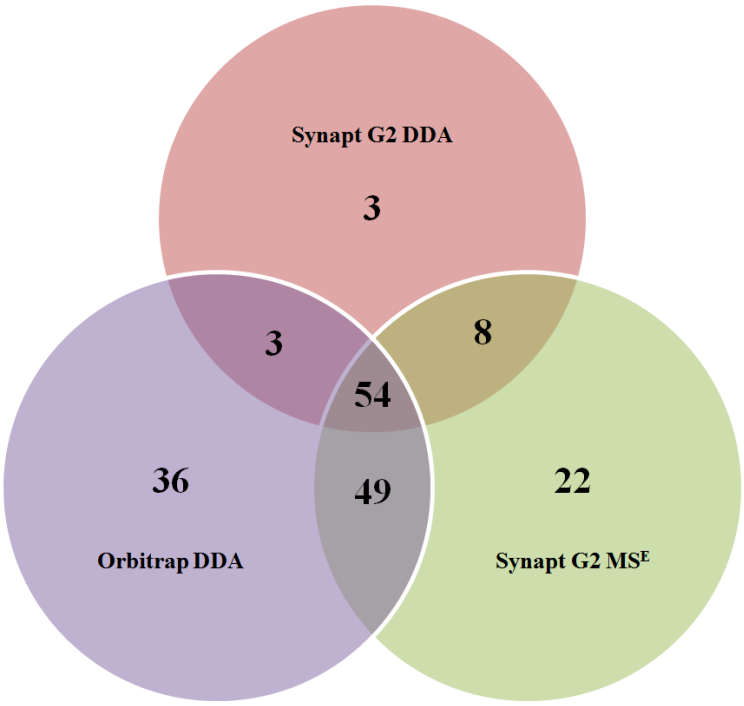


Figure 3.22: Venn diagram of the three mass spectrometry methods. Three samples of superpool control plasma were used for each sample run and each sample was run in triplicate.

The primary objective of this study was to identify whether either form of data acquisition offered an advantage in protein profiling and identification. The secondary aim was to ascertain whether any differences in the protein identifications exist. Using Excel to manually combine proteins and remove duplicate identifications, 172 proteins were identified using all three MS methods i.e. MS^E, DDA with the Synapt G2 and DDA with the orbitrap (Figure 3.22). The Orbitrap identified the most proteins using a DDA method with a total of 142 protein identifications, the Synapt G2 MS^E identified

the second most with 133 identifications and the least were identified by the Synapt G2 DDA with 68 protein identifications. Based on qualitative protein identifications, the orbitrap MS method provides the most extensive protein profile with the Synapt DIA method producing 9 less proteins. The Synapt DDA method identified less than 50% of the total proteins identified by the other two MS methods but did uniquely identify 3 proteins. These were Ig gamma 4 chain C region, Ig lambda 6 chain C region and Ig mu heavy chain disease protein. The difference in protein identification can be attributed to the software used for protein identification. Protein identification for all Synapt experiments was produced by PLGS v2.5.2 whereas Oribtrap data files were not supported by PLGS and thus, MASCOT was used for identification.

The MS^E method identified more peptides per protein than the DDA method. Figure 3.23 compares the peptides identified for 19 proteins common to both the Orbitrap and Synapt methods. More peptides are identified by the Synapt MS^E method than the Orbitrap DDA method for every protein. Therefore, although there may be slightly less protein identification with the MS^E method, more peptide identifications subsequently leads to more confidence in protein identifications, as there are more peptides to ensure proteins are correctly assigned.

This study shows that although there are a few more protein identifications to be gained from the Orbitrap DDA method, the Synapt MS^E method offers greater confidence in protein assignment. Thus, a very small percentage of fewer protein identifications are preferable for a gain in peptide identification and protein assignment confidence.

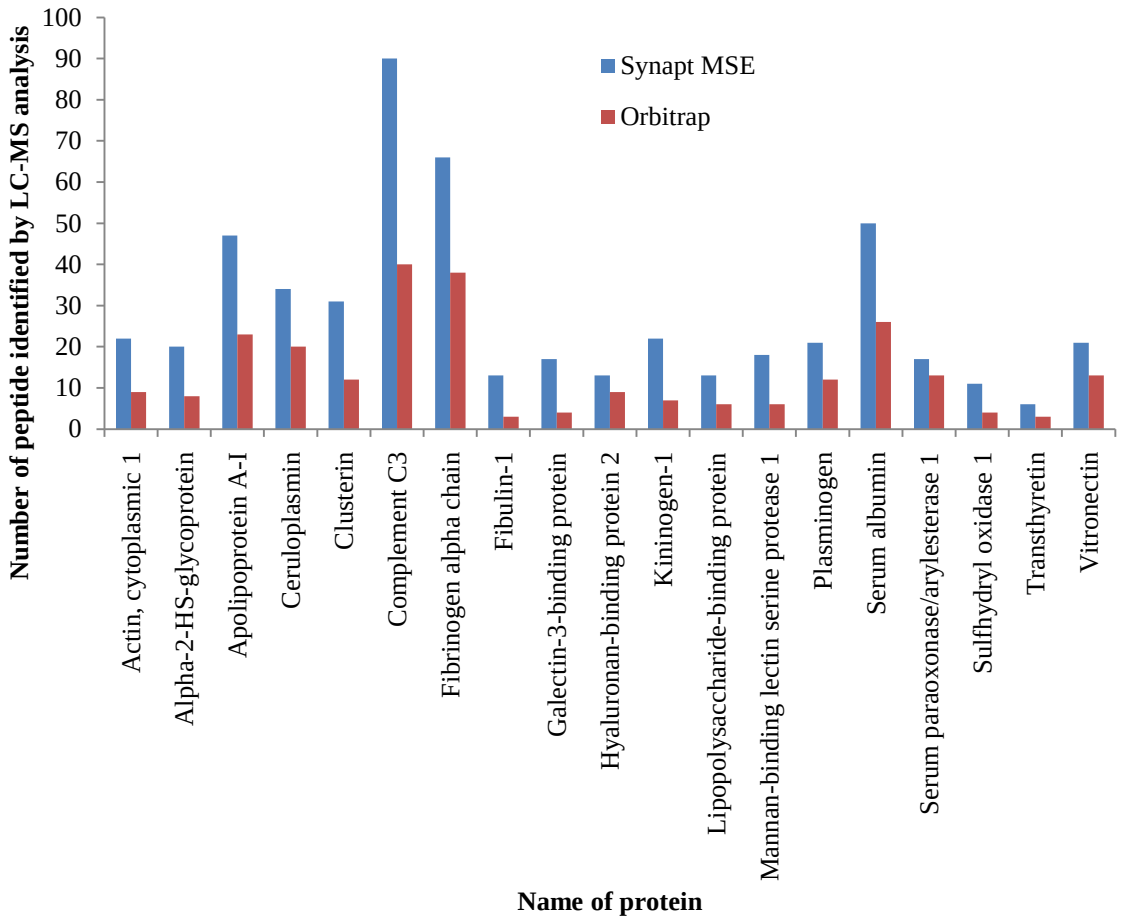


Figure 3.23: This graph shows the number of peptides assigned per protein for the Synapt MS^E and Orbitrap DDA method. The MS^E method clearly identifies a higher number of peptides (N=1).

3.4 Discussion

Plasma proteomics for biomarker discovery projects is challenging, as plasma has a dynamic range which exceeds any mass spectrometer available (Anderson, 2010). The challenging nature of plasma requires sample preparation methods to be robust which can be the most crucial part of a proteomic workflow. As can been seen from the results above, ineffective sample preparation methods can lead to insufficient information of plasma proteins for biomarker studies. In the introduction the three main stages of a proteomic workflow were introduced, samples preparation, sample analysis and data analysis. The aim of this chapter was to evaluate and optimise the sample preparation method and certain elements of the sample analysis (Figure 3.24).

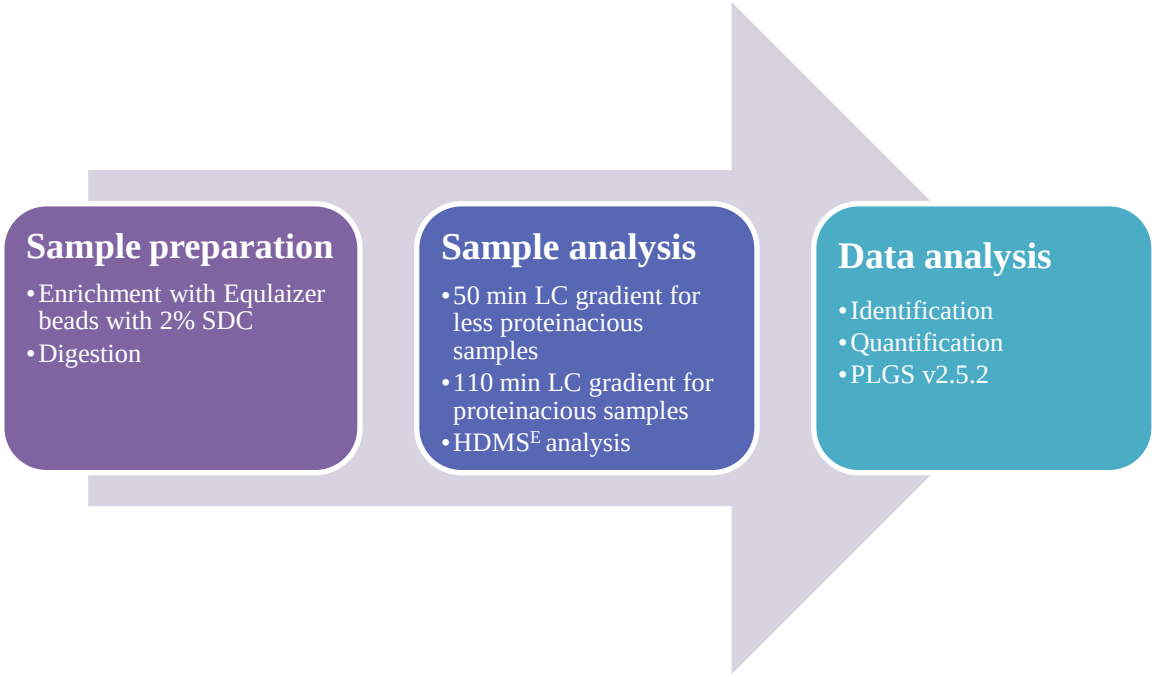


Figure 3.24: Evaluated and optimised proteomic workflow for heart failure biomarker discovery.

In this chapter, method development has clearly shown progression to an optimized method using the equalizer beads with 2% SDC to help produce the most extensive protein profile for discovery phase proteomic projects using healthy control plasma. The nature of SDC precipitation at acidic pH had the advantage of acidifying the peptides for MS analysis. An essential consideration of any proteomics workflow is to ensure any workflow followed must show suitable repeatability and reproducibility (Tabb *et al.*, 2010). The reproducibility offered by the addition of 2% SDC rather than trypsin as the elution reagent, in addition to the more comprehensive protein profile obtained, meant that the 2% SDC was selected for label free LC-MS^{E} clinical proteomic discovery phase studies.

It was found for complex samples such as plasma 110 min LC runs offered a slight advantage over 50 min runs, especially if the sample preparation method produced a proteinacious sample. The longer chromatographic runs also allow peptides that may co-elute over a shorter run to separate and elute over a longer retention time window, increasing chromatographic resolution. Thus, for further clinical proteomic discovery phase studies, the 110 min run was best suited.

Comparison of MS acquisition methods namely DDA and DIA, found the DDA techniques offered an advantage in protein identification, approximately 6% more proteins were identified. However, in MS^E analysis fewer proteins may have been identified but the confidence of the assignment was greater due to data independent acquisition, parameters set for data processing and requirement criteria set for protein identifications. In addition to identify proteins with greater confidence, accurate quantitative data using Hi3 quantitation (Silva *et al.*, 2006) was also obtained which is advantageous in discovery biomarker studies. This balance of quantity and quality of protein data is needed so that results can be discussed and reviewed with confidence (Nilsson *et al.*, 2010).

The use of ion mobility was also tested in the method development. When comparing MS^E runs with and without ion mobility, it was found that ion mobility runs or $HDMS^E$ runs qualitatively improved the data by increasing the number of identifications by 30% in comparison to the MS^E runs. However, the $HDMS^E$ function was unable to correctly quantitate the proteins at the higher abundance range. Despite the higher abundance protein reporting a lower concentration in $HDMS^E$, the protein trend line of the curve (Figure 3.21) mirrors that of the MS^E. Thus, to overcome this limitation a scouting run in MS^E was proposed to help quantitate the protein amount to ensure the same amount of protein is loaded on per sample and proteins can be profiled in $HDMS^E$.

As mentioned previously the dynamic range of plasma is vast (nearly 12 orders of magnitude) and poses a great challenge for any proteomic workflow especially when trying to access the lower abundance proteins where clinically relevant protein biomarkers reside. In this chapter, a sample preparation method has been shown to achieve just over 5 orders of magnitude (Figure 3.16). This is still significantly lower than the proposed dynamic range of plasma however, it is an improvement on the dynamic range achieved by previous workflows (Hortin & Sviridov, 2010). In the proposed workflow using equalizer beads with 2% SDC elution was able to identify pre-existing clinical biomarkers such as PSA and other low abundance proteins that play a role in disease pathology (Table 3.3). This shows that the optimised reproducible method was able to identify proteins that are relevant to disease mechanisms inferring their use as clinical biomarkers.

Table 3.3: Previously determined concentration ranges for proteins identified by equalizer beads with 2% SDC elution analysed by LC-MS^E. This demonstrates that the optimised method is able to identify clinically relevant proteins

Protein Name	Disease	Reference for disease
Mannan binding lectin serine protease 1	Affects the pathogenesis of inflammatory diseases	(Turner, 2003; Kouris *et al.*, 2005; Nagele *et al.*, 1999)
Phosphatidylcholine sterol acyltransferase	Norum disease, familial LCAT deficiency and fish-eye disease	(Vanloo *et al.*, 2000)
Platelet factor 4	Crohns, ulcerative colitis	(Simi *et al.*, 1987)
Prostate Specific Antigen	Prostate cancer marker	(Kuriyama *et al.*, 1981; White *et al.*, 2006)
Cholesteryl ester transfer protein	Heart disease	(Wolfe & Rader, 2004)
Palladin	Pancreatic cancer	(Pogue-Geile *et al.*, 2006)

The plasma samples used in this chapter were all from healthy control patients and differences in plasma protein profiles of healthy and disease patients has been noted (Anderson, 2010). Thus, the findings in this chapter are applicable to plasma from healthy patients. The suitability of the method in disease patients still needs to be determined. This is addressed in the next chapter of this study.

In this chapter a method development to establish the optimum method for clinical proteomic discovery projects has been identified. The equalizer beads with 2% SDC demonstrated reproducibility, was able to cover over 5 orders of magnitude and identify proteins of clinical relevance. The LC gradient length of 110 mins has been determined as ideal for more proteinacious samples however, for less proteinacious samples 50 min gradients are sufficient. The advantages of HDMS^E analysis have been ascertained. Thus, these findings lead to an optimised proteomic workflow on which further analysis of clinical samples could now be based.

Chapter Four

Diagnostic Heart Failure Study

4.1 Introduction

The work discussed in Chapter Three aimed to establish a reliable and reproducible proteomic workflow for a proteomics discovery project. Once the method had been developed, evaluated and optimised, the next step would be to conduct a preliminary study to ensure the method and workflow was suited to clinical proteomic discovery projects using both healthy and diseased samples.

Traditionally diagnosis of systolic heart failure relies heavily on diagnostic imaging and clinical symptoms, which have 70-90% specificity but only 11-55% sensitivity (McMurray & Pfeffer, 2005; McMurray & Stewart, 2000). In this plasma proteomics study, the main objective was to see if proteins within the plasma proteome can act as diagnostic biomarkers for systolic heart failure. Using plasma biomarkers to diagnose systolic heart failure would mean the method would be non-invasive, independent of imaging interpretation and could potentially provide a rapid turnaround of results.

Plasma is thought to be the most comprehensive proteome in the human body, as the circulatory system is representative of the physiological and pathological processes active at the time of sampling (Anderson & Anderson, 2002). However, biomarker discovery projects using plasma come with some caveats. To date the most successful proteomics discovered biomarker has been OVA1 which is also FDA approved (Zhang *et al.*, 2004; Zhang, 2012). In the last 15 years an average of 1.5 new tests were introduced *per annum (p.a)* in comparison to the pre-1993 era when an average of 6 tests were introduced *p.a* (Anderson, 2010), despite thousands of publications claiming biomarker discovery. There may be many reasons and explanations as to why so few biomarker discoveries fail to achieve FDA approval for clinical use. Firstly, the lack of well-developed technologies used to de-convolute the plasma proteome for successful characterisation may contribute. Secondly, poor experimental design that may lack rigor and thoroughness could cause discovery projects to fall short of FDA approval (Lilley *et al.*, 2011).

This chapter addresses both caveats as objectives for this project by using multidimensional chromatography for further method development on the existing

optimised proteomic platform. An experimental design ensuring accurate sampling and profiling removing any bias is used as well.

4.1.1 Discovery proteomic workflow

There are typically three key stages of proteomic projects; sample preparation, sample analysis and data analysis (Figure 4.1). The aim of sample preparation is to ensure that samples are amenable to mass spectrometery. Sample analysis is conducted via LC separation and MS analysis, once the data has been generated, it is then mined for protein identification, quantification and statistical parameters are applied to highlight proteins that have undergone change between sample sets. This formulates the basis of most MS based proteomic discovery studies.

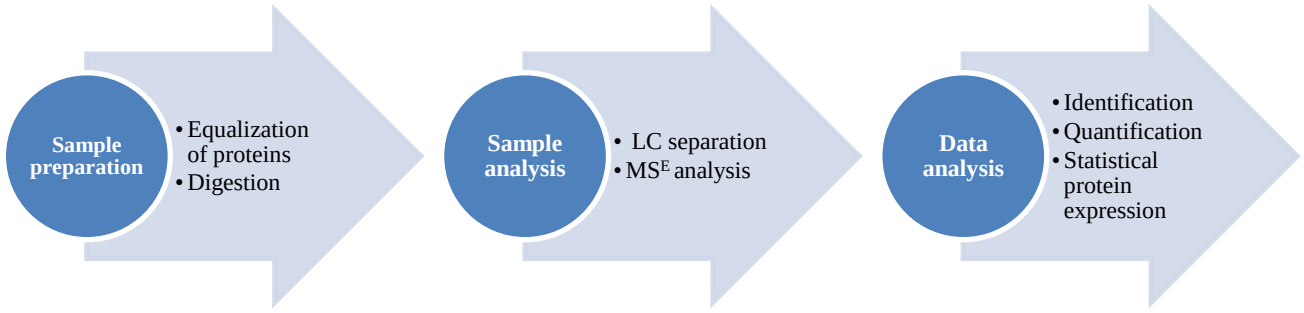


Figure 4.1: Overview of the proteomics workflow and how it is broadly applied within this study.

In this study, a proteomic workflow was designed to determine the protein profile of heart failure patients vs. healthy patients to identify diagnostic markers of disease. The proteomic workflow of this diagnostic study used the sample preparation method developed and optimised on healthy control plasma samples in Chapter Three. An additional separation stage using offline RP chromatography was added for another dimension of peptide separation prior to MS analysis. The samples were analysed using a novel label-free quantitative nanoUPLC-HDMS^E method, where data independent acquisition would ensure that there was no bias or pre-selection of peptides during mass spectrometric acquisition. The data was analysed and quantified using bioinformatic software called PLGS v2.5.2 and quantification was achieved using the Hi3 method

where absolute protein concentration is determined by the use of an internal standard at a known concentration (Silva *et al.*, 2006). This is the first reported use of a label free ion mobility enabled pipeline for the discovery of systolic heart failure biomarkers.

4.1.2 Multidimensional chromatography

Reverse-phase chromatography has been extensively studied and used as a separation technique for peptides prior to mass spectrometry. However, it has been found that many peptides may co-elute in the same chromatographic retention time window, allowing peptides to enter the mass spectrometer simultaneously and causing some peptides to remain un-identified. As mentioned in the introduction, RP chromatography is based on hydrophobic interaction, whereby the hydrophobic stationary phase binds peptides in the presence of a polar solvent. Elution occurs with increasing peptide hydrophobicity with a gradual increase in the proportion of organic solvents.

Multi-dimensional chromatography (in this study two-dimensional liquid chromatography (2D-LC)), has been extensively studied to find the most orthogonal method, increasing separation and peak capacity. Two dimensional-LC consists of either online or offline LC separation over two chromatographic dimensions. Traditionally, a proteomics based approach uses strong cation exchange (SCX) for the first dimension and RP for the second dimension. Gilar *et al*. evaluated a number of different 2D approaches for proteomics to help increase peak capacity, namely SCX-RP and RP-RP (Gilar *et al.*, 2004; Gilar *et al.*, 2012). In their study, Gilar *et al.* found that RP-RP 2D-LC at different pH (pH 2.6 and pH 10) demonstrated high levels of orthogonality in comparison to SCX-RP (Figure 4.2).

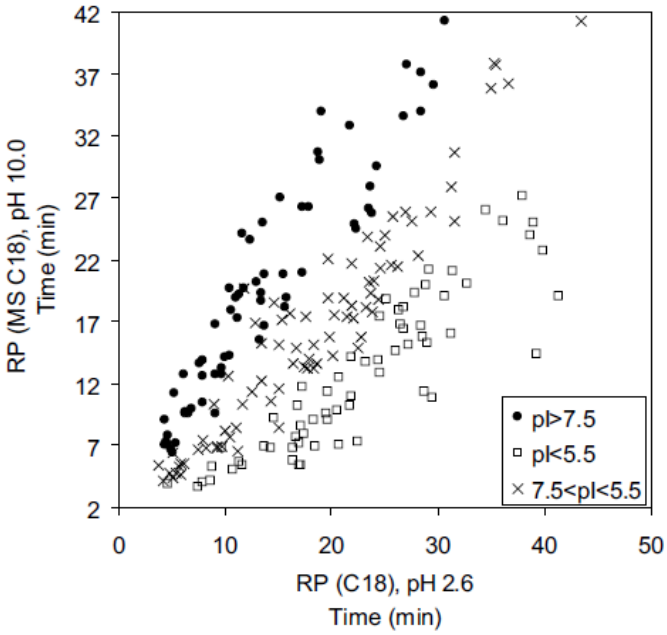


Figure 4.2: Normalised retention time plot taken from (Gilar *et al.*, 2005a; Gilar *et al.*, 2012). This shows the 2D-LC orthogonality of separation of the RP-RP low and high pH.

Additionally Gilar *et al.* observed RP-RP also offered other advantages over SCX-RP; RP-RP mobile phases remain salt free and are compatible with MS, RP-RP offered high peak capacity and better recovery of long and short peptides in comparison to SCX-RP. The study by Gilar *et al.* was further supported by a study published by Zhou *et al.* This study compared RP-RP and SCX-RP analysis of *E.Coli* digest and more peptides and proteins were identified by RP-RP than SCX-RP using the same sample and MS conditions. They also found RP-RP was able to report proteins over a greater dynamic range. The authors suggested this was due to RP-RP higher peak capacity in the first dimension (Zhou *et al.*, 2010).

It is clear that multidimensional chromatography and particularly RP-RP-LC offers an advantage by increasing peak capacity, increasing protein identifications and reducing sample complexity prior to MS analysis (Gilar *et al.*, 2012; Gilar *et al.*, 2005b; Gilar *et al.*, 2009). In this study, an offline 2D RP-RP chromatographic separation approach was utilised, Waters Alliance 2690 HPLC system (Waters Corporation, Milford, MA, USA) was used for the first dimension using high pH RP separation and a low pH RP nanoAcquity system (Waters Corporation, Milford, MA, USA) in the second dimension. Samples were loaded onto the first dimension column at a high pH and

fractions were collected every 1 min for 100 min over a 110 min gradient. The samples were concatenated, lyophilised, reconstituted and applied to the second dimension column at a low pH.

The aims for this study were four-fold:

1. Develop the analytical platform using 2D-LC with HDMS^E for the analysis of pooled plasma samples for biomarker discovery.
2. Compare the number of proteins identified and quantified using 1D and 2D approaches.
3. Assess protein level difference between heart failure and healthy controls for both 1D and 2D approaches.
4. Assess any putative biomarker proteins identified.

4.2 Materials and Methods

4.2.1 Materials

Heart failure plasma was supplied by patients from the British Heart Foundation funded study entitled *"The uroguanylin system in heart failure"*, which recruited patients presenting with heart failure. The samples were centrifuged and plasma was separated from the red blood cells and buffy layer. Ten heart failure samples were selected for this study with age and sex matched control samples from the NHS New and Emerging applications of technology funded study *"Screening for heart failure"* (for sample collection protocol see appendix). The human blood samples were collected in EDTA blood sample tubes donors following informed consent. The blood samples were centrifuged at 1500 x g for 20 minutes at 4°C.The plasma layer was separated from the buffy layer and red blood cells, and stored at -80°C.

ProteoMiner^{TM} Protein Enrichment Kit (Bio-Rad, Hemel Heampstead) was used which contained the PBS wash buffer and 1 mL equalizer bead columns. Digestion reagents dithiothreitol, iodoacetamide and proteomics grade trypsin were supplied by Sigma Aldrich (Poole, UK). Mass spectrometry standards Glu^1-fibrinoginogen peptide, alcohol dehydrogenase and enolase were purchased from Waters (Manchester, UK). Mass spectrometry solvents were all purchased at optima LC-MS grade from Fisher Scientific Ltd (Loughborough, UK).

4.2.2 Sample preparation

Each plasma sample was thawed from -80 °C at room temperature and vortexed to ensure homogeneity and 100 µl plasma aliquots were taken from each sample. The samples were age and sex matched using a propensity score. An aliquot of 100 µl plasma from ten heart failure patients and from ten healthy controls was combined to generate two pools of plasma; one from heart failure patients and one from controls. Each sample was incubated with the equalizer beads and eluted with 2% SDC (see Chapter Two for full details). The elutions were digested and lyophilised and stored at -80 °C prior to analysis.

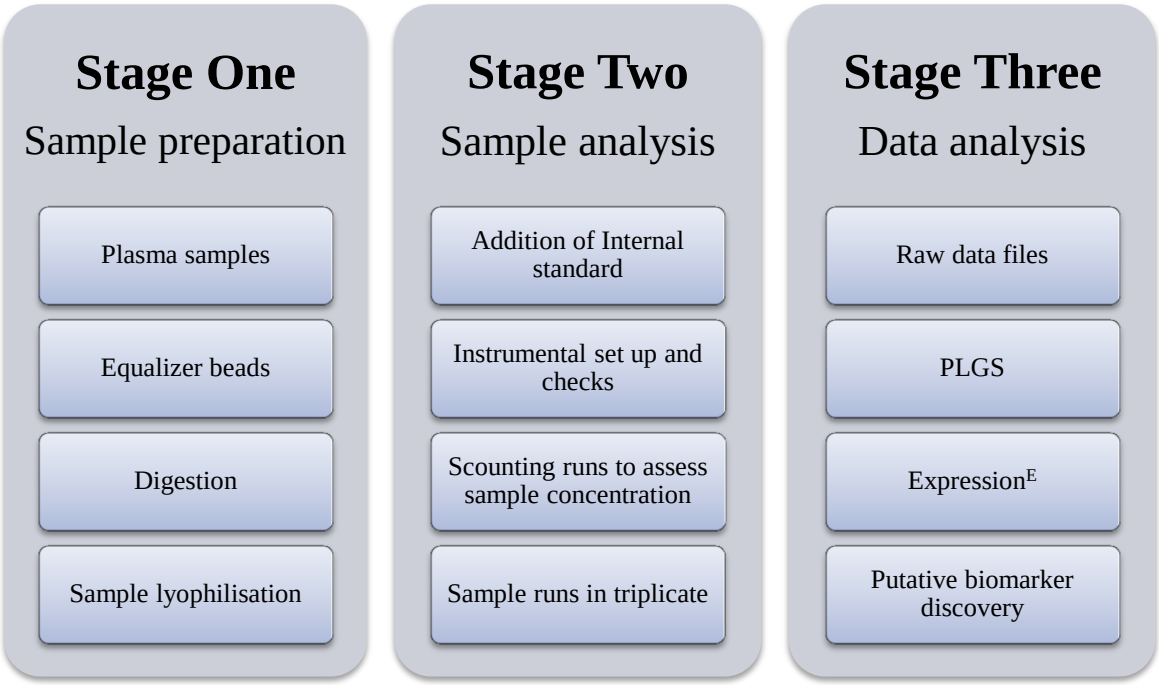
4.2.2.1 Fractionation of pooled plasma with concatenation

An offline method for fractionating the plasma using a RP column with a mobile phase of pH 10 in the first dimension was used. The fractions were collected from the column in 1.5 mL eppendorf tubes and protein detection was monitored at absorbance 214 nm. Fractions were collected at every minute over a 100 min period and concatenated into groups of 10 as described in Table 2.2 in Section 2.8.2.1. A blank chromatographic run of 0.1% formic acid was used between samples to ensure there was no carry over between the heart failure and healthy control samples. The concatenated samples were evaporated for 2 hours and lyophilised overnight. The samples were reconstituted in 0.1% formic acid and spiked with 100 fmol ADH at a 1:1 ratio. For each pooled sample, there were 10 concatenated samples to be run in triplicate in total.

4.2.3 Proteomic workflow

Two proteomic workflows (1D and 2D) were utilised in this study. Each stage of the approach is described in Figure 4.3. To ensure samples were not cross contaminated by proteins from previous runs, blank washes were run between samples of 0.1% formic acid and an enolase was run periodically to check consistency of chromatography.

A: One dimensional analysis



B: Two dimensional analysis

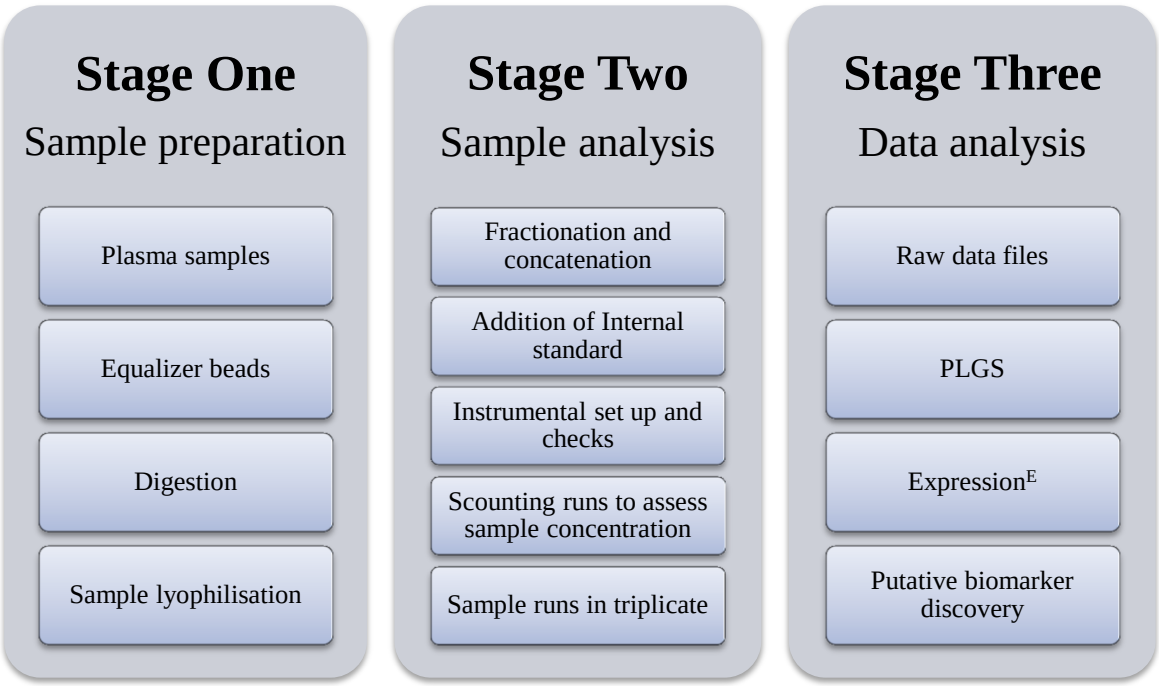


Figure 4.3: This shows the schematic for both the 1D (A) and 2D (B) approaches undertaken for this study. There are three main stages, sample preparation that was developed in Chapter Three, sample analysis that will be developed in this chapter and data analysis. Both workflows will be used in this chapter to assess which is best suited to a discovery proteomics project.

4.2.4 Processing of HDMS^E data

HDMS^E data produced was processed using PLGS v2.5.2. Lockspray (GFP) data was collected during the nanoUPLC-HDMS^E runs through a reference spray. The protein identification and assignment process has been described in detail in section 2.9.4. Precursor and product ions are time aligned and various data interrogation steps take place to ensure protein identifications are confidently assigned. Data was processed using the following processing parameters; 15, 80, 750 for the low, elevated and intensity threshold respectively.

A Uniprot (http://www.uniprot.org/) human reviewed database was used and appended to include ADH (yeast) and ADH was specified as the internal standard with the concentration specified in fmol, this would allow Hi3 quantification. Protein identification criteria were set to determine detection if the following criteria was met, at least two fragment ions per peptide, five fragment ions per protein and a minimum of one peptide per protein with the false discovery rate set at 4%.

Fixed modification of carbamidomethyl cysteine and one missed tryptic cleavage site with variable modifications of oxidation of methionine, acetyl N-terminus, deamidation of asparagines and glutamine were all included in the search parameters of the HDMS^E workflow. All identifications made by PLGS were exported to Microsoft Excel, where further data interrogation and filtering to remove triplicates took place.

4.2.5 Relative protein expression

Relative protein expression was performed within PLGS v2.5.2 using the expression algorithm. The mass spectrometric runs were grouped according their conditions (i.e. heart failure or control) and the data was normalised within each group according to the spiked internal standard ADH.

Only proteins observed in two sample runs reported a ratio for each protein identified. The protein ratio was assigned a probability score between 0 and 1; proteins were reported as up-regulated, down-regulated or unique. The up and down regulation was determined via p values that were either ≥ 0.95 (up regulated) or ≤ 0.05 (down regulated)

and deemed statistically significant. Unique proteins were only identified in either the heart failure or healthy control groups.

4.3 Results

4.3.1 Comparison of 1D and 2D-LC-HDMS^E pooled plasma samples

The pooled samples (ten systolic heart failure and ten matched healthy controls) were treated with the equalizer beads eluted with 2% SDC, tryptically digested and analysed using 1D and 2D-nanoUPLC-HDMS^E in triplicate.

The total number of proteins identified by 1D analysis was 263 and for 2D analysis 521 with data filtered for duplicates and a total of 237 proteins overlapped between the two chromatographic approaches. Ninety percent of the proteins identified in 1D analysis were also identified in 2D analysis and as the proteins were found in both experiments it can be assumed that at these protein identifications were assigned with confidence and the method provided repeatable and reproducible results. These results were achieved with a sample loading of 1.8 µg for 1D and 1.1 µg per fraction for 2D analysis. This equated to a 2 fold increase in the total number of protein observed with ten fraction 2D analysis in comparison to 1D, despite lower sample loading per fraction. However, once all sample loadings for the ten fractions were combined for 2D analysis a total of 11 µg was loaded for 2D in comparison to 1.1 µg for 1D analysis.

The total number of peptides identified on average for 1D and 2D analysis were 2856 and 10370 respectively, there was a 3.6 fold increase in the number of peptides identified by 2D analysis. This demonstrates the separation ability and advantage of the 2D chromatographic separation. An increase in peak capacity has previously been observed in different studies, where an increase in peptides identification leads to an increase in protein identification. Gilar *et al.* compared both 1D and 2D chromatographic separation platforms using undepleted plasma, the 2D chromatographic separation was performed using both RP-RP and SCX-RP and all MS analysis performed using DIA. The study found that there was a 3.6 fold increase in the protein identification using RP-RP and 2.9 fold increase for SCX-RP, RP-RP also lead to 12% more peptides being identified (Gilar *et al.*, 2009). It is clear that the additional

peak capacity and orthogonality of the RP-RP approach greatly improves the total number of peptide and protein identified (Gilar *et al.*, 2005b).

However, the approximately ten-fold protein concentration increase in 2D analysis only resulted in a 50% increase in protein identifications. This can be explained due to the exponential decrease in protein concentration levels in plasma. The number of fractions taken from the first dimension may have been the constraint in this study, as Zhou reported an increase in the number of proteins as a function of fraction depth (Zhou *et al.*, 2010). However, other studies suggest that there is no advantage in excessive fractionation in the first dimension as the gain in protein identification leads to excessive sample analysis time (Patel *et al.*, 2012).

This not only meant a greater number of proteins were identified with 2D analysis, the dynamic range achieved with 2D analysis of pooled sample groups was much greater (Figure 4.4). The sample loadings were similar for all HDMS^E analysis in this study (1.8 µg for 1D runs and 1.1 µg for 2D runs), notably lower abundance proteins were observed with 2D analysis. As figure 4.4 demonstrates, 2D analysis identifies proteins from just over 5 orders of magnitude in comparison to 4 orders of magnitude for 1D analysis. The graph clearly demonstrates that 2D analysis was able to access proteins of lower abundance with a ten-fraction approach. These results are promising for biomarker studies as lower abundance proteins are where clinically relevant biomarkers are found. Thus, a more comprehensive protein profile of the samples is achieved with 2D analysis than with 1D.

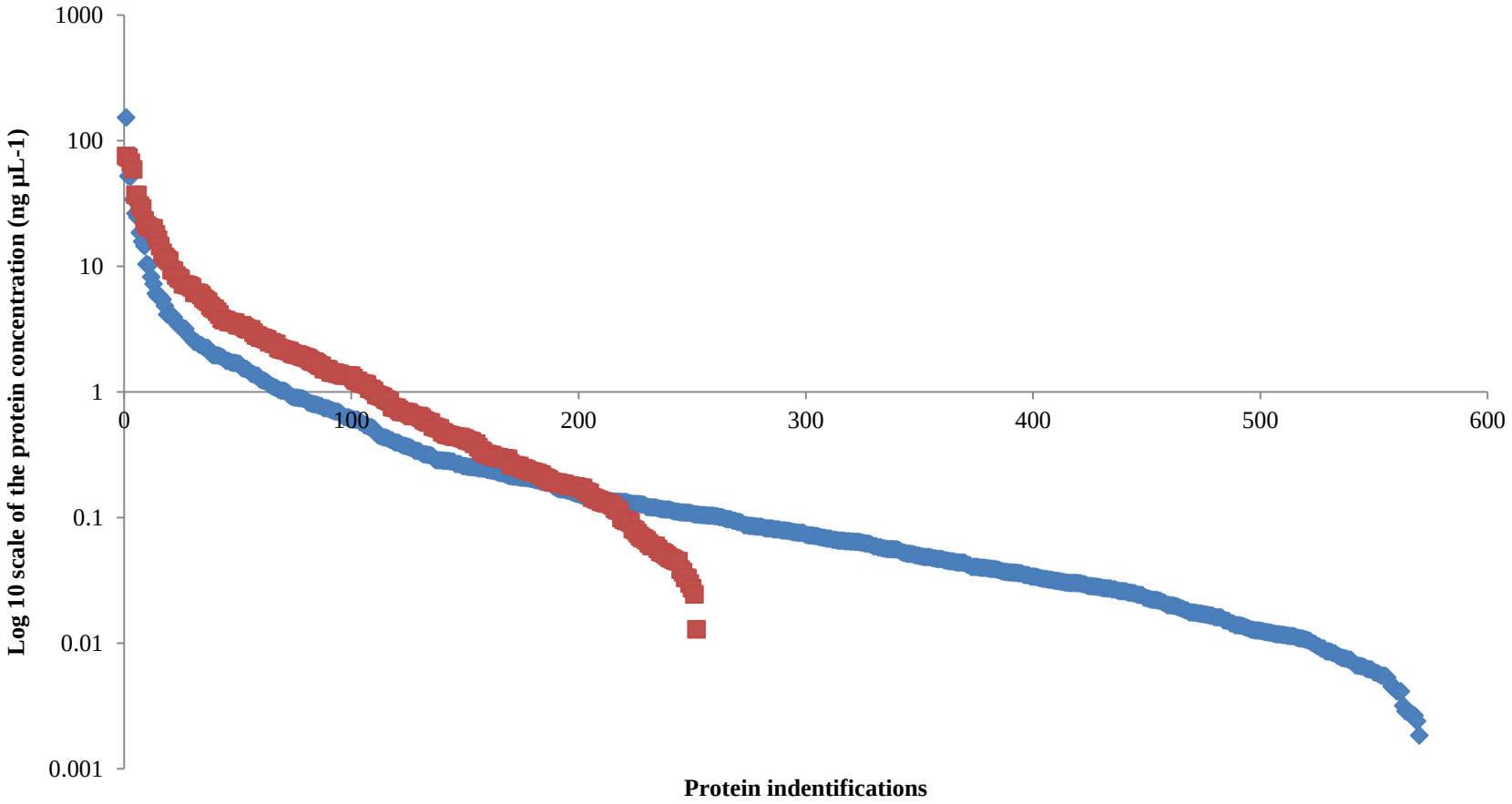


Figure 4.4: This graph shows the dynamic range achieved for the pooled patient samples using 1D and 2D analysis. The RED line indicates the dynamic range achieved using 1D analysis, here it shows 4 orders of magnitude are achieved for the pooled sample. The BLUE line indicates the dynamic range achieved using 2D analysis, here it shows 5 orders of magnitude are achieved for the pooled samples. The 2D protein abundances were normalised to represent an average protein concentration per protein.

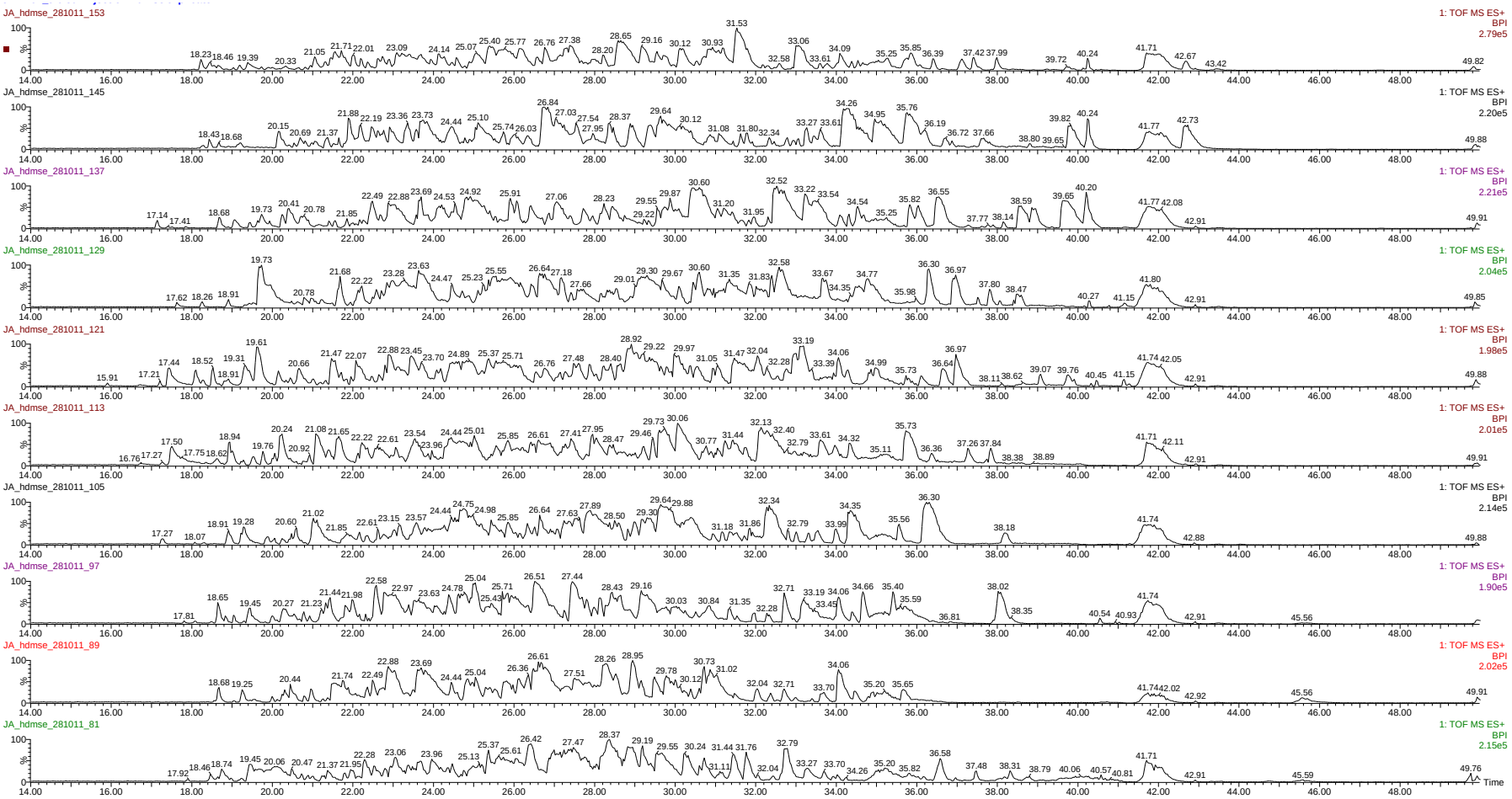


Figure 4.5: Base peak intensity chromatograms for ten concatenated fraction RP-RP separation of equalized plasma peptides for the pooled patient plasma.

4.3.1.1 Concatenation of the first dimension

As mentioned previously, the use of 2D-LC for this study managed to produce 50% increase in protein identifications in comparison to 1D, with a 10 fold greater mass load. The use of RP-RP with low and high pH has also been advantageous for protein numbers in other studies. The use of concatenation could account for the increase in protein numbers as well as the increase due to greater peak capacity. Concatenation combines RP fractions from the beginning, middle and end over a LC gradient in the first dimension. The advantages of this technique are twofold; firstly, the number of peptide fractions in the first dimension is reduced and secondly, concatenation adds to the orthogonality of RP-RP separation (Yang *et al.,* 2012). The increase in orthogonality could attribute to the increase in protein numbers obtained with 2D analysis.

Previous studies have shown that concatenation improves protein numbers with a 2D-LC platform. Wang *et al.* also found that concatenation of the high pH RP fractions prior to low pH RP resulted in a 2 fold increase in both peptide and protein identifications from tryptically digested human cells. It has been suggested that using concatenation, the second dimension of separation space is better utilised; this is the case with this study as shown in Figure 4.5. Peptides elute throughout the gradient in the second dimension and elution is spread wide across the LC elution time window, rather than eluting in peptide clusters (Wang *et al.*, 2011).

4.3.2 Identification of a current biomarker

C-reactive protein (CRP) was identified in pooled healthy control and heart failure plasma samples analysed using 1D and 2D-LC-HDMS^E. CRP was reported as higher in heart failure patients in both forms of analysis but was only deemed statistically significant in 2D-LC analysis with a p value of <0.01, as reported by PLGS.

4.3.3 Relative protein analysis using 2D-LC-HDMS^E

Expression^E analysis was performed to compare relative protein levels in control and heart failure equalized pooled plasma. The Expression^E algorithm is integrated to PLGS v2.5.2 and determines any difference in protein abundance and reports any up or down regulation. All identifications were normalised to the internal standard (ADH).

The ADH normalised expression analysis table indicated that there were a total of 521 proteins, 81 down regulated, 136 up regulated, 75 unique to heart failure and 142 unique to healthy controls using 2D-LC-HDMS^E analysis. A number of proteins reported were either contaminants such as keratin or immune response proteins such as immunoglobulins and human leukocyte antigens. These in total accounted for 120 proteins reported. Once these proteins were removed from the reported list, a total of 401 proteins remained, of which 118 reported ≥1.5 fold change between the two conditions. Table 4.1 displays 10 proteins, 5 from each biological group (see full table in appendix).

Table 4.1: A selection of the proteins identified from 2D-LC-HDMS^E analysis of pooled plasma. The protein denoted as down regulated were lower in healthy controls therefore, indicative of disease. The proteins denoted as up regulated were higher in healthy controls therefore, protective.

Protein Description	**P value (calculated by ExpressionE)**	**Pooled control plasma : Pooled heart failure plasma**
C reactive protein	0	Down regulated
Galectin 3 binding protein	0	Down regulated
Insulin like growth factor binding protein 2	0	Down regulated
Platelet factor 4	0	Down regulated
Serum amyloid P component	0	Down regulated
A disintegrin and metalloproteinase with thrombospondin motifs 13	0.95	Up regulated
Actin aortic smooth muscle	1	Up regulated
Apolipoprotein A II	1	Up regulated
Mannan binding lectin serine protease 1	1	Up regulated
Vitamin D binding protein	1	Up regulated

To ensure that all relative quantitative analysis was accurate, the relative change in the internal standard or alcohol dehydrogenase was cross examined against the calculated relative change. Expression^E analysis reported ADH to have a 1.0 fold change between the heart failure and control plasma. The calculated difference between ADH between the heart failure and control samples was 1.1 fold, this confirmed that the Expression^E analysis could be used with confidence for discovery biomarker projects.

4.3.4 Relative protein analysis using 1D-LC-HDMS^E

Ten individual samples were combined according to their conditions to give two pooled samples representing heart failure and matched healthy controls. The protein abundance for these samples was estimated using the Hi3 algorithm in PLGS and samples were grouped according to their conditions for relative abundance analysis using the Expression^E algorithm in PLGS v2.5.2. All identifications were normalised to the internal standard (ADH).

The ADH normalised expression analysis table indicated that there were a total of 4460 peptides were identified with 24 proteins showing decreased abundance , 61 increased abundance, 109 proteins unique to heart failure and 63 unique to healthy controls using 1D-LC-HDMS^E analysis. Once again, a selection of proteins reported were either contaminants such as keratin or immune response proteins such as immunoglobulins and random protein sequences that were unidentifiable. These in total accounted for 47 of the proteins reported.

Upon comparing the 1D-HDMS^E protein expression list with the 2D-HDMS^E protein expression list, a number of observations were made. All of the proteins identified in the 1D analysis of the pooled samples through expression analysis were identified by the 2D expression analysis unlike a manual comparison where merged protein lists for 1D and 2D were compared in Excel without filtering. A number of immune response proteins were identified in the pooled heart failure group more so than the pooled healthy control sample. CRP is thought to play an immune-responsive role in heart failure and is a marker of inflammation. In contrast to 2D-LC-HDMS^E where CRP showed decreased abundance in healthy patients and was highly significant, CRP was not identified as a significant marker of disease using 1D-LC-HDMS^E analysis.

Mannose binding serine protease 1 (MASP1) was found be significantly up regulated in healthy control patients with both 1D and 2D-LC-HDMS^E analysis. Similarly, platelet factor 4 was found to be significantly down regulated in healthy patient's samples. This would suggest that these proteins could be markers of the different conditions (Table 4.2).

Table 4.2: A selection of the proteins identified from 1D and 2D-LC-HDMS^E analysis of pooled plasma. This table shows the protein ratio's identified for each workflow. As the table shows that differences in ratio for 1D analysis is much greater than 2D analysis.

Protein description	1D Control: Heart Failure Ratio	2D Control: Heart Failure Ratio	1D Control: Heart failure _Log(e)Ratio	2D Control: Heart failure _Log(e)Ratio
Clusterin	0.05	0.88	-2.95	-0.13
Fibrinogen alpha chain	0.11	0.76	-2.21	-0.28
Fibulin 1	0.07	0.58	-2.67	-0.55
Protein AMBP	0.10	0.53	-2.32	-0.63
Apolipoprotein C I	0.11	0.66	-2.25	-0.42
Platelet factor 4	0.44	0.12	-0.82	-2.11
Apolipoprotein D	3.71	2.36	1.31	0.86
Beta 2 glycoprotein 1	3.10	1.25	1.13	0.22
Complement factor H	4.26	1.51	1.45	0.41
Mannan binding lectin serine protease 1	6.05	1.86	1.8	0.62
Phosphatidylinositol glycan specific phospholipase D	11.59	1.82	2.45	0.6
Vitamin K dependent protein S	4.01	1.54	1.39	0.43

Of the 24 proteins with decreased abundance identified by 1D-HDMS^E analysis only 8 of the proteins showed the same relationship with 2D-HDMS^E analysis. Of the 61 up-regulated proteins, only 30 shared the same relationship. When comparing the proteins that followed the same trend in both expression lists for the two chromatographic separations, the ratio between the control group and heart failure was greater for the 1D separations. That is to say a greater difference in protein abundance and thus ratio was observed with 1D analysis than the 2D analysis, as demonstrated in Table 4.3

4.3.5 Comparison of the diagnostic biomarkers identified

In this study, two different chromatographic platforms have been tested against one another using two biological patient sample groups. As mentioned previously, all proteins identified in the 1D expression analysis of the two patient groups were also identified in the 2D expression analysis. It would be expected that proteins identified in both types of analysis would follow the same trend but this was not true as highlighted above.

Table 4.3: This table shows all proteins identified at differing levels in control and heart failure plasma.

Protein Description	1D control: heart failure _Log(e)Ratio	2D control: heart failure _Log(e)Ratio
14 3 3 protein zeta delta	3.12	0.93
Actin cytoplasmic 2	2.07	0.93
Alpha 2 antiplasmin	2.13	0.29
Alpha actinin 1	2.49	0.99
Apolipoprotein D	1.31	0.86
Beta 2 glycoprotein 1	1.13	0.22
Beta Ala His dipeptidase	0.86	0.4
Cadherin 5	1.88	0.39
Carboxypeptidase N subunit 2	2.35	0.39
Cartilage oligomeric matrix protein	0.86	0.33
Glutathione peroxidase 3	1.5	0.11
Hemopexin	1.4	0.19
Mannan binding lectin serine protease 1	1.8	0.62
Phosphatidylinositol glycan specific phospholipase D	2.45	0.6
Platelet factor 4	-0.82	-2.11
Protein AMBP	-2.32	-0.63
Protein Z dependent protease inhibitor	1.88	0.57
Proteoglycan 4	2.14	0.25
Serum paraoxonase arylesterase 1	-2.59	-0.07
Thrombospondin 1	2.59	0.33
Tubulin beta 1 chain	1.76	0.69
Vitamin K dependent protein S	1.39	0.43

A comparison of the differences in the proteins identifications for 1D and 2D analysis was made. Proteins that were expressed at differing levels (i.e. up or down regulated) in the 1D-HDMS^E analysis were selected and the log(e) ratio for both 1D and 2D-LC-HDMS^E analysis for these proteins were compared. The results of the comparison of potential biomarkers are displayed in Figure 4.6.

Of the 69 proteins displayed, 32 proteins showed opposing relationships from 1D and 2D analysis and in most cases the ratio was not comparable between the two analyses. The remaining 37 proteins displayed a similar trend in the plasma ratio with weak correlation seen with the vast majority of the proteins. Platelet factor 4 was the only protein that showed a greater change in ratio for 2D-HDMS^E analysis rather than 1D.

After removing all high abundance proteins, 22 putative biomarkers were identified as markers of heart failure, using the expression analysis of both 1D and 2D analysis (Table 4.3).

Clearly both 1D and 2D-nanoUPLC-HDMS^E can be used as analytical platforms to aid in biomarker discovery and although there are differences in the protein profiles produced by both, similarities do exist as Table 4.3 shows. Software programmes are available that allow protein identification, relative quantitation and statistical analysis within the same data analysis process. Progenesis LC-MS (Nonlinear Dynamics, Newcastle upon Tyne, UK) is specifically designed to process data from multi-fraction experiments and allows comparisons to be made between samples at the peptide and protein level, reducing the level of human error in the data processing.

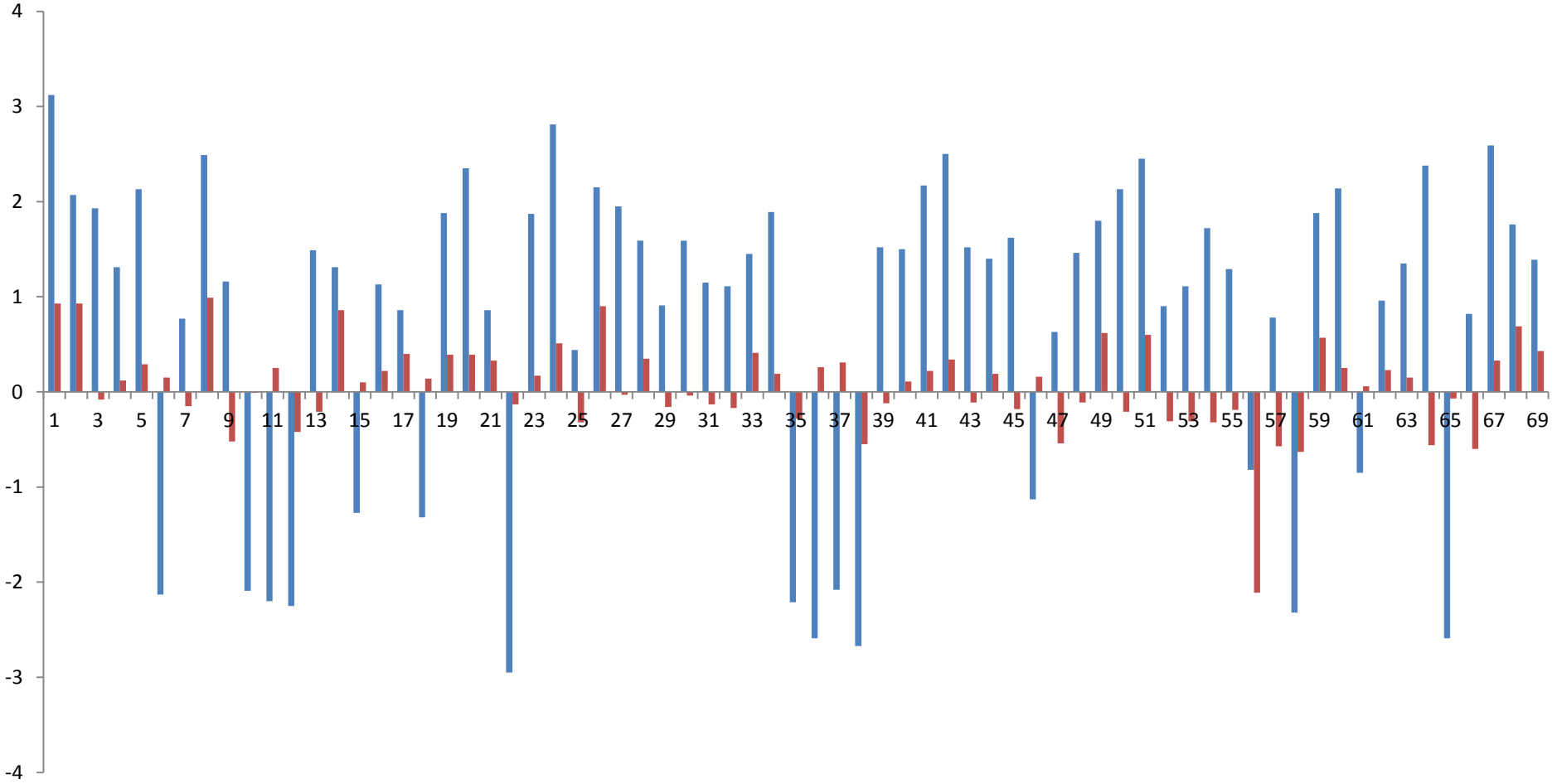


Figure 4.6: This graph shows the Log(e) ratio of both the 1D and 2D-LC-HDMS^{E}. The data displayed in the blue is the Log(e) ratio for 1D analysis, this is the calculated abundance ratio of pooled control to heart failure samples. The data displayed in the red is the Log(e) ratio for 2D analysis, this is the calculated abundance ratio of pooled control to heart failure samples.

4.4 Discussion

This study used two different chromatographic platform methods with two pooled samples from different biological groups: control and heart failure. Each pooled sample combined ten individual samples, once pooled the samples were equalized using the Proteominer^{TM} beads. The pooled samples were analysed using both single and multidimensional chromatography with high/low pH RP-RP separation. For the multidimensional chromatography a total of one hundred fractions were collected and concatenated into ten fractions to maximise the second dimension chromatographic space. This study found that the 2D platform outperformed the 1D with regards to peptide and protein coverage due to the increased separation space in 2D analysis, with similar sample loadings.

Although both the 1D and 2D platforms are suitable for biomarker discovery, 2D analysis offers an advantage in proteome coverage, there was a 3.6 fold increase in the number of peptides identified by 2D analysis in comparison to 1D. An important aspect of proteomic analysis is minimal introduction of technical variation. A comparison of the protein identifications found that 90% of the proteins identified by the 1D approach were also seen in the 2D approach. This not only meant a greater number of proteins were identified with 2D analysis, the dynamic range achieved with 2D analysis of pooled samples was much greater (Figure 4.4). The sample loadings were similar for all chromatographic experiments in this study, notably lower abundance proteins were observed with 2D analysis than 1D. The 2D experiment most definitely enabled a greater number of lower abundant proteins to be identified, as highlighted by Figure 4.4 and the 2 fold increase in the number of proteins identified.

The increased dynamic range seen in this study with the 2D approach has also been seen in previous studies. Patel *et al.* found an increase from 3 to 4 orders of magnitude with an 11 fraction 2D RP-RP approach by comparison with a 1D approach (Patel *et al.*, 2012).

Both the 1D and 2D approaches identified CRP however, only the 2D approach was able to identify CRP as a potential biomarker for heart failure. The 1D analysis did not deem the relative protein abundance change significant between the two biological

groups. Similarly, leucine-rich alpha 2 glycoprotein (LRG) was unidentified in the 1D pooled plasma but identified in the 2D analysis. A recent study by Watson *et al.* (Watson *et al.*, 2011) followed the workflow where-by affinity depleted serum was used and separated using two-dimensional difference gel electrophoresis. Bands which showed greatest difference in protein expression between the groups were excised from the gel, enzymatically digested and then analysed using a nano-LTQ orbitrap mass spectrometer. The data found LRG was over expressed in heart failure patients. Although the 2D analysis carried out in this study was able to identify it, the relative change in abundance was not deemed significant to suggest it may be a biomarker for this study. However, without the 2D analysis this protein would not have been identified at all. Again this shows the importance of increasing the separation space to increase the number of proteins identified in discovery proteomic projects.

The relative quantitative analysis of the internal standard ADH for the pooled plasma samples for in both 1D and 2D approaches was closely related to the calculated relative analysis, instilling confidence in the output.

There was good agreement with the protein identification observed between the 1D and 2D analysis. A number of proteins from both chromatographic analyses were identified unique to either biological group. There were also a large number of proteins which were identified at differing levels and many of these were at a statistically significant level. However, upon close inspection the results from the 1D and 2D analysis of the relative abundance did not closely correlate across the biological groups. A smaller number of proteins did correlate across the biological groups in both 1D and 2D analysis however; few of these were biologically significant. The 2D analysis was able to identify current biological markers used for heart failure diagnosis and also identify markers that have recently been proposed, highlighting the importance of multidimensional chromatography.

The biggest limitation of this study was the pooling of ten samples of each biologically group for 1D and 2D comparative studies which effectively gave an n = 1. The use of pooling samples in proteomic workflows has come into question. Pooling of samples could cause biological variations in individual patient samples to be lost and as a result

pooling relies on biological averaging where protein abundance averages out when pooled (Karp & Lilley, 2007). The main reason for sample pooling in proteomic studies is throughput, as analysis of individual samples using LC-MS/MS can be slow especially with proteomic workflows as one sample needs a 50-110 min LC gradient per sample, in triplicate this would be 2.5-5.5 hours per sample respectively. The slow throughput could place constraints on the number of biological replicates that can be processed reducing the statistical power of a study (Diz *et al.*, 2009).

In this study, Figure 4.5 compared the putative biomarkers identified in both 1D and 2D analysis. Only 37 of the 69 proteins agreed in the expression of the putative protein biomarkers between 1D and 2D analysis. It is possible that pooling of ten samples from each biological group to an n = 1 explains the opposing expression of the proteins identified. However, pooling of samples also enabled the throughput of this study to be reduced 10 fold. A compromise of throughput and pooling size has to be made. Pooling of samples in large scale proteomics studies is needed to increase the throughput. To decrease the biological averaging effects of pooling, several small pools of samples rather than one big pool of n = 1 is a better approach, as variance amongst pooled sample and outliers can still be identified (Karp & Lilley, 2007). Additionally, pooling of samples becomes necessary when large number of samples are analysed, as a single nanoLC-HDMS^E produces 8 GB files, making a small study such as this analysed in single samples rather than a pool would produce of total of 480 GB of data. This can be demanding on processing software.

A label free quantitative method has been tested using the same pooled samples from two biologically different groups on two different platforms, 1D and 2D RP-RP-HDMS^E. Both the chromatographic methods have been tested and compared and good qualitative agreement between the two has been observed. The quantitative agreement between the same sets was a lot lower than the qualitative agreement. Upon close inspection, it has been observed that the 1D approach missed some important known identifications in relative quantitative analysis and a greater dynamic range was achieved with 2D chromatography. The label free 2D-RP-RP-HDMS^E approach has shown that the use of the additional separation space has the potential to identify a greater number of proteins, confidently quantitate relative measurements over a wider

dynamic range. The limitation of n = 1 shall be addressed in Chapter Five, where a greater number of samples will be analysed using the label free 2D-RP-RP-HDMS^{E} proteomic workflow.

Chapter Five

Prognostic Heart Failure Study

5.1 Introduction

The previous chapters helped establish the proteomic workflow for this prognostic study. There are many different biomarkers that are used to help diagnose and monitor a patient's prognosis in heart failure, some of which are highlighted in Figure 5.1. Prognosis in heart failure relies heavily on recently discovered biomarkers derived from natriuretic peptides (Mukoyama *et al.*, 1991). BNP is an endogenous polypeptide hormone that is released upon cardio myocyte injury, stretch and hypoxia. This is then cleaved at the amino terminus to produce its inactive counterpart NT-proBNP. Both of these polypeptides are derived from the same precursor, proBNP. B-type natriuretic peptide (BNP) and N-terminal pro-hormone brain natriuretic peptide (NT-proBNP) are currently the most extensively studied and developed for clinical use. The plasma concentration of these biomarkers significantly increases above baseline levels upon myocardial stress. This makes them ideal biomarkers to measure for diagnosis and prognosis (Isaac, 2008; van Kimmenade & Januzzi, 2012).

Despite the success of the natriuretic peptides, the use of BNP and NT-proBNP as effective and accurate markers of heart failure has been criticised (Berliner *et al.*, 2009). All conditions/processes that cause an increase in ventricular myocyte stress cause an increase in levels of BNP. For example; atrial fibrillation, hypertension, aortic stenosis, unstable ischemic heart disease, right heart failure, diabetic nephropathy, hyperthyroidism, anaemia, declining renal function and severe sepsis are all known to cause an increased circulating BNP and NT-proBNP levels. The levels of circulating natriuretic peptide do not necessarily provide information on the mechanism, aetiology and intensity of myocardial stress (Lok *et al.*, 2013). There is an observed intra-individual variation in natriuretic peptides; for example, in a healthy population free of cardiac disease the background circulating levels of BNP are higher in women than in men, although this is not reflected in the patient population who present with cardiac disease. Obesity also affects the levels of BNP with those who have a higher BMI seem to have lower levels of BNP than those who are classed as having a healthy BMI (Chen *et al.*, 2010; Rocchiccioli *et al.*, 2010).

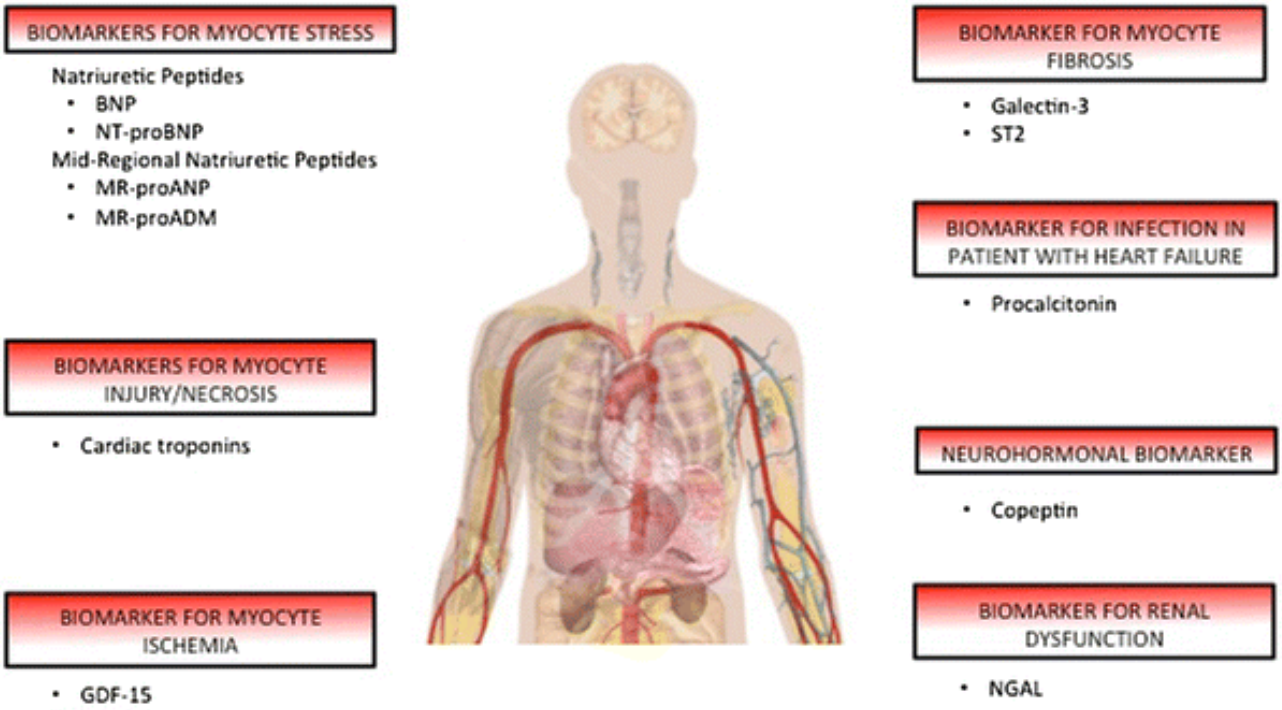


Figure 5.1: Systemic overview of heart failure biomarkers taken from (Choudhary *et al.*, 2013). The biomarkers displayed here are classified according to their function.

Heart failure is a multi-organ disease and the condition often emerges with impairments in multiple regulatory functions, from the renin-angiotensin-aldosterone system to cardiac response to inflammatory pathway activation (Ketchum & Levy, 2011). Therefore, a multi-marker strategy incorporating a variety of biomarkers discovered using proteomics, to understand the prognosis of a patient would lead to greater understanding of the disease and tailored medicinal approach.

Within this chapter, the need for a multi-marker strategy in understanding the prognosis of patients with acute heart failure is addressed by use of a label free 2D-RP-RP-HDMS^{E} proteomics approach. The objective of this study is to use the sample preparation method developed and refined in Chapter Three and the platform method developed and refined in Chapter Four, and apply this to a large acute heart failure sample cohort. The cohort contains samples from two distinct acute heart failure populations, survivors, and non-survivors. The survivors consisted of patients who remained event free for 1-year post initial diagnosis and non-survivors who *did not* remain event free for 1-year post initial diagnosis.

5.1.1 Prognostic biomarker discovery proteomic workflow

In this study, the proteomic workflow was designed to determine the protein profiles of acute heart failure survivors vs. acute heart failure non-survivors, to identify prognostic biomarkers in plasma. The samples were prepared using the equalizer beads and trypsinised, an additional separation stage using offline RP chromatography was added for another dimension of protein separation. The samples were analysed using a novel quantitative label free nanoUPLC-HDMS^E method. The data was analysed and quantified using bioinformatic software PLGS v2.5.2 and quantification was achieved using the Hi3 method where absolute protein concentration is determined by the use of an internal standard at a known concentration (Silva *et al.*, 2006).

One of the limitations of the previous diagnostic study in Chapter Four was the lack of relative quantitative agreement between the two chromatographic analyses using Expression^E. In this study, a new software package was used for the purposes of relative quantitation, Progenesis LC-MS as this programme came with the ability to align data at a peptide and protein level and was designed to process data from fractionated experiments. Alignment at a peptide and protein level was a novel method used for data processing in the discovery of prognostic biomarkers in acute heart failure using a discovery proteomics approach.

The aim for this study was four-fold:

1. Use plasma from two acute heart failure patients groups which had not been previously tested with the optimised sample method and platform method detailed in chapter 3 and 4 respectively to identify prognostic biomarkers.
2. Introduce the use of quality controls into the workflow to ensure machine performance remained consistent throughout analysis.
3. Characterise and quantify the equalised plasma proteins from the two acute heart failure conditions using a 2D-RP-RP-HDMS^E approach.
4. Identify novel prognostic biomarkers generated by this workflow.

5.2 Methods and Materials

5.2.1 Materials

Heart failure plasma was supplied by patients from the British Heart Foundation funded study entitled *"The uroguanylin system in heart failure"*, which recruited patients presenting with heart failure (for sample collection protocol see appendix). The samples were centrifuged and plasma was separated from the red blood cells and buffy layer. One hundred acute heart failure samples were selected for this study, age and sex matched. The samples formed two cohorts of patients, those whom remained event free for one year post initial diagnosis (survivors) and those who did not (non-survivors) who either had died or were re-hospitalised with heart failure within 1 year, labeled as the MACE (major adverse cardiac event). The human blood samples were collected in EDTA blood sample tubes donors following informed consent. The blood samples were centrifuged at 1500 x g for 20 minutes at 4 °C. The plasma layer was separated from the buffy layer and red blood cells, and stored at -80 °C.

ProteoMiner^{TM} Protein Enrichment Kit (Bio-Rad, Hemel Hempstead) was used which contained the PBS wash buffer and 1 mL equalizer bead columns. Digestion reagents dithiothreitol, iodoacetamide and proteomics grade trypsin were supplied by Sigma Aldrich (Poole, UK). Mass spectrometry standards Glu^{1}-fibrinoginogen peptide, alcohol dehydrogenase and enolase were purchase from Waters (Manchester, UK). Mass spectrometry solvents were all purchased at optima LC-MS grade from Fisher Scientific Ltd (Loughborough, UK).

5.2.2 Sample pooling

Samples from each patient cohort were pooled into groups of ten to reduce the sample analysis and data processing time. A total of a hundred samples were used, fifty belonging to the survivor and 50 to the non-survivor cohort. The pools of samples are detailed in Figure 5.1. The one hundred samples were randomly selected from a sample database using the statistical programme R, which generated fifty samples for each biological group.

Table 5.1: Acute heart failure prognostic pooling plasma sample study design. The sample lists of 50 survivors vs. 50 non-survivors was generated using a statistical programme (R) which provided the list of 50 sex and aged matched samples in each group using a propensity score. These have been further sub-divided into populations of 5 pooled groups of sample per biological condition.

Survivor group number	Sample Number	MACE		Non-survivor group number	Sample Number	MACE
Survivor 1	23	0		**Non-survivor 1**	10	1
	32	0			13	1
	43	0			18	1
	57	0			25	1
	81	0			38	1
	84	0			45	1
	87	0			49	1
	118	0			52	1
	129	0			53	1
	142	0			75	1
Survivor 2	150	0		**Non-survivor 2**	95	1
	163	0			106	1
	165	0			108	1
	168	0			112	1
	180	0			116	1
	199	0			124	1
	201	0			128	1
	203	0			131	1
	225	0			141	1
	228	0			151	1
Survivor 3	229	0		**Non-survivor 3**	166	1
	256	0			167	1
	258	0			169	1
	260	0			173	1
	261	0			175	1
	263	0			179	1
	272	0			183	1
	275	0			205	1
	282	0			208	1
	283	0			209	1
Survivor 4	284	0		**Non-survivor 4**	211	1
	285	0			214	1
	287	0			216	1
	289	0			226	1
	291	0			231	1
	292	0			235	1
	294	0			245	1
	297	0			246	1
	299	0			247	1
	300	0			255	1
Survivor 5	301	0		**Non-survivor 5**	277	1
	302	0			280	1
	304	0			286	1
	306	0			288	1
	307	0			295	1
	310	0			296	1
	311	0			305	1
	323	0			314	1
	324	0			316	1
	326	0			325	1

5.2.3 Sample preparation

Each plasma sample was thawed from -80 °C at room temperature and vortexed to ensure homogeneity and 100 µl aliquots were taken from each sample. The plasma aliquot taken from the survivor cohort was combined to give a 1 mL pooled sample; the same was repeated for the matched non-survivor cohort. A total of ten pooled samples were generated, five pooled survivor samples and five pooled non-survivor samples. Each sample was incubated with the equalizer beads and eluted with 2% SDC, as this was determined to be the optimum sample preparation method from chapter 3. The elutions were digested and lyophilised ready for analysis.

5.2.3.1 Fractionation of pooled plasma with concatenation

An offline method for fractionating the plasma using a RP column with a mobile phase of pH 10 in the first dimension was used. The fractions were collected from the column in 1.5 mL eppendorf tubes and protein detection was monitored at absorbance 214 nm. Fractions were collected at every 1 min over a 110 min period and concatenated into groups of ten as described in Table 2.2 in Section 2.8.2. A blank chromatographic run of 0.1% formic acid was used between samples to significantly reduce carry over between the heart failure and healthy controls samples. The concatenated samples were evaporated for 2 hours and lyophilised overnight. The samples were reconstituted in 0.1% formic acid and spiked with 100 fmol ADH at a 1:1 ratio. For each pooled sample, there were ten concatenated samples to be run in triplicate in total. Therefore, every pooled sample was fractionated into ten concatenated samples, leaving ten samples to be analysed in triplicate per pooled sample. Using this fractionation and concatenation method, each biological group had a total of 50 samples to be analysed by nanoUPLC-HDMS^E, in triplicate this was a total of 150 runs per biological group. The total number of nanoUPLC-HDMS^E runs for this entire study was 300, excluding all quality controls and blank runs.

5.2.4 Proteomic workflow

The proteomic workflow utilised in this study, is a slightly modified workflow from chapter four. Each stage of the approach is described in Figure 5.2. Blank washes were run between samples for 0.1% formic acid and an enolase was run periodically to check consistency of chromatography.

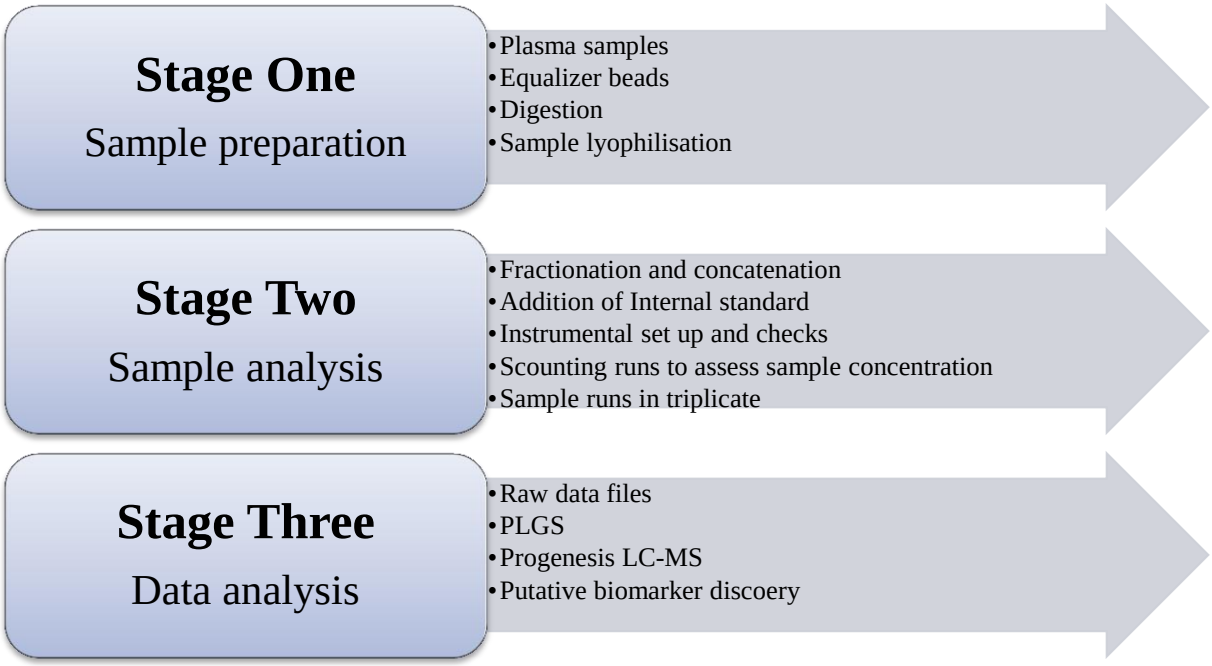


Figure 5.2: The schematic for the proteomic workflow undertaken for this study. There are three main stages, sample preparation that was developed in Chapter Three, sample analysis which was developed in Chapter Four and data analysis with the introduction of Progenesis LC-MS.

5.2.4.1 Quality controls

Implementation of QCs in proteomics studies is becoming increasingly important as the quantitative nature of such experiments increases and a standardised QC practice is yet to be established. The use of QCs can highlight the source of variation, which can be either accounted for or rectified across the proteomic workflow and ensure that the data produced from every study is accurate (Cairns, 2011; Kocher *et al.*, 2011).

In this study, a 5 µl aliquot of every sample produced for nanoLC-HDMS^E analysis was taken to produce a pooled QC sample. The QC sample was run at the start and end of each batch of samples. A batch included ten samples with ten 0.1% formic acid blank

runs. The QC samples were monitored throughout the study to ensure the number of peptides and proteins identified did not vary ±>10%. Assessment of the chromatographic performance was checked by running enolase every 24 hrs.

5.2.5 Processing of HDMS^{E} data

5.2.5.1 PLGS v2.5.2

LC-HDMS^{E} data was processed using PLGS v2.5.2 and the lockmass data was collected during the LC-HDMS^{E} runs through a reference spray for lockspray. The protein identification and assignment process has been described in detail in section 2.9.4, essentially preliminary precursor and product ions are time aligned and various data interrogation steps take place to ensure confident protein identifications. Data was processed using the following parameters: 15, 80, 750 for the low, elevated and intensity threshold respectively.

A uniprot human reviewed database was used and appended to include ADH (yeast). ADH was specified as the internal standard with the concentration specified in fmol, allowing for Hi3 quantification. Protein identification criteria was set to determine detection if the following criteria was met, at least two fragment ions per peptide, five fragment ions per protein and a minimum of one peptide per protein and the false discovery rate was set at 4%.

Fixed modification of carbamidomethyl cysteine and one missed tryptic cleavage site with variable modifications of oxidation of methionine, acetyl N-terminus, deamidation of asparagines and glutamine were all included in the search parameters of the HDMS^{E} workflow. All identifications made by PLGS were exported to Microsoft Excel, where further data interrogation and filtering of data to remove triplicates took place.

5.2.6 Relative protein expression

5.2.6.1 Expression^{E}

Relative protein expression was performed within PLGS v2.5.2 using the expression algorithm. The mass spectrometric runs were grouped according to their conditions (i.e. survivor or non-survivor), and the data was normalised within each group according to the spiked internal standard ADH.

PLGS removed all dubious proteins (i.e. proteins only observed in one sample run) and reported a ratio for each identification. The protein ratio was assigned a probability score between 0 and 1. Proteins were reported as up regulated, down regulated or unique. The up and down regulation was determined via p values that were either ≥0.95 (increased expression) or ≤0.05 (decreased expression) and deemed statistically significant. Unique proteins were only identified in either the heart failure or healthy control groups.

5.2.6.2 Scaffold

Raw data files were processed using PLGS v2.5.2 with all parameters specified above, the FDR was in this case set to 100%. All processed files were exported into scaffold and the data was analysed against the same database used for PLGS.

5.2.6.3 Progenesis LC-MS

Progenesis LC-MS was used for relative protein expression analysis. Raw data files for all samples were imported into Progenesis LC-MS and a reference raw data file was selected. The data file was set as the reference run for typical retention time and *m/z* profile of peptides from the study and all other chromatograms were aligned to this selected run. The quality of the data was reviewed and visually assessed by the alignment scores. Multiple filtering criteria could be utilised within the programme, for example peptides $< 1^+$ or $> 5^+$ were excluded, statistical analysis at the peptide level was also conducted. The peptide and protein identifications were imported from PLGS v2.5.2. The peptide identifications were reviewed and where there was a conflict in peptide identification (i.e. identified for multiple proteins), these peptides could be excluded. Then using the filtered peptide data, the protein identifications were assigned and protein quantification was based solely on unique peptides identified for that protein. Further filtering of the relative quantitative data was conducted based on p-values and fold change. Finally, a report of protein identification with quantitation was generated.

5.3 Results

5.3.1 Quality controls

The quality controls were run in two batches in this project, as during the mid-point of this study, a series of analytical platform checks were run to ensure consistency of performance. A QC was run after approximately 25 hours of clinical sample injections, using the same inlet method and MS method. The QC data was processed using the same PLGS threshold values (15, 80, 500 for the low energy, high energy and intensity threshold respectively).

The typical chromatogram obtained from a QC sample is shown in Figure . The figure displays a representative of the chromatograms seen throughout the study. The chromatograms and signal intensities were compared to assess the reproducibility of the mass spectrometer throughout the study and the retention time and intensity of a randomly selected ion (*m/z* 737.716) was compared using an extracted ion chromatogram.

The chromatograms for the QC's showed good reproducibility as the peptide distribution along the retention time window was similar for the QC's across the study, the peptides began to elute around 16.5 minutes and ended at approx. 44 minutes. The signal intensities for the base peak chromatograms were similar with an average intensity of $1e^5$ (Figure). Further analysis of the reproducibility was conducted by choosing a peak at random, in this case ion with *m/z* 737.716. The ion eluted at 28 minutes ± 9 seconds and the ion intensity $1.1e^6$. This shows good platform stability and performance throughout the study.

As mentioned above, the data from the QC's was processed using PLGS v2.5.2 and the same processing thresholds and parameters were set for samples as well as QC's. The first batch of QC's produced an average of 81 proteins and the second batch of QC's produced an average of 83 proteins with an RSD of 5.4% and 8.8% respectively. As the RSD is <10% this shows high level of reproducibility and this can also be seen in Figure 5.3.

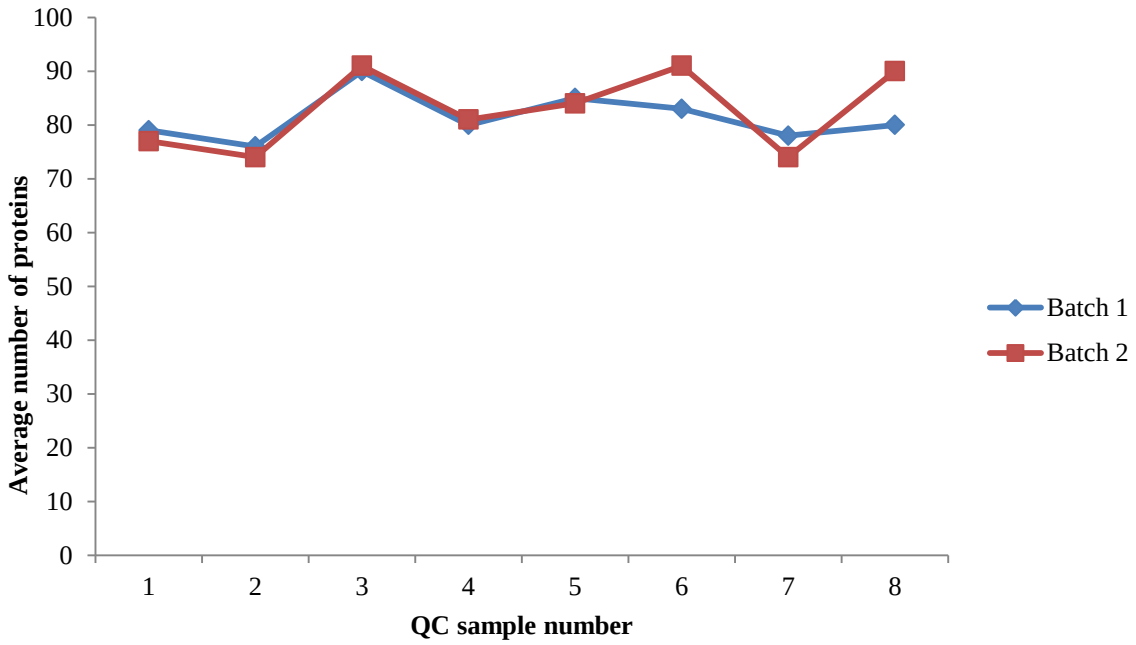


Figure 5.3: This graph show the number of proteins identified in the quality control sample for this study. The blue line indicates the first batch of QC's and the red the second batch of QC's

The proteins identified for all the QC's were compared manually in Excel and the data was filtered to remove proteins that were only seen in one QC sample. The first QC batch had 79 proteins that were identified in more than one injection and the second batch of QC's had 78 proteins that were identified in more than one injection.

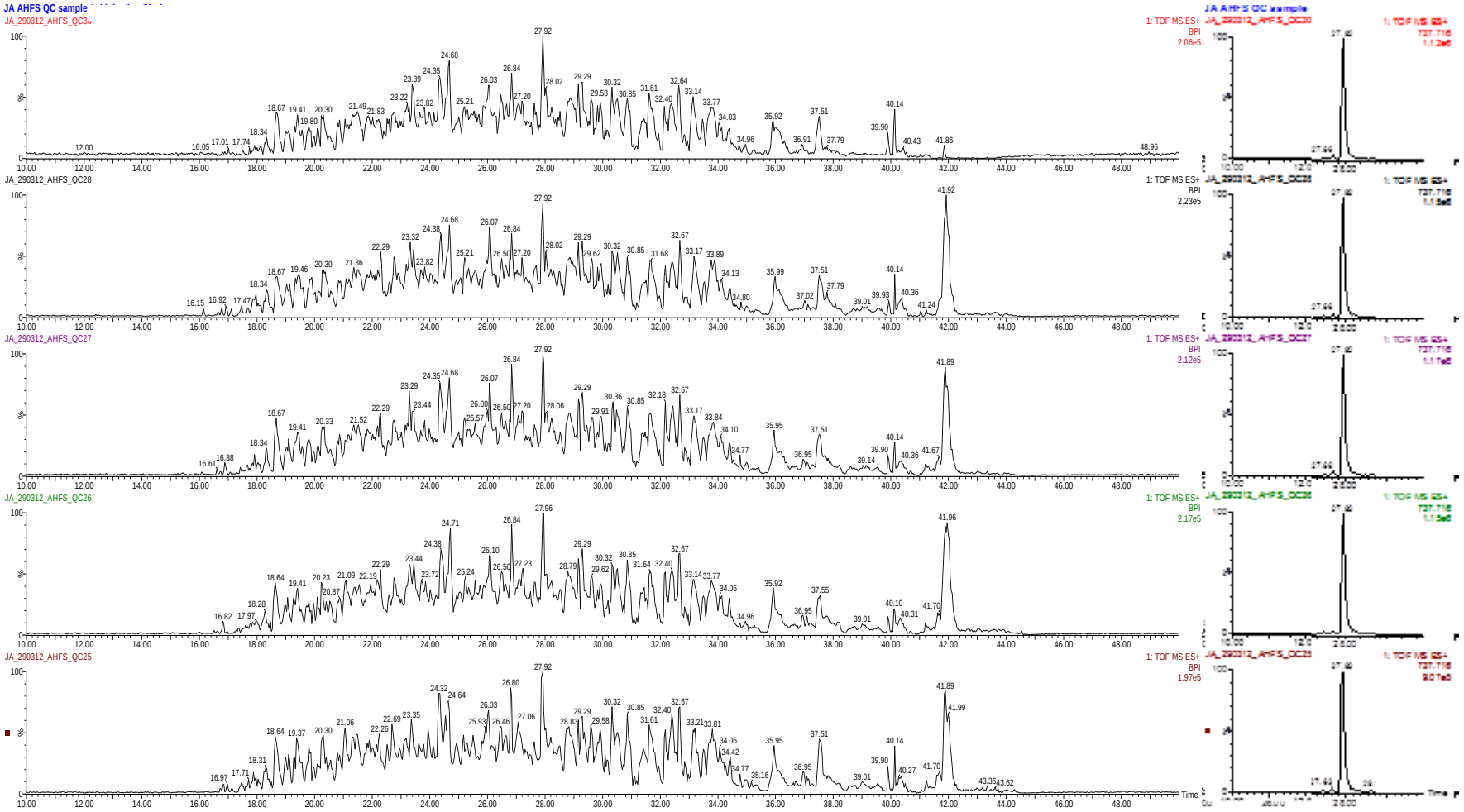


Figure 5.4: This figure shows an example of the chromatograms for the QC samples for this study. On the Left are the base peak intensity chromatograms for 2 QC samples and on the right is an extracted ion chromatogram for the ion 737.716 *m/z* The signal intensities for the chromatograms is $2e^5$ and $1e^5$ respectively.

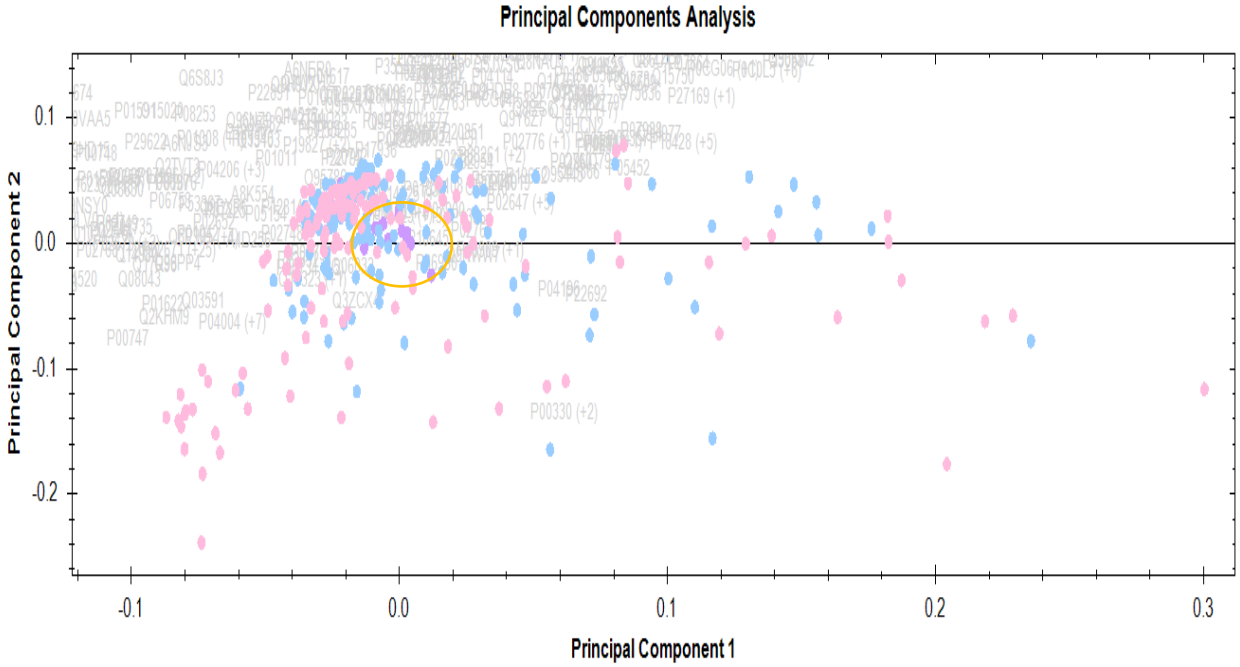


Figure 5.5: This shows a PCA plot of the two biological conditions; survivors in pink and non-survivors in blue and the QC's in purple. As can be seen, the QC's cluster in the same small area of the PCA plot, indicating high reproducibility.

A PCA plot of the QC samples generated in Progenesis-LC MS shows the QC's clustering in the middle of the two biological samples (Figure 5.5). The PCA plot shows the QC samples clustering within the same area of the graph, providing strong evidence of high reproducibility in data produced and proteins identified. The average number of peptides for both batches of QC's was investigated to identify if the same pattern seen at the protein level was repeated at a peptide level. There was an average of 1321 peptides for batch one and 1355 peptides for batch two. As Figure 5.6 shows the peptide level shows a very similar pattern of reproducibility at the protein level.

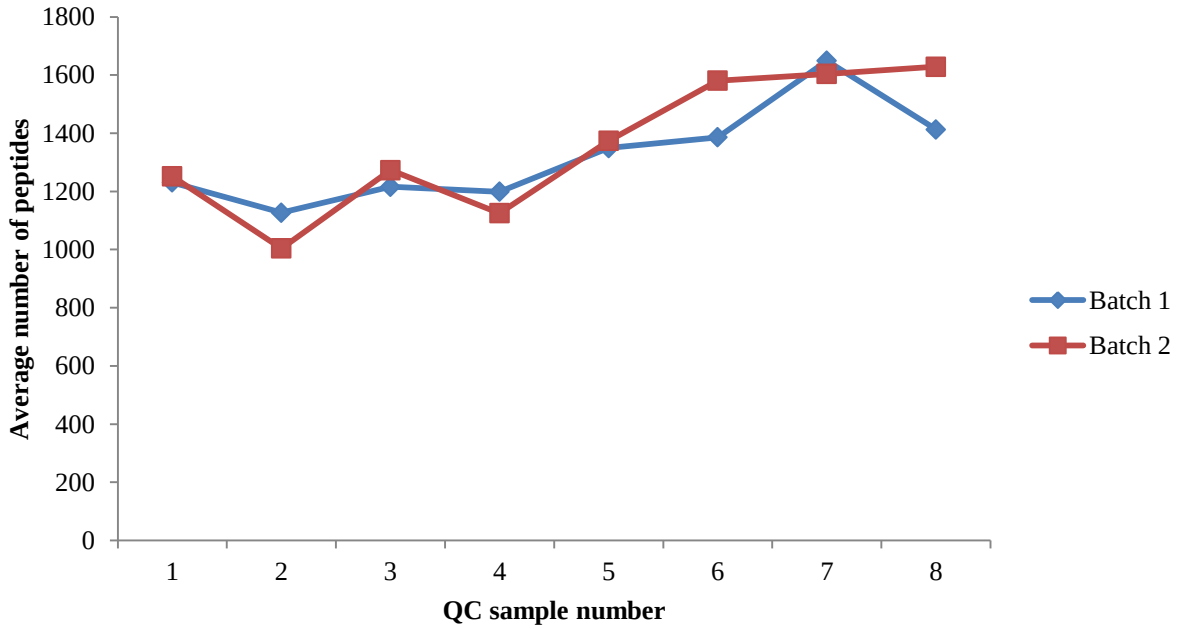


Figure 5.6: This graph show the number of peptides identified in the quality control sample for this study. The blue line indicates the first batch of QC's and the red the second batch of QC's

Seventeen proteins identified in all QC samples were compared. All seventeen proteins were selected at random and the number of peptides identified for the proteins across the sixteen QC samples was averaged and the standard deviation calculated (Figure 5.7). The graph demonstrates consistency in the platform performance and thus, the consistency throughout the study. The standard deviations show some degree of variation, this variation is due to technical (platform) variation but this platform variation still falls within 20% RSD.

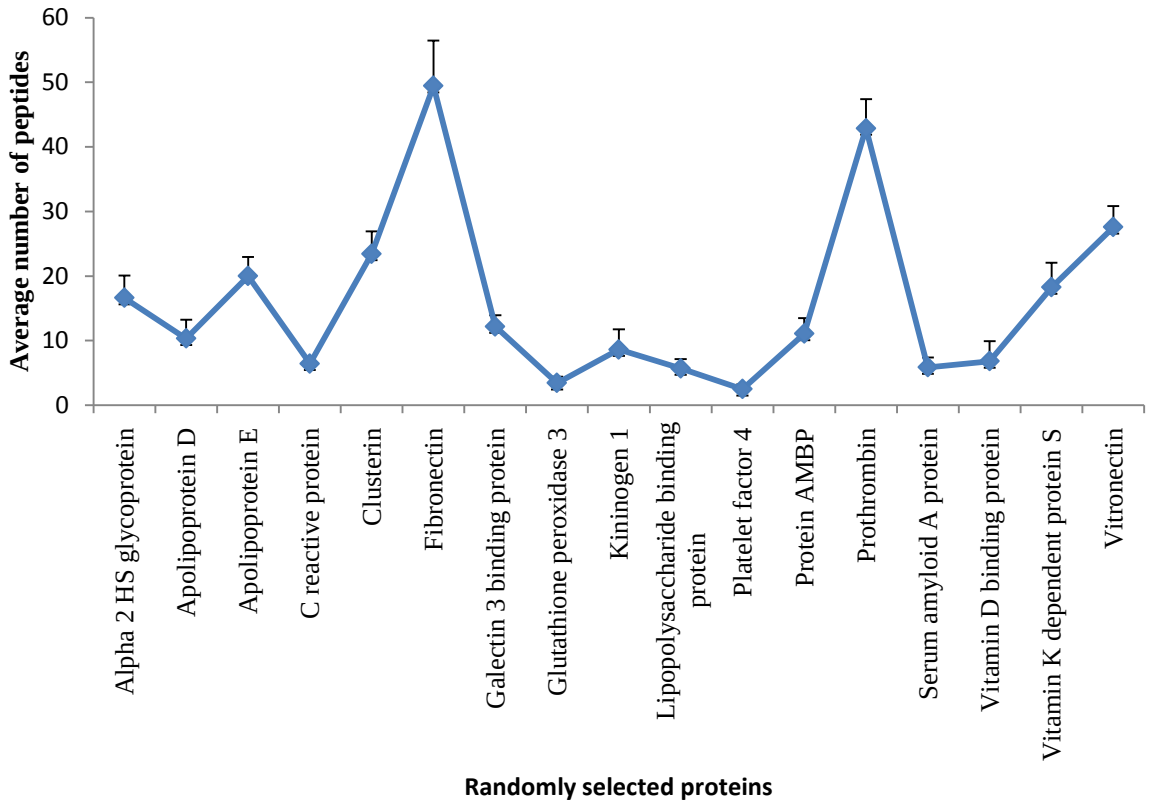


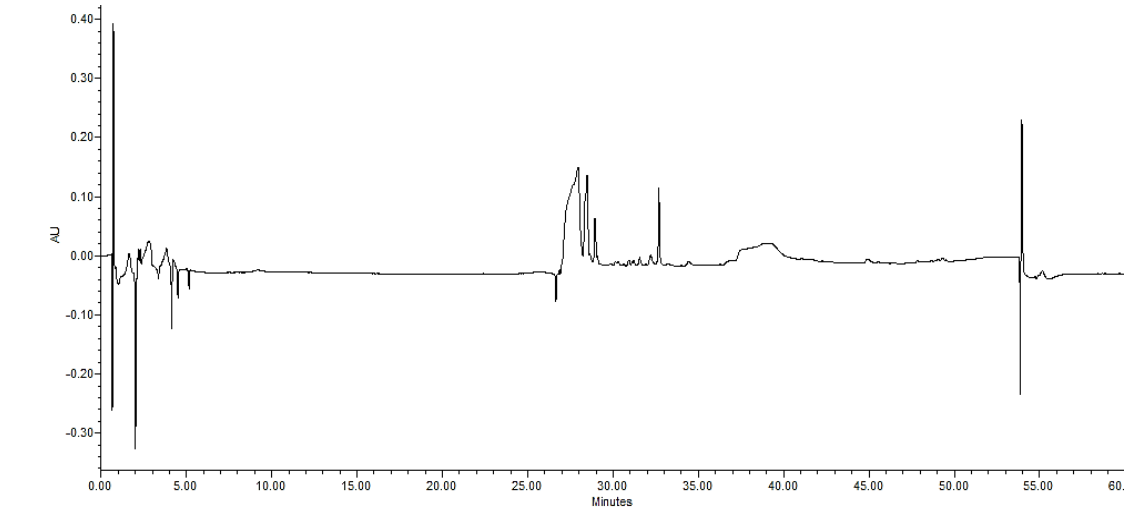
Figure 5.7: The graph shows the average number of peptides identified in 17 randomly selected protein in the sixteen QC runs. The error bars indicate the standard deviation of peptide identifications indicating the reproducibility of the method.

5.3.2 Protein identification of acute heart failure plasma using 2D-RP-RP-HDMS^E

A hundred acute heart failure plasma samples from two different conditions were pooled into groups of five. The samples were treated with the equalizer beads and eluted with 2% sodium deoxycholate. The samples were tryptically digested and fractionated over 110 min high pH RP HPLC run and 100 fractions were collected. The fractions were concatenated and analysed using a nanoLC-HDMS^E method. A typical chromatogram seen in the first and second RP separation is shown in Figure 5.8.

For each of the ten concatenated samples obtained from the pooled samples of the two biological conditions, an aliquot was taken and analysed using a LC-MS^E 'scouting' run.

A



B

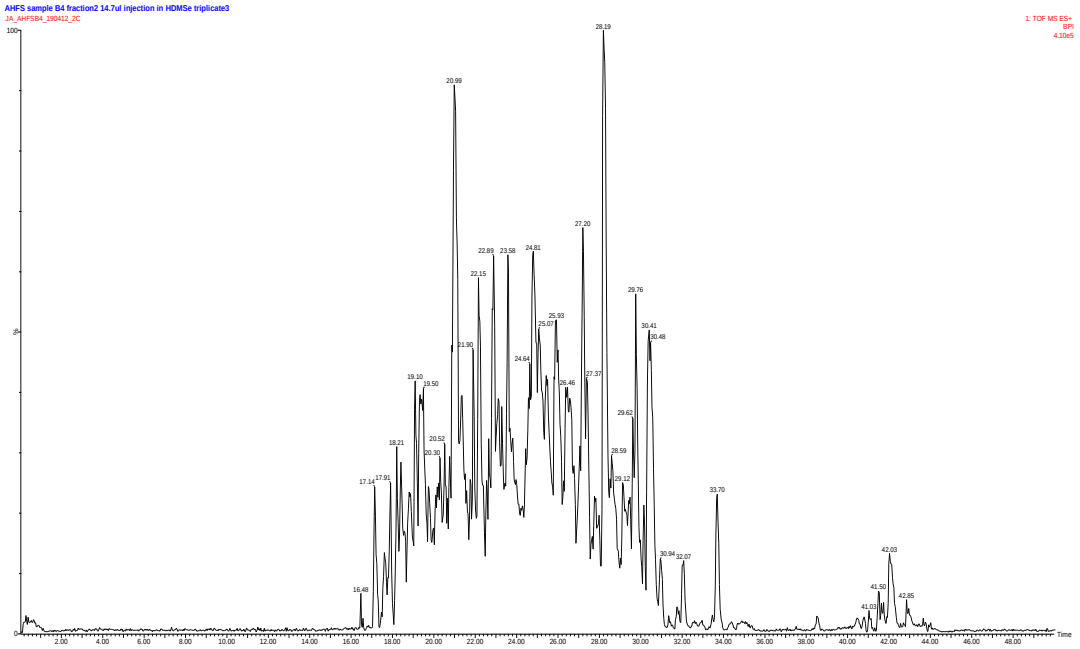


Figure 5.8: Typical chromatograms obtained from the high pH fractions (A) and then run on the conventional low pH RP LC-MS HDMS^{E} (B). A) The chromatogram on top shows the separation in the first high pH dimension at absorbance 214 nm. B) The chromatogram on the bottom shows the separation in the second low pH dimension (B).

Each aliquot was spiked with a known amount of internal standard (ADH typically at 50 fmol on column) and run at 1 µl of sample on column. The protein concentration was determined for each concatenated fraction per pooled sample using the Hi3 approach. Samples loading volumes for nanoUPLC-HDMS^E analysis were adjusted in accordance to the LC-MS^E protein estimation to ensure 1 µg of protein was loaded on column for all samples analysed.

A total of 300 data files were collected from the mass spectrometry analysis, amounting to 1.2 TB of data. This data was analysed for protein identification in PLGS v2.5.2 and relative protein quantitation in Progenesis-LC MS. A total of 25,336 proteins were identified by PLGS from 300 data files using a Uniprot database including all duplicates. To remove duplicate identifications fraction data files were merged to represent one file per pool using the merge function in PLGS v2.5.2, essentially leaving five merged protein identification files for the five pooled survivor population and five merged protein identification files for the five non-survivor population.

Using the merged data files in PLGS, the average number of proteins for each merged fraction in triplicate was calculated, on average 132 to 207 proteins were identified in each merged pooled sample (Figure 5.9). An average of 3,469 peptides were identified per merged pooled sample with an average of 19 peptides were per protein and an average sequence coverage of 30.6%. The relative standard deviation (RSD) was calculated for the number of proteins and peptides identified in each pooled sample in triplicate (shown in Figure 5.10). All pooled samples fall below 20% RSD, which demonstrates the study is reproducible and reliable.

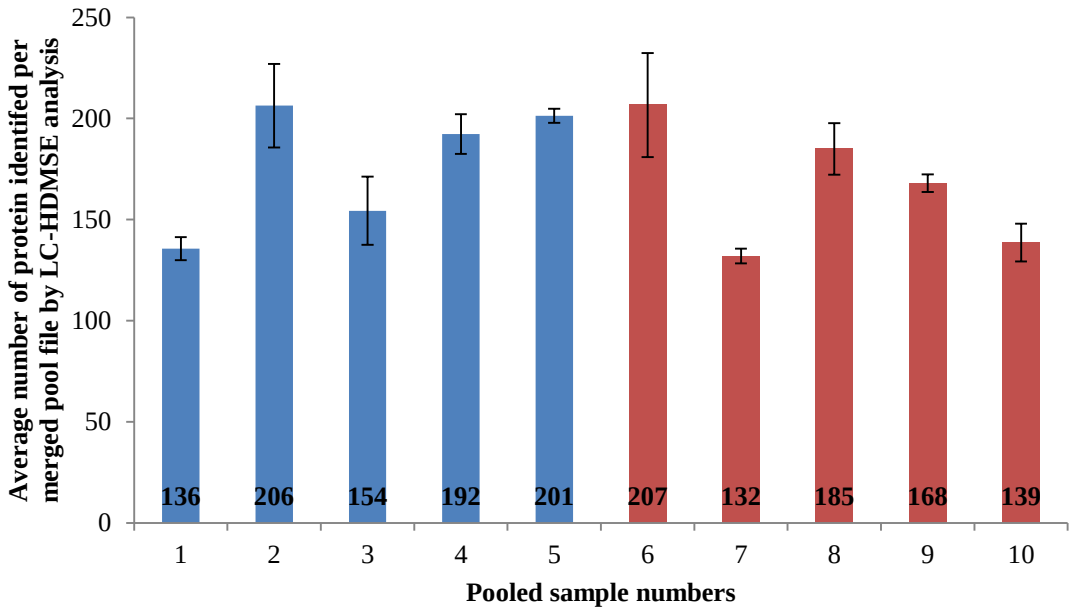


Figure 5.9: This graph shows the average number of proteins identified in the ten pooled samples using PLGS merged data files for each fraction. Samples 1-5 represent the fifty pooled survivor samples (BLUE) and samples 6-10 represent the fifty pooled non-survivor samples (RED). The error bars indicate the standard deviation of protein identifications indicating the reproducibility of the method.

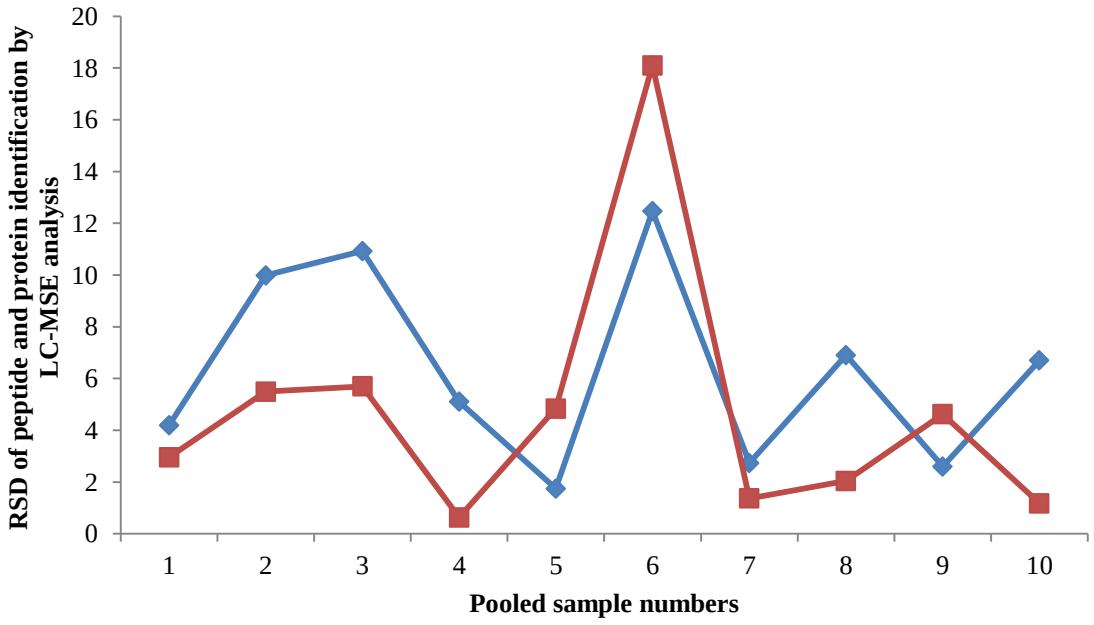


Figure 5.10: This graph shows the RSD for the number of proteins and peptides identified in the ten pooled samples. The BLUE represents the RSD of proteins and the RED line represents the RSD of peptides. Samples 1-5 (n=50) represent five pooled acute heart failure survivors and sample 6-10 (n=50) represent five pooled acute heart failure non-survivors.

5.3.2.1 Relative protein quantification of fractionated acute heart failure plasma

Expression analysis and Scaffold analysis was attempted to carry out relative quantitation on 300 data files produced from this study. However, due to the large amount of data produced (1.2 TB), both programmes were unable to analyse the data. Thus, Progenesis LC-MS was used to perform relative quantitation.

5.3.2.1.1 Progenesis LC-MS analysis of survivor vs. non-survivor

Three hundred samples were analysed for protein identifications in PLGS v2.5.2. Progenesis LC-MS has the ability to analyse fractionated samples and collate all the data from the individually aligned fractions. Briefly, a series of experiments within Progenesis LC-MS were created for each fraction i.e. for this study ten. Each experiment contained the sample runs from that particular fraction. These fractions were aligned at a peptide level within their own fraction experiments and protein identifications from PLGS were imported into Progenesis LC-MS for that fraction. Then a multi-fraction experiment was created where all the individual fraction experiments were collated into a single big experiment and allowed the overall comparison to be made.

In this study, ten fraction experiments were created for the ten concatenated fractions, these were analysed as ten individual experiments. Then an overall multi fraction experiment was created incorporating all ten fraction experiment data and samples were recombined according to their origin. The experimental design was described, survivors vs. non survivors, and the relative protein expression was obtained with the protein statistics function Figure 5.11.

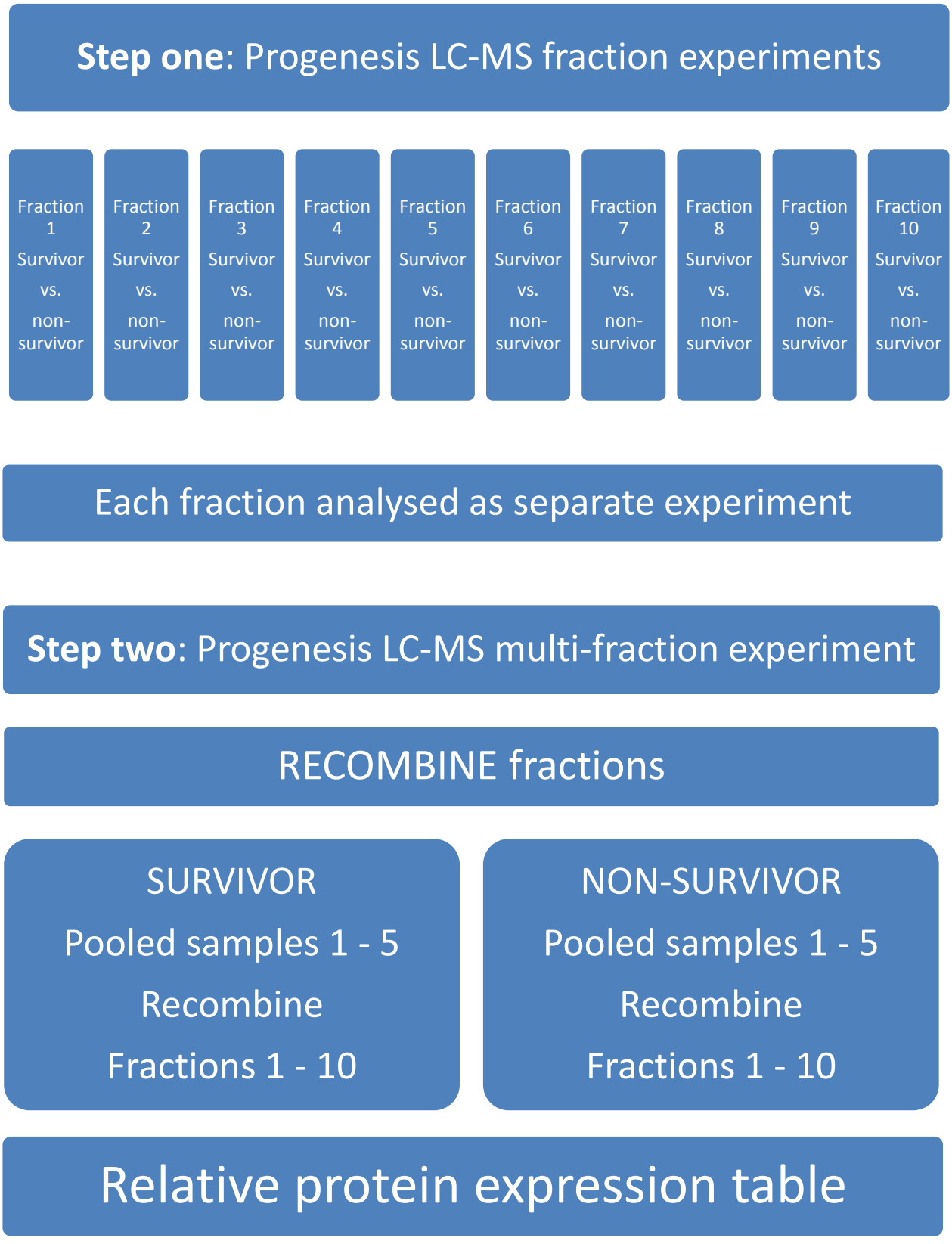


Figure 5.11: Informatic workflow for Progenesis LC-MS multi-fraction experiment. The data analysis pathway was used for this study.

All results in Progenesis were filtered to show proteins identified with ≥2 peptides per protein and exclude any protein identifications with a mass error greater than 10 ppm. After filtering, a total of 579 proteins were identified by Progenesis LC-MS, of which 220 were expressed higher in non-survivors, 287 higher in survivors and 72 were reported with no significant changes in expression (see appendix). The proteins identified from this study were from a variety of different biological function as shown in Figure 5.12. The majority of the proteins originate from cellular processes (15%) followed by proteins from metabolic processes (13%) and biological regulation (12%). PCA analysis in Progenesis LC-MS showed there were small clusters of proteins for each biological group (Figure 5.13). In this prognostic study, the disease pathology is identical but what differs is the severity of the disease. As a result, the overlap of proteins would be expected but it is the degree of expression of these proteins that would change and be indicative of prognosis.

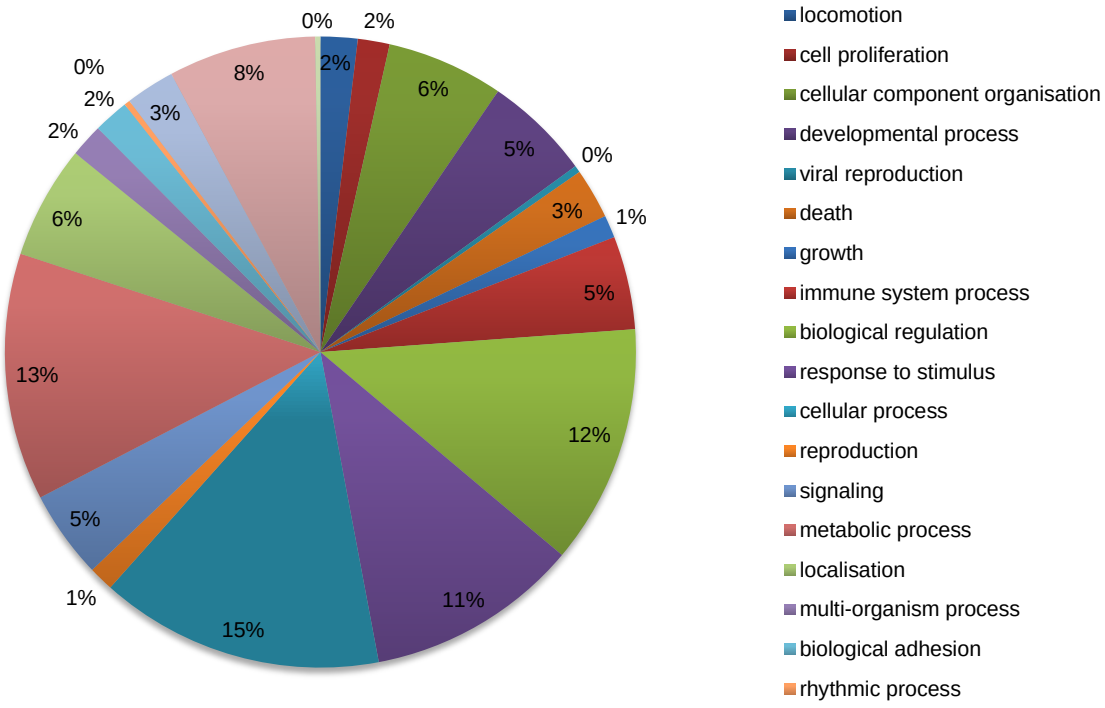


Figure 5.12: Pie chart of the proteins identified in this study and their biological functions. Blast-2-GO was the geneontology programme used (http://www.blast2go.org/).

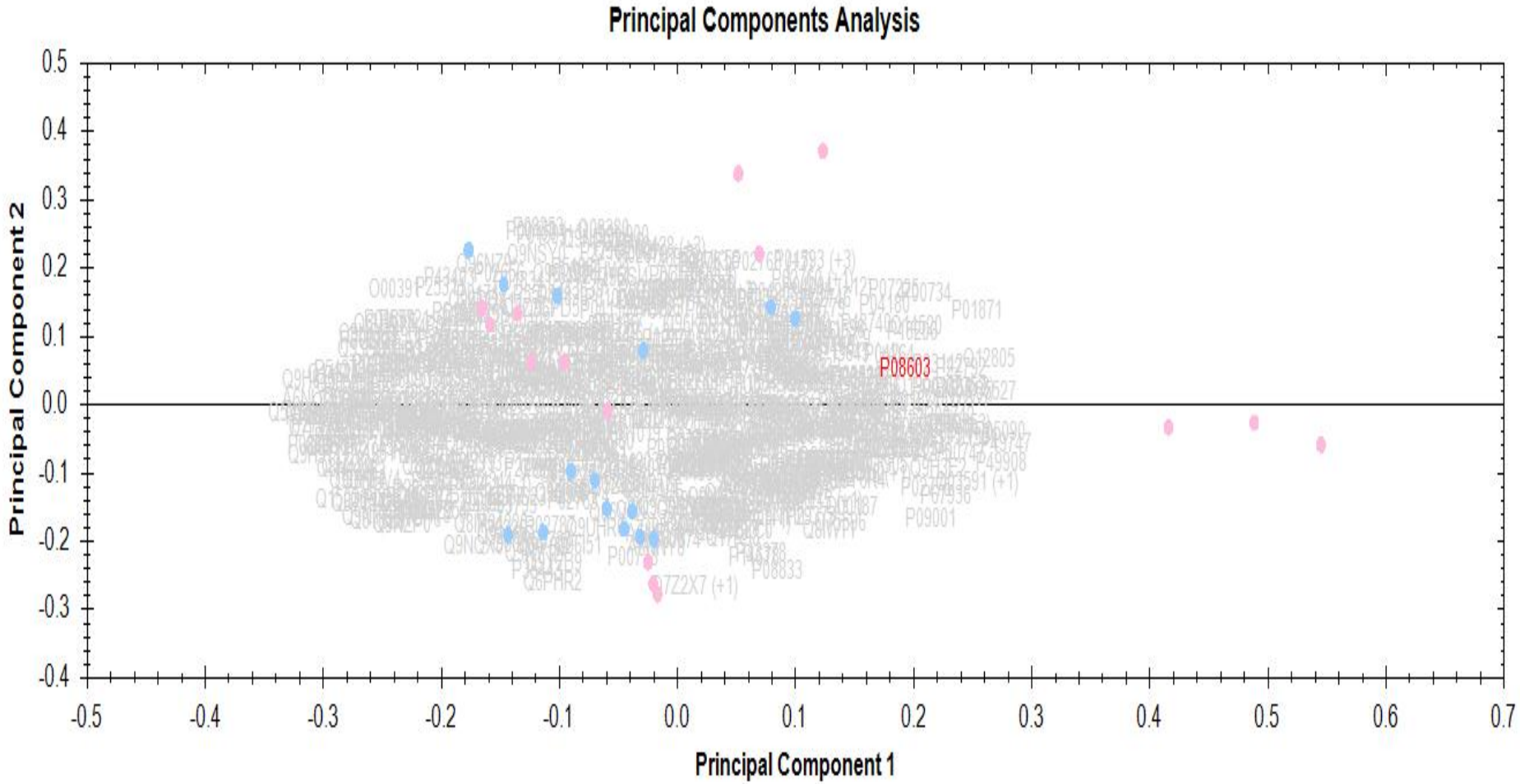


Figure 5.13: PCA analysis of the recombined multi-fraction experiment. The data in pink represents the survivors and the data in blue represents the non-survivors. This PCA analysis shows clustering of both disease states. The survivor group's protein profile shows the greatest difference in protein expression.

The multi-fraction experiment also identified which fractions contained the proteins listed in the final expression analysis. The number of times a protein was identified in a fraction was recorded and the majority of the proteins were identified in only 1 fraction, as shown in Table 5.2. A total of 291 proteins identified by progenesis LC-MS were only found in one fraction. However, the second highest number of proteins was identified in the 10th fraction, with a total of 85 proteins. The use of concatenation was to increase the spread of peptides to ensure maximum use of the second dimension's chromatographic separation space. Approximately 50% of the proteins identified were identified in more than one fraction. The spread of peptides and thus protein identifications implies the concatenation was successful.

Table 5.2: This table shows the number of proteins identified in either one or more fractions.

Number of occurrences	Number of proteins identified
1	291
2	63
3	28
4	22
5	23
6	22
7	17
8	13
9	16
10	85

Further filters were applied within Progenesis LC-MS to refine the protein list. A total of 87 proteins had a p value ≤ 0.05 and a total of 217 proteins showed a ≥ 1.5 max fold change in abundance. When the two protein filters were applied together a total of 36 proteins were identified (p value ≤ 0.05 and a max fold change ≥ 1.5), these are listed in Table 5.3.

Table 5.3: Progenesis produced proteins of significance (p value ≤ 0.05; max fold change ≥ 1.5)

Protein Description	Anova (p) (≤ 0.05)	Max fold change (≥ 1.5)	Highest mean condition	Protein Description	Anova (p) (≤ 0.05)	Max fold change (≥ 1.5)	Highest mean condition
15 hydroxyprostaglandin dehydrogenase	0.03	3.3	Non survivor	Kallistatin	0.02	2.2	Non survivor
A disintegrin and metalloproteinase with thrombospondin motifs 13	0.00	2.3	Non survivor	Leucine rich repeat containing protein 1 1	0.02	2.6	Survivor
Actin aortic smooth muscle	0.01	2.2	Survivor	Mannan binding lectin serine protease 2	0.01	1.9	Survivor
Alpha 1B glycoprotein	0.02	2.4	Non survivor	NADPH adrenodoxin oxidoreductase mitochondrial	0.01	5.0	Non survivor
Apolipoprotein A V	0.00	7.5	Non survivor	Neutrophil defensin 1	0.02	18.0	Non survivor
cAMP dependent protein kinase inhibitor gamma	0.03	3.2	Non survivor	Peptidyl prolyl cis trans isomerase H	0.02	2.9	Survivor
Caspase 3	0.00	2.1	Survivor	Phospholipid transfer protein	0.00	1.7	Non survivor
Cell differentiation protein RCD1 homolog	0.00	3.7	Non survivor	Putative histone lysine N methyltransferase PRDM6	0.05	2.4	Non survivor
Centrosomal protein of 95 kDa	0.00	9.5	Survivor	Rab GTPase binding effector protein 1	0.00	3.7	Survivor
Cholesteryl ester transfer protein	0.01	2.0	Non survivor	RAS guanyl releasing protein 2	0.01	10.0	Non survivor
Collagen alpha 1 XXII chain	0.03	4.5	Survivor	Ribulose phosphate 3 epimerase	0.03	2.6	Non survivor
Cystatin D	0.00	1.9	Non survivor	Serine threonine protein kinase Nek9	0.03	2.1	Survivor
Dual specificity protein phosphatase CDC14C	0.02	28.9	Non survivor	Serine threonine protein kinase ULK3	0.02	4.1	Non survivor
E3 ubiquitin protein ligase BRE1B	0.04	2.2	Survivor	Serum amyloid P component	0.00	2.1	Survivor
H ACA ribonucleoprotein complex subunit 1	0.02	2.4	Non survivor	Sulfatase modifying factor 2	0.04	2.2	Non survivor
Heat shock 70 kDa protein 1 like	0.03	3.1	Survivor	tRNA methyltransferase 112 homolog	0.03	2.6	Non survivor
Homeobox protein DBX2	0.02	4.2	Non survivor	Vesicle transport through interaction with SNAREs homolog 1A	0.01	11.6	Non survivor
Insulin like growth factor II	0.00	4.4	Survivor	Williams Beuren syndrome chromosomal region 16 protein	0.04	2.9	Non survivor

Table 5.4: The target proteins chosen from the multi-fraction experiment from Progenesis LC-MS. All proteins are thought to play a role in the disease pathology.

Protein Description	Anova (p) (≤ 0.05)	Max fold change (≥ 1.5)	Highest mean condition	Lowest mean condition
15 hydroxyprostaglandin dehydrogenase	0.03	3.3	Non-survivor	Survivor
Cholesteryl ester transfer protein	0.01	2.0	Non-survivor	Survivor
Cystatin D	0.00	1.9	Non-survivor	Survivor
Insulin like growth factor II	0.00	4.4	Survivor	Non-survivor
Mannan binding lectin serine protease 2	0.01	1.9	Survivor	Non-survivor
Phospholipid transfer protein	0.00	1.7	Non-survivor	Survivor
Serum amyloid P component	0.00	2.1	Survivor	Non-survivor

To refine the list of proteins in table 5.2, the role of the proteins were examined for biological relevance in acute heart failure disease pathology. The proteins chosen as potential candidate biomarkers for prognosis in acute heart failure are listed in Table 5.4. To ensure the chosen proteins were truly prognostic, the seven proteins chosen were also identified in all ten pooled samples.

5.4 Discussion

5.4.1 Identification of putative prognostic biomarkers

Seven proteins were identified as potential biomarkers for the prognosis of acute heart failure using a 2D-LC-HDMS^E approach; Progenesis LC-MS identified all proteins that were of significance (listed in Table 5.4). These proteins were all thought to relate to disease pathology, in this section the suitability of these proteins as prognostic acute heart failure biomarkers is discussed.

5.4.2 15 hydroxyprostaglandin dehydrogenase

In this prognostic study, 15 hydroxyprostaglandin dehydrogenase (15-PGDH) was found to be higher in abundance in the non-survivor group. NAD^+ dependant 15-PGDH is a catabolic enzyme and controls the biological activity of prostaglandins. The enzyme catalyses the oxidation of the 15-hydroxyl group found on prostaglandins, producing a

metabolite (15-keto) with reduced biological activity (Ensor & Tai, 1995). Prostaglandins play an important role in modulating homeostasis in cardiovascular disease and in acute inflammatory response (Ricciotti & FitzGerald, 2011). Prostaglandins have been shown to play a role in a variety of cardiovascular disease such as myocardial ischemia, hypertrophy fibrosis and atherosclerosis, all of which play a role in acute heart failure pathology (Yuhki *et al.*, 2011; Jabbour *et al.*, 2006).

Changes in 15-PGDH activity have been recorded in hormonal changes and hypertension, both apart of the biological process leading to heart failure (Ensor & Tai, 1995). 15-PGDH has been found to be down regulated in other biological conditions such as lung cancer. Ding *et al.* used quantitative RT-PCR analysis on tumour tissue and found a 2 fold decrease in 15-PGDH expression and in 61% of the samples a 10 fold decrease was observed (Ding *et al.*, 2005). In this study the opposite was observed, 15-PGDH protein expression was 3.3 fold higher in non-survivors in this study.

The relationship observed from this study was not what was observed in literature. We hypothesise the following role of 15-PGDH as an enzyme that acts to breakdown circulating prostaglandin. Prostaglandin has been observed as a vasodilator thus, lower levels of prostaglandin due to increased levels of 15-PGDH may lead to ineffective vasodilation in acute events such as acute heart failure (Kaufman Jr *et al.*, 1987).

5.4.3 Cholesteryl ester transfer protein

In this prognostic study, cholesteryl ester transfer protein (CETP) was found to be higher in abundance in the non-survivor group. CETP is a plasma glycoprotein synthesised mainly in the liver, that facilitates the transfer of cholesteryl ester from high density lipoproteins (HDL) to low density lipoproteins (LDL) and very low density lipoproteins (vLDL) (Weber *et al.*, 2010). Many studies carried out on large populations have suggested increased risk of a cardiovascular event with higher circulating levels of LDL and vLDL and studies have shown reducing LDL levels by statin treatment is cardio-protective (Cholesterol Treatment Trialists' (CTT) Collaboration *et al.*, 2010; Emerging Risk Factors Collaboration *et al.*, 2009). It has been suggested that inhibition of CETP could increase levels of circulating HDL which would in turn reduce the development of atherosclerosis, which pathologically leads to heart failure.

There has been conflicting evidence for the relationship between CETP and cardiovascular diseases such as heart failure. CETP is expressed in very few species such as humans and rabbits but is absent in rodents. Studies involving transgenic mice have shown CETP to be both pro-atherogenic and anti-atherogenic. However, rabbit lead studies have shown CETP inhibitors to markedly reduced atherosclerosis (Barter & Rye, 2012).

Vasan *et al.* measured plasma CETP levels in the 1978 Framingham Heart Study participants using a radio isotopic CETP Activity Assay Kit. Multivariable statistical analysis of the data found an inverse association of CETP activity with heart failure (Kenchaiah *et al.*, 2004). Alssema *et al.* measured CETP plasma levels by immunoassay in 566 patients and found that high CETP levels was associated with increased cardiovascular disease in women with type 2 diabetes (Alssema *et al.*, 2007). Boekholt *et al.* conducted a similar study on plasma obtained from the EPIC study. The cohort comprised of healthy men and women who developed coronary artery disease (associated with heart failure pathogenesis), 755 plasma samples were tested for CETP plasma levels via immunoassay. This study found that patients who presented with elevated triglycerides levels confirmed that increased concentrations of CETP were associated with increased risk of coronary heart disease (Boekholdt *et al.*, 2004).

In this study, CETP was found to be 2 fold higher in non-survivor plasma. Although the role of CETP and heart failure remains unconfirmed, this study suggests that an increased level of CETP is associated with worsening prognosis.

5.4.4 Cystatin D

In this prognostic study, cystatin D was found to be higher in abundance in the non-survivor group. Cystatins are proteolytic enzymes that cleave peptide bonds with an active cysteine residue at their catalytic sites; they regulate the cysteine proteinases secreted by diseases cells inhibiting harmful effects of cysteine proteinases (Shah & Bano, 2009). Cystatins inhibit cathepsins such as B, H, K, L and S, which are implicated in many different disease states, including tissue destruction due to

inflammation (Alvarez-Diaz *et al.*, 2009). Thus, cystatins have an important role in regulating cysteine proteinases and are important in pathological diseases.

A member of the cystatin family, cystatin C, has been identified as a prognostic marker in acute heart failure. Lassus *et al.* studied the plasma of 480 patients admitted with acute heart failure and cystatin C was measured using an immunoturbidimetric assay. Multivariable statistical analysis was done on the data produced by the study and it was found that high levels of cystatin C was associated with a three-fold increase in a 1 year mortality risk (Lassus *et al.*, 2007). Cystatin C has also been shown at higher levels in patients with heart failure in other studies (Carrasco-Sanchez *et al.*, 2011). The role of cystatins in heart failure as has been speculated. It has been suggested that in a disease state, cystatins are released to moderate the cathepsins derived from inflammatory cells (Shah & Bano, 2009).

Cystatin D was 1.9 fold higher in non-survivors, predicting an adverse prognosis. Due to the strong correlation of cystatin C and heart failure, and the role of cystatin's in maintaining homeostasis, cystatin D may also play a role in acute heart failure pathology. As our results suggest, cystatin D could be a prognostic indicator of adverse outcome.

5.4.5 Insulin like growth factor II

In this prognostic study, insulin like growth factor II (IGF2) was found to be higher in abundance in the survivor group. Insulin like growth factor's play a key role in producing rapid metabolic change and have a long term growth effect on cell proliferation. IGF2 is a polypeptide of 67 amino acids and primarily synthesised in the liver where it is released into circulation (Nielsen, 1992).

Watanabe *et al.* measured the serum level of IGF 1 and IGF binding protein 3 in 142 patients using immunoassay techniques. Statistical analysis found that reduced level of IGF 1 in heart failure patients, especially those with advanced heart failure (Watanabe *et al.*, 2010). The authors suggest IGF1 regulates the glucose metabolism and altered glucose metabolism has been associated with heart failure. They also note that IGF1 plays a regulatory role in cardiomyocyte metabolism, transporting insulin sensitive

GLUT4 to the cell surface, acting as a compensatory mechanism in heart failure. Similar findings of low IGF1 levels with increased heart failure risk were also seen in a study by Niebauer *et al.* (Niebauer *et al.*, 1998).

In this study, IGF2 was found to be a higher in survivors, expressing 4.4 fold more IGF2 than non survivors. It is unclear whether IGF2 may play a similar role as IGF1 in heart failure, but IGF's role in producing rapid metabolic changes may explain the over expression in survivors. Thus, the increase in IGF2 in the survivor population may be a marker of metabolic compensation better achieved in the survivor population than the non-survivor.

5.4.6 Mannan binding lectin serine protease 2

In this prognostic study, Mannan binding lectin serine protease 2 (MASP2) was found to be higher in abundance in the survivor group. MASP-2 is a serum protease that plays an important role in the activation of the complement system via mannose-binding lectin (MBL). After activation by auto-catalytic cleavage it cleaves complement C2 and C4, leading to their activation and to the formation of C3 convertase, activating the complement cascade (Sorensen *et al.*, 2005).

Chen *et al.* found that MASP-2 could only interact with C4 once it complexes with MBL, in a sense MBL acts as an inhibitory function of MASP-2, leaving it unable to bind C4 and initiating the complement cascade (Chen & Wallis, 2004). Therefore, higher MASP-2 concentrations in circulation may indicate the MBL regulation and inhibition of complement activation. Frauenknecht *et al.* explored the role of MASP-2 in cardiovascular disease, due to the role of complement in pathogenesis. The authors measure MASP-2 levels in the plasma of patients with coronary artery disease, myocardial infraction and acute ischemic stroke with matched controls. MASP-2 levels were found to be lower in myocardial infarction and stroke patients in comparison to healthy controls and coronary artery disease patients (Frauenknecht *et al.*, 2013). The study concluded that further research was needed to fully understand the role of MASP-2 in cardiovascular diseases, especially due to differing significance of MASP-2 in different cardiovascular diseases. For example, Schweable *et al.* reported that MASP-2 deficient mice conferred protection from myocardial ischemia (Schwaeble *et al.*, 2011).

In this study, we found that MASP-2 was expressed higher in survivors. A 1.9 fold increase was observed in comparison to non-survivors. MASP-2 plays different roles in cardiovascular diseases; clearly the underlying mechanism of the protein's function needs to be determined. Lower levels of MASP-2 have been shown to be indicative of myocardial infarction and stroke patients implying higher levels are protective. However, the opposite has been found to be true of myocardial ischemia.

5.4.7 Phospholipid transfer protein

In this prognostic study, phospholipid transfer protein (PLTP) was found to be higher in abundance in the non-survivor group. PLTP is a lipid transfer/ lipopolysaccharide binding protein that is ubiquitously expressed. Functions include the modification and metabolism of HDL and removal of phospholipid and cholesterol from peripheral tissue. PLTP has been highly associated with the development of atherosclerosis, which leads to cardiovascular diseases (Jiang *et al.*, 2012; Wirtz, 2006).

Robins *et al.* studied plasma of 2679 participants of the Framingham study and measured the activity of PLTP with CETP using fluorometric assays. Statistical analysis of the data found that high PLTP levels had a significant effect on the development of cardiovascular diseases (Robins *et al.*, 2013).

Cheung *et al.* studied the effect of PLTP in acute inflammation using human blood samples along with other acute inflammatory markers measured in cardiovascular disease. PLTP activity was measured using a radioimmunoassay, along with other acute phase inflammatory markers, CRP, Serum amyloid A (SAA) and fibrinogen. The study found that PLTP, CRP and SAA were strongly correlated in cardiovascular disease patients. The authors also report an increase in PLTP levels in acute inflammation and this is also strongly correlated in the CRP levels of the same patient cohort (Cheung *et al.*, 2006).

These two studies strongly suggest high levels of circulating PLTP are associated with a poor outcome. In this study, we found that PLTP was expressed higher in non-survivors (1.7 fold increase in comparison to survivors). The association of PLTP with increased

cardiovascular disease risk and the strong correlation with current acute inflammatory markers used in clinical monitoring, suggests PLTP could be a significant prognostic marker both on its own and in conjunction with existing clinical biomarkers.

5.4.8 Serum amyloid P component

In this prognostic study, serum amyloid P component (SAP) was found to be higher in abundance in the survivor group. SAP is a plasma glycoprotein and is a part of the pentraxin family of proteins that play a role in innate immunity (Emsley *et al.*, 1994). Similar to CRP, SAP shares 66% homology and is also secreted by the liver with a half-life of 24 hours. Unlike CRP SAP is mildly affected by inflammation and the role of SAP in instigating an inflammatory reaction is debated (Pepys *et al.*, 1997; Ying *et al.*, 1993). Although it has been argued that SAP is an acute phase reactant in mice (Pepys *et al.*, 1979), the lack of serum concentration elevation in humans would suggest SAP does not play a significant role in the inflammatory response (Koenig, 2007).

The exact function of SAP in heart failure or cardiovascular diseases is also highly debated. Many different studies have found different roles of SAP. A study by Jenny *et al.* conducted multivariable analysis on patients with cardiovascular disease and monitored SAP adjusted with cardiovascular risk factors. SAP was found to be associated with angina and myocardial infarction but not stroke or cardiovascular death. Interestingly, the authors found that CRP was significantly associated with cardiovascular death and SAP was not. They conclude that although CRP and SAP are both from the pentraxin family and function in innate immunity, CRP and SAP are very different (Jenny *et al.*, 2007). SAP does not react in the manner as CRP in acute inflammation and both proteins play different roles in lipid binding. SAP has been shown to bind HDL and vLDL but not LDL cholesterol (Li *et al.*, 1998) whereas, CRP binds LDL and oxidised LDL and is thought to play a role in lipid metabolism, clearance and deposition (De Beer *et al.*, 1982).

Based on the lipid binding interactions, a role for SAP in atherosclerosis has been suggested. SAP binds to amyloid like structures in LDL, blocking macrophage uptake of SAP bound LDL, preventing atherosclerosis (Stewart *et al.*, 2005). Haudek *et al.* conducted a study using cardiac fibroblasts from mice and administered SAP *in vivo*.

The authors attempted to identify if SAP had an effect on the pathology of cardiac fibrosis. They found that SAP totally inhibited cardiac dysfunction in fibrosis (Haudek *et al.*, 2006).

In this study, SAP was identified in pooled survivor and non-survivor acute heart failure plasma samples analysed using 2D-LC-HDMS^{E}. SAP was reported as higher in survivors than non-survivors, with max fold change of 2.1 between the two conditions. As various studies have shown SAP may play a protective role in atherosclerosis, a process which leads to syndromes such as heart failure. The lack of association of SAP with cardiac death suggests elevated SAP could be a compensatory mechanism, which is associated with good prognosis.

5.4.9 Identification of current biomarkers

C-reactive protein (CRP) is a nonspecific inflammatory marker released from the liver in acute inflammatory response to cytokines, namely interleukin 6. The levels of CRP can be raised up to 100 fold within 48 hours of an inflammatory response and is also associated with atherosclerosis (Casas *et al.*, 2008).

Elster *et al.* first linked heart failure and CRP levels in 1956, whereby 40 of the 30 patients tested showed increased CRP levels and those with high levels of CRP, heart failure appeared to be more severe (Elster *et al.*, 1956). A more recent study by Muller *et al.* studies the plasma of 214 heart failure patients using immunoassays and found that increased CRP levels was a long term mortality marker of acute heart failure patients (Mueller *et al.*, 2006). Similar findings were also seen in a study by Siirila-waris *et al.* The authors found a 27.4% mortality rate in 1 year mortality study that in 620 patients whom presented with acute heart failure. The authors concluded that age, male gender, low systolic blood pressure on admission, renal dysfunction and CRP were all independent prognostic risk factors. In this study elevated CRP was associated with worse prognosis (Siirilä-Waris *et al.*, 2006).

However, the reactionary nature of CRP to all acute inflammation calls into question the specificity of the marker for prognostic value. As acute heart failure is a multi-organ disease it is hard to distinguish elevated levels of CRP due to cardiac decompensation

or acute damage to other organs. CRP's response to IL-6 causes a knock on effect on other proteins involved in the inflammatory response and can cause cardiac tissue damage (Araujo *et al.*, 2009).

In this study, CRP was identified in pooled survivor and non-survivor acute heart failure plasma samples analysed using 2D-LC-HDMS^E. CRP was identified in all 10 fractions therefore, the peptides must have eluted across the length of the chromatographic run. CRP was reported as higher in non-survivors than survivors, which would be expected however, the progenesis LC-MS analysis found CRP to not be significant (p value 0.29) and the maximum fold change was only 1.25 between the two conditions. As CRP is one of the routine biomarkers used in clinical laboratories to monitor patients, the Progenesis LC-MS analysis shows the need for new prognostic biomarkers that are more sensitive and specific to the disease pathology. This also advocates the use of multiple protein biomarkers in such complex multi-organ syndromes.

5.4.10 Comparison of Diagnostic and Prognostic Biomarkers

Chapter Four focussed on identifying diagnostic biomarkers of acute heart failure using two pooled samples (n=1 for each biological group) and Chapter Five aimed to identify biomarkers of prognosis in acute heart failure. In total five proteins were identified as putative biomarker in both the diagnostic and prognostic studies. These proteins included *a disintegrin and metalloproteinase with thrombospondin motifs 13, Actin aortic smooth muscle, C-reactive protein, Insulin like growth factor II, and Serum amyloid P component* (See Table 5.5).

Four of the five proteins showed correlation of results. The only protein that showed a different protein abundance pattern was a disintegrin and metalloproteinase with thrombospondin motifs 13. The discrepancy in results can be explained due to differences in the proteomics workflow and samples. Although the platform methods for both the diagnostic study and prognostic study were the same (2D-LC-HDMS^E), the number of samples used differed. Ten samples were pooled to give and n = 1 for two biological groups for the diagnostic study, whereas fifty samples were pooled to give an n = 5 for two biological groups for the prognostic study. The software used for both

studies for protein abundance comparison also differed, for the diagnostic study Expression^{E} was used to compare protein abundance whereas for the prognostic study Progenesis was used. This could explain the discrepancy in results.

Table 5.5: This table compare the proteins identified by both the diagnostic and prognostic acute heart failure studies in Chapters Four and Five. The table compares the highest mean conditions or the group with the highest abundance of that particular protein for both studies.

	Diagnostic study	**Prognostic study**
Protein name	**Highest mean condition**	
A disintegrin and metalloproteinase with thrombospondin motifs 13	Control	Non-survivor
Actin aortic smooth muscle	Control	Survivor
C-reactive protein	Disease	Non-survivor
Insulin like growth factor II	Control	Survivor
Serum amyloid P component	Disease	Survivor

5.5 Summary

Poor prognosis is a feature of heart failure. The Framingham study (US) and the Hillingdon heart failure study (UK) both calculated a 1-year survival rate for heart failure was 70%, from initial diagnosis. Only 35% of patients showed a 5-year survival rate following heart failure diagnosis in the Framingham study (Mosterd & Hoes, 2007). The poor prognosis associated with acute heart failure and the ineffective treatment strategies based on current biomarkers necessitates the introduction of new more specific biomarkers of prognosis. The biomarkers will allow clinicians the ability to grade the severity of disease and tailor therapy to the patient's needs for a more personalised medicine approach. The use of proteomics for plasma protein profiling is increasing and platform performance enhancements such as ion mobility and the use of fractionation with 2D-LC systems is allowing more accurate and reproducible quantitative data to be produced (Giles *et al.*, 2011; Gilar *et al.*, 2012).

While proteomic methodology has developed in many areas to achieve entire proteome profiling, proteins remain complex by nature. Proteins undergo post-translation modifications, vary in concentration and expression. A study by the Human Plasma Proteome Project (HUPO) sent a sample of 20 recombinant human proteins with at least one unique peptide to 27 laboratories for identification. Only seven of the 27 laboratories were able to identify all 20 proteins correctly. Further analysis of the results showed that the sources of error comprised of missed identification, contamination, database matching, and curation of protein identifications (Bell *et al.*, 2009). Thus, when conducting such large quantitative proteomic studies, it is essential to maintain quality control throughout the study and generate reproducible results.

The aim of this chapter was to put into place the method development of chapter 3 and 4, and apply this to a large cohort of patient samples for a prognostic acute heart failure study to discern any changes in protein expression. This was to be achieved without the loss of data quality.

Quality control samples were used to ensure all results were consistent throughout. This was achieved by closely monitoring the peptide retention times, the signal intensity and number of proteins identified in each run. This process was made more thorough by monitoring the peak shape and width of an extracted ion (*m/z* 737.716). A novel method of equalizer bead technology with 2% sodium deoxycholate elution and offline 2D high pH - low pH RP-RP separation with $HDMS^E$ was used to profile the proteome of two acute heart failure populations (survivor and non-survivor). Protein identifications were achieved with PLGS v2.5.2 using the Hi3 method and relative quantitation was calculated by Progenesis LC-MS.

This method was able to identify current markers of acute heart failure prognosis CRP, and our study concluded that this marker was not statistically significant to be considered as a biomarker of prognosis. Seven new biomarkers were identified (summarised in Table 5.4) that have yet to be used as clinical prognostic biomarker of acute heart failure. As heart failure is a syndrome with multi-organ involvement, a

multi-marker approach could provide a better answer to the clinical question. In addition to the identification of the biomarkers, gene ontology analysis for the function of all the proteins identified suggest that the second largest group of proteins were derived from metabolic processes (Figure 5.12).

There were some limitations of the study. The identification of low abundance proteins is hindered by the dynamic range capabilities of the instrumentation. The dynamic range achieved by the Synpat G2 is five orders of magnitude with ion mobility analysis enabled. Lower abundance proteins are still left undetected, where most biomarkers are suggested to be.

The use of pooled samples for analysis was also a limitation. Using pooled samples meant that changes in protein abundance were observed in pooled groups and not individuals. Although verification of the proteins identified in individual samples could overcome this limitation. In this study, the choice of samples and pooling was determined by a statistical programme 'R'. Analysis of the samples was randomised to ensure no mass spectrometric bias was introduced and Progenesis LC-MS was used for relative quantitation as it is designed for fractionated studies.

The size of the study was another limitation of data analysis. The study produced a total of 1.2 TB of data for individual analysis and then combined analysis. The size of the data produced meant software packages available for relative quantitation alongside Progenesis LC-MS were not used as they simply could not recombine the data for an overall comparison of biological groups. Using more than one piece of software for relative quantitation would allow comparisons between the proteins assigned and identified and their statistical value to be further scrutinised.

Chapter Six
Verification

6.1 Introduction

The preceding chapters have discussed quantitative discovery based proteomics comparing the entire proteome of different populations in the hope of identifying biomarkers. Verification aims to confirm if the putative biomarkers identified from discovery projects are valid discoveries, and only the most robust biomarkers are put forward for pre-clinical validation. Despite this clear pipeline, many discoveries in plasma proteomics have failed to enter the clinical laboratories.

The number of FDA approved biomarkers has declined to less than one new protein diagnostic biomarker per year (Anderson & Anderson, 2002). The only FDA approved biomarker discovered by proteomics is OVA1 for ovarian cancer that was granted FDA status in 2009 (Zhang, 2012). Pan *et al.* suggest that the main bottleneck of validated markers using proteomic approaches is the verification methods used for assessing the selected biomarkers performance in a large number of samples (Pan *et al.*, 2009). The effort and expense incurred for verification of biomarkers is considerable and can be the rate limiting step in achieving the goal of FDA biomarker approval.

A variety of methods are used in biomarker verification such as western blotting, selected reaction monitoring and enzyme linked immunosorbent assays (ILMA/immunoassay). All three approaches have advantages and disadvantages. Western blots are unable to resolve proteins with clarity especially low abundance proteins. SRM approaches need plasma to be depleted to achieve quantification in the ng/ml range. MS-based approaches are also unable to fulfil the high throughput and precision required for FDA approval for biomarkers (Tambor *et al.*, 2010). ILMA methods are able to meet the required sensitivity of these FDA approaches; approximately 80% of the FDA approved protein biomarkers are measured by immunoassays clinically (Anderson, 2010).

ELISA methods are described as the ‘gold standard’ technique for targeted quantification of a protein due to good sensitivity and throughput and can be performed using a single antigen specific antibody (indirect assay) or two antibodies in a sandwich assay. ELISA assays are able to measure proteins <1 pg/mL in plasma, have a CV of 5-20% between replicate measurements and have extremely high throughput in 96 well

formats (Kingsmore, 2006). To exemplify this, interleukin 6 is one of the lowest abundance proteins known in plasma and is measured by ELISA at a concentration as low as 0.15 pg/mL, with a CV of 5% (Rifai *et al.*, 2006).

ELISA was the method of choice to confirm the putative biomarkers chosen in Chapter Five and verify the protein expression in the two populations (survivor and non-survivor). We have used a variant of ELISA, immunoluminometric assay or ILMA, as the use of flash chemiluminescence affords even greater sensitivity of detection, without any temperature dependence or chemical inhibition of enzymes that could affect ELISA results. The high throughput nature of the method would enable fast analysis of the 100 plasma samples used in the prognostic study of Chapter Five. The use of a different platform to verify the discovery of biomarkers using mass spectrometry would give a multiplatform agreement and verification.

The aims for this study were:

1. Develop an ILMA assay for the measurement of plasma proteins.
2. Verify the results from the prognostic study (Chapter Five) in individual plasma samples using ILMA analysis.
3. Identify any statistical significance of ILMA confirmed biomarkers.

6.2 Methods and Materials

6.2.1 Materials

Heart failure plasma was supplied by patients from the British Heart Foundation funded study entitled *"The uroguanylin system in heart failure"*, which recruited patients presenting with heart failure (for sample collection protocol see appendix). The samples were centrifuged and plasma was separated from the red blood cells and buffy layer. A hundred acute heart failure samples were selected for this study, age and sex matched. The samples formed two cohorts of patients, those whom remained event free for one year post initial diagnosis (survivors) and those who did not (non-survivors). The human blood samples were collected in EDTA blood sample tubes donors following informed consent. The blood samples were centrifuged at 1500 x g for 20 minutes at 4 °C. The plasma layer was separated from the buffy layer and red blood cells, and stored at -80 °C.

All antibodies for ILMA methods were supplied by R&D (R&D systems, UK), Millipore (Chembiotech, UK), Abnova (Abnova, Taiwan) or Pierce (Thermo Scientific, UK). 96 well microtitre-plates were supplied by Dynex Technologies (West Sussex, UK), Methyl Acridinium Ester (MAE) by Molecular Light Technology (Cardiff, UK.) and Streptavidin was supplied by Chemicon International (Harrow, UK) (See Chapter 2 Tables 2.3-2.5).

6.2.2 ILMA

All ILMA assays were optimised in house. Assays for IGF2, CYS D and SAP were performed using a sandwich assay. A capture antibody was coated onto the 96-well plate, followed by the antigen and a biotinylated antibody was used for detection. The capture antibody was coated at 100 ng in 100 µl per well made in PBS and the detection antibody was coated at 10 ng in 100 µl per well made in ILMA. The plasma dilutions were 50 µl plasma in 100 µl ILMA for IGF2, 10 µl plasma in 100 µl ILMA for CYS D and 1 µl plasma in 100,000,000 µl ILMA for SAP. The standard curve using recombinant protein for IGF2, CYS D and SAP were 0 – 500 pg, 0 – 100 pg and 0 – 50 pg respectively. Approximately 100 µl of MAE – labelled streptavidin in ILMA was added per well for fluorescence and the fluorescence was measured using a luminometer (Dynex MLX Luminometer, Virginia, USA).

Assays for 15-PGDH, CETP, MASP2 and PLTP were all performed initially using a sandwich assay however, these all required further method development (discussed in results section). An indirect ILMA assay was performed for these four proteins, in this ILMA technique only a detection antibody was required. The 96-well plate was coated with the 100 µl of antigen containing plasma after a 1 in 20,000 dilution in PBS. After overnight incubation the detection antibody in ILMA for all four proteins was pipetted on to the plates at 10 ng in 100 µl per well made. The standard curve was measured using a pooled sample with various dilutions. The four dilutions were as follows: 1 in 20,000, 1 in 10,000, 1 in 5000, and blank.

After overnight incubation, approximately 100 µl of MAE – labelled streptavidin in ILMA was added per well for fluorescence and the fluorescence was measured using a luminometer (Dynex MLX Luminometer, Virginia, USA).

6.2.3 Protein concentrations in plasma

Table 6.1: Protein concentration of the seven protein biomarkers to be measured by ILMA, those with ** indicates plasma protein concentrations that remain unpublished.

Protein name	Plasma protein concentration (ng/mL)	Reference
Cystatin D	**	**
Insulin like growth factor II	294-527	(Polanski & Anderson, 2007)
Mannan binding lectin serine protease 2	321-751	(Moller-Kristensen *et al.*, 2003)
Phospholipid transfer protein	2.3-33.4	(Huuskonen *et al.*, 2000)
Serum amyloid P component	1000-9000	(Hortin *et al.*, 2008)
Cholesteryl ester transfer protein	1.6-6.2	(Boekholdt *et al.*, 2004)
15 hydroxyprostaglandin dehydrogenase NAD	**	**

6.2.4 Statistical analysis of results

All statistical analysis was carried out using Statistical Package for Social Sciences version 20.0 (SPSS) for this chapter. ANOVA tests were used to calculate significance, a p value of < 0.05 was considered statistically significant. A box plot was used to illustrate the median and inter-quartile ranges with whiskers representing the 2.5 and 97.5 percentile. Spearman Rho tests were performed to assess the relationship between continuous variables and highlight correlations. A Cox regression analysis was used to assess the univariate and multivariate analysis of multiple variables and test the independent prognostic value of plasma SAP concentrations. Kaplan Meier survival curves were constructed to illustrate the relationship of plasma SAP levels and composite endpoint of death and death and/or heart failure.

6.3 Results

6.3.1 Patient characteristics

Table 6.2: This table shows the patient characteristics of the survivor and non-survivor population of the cohort chosen. The variables are displayed as median [minimum – maximum] ranges or n (percentage).

Variables	Survivors	Non-Survivors	p-values
Age (years)	76 [50 - 90]	78 [55 - 90]	0.410
Male Gender	46 (90%)	40 (82%)	0.258
Body Mass Index	31.9 [17.4 - 85.7]	31.0 [21.2 - 52.6]	0.550
Heart rate (beats min^{-1})	86 [46 - 162]	80 [47 - 160]	0.139
Systolic BP (mm Hg)	145.5 [100 - 203]	124 [92 - 184]	0.001
Diastolic BP (mm Hg)	82.5 [47 - 128]	70 [46 - 97]	0.000
Resp. rate (breaths min^{-1})	20 [16 - 30]	21.5 [12 - 40]	0.427
eGFR (ml min^{-1} /1.73m^2)	53 [16 - 99]	50 [11 - 90]	0.115
Urea (mg/dL)	8.6 [3.4 - 37.2]	12.8 [4.6 - 41.4]	0.001
Sodium (mg/dL)	139 [123 - 147]	137.5 [119 - 144]	0.554
White cell count (mol/L)	8.4 [2.7 - 17.4]	7.9 [4.0- 17.9]	0.565
Glucose (mmol/L)	7.1 [4.2 - 21.3]	6.6 [3.8 - 23.2]	0.646
Haemoglobin (g/L)	13 [9.1-16.1]	12.1 [7.9 - 16.9]	0.089
Past History			
Heart failure	17 (33%)	31 (63%)	0.005
Ischemic heart disease	0 (0%)	4 (8%)	0.054
Hypertension	33 (65%)	24 (49%)	0.157
Diabetes	19 (37%)	22 (45%)	0.542
Treatment			
Beta Blockers	33 (65%)	26 (53%)	0.310
ACE inhibitor/ARB	36 (71%)	26 (53%)	0.099
Diuretics	44 (86%)	35 (71%)	0.087
Biomarkers			
NTproBNP(pmol/L)	2833.3 [0.3 - 13112.6]	3850.1 [60.9 - 15902.6]	0.074
SAP (ng/mL)	158.3 [0.4 - 861.6]	62.8 [0.1 - 821.5]	0.011

6.3.2 Indirect ILMA

There were four proteins measured by an indirect ILMA, these were 15-PGDH, CETP, MASP2 and PLTP. Initially all four proteins were measured using a sandwich ILMA technique, but these assays failed to establish a standard curve according to antibody dilutions (Figure 6.1).

As can be seen in Figure 6.1 all four sandwich assays failed to produce a response with increasing antibody concentration. This could be due to the quality of the either the capture or detection antibody or both. Due to the time constraints of this project, this could not be further evaluated in detail. However, an alternative strategy using an indirect ILMA approach was attempted, removing the need of the capturing antibody. The ILMA assays only needed a detection antibody, the biotinylated antibody. The standard curves for the four proteins are displayed in Figure 6.2. This method improvement lead to a positive antigen response with increasing concentration, showing the assay was able to detect the protein antigens in plasma.

Based on the indirect ILMA method, patient plasma samples were diluted 1 in 20,000 in accordance to plasma known concentration (Table 6.1). The concentrations of the four proteins in individual plasma samples from the pooled groups were tested via an indirect ILMA. This formed a total of one hundred samples, fifty survivor patient plasma samples and fifty non-survivor patient plasma samples. Statistical analysis of the data obtained from the two groups demonstrated no significant difference in the expression of the four proteins when measured in individual plasma samples (Table 6.3).

Table 6.3: ANOVA calculated p-values for measuring the significance of the protein intensities when comparing survivors vs. non-survivors. As no standards were available, the units are in relative light units from the luminometer readings.

Candidate protein name	Biological group. Survivor (0) vs. non-survivor (1)	Mean	Standard Deviation	ANOVA calculated p-value
Mannan binding lectin serine protease 2	0	300.63	66.21	0.677
	1	294.03	89.98	
	Total	297.33	78.67	
Cholesteryl ester transfer protein	0	310.53	51.51	0.322
	1	296.89	82.17	
	Total	303.71	68.57	
Phospholipid transfer protein	0	315.91	62.53	0.538
	1	305.70	98.57	
	Total	310.80	82.28	
15 hydroxyprostaglandin dehydrogenase	0	332.12	57.40	0.773
	1	320.02	49.82	
	Total	321.57	53.50	

To identify whether there was any correlation between the data sets, correlation was measured with spearman's rho test

Table 6.4. A strong correlation was identified between PLTP and CETP; this correlation could be due to both products being heavily involved in lipid metabolism (Weber *et al.*, 2010; Jiang *et al.*, 2012). MASP 2 was also correlated to both CETP and PLTP, the reason for this correlation remains unclear.

Table 6.4: Spearman Rho correlation for the four proteins measured via indirect ILMA.

Correlations

			PGDH	CETP	MASP2	PLTP
Spearman's rho	PGDH	Correlation Coefficient	1.000	.313**	.258**	.321**
		Sig. (2-tailed)	.	.002	.010	.001
		N	100	100	100	100
	CETP	Correlation Coefficient	.313**	1.000	.898**	.934**
		Sig. (2-tailed)	.002	.	.000	.000
		N	100	100	100	100
	MASP2	Correlation Coefficient	.258**	.898**	1.000	.897**
		Sig. (2-tailed)	.010	.000	.	.000
		N	100	100	100	100
	PLTP	Correlation Coefficient	.321**	.934**	.897**	1.000
		Sig. (2-tailed)	.001	.000	.000	.
		N	100	100	100	100

**. Correlation is significant at the 0.01 level (2-tailed).

6.3.3 Sandwich ILMA

The remaining three proteins, IGF2, CYS D and SAP, were measured by a sandwich ILMA. All three sandwich assays produced a response to increasing antibody concentrations, with positive standard curves (see appendix, Figure C-1).

Based on the sandwich ILMA method, patient plasma samples were diluted in accordance to their plasma protein concentrations (Table 6.1), the concentrations of the three proteins in individual plasma samples from the pooled groups were tested. In total, this formed one hundred samples, fifty survivor patient plasma samples and fifty non-survivor patient plasma samples. Statistical analysis of the data obtained from the two groups (survivor vs. non-survivor) demonstrated no significant difference in the expression in IGF2 and CYS D when measured in individual plasma samples but SAP was highly significant (Table 6.5).

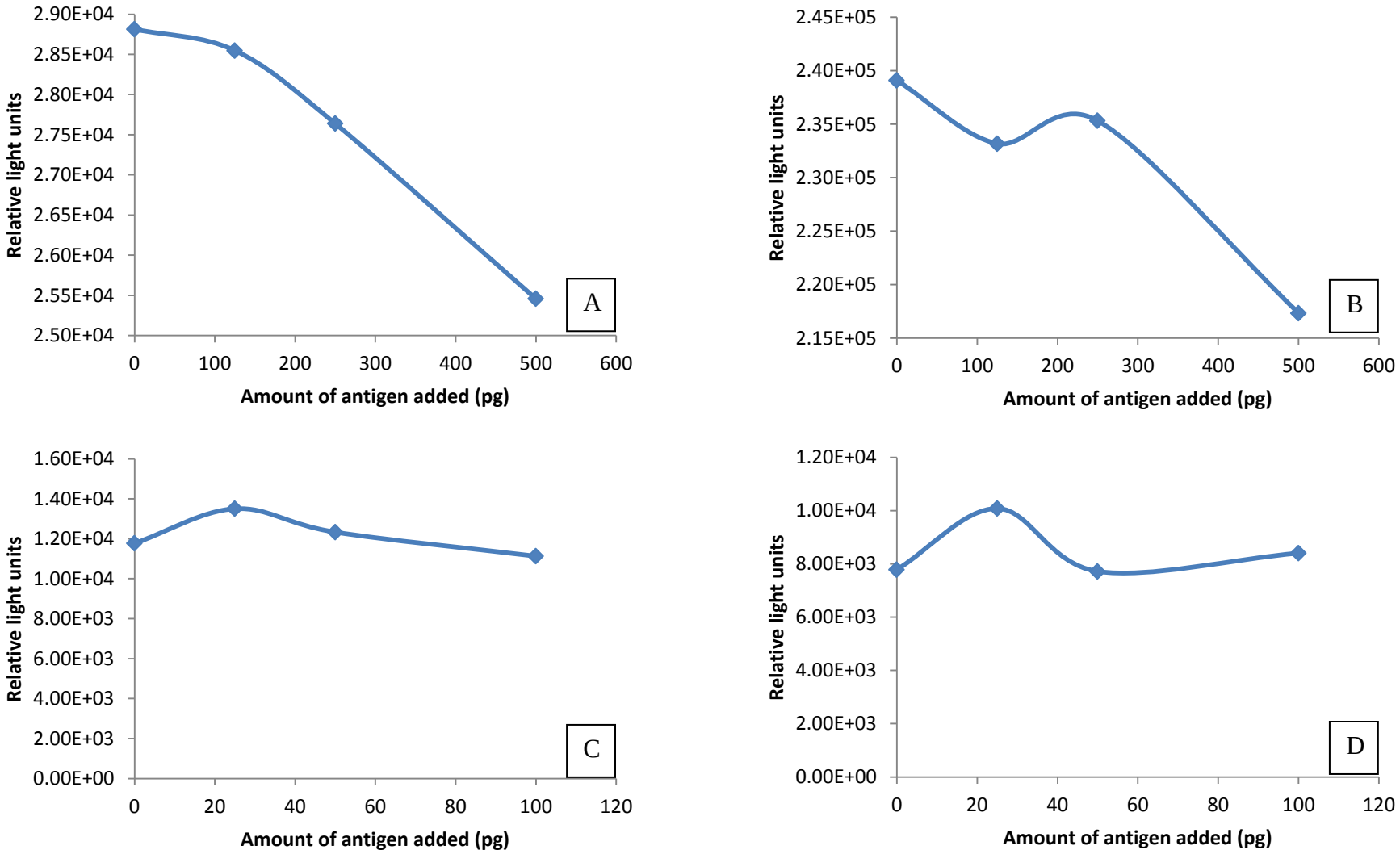


Figure 6.1: The standard curve for MASP 2 (A), CETP (B), PLTP (C) and 15-PGDH (D) using a sandwich ELISA method. These graph show the relationship of the chemiluminescence of each antigen at different concentrations

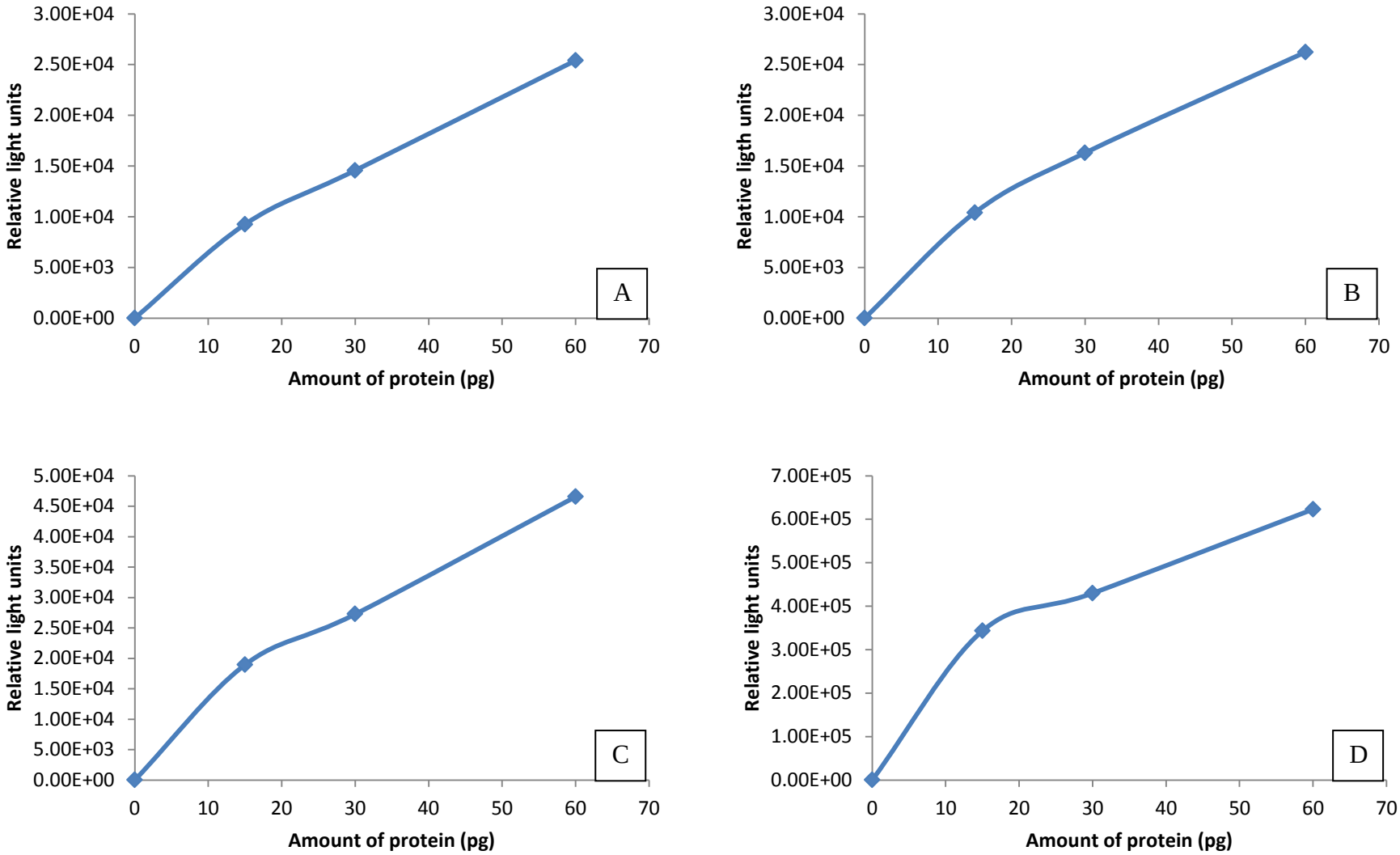


Figure 6.2: The standard curve for MASP 2 (A), CETP (B), PLTP (C) and 15-PGDH (D) using an indirect ELISA method. These graphs show the relationship of the chemiluminescence of each antigen using different plasma protein concentrations.

Table 6.5: ANOVA calculated p-values for measuring the significance of the protein intensities when comparing survivors vs. non-survivors for logged ng/ mL values of IGF2, CYS D and SAP.

Candidate protein name	Biological group. Survivor (0) vs. non-survivor (1)	Mean	Standard Deviation	ANOVA calculated p-value
Insulin like growth factor 2	0	2.22	0.95	
	1	2.29	1.11	0.677
	Total	2.26	1.02	
Cystatin D	0	2.22	0.78	
	1	2.45	0.92	0.143
	Total	2.32	0.86	
Serum amyloid P component	0	2.10	0.56	
	1	1.32	1.22	**<0.0005**
	Total	1.72	1.02	

To identify whether there was any correlation between the data sets, correlation was measured with spearman's rho test. No strong correlation was identified between IGF 2, CYS D and SAP.

6.3.3.1 Serum amyloid P component

The ANOVA analysis shown in table 6.4 found serum amyloid P component (SAP) to be significant when measured in the patient samples using ILMA; this confirms the results obtained in chapter six. Plasma SAP concentration was measured using a sandwich ILMA in fifty acute heart failure survivor patients and fifty acute heart failure non-survivor patients, a hundred samples in total. The measurement of SAP was in ng/ mL in plasma. The total median for all samples was 137.2 [0.12-861.59] (median [range]), the median plasma SAP content was higher for survivors 158.3 [0.4-861.59] and lower for non-survivors 62.8 [0.12-821.5] and p value of 0.011 (calculated by Mann Whitney U test). The box plot below shows the significant difference between the two different groups. The distribution of the data shows the plasma concentration of SAP is much higher in survivors than non-survivors (Figure 6.3). SAP levels were also higher in patients who were discharged compared to those who died as in patients (see boxplot in appendix).

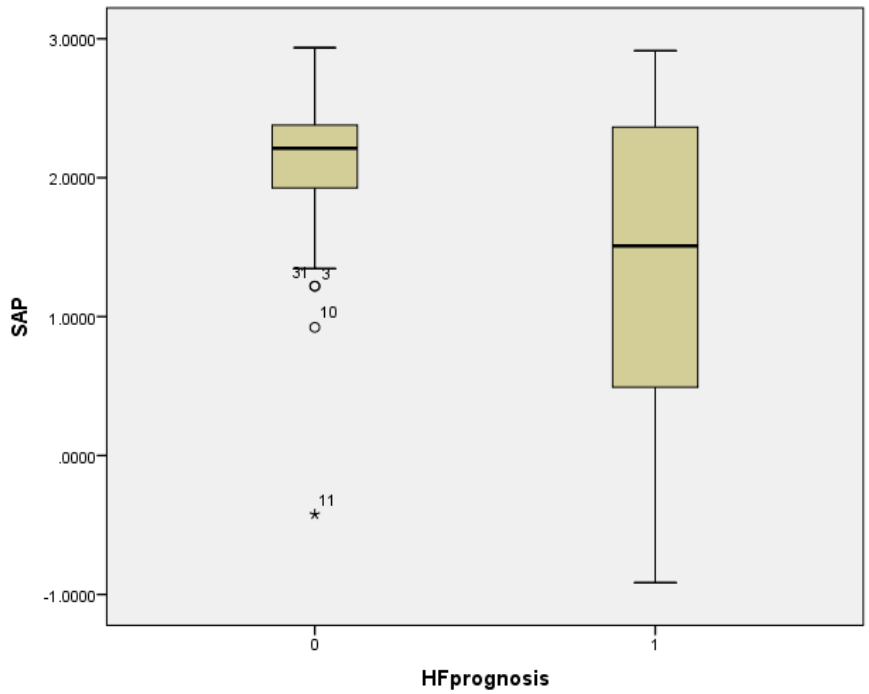


Figure 6.3: Boxplot comparing serum amyloid P component (SAP) in acute heart failure patients. The Y axis shows the Log plasma SAP concentration (ng/ mL) and the X axis shows the patient groups (0 = survivors and 1 = non-survivors with end point of 1 year).

Spearman Rho correlations were carried out to assess the relationships between continuous variables in plasma SAP levels (See Table 6.6). SAP was found to be correlated to plasma glucose levels with a Spearman's Rho correlation coefficient of 0.282 and p value of 0.026. No correlation was observed with age, sex, body mass index, estimated glomerular filtration rate, NTproBNP, CRP and troponin. Thus, plasma measured SAP levels were independent of these variables.

Cox regression analysis was carried out to identify composite prognostic endpoints in this study, all data was censored at the end point of death and death and/or heart failure. Multiple parameters were included in the Cox model including age, sex, NYHA class, heart rate systolic blood pressure (BP), respiratory rate, glomerular filtration rate (eGFR), urea, sodium, haemoglobin, past history of heart failure, ischemic heart disease, hypertension or diabetes, treatment with beta blockers, ACE inhibitor or diuretics, and biomarkers NTpro-BNP with SAP (Table 6.7 and Table 6.8). The Cox analysis shows that SAP is a statistically significant independent predictor of death and

death and/or heart failure in both univariate and multivariate analysis. The established biomarker of prognosis NTproBNP was no longer an independent predictor in multivariate analysis.

Kaplan Meier curves were produced to illustrate the estimated proportion of patients who survived over 1 year under the same conditions. The Kaplan Meier curve is a step function; the estimated survival probability remains constant between the two populations of patients, the survival probability only decreases with each event.

Figure 6.4 and Figure 6.5 show the Kaplan Meier curve for patients with above and below median plasma SAP levels. There is a clear separation between the curves indicating a poor outcome is associated with below median SAP levels for both death ($p = 0.007$) and death and/or heart failure ($p = 0.004$) as the end points. The Kaplan Meier curves for both plateau after approximately 300 days. This indicates that plasma SAP levels could be a valuable short-term marker post initial presentation and diagnosis of acute heart failure.

Table 6.6: Spearman rho correlations of continuous variable in plasma SAP levels

SAP	Correlation Coefficient	1.000	NYHA_Admission	Correlation Coefficient	.035
	Sig. (2-tailed)			Sig. (2-tailed)	.735
	N	100		N	97
Age	Correlation Coefficient	-.096	Admission haemaglobin	Correlation Coefficient	.038
	Sig. (2-tailed)	.343		Sig. (2-tailed)	.709
	N	100		N	100
Oedema	Correlation Coefficient	.010	Admission White cell count	Correlation Coefficient	.185
	Sig. (2-tailed)	.925		Sig. (2-tailed)	.066
	N	100		N	100
Past history of HF	Correlation Coefficient	-.088	Admission Urea	Correlation Coefficient	-.072
	Sig. (2-tailed)	.385		Sig. (2-tailed)	.476
	N	100		N	100
Past history of Ischemic Heart Disease	Correlation Coefficient	.131	Admission eGFR	Correlation Coefficient	.008
	Sig. (2-tailed)	.194		Sig. (2-tailed)	.939
	N	100		N	100
Past history of hypertension	Correlation Coefficient	.065	Admission Na	Correlation Coefficient	-.053
	Sig. (2-tailed)	.522		Sig. (2-tailed)	.597
	N	100		N	100
Past history of TIA	Correlation Coefficient	-.147	Admission Glucose	Correlation Coefficient	.282*
	Sig. (2-tailed)	.145		Sig. (2-tailed)	.026
	N	100		N	62
Past history of CVA	Correlation Coefficient	-.103	Admission Cholesterol	Correlation Coefficient	.212
	Sig. (2-tailed)	.308		Sig. (2-tailed)	.556
	N	100		N	10
Past history COPD	Correlation Coefficient	-.176	Admission Triglycerides	Correlation Coefficient	.084
	Sig. (2-tailed)	.079		Sig. (2-tailed)	.844
	N	100		N	8
Past history Hyperlidaemia	Correlation Coefficient	-.089	Admission CRP	Correlation Coefficient	-.126
	Sig. (2-tailed)	.378		Sig. (2-tailed)	.325
	N	100		N	63
Heart Rate	Correlation Coefficient	.026	Admission Creatinine Kinase	Correlation Coefficient	
	Sig. (2-tailed)	.797		Sig. (2-tailed)	
	N	97		N	0
Systolic Blood Pressure	Correlation Coefficient	.038	Admission Troponin	Correlation Coefficient	.085
	Sig. (2-tailed)	.709		Sig. (2-tailed)	.706
	N	97		N	22
RR	Correlation Coefficient	-.155	Body mass index	Correlation Coefficient	.076
	Sig. (2-tailed)	.156		Sig. (2-tailed)	.559

Table 6. 7: Cox regression analysis for death at 1 year post-hospital admission and acute heart failure diagnosis. Multivariable analysis results are reported for model 1 which included variables and biomarkers (except SAP) which were significant on univariable analysis. Multivariable Model 2 used the variables in model 1 with the addition of SAP.

Parameters	Univariable HR (95% CI)	P value	Multivariable Model 1 HR (95% CI)	P value	Multivariable Model 2 HR (95% CI)	P value
Age (years)	1.044 (1.007 – 1.082)	0.020	1.026 (0.984 – 1.070)	0.231	1.027 (0.984 – 1.072)	0.222
Male Gender	1.072 (0.419 – 2.743)	0.885				
NYHA class	3.994 (1.449 – 11.012)	0.007	1.138 (0.421 – 3.079)	0.799	1.377 (0.521 – 3.639)	0.518
Heart rate (beats min^{-1})	0.986 (0.972 – 1.000)	0.048	0.986 (0.969 – 1.002)	0.082	0.992 (0.976 – 1.008)	0.338
Systolic BP (mm Hg)	0.981 (0.967 – 0.996)	0.013	0.979 (0.962 – 0.997)	0.020	0.982 (0.966 – 0.998)	0.031
Resp. rate (breaths min^{-1})	1.088 (1.010 – 1.171)	0.025	0.996 (0.919 – 1.078)	0.913	0.967 (0.889 – 1.051)	0.432
eGFR (ml min^{-1} /1.73m^2)	0.984 (0.968 – 1.000)	0.046	1.005 (0.979 – 1.031)	0.725	1.006 (0.981 – 1.031)	0.647
Urea (mg/dL)	1.061 (1.025 – 1.098)	0.001	1.070 (1.012 – 1.131)	0.017	1.056 (1.000 – 1.115)	0.052
Sodium (mg/dL)	0.983 (0.922 – 1.047)	0.588				
Haemoglobin (g/L)	0.863 (0.714 – 1.044)	0.129				
Past History						
Heart failure	2.413 (1.252 – 4.651)	0.008	0.953 (0.390 – 2.329)	0.915	1.363 (0.569 – 3.265)	0.487
Ischemic heart disease	6.367 (2.176 – 18.631)	0.001	2.666 (0.695 – 10.219)	0.153	3.500 (0.905 – 13.540)	0.070
Hypertension	0.899 (0.479 – 1.688)	0.741				
Diabetes	1.471 (0.785 – 2.757)	0.229				
Treatment						
Beta Blockers	0.650 (0.347 – 1.218)	0.179				
ACE inhibitor/ARB	0.515 (0.275 – 0.966)	0.039	1.003 (0.368 – 2.732)	0.996	0.697 (0.243 – 1.998)	0.502
Diuretics	0.361 (0.185 – 0.703)	0.003	0.220 (0.073 – 0.664)	0.007	0.317 (0.100 – 1.004)	0.051
Biomarkers						
Log NTproBNP(pmol/L)	20.33 (0.976 – 4.236)	0.058	1.500 (0.691 – 3.253)	0.305	1.212 (0.581 – 2.530)	0.608
Log SAP (ng/mL)	0.541 (0.413 – 0.708)	<0.0005			0.572 (0.399 – 0.819)	0.002

Table 6.8: Cox regression analysis for death and/or heart failure at 1 year post-hospital admission and acute heart failure diagnosis. Multivariate analysis results are reported for model 1 which included variables and biomarkers (except SAP) which were significant on univariate analysis. Multivariate Model 2 used the variables in model 1 with the addition of SAP.

Parameters	Univariable HR (95% CI)	P value	Multivariable Model 1 HR (95% CI)	P value	Multivariable Model 2 HR (95% CI)	P value
Age (years)	1.018 (0.989 – 1.047)	0.233	1.013 (0.978 – 1.048)	0.479	1.008 (0.972 – 1.046)	0.657
Male Gender	0.611 (0.298 -1.255)	0.180				
NYHA class	2.261 (1.117 – 4.577)	0.023	1.853 (0.726 – 4.733)	0.197	2.229 (0.876 – 5.671)	0.093
Heart rate (beats min^{-1})	0.991 (0.979 – 1.002)	0.110				
Systolic BP (mm Hg)	0.977 (0.964 – 0.990)	0.001	0.982 (0.966 – 0.999)	0.032	0.986 (0.970 – 1.002)	0.081
Resp. rate (breaths min^{-1})	1.023 (0.956 – 1.095)	0.507				
eGFR (ml min^{-1} /1.73m^2)	0.985 (0.970 – 0.999)	0.036	1.002 (0.979 – 1.026)	0.872	1.000 (0.976 – 1.024)	0.978
Urea (mg/dL)	1.059 (1.027 – 1.092)	0.000	1.064 (1.009 – 1.121)	0.021	1.054 (1.000 – 1.110)	0.048
Sodium (mg/dL)	0.972 (0.919 – 1.029)	0.331				
Haemoglobin (g/L)	0.859 (0.729 – 1.013)	0.070	0.994 (0.829 – 1.192)	0.950	1.044 (0.876 – 1.242)	0.632
Past History						
Heart failure	2.347 (1.345 – 4.094)	0.03	1.443 (0.678 – 3.071)	0.341	1.675 (0.799 – 3.511)	0.172
Ischemic heart disease	3.856 (1.370 – 10.849)	0.011	1.829 (0.567 – 5.904)	0.312	3.321 (0.968 – 11.392)	0.056
Hypertension	0.619 (0.359 – 1.068)	0.085	0.471 (0.238 – 0.932)	0.031	0.522 (0.253 – 1.075)	0.078
Diabetes	1.430 (0.830 – 2.465)	0.198				
Treatment						
Beta Blockers	0.653 (0.378 – 1.127)	0.126				
ACE inhibitor/ARB	0.577 (0.334 – 0.999)	0.050	0.775 (0.332 – 1.809	0.555	0.587 (0.236 – 1.457)	0.251
Diuretics	0.483 (0.261 – 0.892)	0.020	0.333 (0.134 – 0.828)	0.018	0.491 (0.187 – 1.293)	0.150
Biomarkers						
Log NTproBNP(pmol/L)	1.785 (1.003 – 3.175)	0.049	1.277 (0.631 – 2.585)	0.497	1.143 (0.573 – 2.280)	0.705
Log SAP (ng/mL)	0.561 (0.437 – 0.721)	<0.0005			0.506 (0.373 – 0.688)	<0.0005

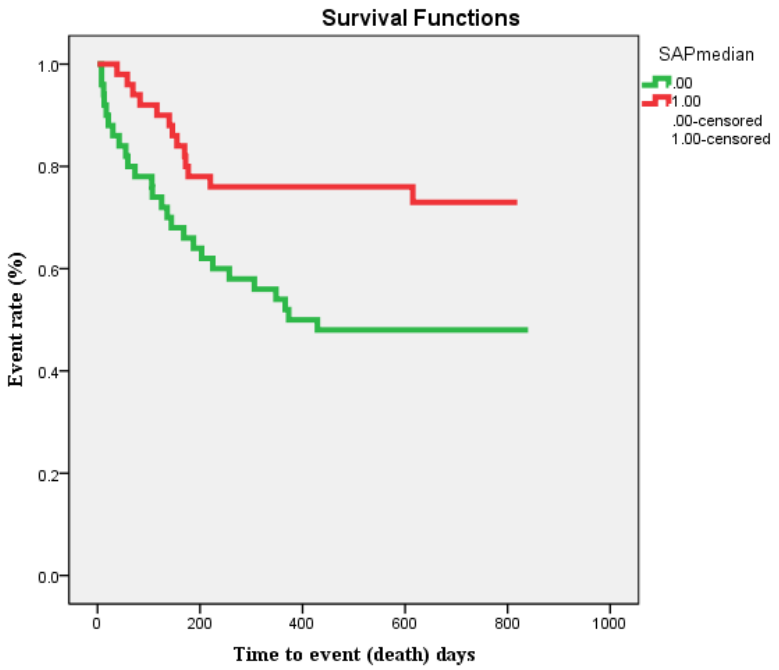


Figure 6.4: Kaplan Meier survival curve for patients with plasma SAP concentrations below or above median, in acute heart failure. The green line denotes the infra-median and the red the supra-median SAP group, for the endpoint of death.

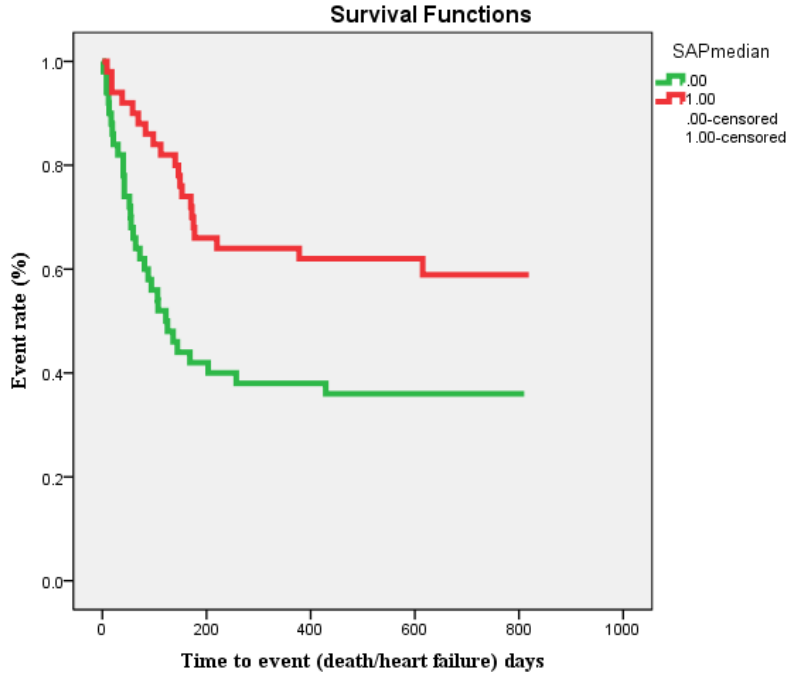


Figure 6.5: Kaplan Meier survival curve for patients with plasma SAP concentrations below or above median, in acute heart failure. The green line denotes the infra-median and the red the supra-median SAP group, for the endpoint of death and/or heart failure.

6.4 Discussion

The aim of all proteomics discovery biomarker projects is to eventually identify biomarkers that have clinical utility. The increasing number of biomarkers identified for a variety of diseases requires all biomarker discoveries to be verified to assess the specificity of the putative biomarkers in a small cohort of patients and validated to establish sensitivity, specificity and optimisation in a larger cohort of patients (Rifai *et al.*, 2006). The ELISA technique has been long established and is often quoted as the 'gold standard' technique for protein measurement, with high throughput and well established protocols in clinical laboratories (Kingsmore, 2006).

In Chapter Five, seven proteins were identified as putative biomarkers for acute heart failure prognosis using discovery proteomic techniques. A hundred plasma samples from two acute heart failure populations (survivor and non-survivor) were pooled and analysed using an optimised proteomics workflow of label free 2D-LC-HDMS^E and data analysis for protein quantitation and identification. Using statistical analysis available within Progenesis LC-MS, the seven proteins, IGF2, CYS D, SAP, MASP2, PLTP, CETP and 15-PGDH, were expressed at different levels in both of the acute heart failure groups and deemed statistically significant.

In this study, the same one hundred samples were tested singularly using two different ILMA techniques to verify the plasma protein expression of the seven candidate proteins. Six of seven proteins did not show statistically significant plasma protein changes when measured individually rather than in pooled samples. The explanation for the discrepancy between the discovery phase and verification phase measurements is potentially three-fold.

The discovery phase analysis was conducted on a hundred pooled plasma samples, with ten individual samples pooled to produce one pooled sample. The verification ILMA analysis was conducted on the same samples but the samples were un-pooled and untreated plasma (i.e. did not undergo any sample preparation prior to analysis). Verification was conducted in untreated plasma samples as verification should be performed on samples that would be used for clinical testing and allows marker specificity for the given disease to be truly tested.

The ILMA methods rely heavily on antibody specificity for the antigen epitope binding (Pan *et al.*, 2009). Although all the proteins identified had antibodies available, the specificity of these antibodies could be a reason for the discrepancies seen in the protein expression levels. This is clearly demonstrated with MASP2, PLTP, CEPT and 15-PGDH, where ILMA method development was required, removing the use of the capturing antibody as it did not successfully bind the antigen.

ELISA assays are also susceptible to a variety of conditions other than binding such as incubation time and temperature, sample volume and dilution, pH, composition and concentration of the diluents, enzyme and substrate type and quality of the detector (Rifai *et al.*, 2006). Hence, we used more reliable chemiluminescent ILMA assays rather than enzyme based assays (ELISA). Great care was taken to control all variables and achieve consistency throughout the ILMA analysis. A limitation of the ILMA was the need for extensive dilution of plasma samples prior to analysis, introducing the potential of pipetting/human error.

An alternative protein targeting method is selected reaction monitoring (SRM) and is becoming a popular alternative for verification (Method of the Year 2012. 2013; Domon & Aebersold, 2010). SRM mass spectrometry methods can be used to target a specific protein and offer an alternative to antibody dependent techniques, removing the variables that affect protein measurements with ILMA. This method could be used as an alternative to limit the variation in technique for the six proteins.

While the six proteins, IGF 2, CYS D, MASP2, PLTP, CETP, 15-PGDH, after ILMA analysis were not as significant as suggested by Progenesis and proteomic analysis in the discovery phase, SAP was found to be highly significant in both discovery and verification phases of this study.

SAP was expressed 2.1 times higher in acute heart failure survivors than non-survivors and was highly significantly, with a p value < 0.0005, in the discovery phase of this study. A sandwich ILMA method was used to verify the protein expression of SAP in individual plasma samples and found to remain significant, p value <0.0005 and

expression was 1.6 times higher in acute heart failure survivors. The difference in SAP protein concentration in the two populations is exemplified in the box plot in Figure 6.3.

Further statistical analysis of SAP in plasma using Cox regression analysis confirmed SAP was an independent prognostic indicator in acute heart failure patients and Kaplan Meier curves show the short term prognostic value (under 300 days) of plasma SAP concentration with death and death and/or heart failure as eventual outcomes. The Spearman's Rho analysis found SAP to be independent of variables such as age, sex and body mass index, which are known to affect base levels of current prognostic biomarkers such as NTproBNP (Murray *et al.*, 2006).

The results for SAP have shown discovery proteomic strategies are able to identify biomarkers for acute heart failure prognosis and can be further verified, as the same plasma SAP expression was seen in both phases of the study and SAP remained significant. The SAP results also demonstrate the ability of two different analytical techniques to work within a proteomic pipeline and identify a putative marker. The results from the ILMA for SAP as a prognostic marker in acute heart failure are promising. For large scale significance to be measured, the ILMA method would need to be replicated in a larger validation cohort.

Chapter Seven

Conclusions and Future work

7.1 Conclusions

Currently 14 million people in Europe suffer from heart failure, with 3.6 million people being diagnosed each year (HFM, 2007). In the UK, 900,000 people have heart failure with equally as many with myocardial damage but are symptomless. The prognosis of heart failure is worse than that of breast and prostate cancer, 30-40% of patients who are diagnosed with heart failure die within 1 year of diagnosis. High NHS care costs make heart failure an economic burden, accounting for 2% of the total NHS budget (National Institute for Health and Clinical Excellence, 2010).

Since Mark Wilkins first introduced the term proteome in 1994 as a complement to the genome (Wilkins *et al.*, 1996), proteomics has undergone rapid change and has found itself at the forefront of biomarker research. Petricoin *et al.* first described the use of proteomic patterns in serum to identify ovarian cancer (Petricoin *et al.*, 2002). Since then, an FDA approved panel of biomarkers for ovarian cancer, OVA1, has been developed (Zhang & Chan, 2010). A serum based proteomic test, VeriStrat has been shown to act as a prognostic indicator in advanced stage non-small cell lung cancer using MALDI-TOF MS signatures of eight protein or peptides (Taguchi *et al.*, 2007; Kuiper *et al.*, 2012). Schwarz *et al.* have discovered a serum based multiplex assay of 51 proteins for schizophrenia, which is widely used to distinguish control and disease populations. The authors report 83% sensitivity with 83% specificity of the multicentre immunoassay based test (Craddock *et al.*, 2008). A study of the plasma proteome for cardiovascular biomarkers by HUPO found 345 proteins with cardiovascular related functions out of a total 3020 protein identified in plasma (Berhane *et al.*, 2005). Mebazza *et al*. have recently found plasma Quiescin 6 concentration accurately identifies acute decompensated heart failure in combination with BNP (Mebazaa *et al.*, 2012). Watson *et al.* identified leucine rich alpha-2-glycoprotein as a biomarker of heart failure and ventricular dysfunction using serum samples (Watson *et al.*, 2011). It is clear, progress is being made and proteomics has affected diagnosis and prognosis of certain diseases.

Plasma is commonly used in proteomics as it can be obtained non-invasively and is arguably the most representative fluid of the human proteome. This thesis used plasma to search for biomarkers of acute heart failure. The rationale for using plasma rather

than other biofluids was due the disease's pathology as heart failure is a multi-organ disease with complex inter-connected pathophysiology. It was therefore thought that plasma could be used to identify a panel of proteins, which, in conjunction would be more disease specific. However, plasma is a challenging biofluid to profile using mass spectrometry as the dynamic range of plasma spans over 12 orders of magnitude, whereas most current mass spectrometers can profile proteins of 4-5 orders of magnitude (Anderson & Anderson, 2002).

The primary aim of this thesis was to optimise a plasma proteomic sample preparation method that was compatible with mass spectrometry and reduce the dynamic range of plasma proteins (Chapter 3). Immunodepletion is a commonly used method for plasma proteomics. However, this method can cause co-depletion of smaller low abundance proteins when depleting carrier molecules such as albumin (Tu *et al.*, 2010). Due to this limitation, many different methods were attempted and eventually equalizer beads were found to reduce the dynamic range of plasma, identifying proteins of varying plasma protein concentrations reproducibly using nanoUPLC-MS^E (Chapter 3 section 3.3.3.5.1).

The reproducibility of a chosen proteomic method is essential to ensure all results obtained are reliable and suitable for biomarker studies. Using healthy control samples and the optimised sample preparation method, technical reproducibility and precision was demonstrated over 5 orders of magnitude (Hakimi *et al.*, 2013). Although lower than the 12 orders of magnitude of plasma, the 5 orders of magnitude achieved is greater than other assessed workflows.

In addition to plasma sample preparation methods, mass spectrometry method development was also conducted. In this thesis, a novel proteomic mass spectrometry workflow was developed for biomarker discovery (chapter 3 section 3.3.3.5.3). Nano-UPLC-MS^E and nano-UPLC-$HDMS^E$ workflows were compared, nano-LC-$HDMS^E$ incorporated the ion mobility separation within the analysis. Thus, samples were separated according to mass, charge and size with the inclusion of drift time. Using label free Hi3 absolute quantitation (Silva *et al.*, 2006), $HDMS^E$ out performed MS^E, profiling the plasma proteome in greater depth.

Chapter Three focussed on the development of sample preparation and mass spectrometry methods using healthy control plasma. Chapter Four took the developed proteomic methods, applied them to a population of acute heart failure plasma, and matched healthy controls. Multidimensional chromatography was also introduced as a further platform development. This thesis demonstrated multidimensional chromatography was able to increase peak capacity, increase the number of protein identifications for the same samples and provide confident protein identifications. Chapter Four showed that increasing the number of fractions per sample and optimising sample loading in the second dimension can increase the dynamic range of protein identifications but at the expense of analysis time. Chapter Four also demonstrated the problem of pooling samples to provide a super-sample of n=1. Where possible, n=1 should be avoided as a large group of pooled samples could lead to a loss of information of biological variation in plasma protein concentrations. However, when time is limited and maximum throughput is required, n=1 is the only reasonable option. A fine balance has to be struck between high throughput and sample pooling. Despite this limitation, high/low pH RP-RP chromatography offers an overall advantage to a proteomic workflow and is a time efficient method of analysing a large number of samples.

The method and platform development in Chapters Three and four culminated in a prognostic biomarker study for acute heart failure in Chapter Five. A total of one hundred samples from survivor and non-survivor groups with an end point of 1 year were pooled and using high/low pH LC-HDMS^{E}, the proteome of both patient sample groups were investigated. Using proteomic software, proteins were identified, quantified and expression analysis of the two groups was compared. The quality of the data was assured by the judicious use of relevant QC samples, which demonstrated high levels of uniformity.

A comparison of the two conditions (survivor vs. non-survivor) identified proteins that were differentially expressed between both groups and research within the thesis demonstrated that biomarkers could be identified. However, the size of the study produced a large amount of data, which was only processed by a single proteomic software package. Proteomic studies are hampered by the length of sample analysis and

the amount of data produced, even with sample pooling. Pooling of samples should be carefully considered when undertaking a study, even though pooling has limitations it is still a valid and valuable time and data size saving procedure that could and should be considered (Diz *et al.*, 2009).

A total of seven disease pathology related and statistically significant proteins identified by the prognostic study were further verified using an antibody based method, ILMA. SAP, which was identified in the pooled samples as a putative biomarker, remained significant even when measured in the same samples individually (Chapter 6). The identification of SAP as a putative biomarker using a pooled experimental design proves that pooling is not completely disadvantageous. Using the proposed experimental pipeline, we have managed to yield a putative biomarker for acute heart failure prognosis.

SAP has been shown in this study to be an independent predictor of death and /or heart failure as end points within one year. This confirms that the discovery proteomic workflow can lead to the successful identification of putative biomarkers and identify biomarkers that have ordinarily been over looked in acute heart failure pathology. The role of SAP in heart failure is still highly debated and has been found to be associated with many different co-morbidities but not with cardiovascular death by other researchers (Ying *et al.*, 1993; Pepys *et al.*, 1979; Jenny *et al.*, 2007; De Beer *et al.*, 1982; Haudek *et al.*, 2006). Although CRP and SAP come from the same family of proteins (pentraxin family), their roles in disease pathology seem to differ. Further investigations of SAP's role in acute heart failure would need to be determined to understand its function, although it has been hypothesised that SAP plays a role in preventing atherosclerotic progression of disease by blocking macrophage uptake of SAP bound LDL (Stewart *et al.*, 2005).

Our study has found that SAP outperforms NTproBNP, an established prognostic biomarker for acute heart failure (Gardner *et al.*, 2003), within the selected patient cohort for predicting adverse outcome in acute heart failure patients. The use of SAP as an independent prognostic biomarker would enable clinicians to categorise patients into high risk and low risk post-discharge groups, this could lead to a more tailored

approach for personalised medicine strategies, as patients deemed as high risk of a post-discharge event could be eligible for more intensive treatments or increased monitoring by hospital or community nurses. The high significance of SAP over NTproBNP in the same patient cohort suggests that SAP could be used in place of the NTproBNP test or in conjunction for further clarity of prognosis. Further investigations of SAP with existing and clinically used acute heart failure biomarkers is required. SAP would also need to be tested for specificity at different stages of acute heart failure with all co-morbidities associated with the disease, alongside established biomarkers such as NT-proBNP and the effects of therapy on SAP should be investigated.

The relationship of SAP to acute heart failure discovered in this proteomic study could lead to SAP being targeted for therapeutic purposes. SAP is known to bind LDL reducing its function in atherosclerosis (Stewart *et al.*, 2005) and SAP injections have been shown to inhibit fibrosis in animal models (Crawford *et al.*, 2012). As high levels of SAP seem to confer better prognosis in this study, the molecular mechanisms involved could be explored and exploited for new targets for therapeutics.

SAP was identified with >2 fold change in protein expression between the two acute heart failure conditions. This was the only protein above >2 fold change short listed for the target proteins and was the only successful protein to be identified in verification tests. Additional stringency efforts when using pooled samples may need to be put in place, as the subtle changes in plasma protein concentrations may be lost. Thus, higher fold changes may need to be used as a positive biomarker identification filter when comparing pooled samples than individual samples. This may mean a smaller number of putative markers are identified but they will be identified with a higher degree of confidence prior to verification tests.

This thesis aimed to identify biomarkers in acute heart failure patient plasma, using a novel platform combining 2D-LC and nanoUPLC-HDMS^E with bioinformatic approaches. Biomarkers of prognosis were identified and SAP was successfully verified in further antibody analyses. To our knowledge, this is the first acute heart failure proteomic biomarker study to combine a novel sample preparation and analysis method of equalizer beads with 2D-RP-RP-HDMS^E to identify prognostic markers of acute

heart failure. Furthermore, this is the only plasma proteomic biomarker study to identify SAP as a biomarker using this workflow. The utility of such an approach is strongly demonstrated and justified.

7.2 Future work

Mass spectrometry based biomarker discovery proteomics has greatly improved since it was first used for searching for disease biomarkers in 2002 (Petricoin *et al.*, 2002). However, proteomics remains a new field and will require significant development to search for proteins present in and below the ng/mL range. The greatest challenge in plasma proteomics is overcoming the complexity of the samples and identifying every single protein in plasma. Collaborative efforts have been made to ensure all proteins, corresponding peptides, post-translational modification and their concentrations are identified in a shared common repository. HUPO's Plasma Proteome Project and Proteomics IDEntifications database (PRIDE) are examples of such efforts (Omenn *et al.*, 2005; Martens *et al.*, 2005).

In addition to the challenges faced in sample analysis, data analysis also requires further enhancement. As shown in this thesis, 1.2 TB of data was collected from only one hundred pooled and fractionated samples using the optimum workflow. Many proteomic software packages available commercially such as PLGs v2.5.2 and Scaffold were unable to cope with this volume and re-combining required when fractionated data was used. Development in this area of proteomics is essential to confidently interrogate data from mass spectrometers.

The data presented in this thesis has identified protein biomarkers of acute heart failure. However, additional interrogation of data could potentially identify more information on the biological process that takes place in acute heart failure.

1. Post translational modifications of proteins in different disease states such as glycosylation and phosphorylation could be investigated, as these modifications have been targets for understanding disease biology.
2. Investigation of the effect of changing the stringency on the data filters to identify biomarkers could yield new or slightly different information.

3. Investigation of protein functions of the identified proteins could provide an insight into disease state pathology.

In addition to investigating the data that has already been presented for further information, this has also highlighted other areas for method development and potentially new areas of research and further work.

The proteins identified in the prognostic study (chapter 5) were analysed using gene ontology programs and proteins were separated according to molecular function. One of the largest groups of molecular function identified was proteins related to metabolism. A metabolomic study of the plasma could identify further markers of acute heart failure prognosis and cross-examination of proteomic and metabolomic data to search for similarity would increase confidence in protein and metabolite identifications. As only one of the seven identified biomarkers were verified using ILMA assays, SRM could be performed as an alternative for verification of putative biomarkers. Although SRM verification is time intensive, as the method requires optimisation and evaluation (Method of the Year 2012. 2013).

Although this study found SAP was a positive marker of acute heart failure, the verification procedure has some limitations. The cohort tested only comprised of a hundred samples. However, it had sufficient power to outperform other established markers and clinical features of acute heart failure. SAP levels need to be evaluated further with a larger acute heart failure cohort of samples (approximately 1000 samples). The SAP biomarker needs to be cross-validated against other existing biomarkers such as NT-proBNP and clinical features/co-morbidities of acute heart failure to determine its independent prognostic power. Mechanistic studies need to be undertaken to understand the role of SAP in acute heart failure pathology and therapeutic effect of SAP targeted treatment.

Chapter Eight

Appendix

8.1 Appendix A

Lists of the proteins identified for proteins from the equalizer beads with sodium deoxycholate method

Table A-1: A list of the 169 proteins identified by the equalizer beads with SDC method

Actin cytoplasmic 1
Actin cytoplasmic 2
Adipocyte plasma membrane associated protein
Alcohol dehydrogenase 1
Alpha 1 acid glycoprotein 1
Alpha 1 acid glycoprotein 2
Alpha 1 antichymotrypsin
Alpha 1 antitrypsin
Alpha 1B glycoprotein
Alpha 2 antiplasmin
Alpha 2 HS glycoprotein
Alpha 2 macroglobulin
AN1 type zinc finger protein 2B
Angiotensinogen
Antithrombin III
Apolipoprotein A I
Apolipoprotein A II
Apolipoprotein A IV
Apolipoprotein C I
Apolipoprotein C II
Apolipoprotein C III
Apolipoprotein D
Apolipoprotein E
Apolipoprotein F
Apolipoprotein L1
Apolipoprotein M
Beta 2 glycoprotein 1
Beta Ala His dipeptidase

Brain acid soluble protein 1
C reactive protein
C type lectin domain family 4 member C
C4b binding protein alpha chain
C4b binding protein beta chain
Cadherin 5
Calmodulin like protein 3
Carboxypeptidase N subunit 2
Cartilage oligomeric matrix protein
CD5 antigen like
Cdc42 effector protein 3
Ceruloplasmin
Clusterin
Coagulation factor IX
Coagulation factor VII
Coagulation factor X
Coagulation factor XII
Coagulation factor XIII A chain
Coagulation factor XIII B chain
Colipase like protein C6orf127
Collectin 11
Complement C1q subcomponent subunit A
Complement C1q subcomponent subunit B
Complement C1q

subcomponent subunit C
Complement C1r subcomponent like protein
Complement C1r subcomponent
Complement C1s subcomponent
Complement C3
Complement C4 A
Complement C4 B
Complement component C9
Complement factor B
Complement factor H
Complement factor H related protein 1
Complement factor H related protein 3
Complement factor H related protein 4
Complement factor H related protein 5
DTW domain containing protein 1
EGF containing fibulin like extracellular matrix protein 1
Enoyl CoA hydratase mitochondrial
Exocyst complex component 1
Fibrinogen alpha chain
Fibrinogen beta chain
Fibrinogen gamma chain
Fibulin 1

Fibulin 5
Ficolin 3
Four and a half LIM domains protein 3
Galectin 3 binding protein
Gap junction beta 6 protein
Gelsolin
Glutathione peroxidase 3
Haptoglobin
Haptoglobin related protein
Hemoglobin subunit beta
Hemoglobin subunit delta
Hemoglobin subunit gamma 1
Hemopexin
Histidine rich glycoprotein
Hyaluronan binding protein 2
Ig alpha 1 chain C region
Ig alpha 2 chain C region
Ig gamma 1 chain C region
Ig gamma 2 chain C region
Ig gamma 3 chain C region
Ig gamma 4 chain C region
Ig heavy chain V III region BR
Ig heavy chain V III region WAS
Ig kappa chain C region
Ig kappa chain V I region EU
Ig kappa chain V II region TEW
Ig kappa chain V III region SIE

Ig kappa chain V III region VG Fragment
Ig kappa chain V III region WL
Ig kappa chain V IV region Len
Ig lambda 2 chain C regions
Ig lambda 3 chain C regions
Ig lambda chain V I region HA
Ig lambda chain V I region WAH
Ig lambda chain V III region LI
Ig lambda chain V IV region Hil
Ig mu chain C region
Ig mu heavy chain disease protein
Immunoglobulin J chain
Insulin like growth factor binding protein 4
Inter alpha trypsin inhibitor heavy chain H1
Inter alpha trypsin inhibitor heavy chain H2
Inter alpha trypsin inhibitor heavy chain H3
Inter alpha trypsin inhibitor heavy chain H4
Intercellular adhesion molecule 3
Keratocan
Kininogen 1
Lipopolysaccharide binding protein
Low affinity immunoglobulin gamma Fc region receptor II b
Lumican
Mannan binding lectin

serine protease 1
N acetylmuramoyl L alanine amidase
NudC domain containing protein 2
Pallidin
Phosphatidylcholine sterol acyltransferase
Phospholipid transfer protein
Plasminogen
Platelet factor 4
Platelet factor 4 variant
Pregnancy zone protein
Progesterone induced blocking factor 1
Protein AMBP
Protein Z dependent protease inhibitor
Prothrombin
Putative pre mRNA splicing factor ATP dependent RNA helicase DHX32
Putative trypsin 6
Putative uncharacterized protein L C100506887
Putative uncharacterized protein L C388882
Ras related protein Rab 44
Selenoprotein P
Serotransferrin
Serum albumin
Serum albumin
Serum amyloid A 4 protein
Serum amyloid A protein
Serum paraoxonase arylesterase 1
Serum paraoxonase lactonase 3
Sex hormone binding globulin

Sulfhydryl oxidase 1
Thrombospondin 1
Thrombospondin 4
Tigger transposable element derived protein 4
Transcobalamin 2
Transthyretin
Trypsin 1
Tubulin alpha 3C D chain

Uncharacterized protein C21orf58
Uncharacterized protein C9orf153
Vacuolar protein sorting associated protein 37A
Vasodilator stimulated phosphoprotein
Vitamin D binding protein
Vitamin K dependent

protein C
Vitamin K dependent protein S
Vitamin K dependent protein Z
Vitronectin
Zinc finger protein 143

Table A-2: List of the 271 proteins identified in the reproducibility study using the equalizer beads with SDC method

Actin cytoplasmic 1
Actin cytoplasmic 2
Adiponectin
Afamin
Alcohol dehydrogenase 1
Alpha 1 antichymotrypsin
Alpha 1 antitrypsin
Alpha 1B glycoprotein
Alpha 2 antiplasmin
Alpha 2 HS glycoprotein
Alpha 2 macroglobulin
Angiotensinogen
Ankyrin repeat and MYND domain containing protein 2
Antithrombin III
Apolipoprotein A I
Apolipoprotein A II
Apolipoprotein A IV
Apolipoprotein C I
Apolipoprotein C II
Apolipoprotein C III
Apolipoprotein D
Apolipoprotein E
Apolipoprotein L1
Apolipoprotein M
Attractin
Beta 2 glycoprotein 1

Beta Ala His dipeptidase
C4b binding protein alpha chain
Calcyclin binding protein
Carboxypeptidase N catalytic chain
Carboxypeptidase N subunit 2
Cathepsin B
CD5 antigen like
Clusterin
Coagulation factor IX
Coagulation factor X
Coagulation factor XII
Complement C1r subcomponent like protein
Complement C1r subcomponent
Complement C1s subcomponent
Complement C2
Complement C5
Complement component C6
Complement component C7
Complement component C8 alpha chain
Complement

component C8 beta chain
Complement component C8 gamma chain
Complement component C9
Complement factor B
Complement factor H
Complement factor H related protein 1
Complement factor H related protein 2
Complement factor H related protein 3
Complement factor H related protein 4
Complement factor I
Corticosteroid binding globulin
Extracellular matrix protein 1
Fibrinogen alpha chain
Fibrinogen beta chain
Fibrinogen gamma chain
Fibulin 1
Ficolin 3
Galectin 3 binding protein
Gap junction beta 6 protein
Gelsolin

Glutathione peroxidase 3
Golgi SNAP receptor complex member 2
Guanine nucleotide binding protein G I G S G subunit gamma 10
Haptoglobin
Haptoglobin related protein
Hemopexin
Heparin cofactor 2
Histidine rich glycoprotein
Hyaluronan binding protein 2
Ig alpha 1 chain C region
Ig alpha 2 chain C region
Ig gamma 1 chain C region
Ig gamma 2 chain C region
Ig gamma 3 chain C region
Ig gamma 4 chain C region
Ig kappa chain C region
Ig lambda 2 chain C regions
Ig mu chain C region
Ig mu heavy chain disease protein
Insulin like growth factor binding protein complex acid labile subunit
Inter alpha trypsin inhibitor heavy chain H1
Inter alpha trypsin inhibitor heavy chain H2
Inter alpha trypsin inhibitor heavy chain H3
Inter alpha trypsin inhibitor heavy chain H4
Kallistatin
Keratocan
Kininogen 1
L lactate dehydrogenase C chain
Leucine rich alpha 2 glycoprotein
Lumican
Matrix metalloproteinase 26
Monocyte differentiation antigen CD14
N acetylmuramoyl L alanine amidase
Paralemmin 2
Peptidase inhibitor 16
Phosphatidylinositol glycan specific phospholipase D
Photoreceptor specific nuclear receptor
Pigment epithelium derived factor
Plasma protease C1 inhibitor
Pregnancy zone protein
Programmed cell death protein 2
Properdin
Protein AMBP
Protein CutA
Protein ripply3
Prothrombin
Putative zinc alpha 2 glycoprotein like 1
Retinol binding protein 4
RNA pseudouridylate synthase domain containing protein 2
Serotransferrin
Serum albumin
Serum amyloid A 4 protein
Serum amyloid P component
Serum paraoxonase arylesterase 1
Sex hormone binding globulin
Sodium potassium transporting ATPase subunit beta 1 interacting protein 2
Tetranectin
Thyroxine binding globulin
Vasodilator stimulated phosphoprotein
Vesicular integral membrane protein VIP36
Vitamin D binding protein
Vitamin K dependent protein C
Vitronectin
Zinc alpha 2 glycoprotein
Zinc finger HIT domain containing protein 2
Adenine phosphoribosyltransferase
Adipocyte plasma membrane associated protein
Alpha 1 acid glycoprotein 1
Alpha 1 acid glycoprotein 2
AN1 type zinc finger protein 2B
Antigen presenting glycoprotein CD1d
Apolipoprotein F
Brain acid soluble protein 1
C reactive protein
C type lectin domain family 4 member C
C4b binding protein

beta chain
Cadherin 5
Calmodulin like protein 3
Cartilage oligomeric matrix protein
Caytaxin
Cdc42 effector protein 3
Cell differentiation protein RCD1 homolog
Cell surface A33 antigen
Centrin 2
Ceruloplasmin
Charged multivesicular body protein 4a
Cholesteryl ester transfer protein
Coagulation factor VII
Coagulation factor XIII A chain
Coagulation factor XIII B chain
Colipase like protein C6orf127
Collectin 11
Complement C1q subcomponent subunit A
Complement C1q subcomponent subunit B
Complement C1q subcomponent subunit C
Complement C3
Complement C4 A
Complement C4 B
Complement factor H related protein 5
Decorin
DNA directed RNA polymerase I subunit RPA43
DNA directed RNA polymerases I II and III subunit RPABC2

DTW domain containing protein 1
EGF containing fibulin like extracellular matrix protein 1
Enoyl CoA hydratase mitochondrial
Exocyst complex component 1
Fibulin 5
Four and a half LIM domains protein 3
G antigen 1
GatC like protein
Granzyme M
Hemoglobin subunit alpha
Hemoglobin subunit beta
Hemoglobin subunit delta
Hemoglobin subunit epsilon
Hemoglobin subunit gamma 1
Ig heavy chain V III region BR
Ig heavy chain V III region PM
Ig heavy chain V III region TIL
Ig heavy chain V III region WAS
Ig kappa chain V I region AG
Ig kappa chain V I region EU
Ig kappa chain V I region Hau
Ig kappa chain V I region Ni
Ig kappa chain V I region Roy
Ig kappa chain V I region WAT
Ig kappa chain V I region Wes
Ig kappa chain V II

region TEW
Ig kappa chain V III region CLL
Ig kappa chain V III region GL
Ig kappa chain V III region HIC
Ig kappa chain V III region SIE
Ig kappa chain V III region Ti
Ig kappa chain V III region VG Fragment
Ig kappa chain V III region WL
Ig kappa chain V IV region Len
Ig lambda chain V I region HA
Ig lambda chain V I region WAH
Ig lambda chain V III region LI
Ig lambda chain V III region SH
Ig lambda chain V IV region Hil
Immunoglobulin J chain
Insulin like growth factor binding protein 4
Intercellular adhesion molecule 3
Lipopolysaccharide binding protein
Low affinity immunoglobulin gamma Fc region receptor II b
Ly6 PLAUR domain containing protein 2
Lysophospholipase like protein 1
Mannan binding lectin serine protease 1
Methyltransferase like protein 4
NADH dehydrogenase ubiquinone iron sulfur

protein 6 mitochondrial
Neurocan core protein
Noggin
NudC domain containing protein 2
Pallidin
Phosphatidylcholine sterol acyltransferase
Phosphoenolpyruvate carboxykinase GTP mitochondrial
Phospholipid transfer protein
Plasminogen
Platelet factor 4
Platelet factor 4 variant
Probable histone acetyltransferase MYST1
Progesterone induced blocking factor 1
Prostate specific antigen
Protein FAM171B
Protein phosphatase 1 regulatory subunit 14B
Protein Z dependent protease inhibitor
Putative pre mRNA splicing factor ATP dependent RNA helicase DHX32
Putative trypsin 6
Putative uncharacterized protein C21orf77

Putative uncharacterized protein L C100506887
Putative uncharacterized protein L C388882
Ras related protein Rab 44
RWD domain containing protein 4
SAM pointed domain containing Ets transcription factor
Sarcoma antigen 1
Selenoprotein P
Serum amyloid A protein
Serum paraoxonase lactonase 3
Solute carrier organic anion transporter family member 5A1
Sterile alpha motif domain containing protein 13
Sulfhydryl oxidase 1
SWI SNF related matrix associated actin dependent regulator of chromatin subfamily E member 1
Syntaxin 12
Testis expressed sequence 13A protein
TFIIA alpha and beta

like factor
Thrombospondin 1
Thrombospondin 4
Tigger transposable element derived protein 4
Transcobalamin 2
Transthyretin
Trypsin 1
Tubulin alpha 3C D chain
Ubiquitin conjugating enzyme E2 Q2
Uncharacterized protein C1orf212
Uncharacterized protein C20orf196
Uncharacterized protein C21orf58
Uncharacterized protein C4orf32
Uncharacterized protein C5orf52
Uncharacterized protein C9orf153
Vacuolar protein sorting associated protein 37A
Vezatin
Vitamin K dependent protein S
Vitamin K dependent protein Z
Zinc finger protein 143
Zinc finger protein 449

Table A-3: List of the eighteen randomly selected proteins for the reproducibility study

P02790	Hemopexin
P08603	Complement factor H
P02647	Apolipoprotein A I
P02765	Alpha 2 HS glycoprotein
P00734	Prothrombin
P02768	Serum albumin
P04004	Vitronectin

P06727	Apolipoprotein A IV
P02760	Protein AMBP
Q96PD5	N acetylmuramoyl L alanine amidase
P09871	Complement C1s subcomponent
P02652	Apolipoprotein A II

P10909	Clusterin
P04003	C4b binding protein alpha chain
P02671	Fibrinogen alpha chain

P02649	Apolipoprotein E
Q14520	Hyaluronan binding protein 2
P00742	Coagulation factor X

Table A-4: A list of the proteins identified by MS^E (165) and $HDMS^E$ (228) using equalizer beads with sodium deoxycholate method

MS^E:

14 3 3 protein epsilon
14 3 3 protein zeta delta
Actin alpha skeletal muscle
Actin cytoplasmic 1
Actin cytoplasmic 2
Actin gamma enteric smooth muscle
Alpha 1 antitrypsin
Alpha 2 antiplasmin
Alpha 2 HS glycoprotein
Alpha 2 macroglobulin
Angiotensinogen
Ankyrin repeat and SCS box protein 14
Apolipoprotein A I
Apolipoprotein A II
Apolipoprotein A IV
Apolipoprotein C I
Apolipoprotein C II
Apolipoprotein C III
Apolipoprotein D
Apolipoprotein E
Apolipoprotein F
Apolipoprotein L1
Apolipoprotein M
Beta 2 glycoprotein 1
Beta Ala His dipeptidase
C reactive protein
C4b binding protein alpha chain
C4b binding protein beta chain

Cadherin 5
Calcium binding protein 8
Cartilage acidic protein 1
Cartilage oligomeric matrix protein
CD5 antigen like
Ceruloplasmin
Clusterin
Coagulation factor IX
Coagulation factor VII
Coagulation factor X
Coagulation factor XII
Coagulation factor XIII A chain
Coagulation factor XIII B chain
Collectin 11
Complement C1q subcomponent subunit A
Complement C1q subcomponent subunit B
Complement C1q subcomponent subunit C
Complement C1r subcomponent
Complement C1s subcomponent
Complement C3
Complement C4 A
Complement C4 B
Complement

component C9
Complement factor D
Complement factor H
Complement factor H related protein 1
Complement factor H related protein 2
Complement factor H related protein 3
Complement factor H related protein 4
Complement factor H related protein 5
EGF containing fibulin like extracellular matrix protein 1
Extracellular superoxide dismutase Cu Zn
Fibrinogen alpha chain
Fibrinogen beta chain
Fibrinogen gamma chain
Fibulin 1
Fibulin 5
Ficolin 3
G patch domain containing protein 2
Galectin 3 binding protein
Gelsolin
Glutathione peroxidase 3
Haptoglobin
Haptoglobin related protein
Hemopexin

Heparin cofactor 2
Hepatocyte growth factor activator
Histidine rich glycoprotein
HLA class I histocompatibility antigen Cw 17 alpha chain
Hyaluronan binding protein 2
Ig alpha 1 chain C region
Ig alpha 2 chain C region
Ig gamma 1 chain C region
Ig gamma 2 chain C region
Ig gamma 3 chain C region
Ig gamma 4 chain C region
Ig heavy chain V III region BR
M
Ig heavy chain V III region TIL
Ig heavy chain V III region TUR
Ig kappa chain C region
Ig kappa chain V I region AU
Ig kappa chain V I region Gal
Ig kappa chain V I region Lay
Ig kappa chain V I region WAT
Ig kappa chain V III region HIC
Ig kappa chain V III region SIE
Ig kappa chain V III region VG Fragment
Ig kappa chain V III region WL
Ig kappa chain V IV

region Len
Ig lambda 1 chain C regions
Ig lambda 2 chain C regions
Ig lambda 3 chain C regions
Ig lambda chain V III region LI
Ig lambda chain V III region SH
Ig mu chain C region
Ig mu heavy chain disease protein
Immunoglobulin J chain
Immunoglobulin lambda like polypeptide 5
Insulin like growth factor binding protein 4
Inter alpha trypsin inhibitor heavy chain H1
Inter alpha trypsin inhibitor heavy chain H2
Inter alpha trypsin inhibitor heavy chain H3
Inter alpha trypsin inhibitor heavy chain H4
Keratin associated protein 19 5
Kininogen 1
Lipopolysaccharide binding protein
Lumican
Mannan binding lectin serine protease 1
Mannan binding lectin serine protease 2
Membrane spanning 4 domains subfamily A member 18
N acetylmuramoyl L alanine amidase
Neutrophil defensin 3

Nuclear factor of activated T cells cytoplasmic 4
Nuclear interacting partner of ALK
Phosphatidylcholine sterol acyltransferase
Phosphatidylinositol glycan specific phospholipase D
Phospholipid transfer protein
Plasma kallikrein
Plasminogen
Platelet factor 4
Platelet factor 4 variant
Pregnancy zone protein
Proteasome subunit beta type 9
Protein AMBP
Protein CREG2
Protein phosphatase 1M
Protein Z dependent protease inhibitor
Prothrombin
Putative BMS1 like protein ENSP00000383048
Putative uncharacterized protein FLJ45825
Retinoic acid receptor responder protein 2
Retinol binding protein 4
Selenoprotein P
Serotransferrin
Serum albumin
Serum albumin
Serum amyloid A 4 protein
Serum amyloid A protein
Serum paraoxonase arylesterase 1
Sulfhydryl oxidase 1
Tetranectin

Thioredoxin domain containing protein 9
Thrombospondin 4
Transcription cofactor vestigial like protein 1
Transforming growth factor beta induced protein ig h3
Transthyretin
Tubulin alpha 4A chain

Tubulin beta 1 chain
Uncharacterized protein C22orf34
UPF0583 protein C15orf59
Vacuolar protein sorting associated protein 4A
Vitamin D binding protein
Vitamin K dependent

protein C
Vitamin K dependent protein S
Vitronectin
Zinc finger protein 566

HDMS^{E}:

14 3 3 protein epsilon
14 3 3 protein epsilon
14 3 3 protein zeta delta
72 kDa type IV collagenase
Actin alpha skeletal muscle
Actin cytoplasmic 1
Actin cytoplasmic 2
Actin gamma enteric smooth muscle
Adiponectin
Alcohol dehydrogenase 1
Alpha 1 antichymotrypsin
Alpha 1 antitrypsin
Alpha 1B glycoprotein
Alpha 1B glycoprotein
Alpha 2 antiplasmin
Alpha 2 HS glycoprotein
Alpha 2 macroglobulin
Alpha actinin 1
Alpha actinin 4
Angiotensinogen
Antithrombin III
Apolipoprotein A I
Apolipoprotein A II
Apolipoprotein A IV
Apolipoprotein C I
Apolipoprotein C II
Apolipoprotein C III

Apolipoprotein D
Apolipoprotein E
Apolipoprotein F
Apolipoprotein L1
Beta 2 glycoprotein 1
Beta Ala His dipeptidase
C reactive protein
C4b binding protein alpha chain
C4b binding protein beta chain
Cadherin 1
Cadherin 13
Cadherin 5
Carboxypeptidase N subunit 2
Cartilage oligomeric matrix protein
CD5 antigen like
Ceruloplasmin
Clusterin
Coagulation factor IX
Coagulation factor VII
Coagulation factor X
Coagulation factor XI
Coagulation factor XI
Coagulation factor XII
Coagulation factor XIII A chain
Coagulation factor XIII B chain
Collectin 10

Collectin 11
Complement C1q subcomponent subunit A
Complement C1q subcomponent subunit B
Complement C1q subcomponent subunit C
Complement C1r subcomponent
Complement C1s subcomponent
Complement C3
Complement C4 A
Complement C4 B
Complement component C8 alpha chain
Complement component C8 alpha chain
Complement component C8 gamma chain
Complement component C9
Complement factor D
Complement factor H
Complement factor H related protein 1
Complement factor H related protein 2

Complement factor H related protein 3
Complement factor H related protein 4
Complement factor H related protein 5
Cystatin C
Desmoglein 2
DNA binding protein inhibitor ID 1
EGF containing fibulin like extracellular matrix protein 1
Endoplasmin
Endoplasmin
Extracellular matrix protein 1
Extracellular superoxide dismutase Cu Zn
Fibrinogen alpha chain
Fibrinogen beta chain
Fibrinogen gamma chain
Fibulin 1
Fibulin 5
Ficolin 2
Ficolin 3
Forkhead box protein J2
Galectin 3 binding protein
Gelsolin
Glucokinase regulatory protein
Glucokinase regulatory protein
Glutathione peroxidase 3
Haptoglobin
Haptoglobin related protein
Hemoglobin subunit alpha
Hemoglobin subunit beta
Hemopexin
Heparin cofactor 2

Hepatocyte growth factor activator
Hepatocyte growth factor activator
Histidine rich glycoprotein
Hyaluronan binding protein 2
Ig alpha 1 chain C region
Ig alpha 2 chain C region
Ig gamma 1 chain C region
Ig gamma 2 chain C region
Ig gamma 3 chain C region
Ig gamma 4 chain C region
Ig heavy chain V III region BR
Ig heavy chain V III region PM
Ig heavy chain V III region TEI
Ig heavy chain V III region TIL
Ig heavy chain V III region TUR
Ig heavy chain V III region WAS
Ig kappa chain C region
Ig kappa chain V I region AG
Ig kappa chain V I region AU
Ig kappa chain V I region EU
Ig kappa chain V I region Gal
Ig kappa chain V I region Ni
Ig kappa chain V I region Scw
Ig kappa chain V I region WEA
Ig kappa chain V II

region TEW
Ig kappa chain V III region HIC
Ig kappa chain V III region NG9 Fragment
Ig kappa chain V III region VG Fragment
Ig kappa chain V III region WL
Ig kappa chain V IV region Len
Ig lambda 2 chain C regions
Ig lambda 6 chain C region
Ig lambda chain V I region HA
Ig lambda chain V III region LI
Ig lambda chain V III region SH
Ig lambda chain V IV region Hil
Ig mu chain C region
Ig mu heavy chain disease protein
Immunoglobulin J chain
Immunoglobulin lambda like polypeptide 5
Insulin like growth factor binding protein 4
Insulin like growth factor II
Inter alpha trypsin inhibitor heavy chain H1
Inter alpha trypsin inhibitor heavy chain H2
Inter alpha trypsin inhibitor heavy chain H3
Inter alpha trypsin inhibitor heavy chain H4
Kallistatin
Kininogen 1

LEM domain containing protein 2
Lipopolysaccharide binding protein
Lumican
Mannan binding lectin serine protease 1
Mannan binding lectin serine protease 2
Mannan binding lectin serine protease 2
Mannose binding protein C
Matrix Gla protein
Methyltransferase like protein 4
MICAL like protein 2
Myosin light polypeptide 6
N acetylmuramoyl L alanine amidase
N acylneuraminate cytidylyltransferase
Neuferricin
Paladin
Peroxisomal 2 4 dienoyl CoA reductase
Phosphatidylcholine sterol acyltransferase
Phosphatidylinositol glycan specific phospholipase D
Phospholipid transfer protein
Pigment epithelium derived factor
Plasma kallikrein
Plasma protease C1 inhibitor
Plasma protease C1 inhibitor
Plasma serine protease inhibitor

Plasminogen
Platelet basic protein
Platelet factor 4
Platelet glycoprotein Ib alpha chain
Platelet glycoprotein Ib alpha chain
PTE ankyrin domain family member E
Pregnancy zone protein
Properdin
Properdin
Protein ACN9 homolog mitochondrial
Protein AMBP
Protein S100 A9
Protein Z dependent protease inhibitor
Proteoglycan 4
Proteoglycan 4
Prothrombin
Protocadherin beta 12
Retinol binding protein 4
Secernin 1
Selenoprotein P
Selenoprotein P
Serotransferrin
Serum albumin
Serum albumin
Serum amyloid A 4 protein
Serum amyloid A protein
Serum amyloid P component
Serum paraoxonase arylesterase 1
SPARC like protein 1
SPARC
Sulfhydryl oxidase 1

Thrombospondin 1
Thrombospondin 4
Transforming growth factor beta induced protein ig h3
Transforming growth factor beta induced protein ig h3
Transitional endoplasmic reticulum ATPase
Transthyretin
TTD non photosensitive 1 protein
Tubulin alpha 1B chain
Tubulin alpha 1C chain
Tubulin alpha 4A chain
Tubulin beta 1 chain
Tubulin beta 2A chain
Tubulin beta 3 chain
Tubulin beta 6 chain
Tubulin beta 8 chain
Tubulin polymerization promoting protein family member 2
Vitamin D binding protein
Vitamin K dependent protein C
Vitamin K dependent protein S
Vitamin K dependent protein Z
Vitamin K dependent protein Z
Vitronectin
ZAR1 like protein
Zinc finger protein 165
Zinc finger protein basonuclin 2

8.2 Appendix B

List of the up and down regulated proteins identified by Expression^{E} and Progenesis

Table B-1: Proteins identified from 2D-LC-HDMS^{E} analysis of pooled plasma. The protein denoted as down regulated (0 - 0.05) were lower in healthy controls therefore, indicative of disease. The proteins denoted as up regulated (0.95 – 1) were higher in healthy controls therefore, protective.

Apolipoprotein C I	0
Apolipoprotein C III	0
C reactive protein	0
Ceruloplasmin	0
Coagulation factor XII	0
Collagen alpha 1 XVIII chain	0
Collectin 11	0
Complement component C8 beta chain	0
Complement factor H related protein 2	0
EGF containing fibulin like extracellular matrix protein 1	0
Endoplasmin	0
Extracellular matrix protein 1	0
Extracellular superoxide dismutase Cu Zn	0
Fibulin 1	0
Fibulin 5	0
Galectin 3 binding protein	0
Insulin like growth factor binding protein 2	0
Insulin like growth factor binding protein 4	0
Inter alpha trypsin inhibitor heavy chain H3	0
Kininogen 1	0
Lumican	0
Nidogen 1	0
Peroxisomal 2 4 dienoyl CoA reductase	0
Pigment epithelium derived factor	0
Platelet factor 4	0
Polyubiquitin B	0
Pregnancy zone protein	0
Protein AMBP	0
Protocadherin beta 12	0
Selenoprotein P	0
Serum amyloid A 4 protein	0
Serum amyloid A protein	0
Serum amyloid P component	0
Sulfhydryl oxidase 1	0
Vitamin K dependent protein C	0
Desmoglein 2	0.01
SPARC like protein 1	0.01
Antithrombin III	0.02
Alpha 1 acid glycoprotein 2	0.03
TH complex subunit 7 homolog	0.03
A disintegrin and metalloproteinase with thrombospondin motifs 13	0.95
Secreted phosphoprotein 24	0.97
SPARC	0.97
Myosin regulatory light polypeptide 9	0.98
Peripheral plasma membrane protein CASK	0.98
Putative tubulin beta chain like protein ENSP00000290377	0.98
Ficolin 2	0.99
Tubulin beta 3 chain	0.99
Tubulin beta 6 chain	0.99

14 3 3 protein beta alpha	1
14 3 3 protein epsilon	1
14 3 3 protein eta	1
14 3 3 protein gamma	1
14 3 3 protein theta	1
14 3 3 protein zeta delta	1
Actin alpha cardiac muscle 1	1
Actin alpha skeletal muscle	1
Actin aortic smooth muscle	1
Actin cytoplasmic 1	1
Actin cytoplasmic 2	1
Actin gamma enteric smooth muscle	1
Adipocyte plasma membrane associated protein	1
Afamin	1
Alpha actinin 1	1
Alpha actinin 4	1
Apolipoprotein A II	1
Apolipoprotein C IV	1
Apolipoprotein D	1
Beta actin like protein 2	1
Beta Ala His dipeptidase	1
C4b binding protein beta chain	1
Calreticulin	1
Coagulation factor VII	1
Coagulation factor XIII A chain	1
Complement C1q subcomponent subunit B	1
Complement component C8 alpha chain	1
Complement factor H	1
Complement factor H related protein 1	1
Complement factor H related protein 3	1
Fanconi anemia group E protein	1
Insulin like growth factor II	1
Integrin beta 3	1
Intelectin 1	1
Inter alpha trypsin inhibitor heavy chain H2	1
Mannan binding lectin serine protease 1	1

Myosin light chain 6B	1
Myosin light polypeptide 6	1
Myosin regulatory light chain 12A	1
Myosin regulatory light chain 12B	1
Phosphatidylinositol glycan specific phospholipase D	1
Platelet basic protein	1
Platelet factor 4 variant	1
Platelet glycoprotein Ib beta chain	1
POTE ankyrin domain family member E	1
PTE ankyrin domain family member F	1
PTE ankyrin domain family member I	1
PTE ankyrin domain family member J	1
Properdin	1
Protein Z dependent protease inhibitor	1
Serum albumin	1
Serum paraoxonase lactonase 3	1
TFIIH basal transcription factor complex helicase XPB subunit	1
Transthyretin	1
Tropomyosin alpha 1 chain	1
Tropomyosin alpha 3 chain	1
Tropomyosin alpha 4 chain	1
Tropomyosin beta chain	1
Tubulin alpha 3C D chain	1
Tubulin alpha 3E chain	1
Tubulin alpha 4A chain	1
Tubulin beta 1 chain	1
Tubulin beta 2A chain	1
Tubulin beta 2B chain	1
Tubulin beta 8 chain B	1
Tubulin beta chain	1
Vitamin D binding protein	1
Vitamin K dependent protein S	1
Vitamin K dependent protein Z	1

Table B-2: A list of all the proteins identified by progenesis LC-MS for the acute heart failure study. After filtering, a total of 579 proteins were identified by Progenesis LC-MS, of which 220 were expressed higher in non-survivors (A), 287 higher in survivors (B) and 72 were reported with no significant changes in expression (C)

(A)

14 3 3 protein gamma
15 hydroxyprostaglandin dehydrogenase NAD
26S proteasome non ATPase regulatory subunit 7
28S ribosomal protein S34 mitochondrial
40S ribosomal protein S14
40S ribosomal protein SA
A disintegrin and metalloproteinase with thrombospondin motifs 13
Actin cytoplasmic 1
Actin cytoplasmic 2
Acyl CoA binding domain containing protein 5
Acyl CoA dehydrogenase family member 9 mitochondrial
Alcohol dehydrogenase 1
Aldehyde dehydrogenase dimeric NADP preferring
Alpha 1 acid glycoprotein 1
Alpha 1 antichymotrypsin
Alpha 1B glycoprotein

Alpha 2 HS glycoprotein
Anion exchange protein 4
Annexin 2 receptor
Annexin A9
Antithrombin III
Apolipoprotein A V
Apolipoprotein L1
Arf GAP with dual PH domain containing protein 1
Arf GAP with GTPase ANK repeat and PH domain containing protein 5
Axonemal dynein light intermediate polypeptide 1
Beta 1 3 N acetylglucosaminyltransferase lunatic fringe
Beta 2 glycoprotein 1
Beta Ala His dipeptidase
Biogenesis of lysosome related organelles complex 1 subunit 2
Bone marrow proteoglycan
BTB PZ domain containing protein KCTD9
C reactive protein
C4b binding protein beta chain

Cadherin 1
Cadherin related family member 2
Calmodulin
Calpain 7
cAMP dependent protein kinase inhibitor gamma
Carbohydrate sulfotransferase 4
Cathelicidin antimicrobial peptide
CD2 antigen cytoplasmic tail binding protein 2
Cell differentiation protein RCD1 homolog
Ceruloplasmin
Chitotriosidase 1
Cholesteryl ester transfer protein
Chromogranin A
Coagulation factor VII
Coagulation factor XI
Coagulation factor XII
Collagen alpha 1 XXV chain
Collectin 10
Complement C5
Complement component C8 beta chain
Complement component C8 gamma chain

Complement factor H related protein 2
Cullin 4A
Cystatin D
Cytosol aminopeptidase
D tyrosyl tRNA Tyr deacylase 1
Desmoglein 2
Dihydrofolate reductase
Discoidin domain containing receptor 2
DNA apurinic or apyrimidinic site lyase
DNA repair protein RAD51 homolog 1
DnaJ homolog subfamily B member 13
Dual specificity protein phosphatase CDC14C
Endophilin A2
Enolase 1
Extracellular superoxide dismutase Cu Zn
Fermitin family homolog 3
Glutaredoxin like protein C5orf63
Golgi reassembly stacking protein 2
GRB2 related adapter protein
H ACA ribonucleoprotein complex subunit 1
Haptoglobin related protein
Heat shock factor protein 4
Heat shock protein 75 kDa mitochondrial
Heat shock protein HSP 90 beta
Hematopoietic progenitor cell antigen CD34
Hemoglobin subunit beta

Hemoglobin subunit delta
Heterogeneous nuclear ribonucleoprotein A3
Hippocalcin like protein 4
Histidine rich glycoprotein
Homeobox protein DBX2
Homeobox protein Hox A11
Homeobox protein Hox B6
Homer protein homolog 2
Ig alpha 1 chain C region
Ig gamma 1 chain C region
Ig gamma 2 chain C region
Ig heavy chain V I region HG3
Ig heavy chain V II region SESS
Ig heavy chain V III region BR
Ig heavy chain V III region GAL
Ig heavy chain V III region TEI
Ig kappa chain C region
Ig kappa chain V II region TEW
Ig kappa chain V III region NG9 Fragment
Ig kappa chain V III region WL
Ig lambda chain V III region SH
Ig lambda chain V IV region Hil
Ig mu heavy chain disease protein
Insulin like growth factor binding protein complex acid labile

subunit
Intracellular hyaluronan binding protein 4
Intraflagellar transport protein 20 homolog
JmjC domain containing protein 4
JmjC domain containing protein 7
Kallistatin
Keratin type I cytoskeletal 14
Keratin type I cytoskeletal 16
Kininogen 1
Kv channel interacting protein 1
Lactotransferrin
Lamin B1
Lipopolysaccharide binding protein
Lipoprotein lipase
Lumican
Matrix Gla protein
Mesencephalic astrocyte derived neurotrophic factor
Metalloproteinase inhibitor 2
Methyltransferase like protein 8
Mismatch repair endonuclease PMS2
Molybdopterin synthase catalytic subunit
Multiple inositol polyphosphate phosphatase 1
Myogenic factor 6
NADPH adrenodoxin oxidoreductase mitochondrial
NADPH oxidase 4
Neutrophil defensin 1
Nuclear mitotic apparatus protein 1
Olfactomedin like

protein 1
Olfactory receptor 5M8
Oncomodulin 1
Oncoprotein induced transcript 3 protein
Peroxisomal 2 4 dienoyl CoA reductase
PH and SEC7 domain containing protein 3
Phosphatidylinositol glycan specific phospholipase D
Phospholipid transfer protein
Plasma serine protease inhibitor
Platelet factor 4
Potassium voltage gated channel subfamily E member 1
POTE ankyrin domain family member F
Probable palmitoyltransferase ZDHHC5
Procollagen C endopeptidase enhancer 1
Programmed cell death protein 4
Proprotein convertase subtilisin kexin type 9
Protein BANP
Protein ERGIC 53
Protein FAM166B
Protein FAM195B
Protein KTI12 homolog
Protein MEM1
Protein rogdi homolog
Proteoglycan 4
Pulmonary surfactant associated protein B
Putative beta actin like protein 3
Putative histone lysine N methyltransferase PRDM6

Putative hydrolase RBBP9
Putative lipocalin 1 like protein 1
Putative TBC1 domain family member 29
Putative tropomyosin alpha 3 chain like protein
Putative uncharacterized protein FLJ00310
PWWP domain containing protein 2A
RAC gamma serine threonine protein kinase
Ras association domain containing protein 10
RAS guanyl releasing protein 2
Ras related protein Rab 30
Ras related protein Rab 4A
Ras specific guanine nucleotide releasing factor 1
Regulation of nuclear pre mRNA domain containing protein 1B
Ribulose phosphate 3 epimerase
SEC14 domain and spectrin repeat containing protein 1
Serine protease HTRA1
Serine threonine protein kinase 4
Serine threonine protein kinase ULK3
Serotransferrin
Serum amyloid A 4 protein
Serum amyloid A protein
Serum deprivation response protein
Serum paraoxonase

lactonase 3
Sex hormone binding globulin
SH3 containing GRB2 like protein 3 interacting protein 1
Sorcin
Sorting nexin 17
Sulfatase modifying factor 2
Talin 1
Testis expressed sequence 264 protein
Tetranectin
THO complex subunit 7 homolog
Thrombospondin 4
Transducin like enhancer protein 3
Transforming growth factor beta induced protein ig h3
Transmembrane protein 14E
Transmembrane protein 40
tRNA methyltransferase 112 homolog
Tubulin beta 1 chain
Tubulin beta 6 chain
Tubulin beta chain
Tuftelin
Tyrosine protein kinase ZAP 70
Tyrosine protein phosphatase non receptor type 9
Ubiquitin carboxyl terminal hydrolase 27
Ubiquitin conjugating enzyme E2 G1
Uncharacterized protein C10orf131
Uncharacterized protein C16orf55
Uncharacterized protein C1orf122

Uncharacterized protein C2orf42
Uncharacterized protein C7orf62
Uncharacterized protein C8orf78
Uncharacterized protein UNQ511 PRO 1026
Vesicle transport through interaction with t SNAREs homolog 1A

Vitamin K dependent protein Z
Vitelline membrane outer layer protein 1 homolog
WD repeat containing protein 54
Williams Beuren syndrome chromosomal region 16 protein
WW domain binding

protein 5
Zinc alpha 2 glycoprotein
Zinc finger protein 2
Zinc finger protein 568

(B)

14 3 3 protein beta alpha
14 3 3 protein epsilon
14 3 3 protein zeta delta
39S ribosomal protein L3 mitochondrial
72 kDa type IV collagenase
Actin aortic smooth muscle
Actin gamma enteric smooth muscle
Adenomatous polyposis coli protein
Adiponectin
ADP ribosylation factor like protein 14
ADP ribosylation factor like protein 15
Afamin
Alcohol dehydrogenase NADP
Alpha 1 acid glycoprotein 2
Alpha 1 antitrypsin
Alpha 2 antiplasmin
Alpha 2 macroglobulin
Alpha actinin 1
Alpha actinin 2
Alpha actinin 3
Alpha actinin 4
Alpha ketoglutarate dependent dioxygenase FTO

Amyloid beta A4 protein
Anaphase promoting complex subunit 16
Angiotensinogen
Ankyrin repeat and SOCS box protein 7
AP 3 complex subunit delta 1
Apolipoprotein A I
Apolipoprotein A II
Apolipoprotein A IV
Apolipoprotein B 100
Apolipoprotein C I
Apolipoprotein C II
Apolipoprotein C III
Apolipoprotein C IV
Apolipoprotein D
Apolipoprotein E
Apolipoprotein F
Apolipoprotein M
ATP binding cassette sub family F member 2
Basic leucine zipper transcriptional factor ATF like 3
BH3 interacting domain death agonist
BTB POZ domain containing protein KCTD16
C C motif chemokine 17
C C motif chemokine

18
C4b binding protein alpha chain
Cadherin 12
Cadherin 5
Calcium binding protein p22
Calreticulin
Carboxypeptidase N subunit 2
Cartilage oligomeric matrix protein
Caspase 3
CD5 antigen like
Centrosomal protein of 95 kDa
Chloride intracellular channel protein 1
Clusterin
Coagulation factor IX
Coagulation factor V
Coagulation factor X
Coagulation factor XIII A chain
Coagulation factor XIII B chain
Coiled coil domain containing protein 92
Collagen alpha 1 XXII chain
Collectin 11
Complement C1q subcomponent subunit A

Complement C1q subcomponent subunit B
Complement C1q subcomponent subunit C
Complement C1r subcomponent
Complement C1s subcomponent
Complement C3
Complement C4 A
Complement C4 B
Complement component C8 alpha chain
Complement component C9
Complement factor D
Complement factor H
Complement factor H related protein 1
Complement factor H related protein 3
Complement factor H related protein 5
Dickkopf related protein 3
Dynamin like 120 kDa protein mitochondrial
E3 ubiquitin protein ligase BRE1B
E3 ubiquitin protein ligase RBX1
EGF containing fibulin like extracellular matrix protein 1
Endoplasmin
F box only protein 17
Fibrinogen alpha chain
Fibrinogen beta chain
Fibrinogen gamma chain
Fibrinogen like protein 1
Fibronectin
Fibronectin type III

domain containing protein 5
Fibulin 1
Fibulin 5
Ficolin 2
Ficolin 3
Forkhead box protein D4 like 3
Fructose bisphosphate aldolase C
Galectin 3 binding protein
Gelsolin
Glucosidase 2 subunit beta
Glutathione peroxidase 3
Glyceraldehyde 3 phosphate dehydrogenase
Golgin subfamily B member 1
GRIP and coiled coil domain containing protein 1
Guanine nucleotide binding protein subunit alpha 13
Haptoglobin
Heat shock 70 kDa protein 1 like
Hemicentin 2
Hemoglobin subunit alpha
Hemoglobin subunit gamma 2
Hemopexin
Heparin cofactor 2
Hepatocyte growth factor activator
Histone deacetylase 11
HLA class I histocompatibility antigen B 15 alpha chain
HLA class II histocompatibility

antigen DRB1 11 beta chain
Hyaluronan binding protein 2
Ig alpha 2 chain C region
Ig gamma 3 chain C region
Ig gamma 4 chain C region
Ig heavy chain V II region NEWM
Ig heavy chain V III region TUR
Ig heavy chain V III region WEA
Ig kappa chain V I region EU
Ig kappa chain V I region Gal
Ig kappa chain V I region Ni
Ig kappa chain V II region Cum
Ig kappa chain V III region GOL
Ig kappa chain V III region POM
Ig kappa chain V III region SIE
Ig kappa chain V IV region Len
Ig lambda 1 chain C regions
Ig lambda 2 chain C regions
Ig lambda 3 chain C regions
Ig lambda 6 chain C region
Ig lambda chain V III region LOI
Ig mu chain C region
Immunoglobulin J chain
Immunoglobulin lambda like polypeptide 5
Inhibin beta C chain

Insulin like growth factor binding protein 2
Insulin like growth factor binding protein 3
Insulin like growth factor binding protein 4
Insulin like growth factor II
Integrin alpha IIb
Integrin beta 3
Intelectin 1
Inter alpha trypsin inhibitor heavy chain H1
Inter alpha trypsin inhibitor heavy chain H2
Inter alpha trypsin inhibitor heavy chain H3
Inter alpha trypsin inhibitor heavy chain H4
Interferon alpha 6
Interferon induced protein 44 like
Junctophilin 3
Keratin type I cytoskeletal 10
Keratin type I cytoskeletal 9
Keratin type II cytoskeletal 1
Keratin type II cytoskeletal 5
Kinectin
L lactate dehydrogenase C chain
Leucine rich repeat containing protein 1
Ligand of Numb protein X 2
LSM domain containing protein 1
Lys 63 specific deubiquitinase BRCC36
Lysozyme C

MAGUK p55 subfamily member 6
Mannan binding lectin serine protease 1
Mannan binding lectin serine protease 2
Mannose binding protein C
MAPK MAK MRK overlapping kinase
Meiotic recombination protein SPO 11
Mitogen activated protein kinase kinase kinase 4
Monocyte differentiation antigen CD14
Monofunctional C1 tetrahydrofolate synthase mitochondrial
Muscleblind like protein 1
Myosin 9
Myosin light polypeptide 6
N acetylmuramoyl L alanine amidase
Nidogen 1
NTPase KAP family P loop domain containing protein 1
Nuclear protein 1
Nuclear receptor binding protein 2
Nucleoporin GLE1
Organic solute transporter subunit beta
Peptidyl prolyl cis trans isomerase H
Peptidyl prolyl cis trans isomerase like 6
Periodic tryptophan protein 1 homolog
Phosphatidylcholine sterol acyltransferase
PI PLC X domain

containing protein 2
Pigment epithelium derived factor
Plasma kallikrein
Plasma protease C1 inhibitor
Plasminogen
Platelet basic protein
Platelet factor 4 variant
Platelet glycoprotein Ib alpha chain
Pleckstrin homology domain containing family M member 1
Poly rC binding protein 3
POTE ankyrin domain family member E
POTE ankyrin domain family member J
Prefoldin subunit 3
Pregnancy zone protein
Probable ATP dependent RNA helicase DDX58
Proline rich acidic protein 1
Properdin
Protein AMBP
Protein CBFA2T3
Protein disulfide isomerase A6
Protein EAN57
Protein FAM154A
Protein FAM9C
Protein SOX 15
Protein Z dependent protease inhibitor
Prothrombin
Putative DAP 2 like protein C8orf68
Putative protein ZNF720
Putative serum amyloid A 3 protein
Putative uncharacterized protein

C21orf32
Putative V set and immunoglobulin domain containing like protein ENSP00000303034
Putative zinc finger protein 861
Rab GTPase binding effector protein 1
Retinoic acid receptor responder protein 2
Retinol binding protein 4
Retinol dehydrogenase 8
Rho GTPase activating protein 44
RUN domain containing protein 3A
SAM domain and HD domain containing protein 1
Secreted phosphoprotein 24
Selenoprotein P
Semaphorin 3C
Septin 6
Serglycin
Serine threonine protein kinase 38 like
Serine threonine protein kinase Nek9
Serine threonine protein kinase SMG1
Serum albumin
Serum albumin
Serum amyloid P component
Serum paraoxonase arylesterase 1
Small nuclear ribonucleoprotein

associated proteins B and B
Sorting nexin 22
Sorting nexin 25
SPARC like protein 1
SPARC
Spectrin beta chain brain 2
Sprouty related EVH1 domain containing protein 3
Stimulated by retinoic acid gene 8 protein homolog
Sulfhydryl oxidase 1
Susceptibility protein NSG x
Synaptonemal complex central element protein 2
T box transcription factor TBX15
Talanin
TBC1 domain family member 3F
Telomere length regulation protein TEL2 homolog
Testis expressed sequence 22 protein
Tetraspanin 3
Thrombospondin 1
TIR domain containing adapter molecule 2
TP53 target gene 3 protein
Transcriptional repressor CTCFL
Transmembrane protein 198
Transmembrane protein 39B
Transthyretin

Trichoplein keratin filament binding protein
Tropomyosin alpha 3 chain
Tropomyosin alpha 4 chain
Tubulin alpha 4A chain
Tubulin beta 2B chain
Tyrosine protein kinase JAK2
Uncharacterized protein C15orf43
Uncharacterized protein C17orf77
Uncharacterized protein C1orf182
Uncharacterized protein C20orf196
UPF0534 protein C4orf43
Vinculin
Vitamin D binding protein
Vitamin K dependent protein C
Vitamin K dependent protein S
Vitronectin
von Willebrand factor
Zinc finger CCCH domain containing protein 6
Zinc finger matrin type protein 4
Zinc finger protein basonuclin 2
Zinc finger protein with KRAB and SCAN domains 3
Zinc finger protein with KRAB and SCAN domains 4

(C)

1 2 dihydroxy 3 keto 5 methylthiopentene

dioxygenase
14 3 3 protein eta

5 AMP activated protein kinase subunit

beta 2
Activator of 90 kDa heat shock protein ATPase homolog 1
AN1 type zinc finger protein 2B
Arfaptin 1
Beta actin like protein 2
Bifunctional methylenetetrahydrofolate dehydrogenase cyclohydrolase mitochondrial
C terminal binding protein 1
C type lectin domain family 3 member A
Cathepsin G
Centrosomal protein of 290 kDa
Centrosomal protein of 57 kDa
Coiled coil domain containing protein 62
Cyclin dependent kinase 2 interacting protein
Cytochrome b5 type B
Dixin
DNA cross link repair 1A protein
Dynactin subunit 1
Dynein heavy chain 17 axonemal
EF hand domain containing protein D1
Exosome complex component RRP40
FERM domain containing protein 4B
G2 M phase specific E3 ubiquitin protein ligase
Galectin 2
Glucocorticoid receptor DNA binding factor 1
Golgin subfamily A member 3

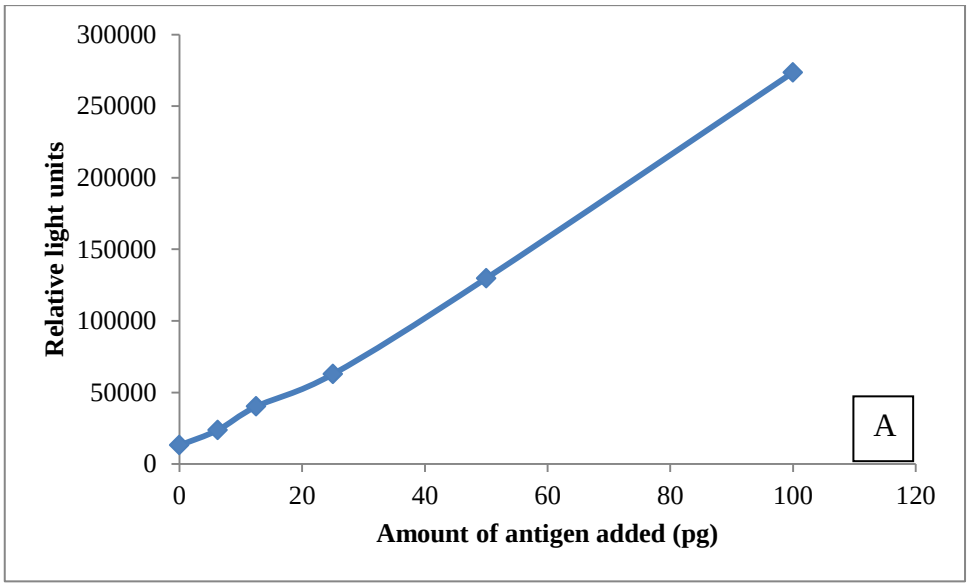
Histone lysine N methyltransferase MLL4
Keratin type I cuticular Ha3 I
Ligand dependent nuclear receptor interacting factor 1
Mitochondrial import inner membrane translocase subunit TIM44
Myosin Vb
Neurturin
Nuclear distribution protein nudE like 1
xytocin receptor
Pantothenate kinase 3
Platelet activating factor acetylhydrolase 2 cytoplasmic
Poly A polymerase gamma
PTE ankyrin domain family member I
Procollagen galactosyltransferase 1
Proteasome subunit beta type 8
Protein disulfide isomerase A4
Protein disulfide isomerase
Protein FAM171B
Protein lin 52 homolog
Protein NipSnap homolog 3B
Putative membrane spanning 4 domains subfamily A member 4E
Putative prolyl tRNA synthetase associated domain containing protein 1
Putative uncharacterized protein C13orf43

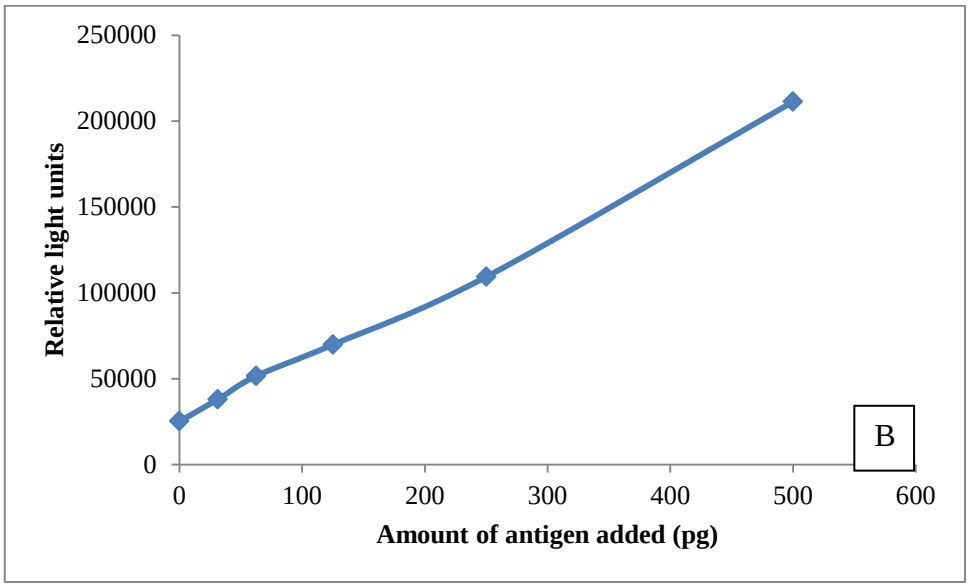
Rho related GTP binding protein RhoF
Selenoprotein H
Serine threonine protein kinase 17B
Serine threonine protein kinase 32A
Sialic acid binding Ig like lectin 16
Sperm protein associated with the nucleus on the X chromosome N1
SWI SNF related matrix associated actin dependent regulator of chromatin subfamily A containing DEAD
T cell receptor gamma 2 chain C region
TAF5 like RNA polymerase II p300 CBP associated factor associated factor 65 kDa subunit 5L
Testis specific gene 13 protein
Tigger transposable element derived protein 3
Transitional endoplasmic reticulum ATPase
Tyrosine protein kinase RYK
Ubiquitin conjugating enzyme E2 C
Ubiquitin conjugating enzyme E2 U
Uncharacterized protein C2orf15
Uncharacterized protein C3orf43
Uncharacterized protein KIAA0753
UPF0568 protein C14orf166
Zinc finger protein 541

Zinc finger protein 593
Zinc finger protein 827
Zinc finger protein with KRAB and SCAN domains 2

8.3 Appendix C

Figure C-1: The standard curve for SAP (A), IGF2 (B) and CYSD (C) usinag a sandwich ELISA method. These graphs show the relationship of the chemiluminescence of each antigen using different plasma protein concentrations.





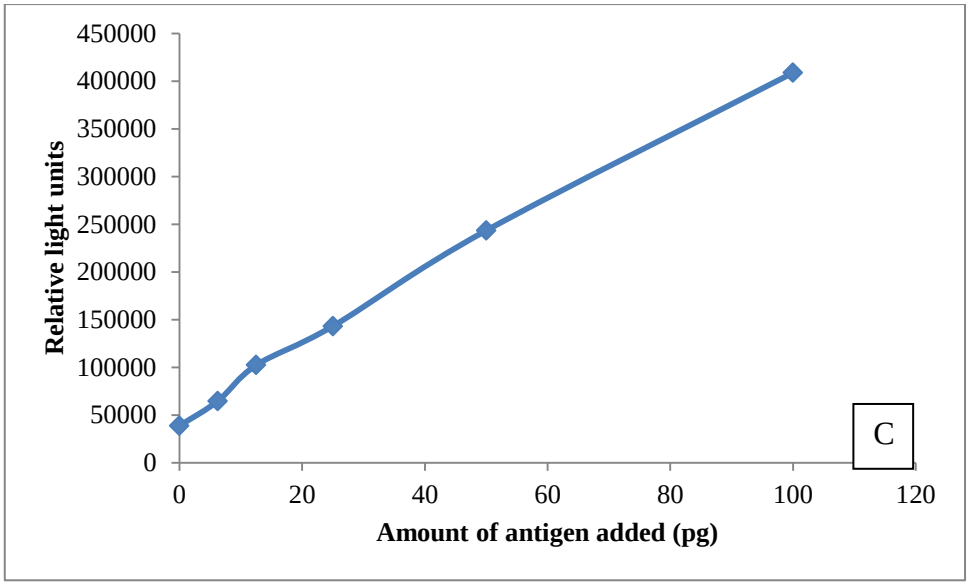
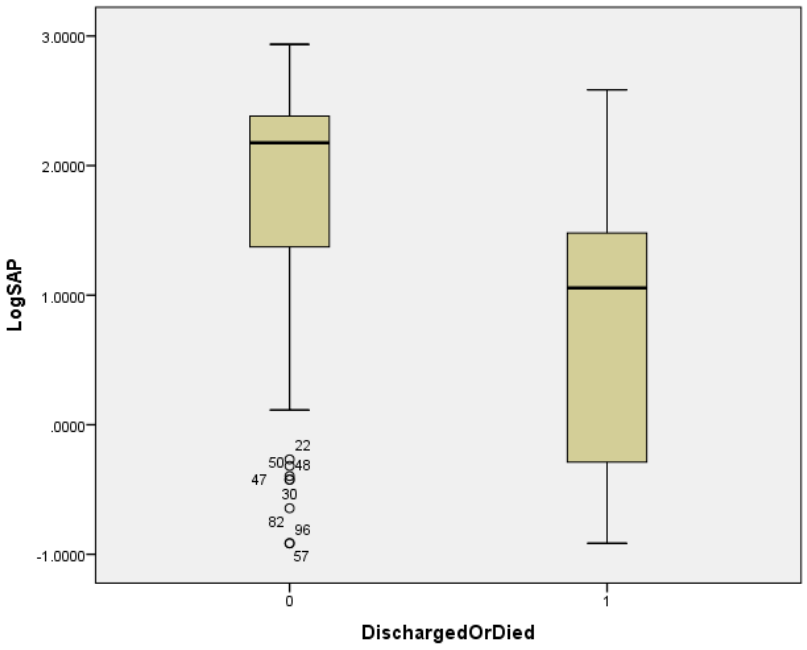

450000
400000
350000
300000
250000
200000
150000
100000
50000
0
Relative light units
0
20
40
60
80
100
120
Amount of antigen added (pg)
C

Figure C-2: Boxplot comparing serum amyloid P component (SAP) in acute heart failure patients. The Y axis shows the Log plasma SAP concentration (ng/ mL) and the X axis shows the patient groups (0 = survivors and 1 = non-survivors with end point of 1 year). This shows that SAP levels were higher in survivors than non-survivors.



8.4 Appendix D

Sample collection protocol for Chapters Four and Five

Biochemical Markers involved in the prediction of future cardiac events in Acute Heart Failure - Research Proposal to COREC for ethical approval

Principal Investigator:

Prof Leong Ng (FRCP. MD), Professor & Honorary Consultant in Medicine & Therapeutics, University of Leicester, Leicester

Co Investigators:

Dr Iain Squire (BSc. FRCP. MD), Senior Lecturer & Honorary Consultant Physician, University of Leicester, Leicester

Dr Derek Chin (FRCP. MD), Consultant Cardiologist, Glenfield Hospital, Leicester

Dr Turab Ali (MRCP), Clinical Research Fellow, University of Leicester, Leicester

Dr. Noor Mohammed (MRCP), Clinical Research Fellow, University of Leicester, Leicester

Dr. Hafid Narayan (MRCP), Clinical Research Fellow, University of Leicester, Leicester

Department Address:

Department of Cardiovascular Medicine

Pharmacology & Therapeutics Division

Level 4, Robert Kilpatrick Clinical Sciences Building

University of Leicester

Leicester Royal Infirmary

Infirmary Road

Leicester, LE2 7LX, UK

Ph: 0116 252 3125

Fax: 0116 252 3108

Background:

Estimates of the incidence of heart failure are difficult due to different methods of diagnosing this condition in populations. However, Hillingdon heart study estimates an incidence of around 140 per 100,000 for males and 120 per 100,000 in females. Other studies have estimated an overall prevalence of around 2% of the population, increasing steeply with increasing age. Again the management of heart failure has improved over recent years with introduction of

agents such as ACE-inhibitors, aldosterone antagonists and beta-blockers. However the degree of morbidity and rates of mortality due to this condition remain immense. Current treatments for heart failure extend the life of the patient but do not stop the progression of the disease process. These treatments may not be addressing the underlying cause of cellular injury.[1]

Arginine Vasopressin (AVP):

Vasoactive neurohormonal systems, e.g. sympathetic nervous system (SNS), renin-angiotensin-aldosterone system (RAAS), and arginine vasopressin (AVP), are normally stimulated under conditions of acute volume depletion, and are activated by the low cardiac output and arterial pressure. However, sustained and chronic activation of these systems, can cause progressive ventricular remodeling. Vasoconstriction, water retention, and increased blood volume then ensues, accelerating progression of congestive heart failure.[2] A posterior pituitary hormone, AVP may directly and adversely affect myocardial function due to the effects on myocardial contractility and cell growth. AVP contributes to hyponatremia, a powerful predictor of poor outcome in heart failure.[3] The hormone has been shown to be a marker for the presence and severity of chronic heart failure,[4] but little is known about the role of AVP in acute heart failure. Improvement in survival in Heart Failure has been demonstrated with the administration of AVP antagonists,[5] and they could be a useful addition to ACEIs in the management of Congestive Heart Failure.[6] AVP receptor antagonist's dual receptor activity may be particularly useful as these pts may benefit from the increased cardiac output, reduced total peripheral resistance, and reduced mean arterial BP that results from V1a receptor blockade as well as the reduced congestion, reduced cardiac preload, and increased Na concentration induced by V2 receptor antagonism.[7]

Adrenomedullin (ADM):

Adrenomedullin is a recently discovered endogenous peptide,[8] secreted from cardiac myocytes and fibroblasts.[9] ADM is elevated in CHF, suggesting a possible role in the pathophysiology of this disease.[8,10,11] ADM has hypotensive and natriuretic actions.[8] It may be an antifibrotic, antihypertrophic, and positive inotropic factor in the failing and hypertrophied heart.[10] In fact ADM is a powerful marker for LV mass,[9] and increase in plasma ADM levels in rats with Heart Failure correlated with ventricular weight.[12] Also plasma ADM levels were correlated with pulmonary artery pressure and pulmonary capillary wedge pressure.[13] Acute administration of ADM has been shown to improve the hemodynamics, renal function, and hormonal parameters in patients with heart failure.[10] Long-term administration of ADM induces pronounced and sustained cardiovascular and renal effects in experimental heart failure, including reductions in cardiac preload and afterload, as well as augmentation of cardiac output, sodium excretion, and

glomerular filtration.[14] Plasma ADM was correlated strongly with Endothelin-1, -2, ANP, and Norepinephrine, and relatively weakly with LV EF. Plasma ADM levels significantly decreased after treatment.[15] It has been established that plasma ADM levels are increased in patients with heart failure in proportion to the severity of the disease.[13,10,11,16,17,12,18] Yu et al found that plasma ADM concentrations in patients with heart failure are determined by the presence of diastolic dysfunction, and are especially raised in the presence of a restrictive filling pattern.[19] Furthermore, recent studies suggest that plasma ADM level is an independent prognostic indicator of cardiac injury[16,17] and heart failure.[10,16,17] In HF, elevated circulating ADM also identifies patients likely to receive long-term benefit from inclusion of additional anti-failure therapy (Carvedilol). Manipulation of the ADM system holds promise as a therapeutic strategy in cardiac disease.[16]

Endothelin-1 (ET-1):

Less than 20 years after its discovery, ET is not only one of the most potent known vasoconstrictors; it also has multiple other actions. It is produced by cardiac myocytes and vascular endothelial cells[20] and mediates pathologic hypertrophy and fibrosis of both ventricular and vascular tissues, and potentiates the effects of other neurohormones, and acts as a proarrhythmic.[21] ET-1 causes cell proliferation[22] and myocardial effects. Also ET-1 contributes to vascular dysfunction, myocardial ischemia and renal impairment in chronic heart failure.[23] Production of endothelin-1 is markedly increased in the myocardium of rats with heart failure.[20] Plasma levels of ET-1 are elevated in chronic heart failure, independent of aetiology,[24] and the magnitude of elevation correlates with disease severity and symptoms.[25,23,24,26,27] ET receptor antagonists indeed have a potential to improve hemodynamics, symptoms, and potentially prognosis in patients with CHF, which still carries a high mortality,[27] though in the recently completed ET-A Receptor Antagonist Trial in Heart Failure (EARTH) trial in patients with chronic heart failure showed lack of beneficial effects of long-term treatment with ET antagonists in patients.[24] Yamamoto et al showed chronic ETA receptor blockade's beneficial effects in diastolic failure through attenuation of the progression of LVH and fibrosis.[28] Plasma levels of ET-1, are strong independent predictors of death after myocardial infarction as well as in congestive heart failure.[24,25,26,27,29] ET antagonists markedly increased survival, reduced LV dilatation, reduced hypertrophy and increased contractility of the noninfarcted LV wall. Chronic treatment with ET antagonists bosentan might be beneficial in CHF and might increase long-term survival in this disease.[30,31]

Guanylin and Uroguanylin (GN and UGN):

These 2 hormones were discovered to be activators of guanylate cyclase in the kidney and could act as powerful natriuretic hormones[32], in addition to their known effects on inhibition of sodium reabsorption in the gut. Their role in heart failure is unknown at present, although preliminary evidence suggests that there are factors in urine that could stimulate guanylate cyclase, which may be related to uroguanylin[33].

Matrix proteinases (MMPs) and Tissue inhibitors of Matrix Proteinases (TIMPs)

Matrix metalloproteinases (MMPs) and their respective inhibitors (TIMPs) have been implicated in the pathogenesis of heart failure. MMPs may lead to degradation of extracellular matrix, leading to ventricular remodeling and TIMPs may also be involved in this process. Our recent studies suggest that MMP9 and TIMP1 may be of importance in cardiac remodeling after myocardial infarction[34,35].

Gene polymorphisms and vascular disease

In a study of MMP-9 plasma concentrations of 1127 patients with documented evidence of coronary artery disease, the T allele of the C-1562T polymorphism was associated with increased MMP-9 levels, there was also a significant association between the R279Q polymorphism and cardiovascular events in patients with stable angina over a mean follow-up period of 4.1 years[36]. The C-1562 polymorphism has also been shown to be associated with increased area of complicated coronary plaques in post mortem studies, implicating this polymorphism in the development of significant coronary disease[37]. In addition, variations in the promoter regions of MMP-9 gene have been related to the severity of atherosclerotic disease as detected in angiographic studies[38,39].

Promotor polymorphisms in the MMP-2 gene (-790T/G and –735C/T) have also been shown to be more frequent amongst patients with chronic heart failure (NYHA class II-IV) as compared to matched control subjects[40]. This provides further evidence as to the importance of genetic factors in ventricular remodelling

Mizon-Gerard et al [41] investigated polymorphisms of the MMP-2 gene (-1306C>T), MMP-3 (-1171 5A > 6A), and MMP-9 (-1562 C > T). They demonstrated that MMP-3 polymorphism had a different impact on survival in heart failure patients with

ischaemic and non ischaemic aetiology, the MMP-3 5A/5A genotype being a predictor of cardiac mortality in patients with non-ischaemic but not ischaemic heart failure.

Recent studies using genome wide scans have revealed loci which are associated with coronary artery disease[42], such as polymorphisms on chromosome 9p21.3, 6q25.1 (rs6922269) and chromosome 2q36.3 (rs2943634).

Hypothesis:

Biochemical Markers like Arginine Vasopressin, Adrenomedullin, Endothelin-1, Guanylin and Uroguanylin, N-terminal Brain Natriuretic Peptide levels, Matrix proteinases and their inhibitors TIMPs during Acute Heart Failure predicts future cardiac events. Gene polymorphisms of these and other proteins contribute to outcome following acute heart failure.

Objectives:

To establish:

1. The role of biochemical markers in diagnosis of 'acute' Heart Failure
2. The role of biochemical and genetic markers in prediction of future cardiac events in heart failure
3. The relation between new biomarkers and established markers of HF (like N-BNP)
4. The relation between biomarkers and severity of heart failure / symptoms
5. The relation between rise and fall in the plasma levels of biomarkers and worsening and improvement respectively of symptoms of heart failure
6. The relation between biomarkers and degree of LV systolic and diastolic dysfunction
7. The relation between the biochemical and genetic markers and the cause of heart failure

***Study Design & Methods:**

Patients with Acute Heart Failure will be recruited from the acute admission wards of Leicester Royal Infirmary and Glenfield Hospital, Leicester. It is a prospective observational cohort study.

Patients will be approached by a medically qualified investigator and formal informed consent will be obtained after giving information about the study. The participation in the study is totally voluntary. Once patient gives consent, blood and urine samples (20ml each) for biomarkers will be obtained and detailed information leaflets will be provided. Then the patient proforma will be filled (obtaining information regarding symptoms, brief examination and

drugs and results of the investigations). Patients with acute coronary syndrome will be excluded from the study. The urine samples do not need any processing before storage, while the blood samples will be centrifuged and the resulting plasma will be clearly labeled and stored at -70^0C until processed (for biochemical markers) at a later stage. The buffy coat from spinning down blood samples will be stored for extraction of DNA for the determination of gene polymorphisms. Plasma, DNA and urine samples will be stored for up to 15 years after the end of the study, with the patient's consent, for future analyses of new proteins (after seeking ethical approval) that may have been discovered which may be of use in predicting prognosis of these patients. During the admission patient's progress will be monitored and Echocardiography (ultrasound scan) will be carried out prior to discharge. Full examination will be done in addition to Left ventricular systolic / diastolic function by 2-Dimensional, M-Mode, Doppler and novel 3-Dimensional Echocardiography techniques. This will be performed by the MD candidate who is experienced and has passed the British Society of Echocardiography (BSE) accreditation examination for Transthoracic Echocardiography. Before discharge, again, blood and urine samples (again 20ml each) will be obtained to see the effect of treatment and improvement of symptoms on biochemical markers.

We will study the relationship between AVP, ADM, ET-1, GN, UGN, N-BNP, MMPs and TIMPs (in blood and urine) and also polymorphisms in genes, and clinical outcome following Acute Heart Failure admission. Other novel cardiovascular markers may also be measured. We will recruit 819 patients in order to be able to analyze event-free survival using the Cox hazard models using the above markers and clinical factors/variables. This figure is based on an estimate of the standardized difference in biochemical levels between those with events and those who survive event-free of 0.208 (with a event free vs event ratio of 2:1) with a power of 80% at $p<0.05$.

In a subset of 100 patients, daily (up to 5 days, including admission & discharge days) blood samples (10ml each for additional samples) will be obtained to determine the time course of the various biomarkers to the effects of treatment of Acute Heart Failure. This sample number was chosen as in previous studies on peptides, we have found that with the standard deviations of measurement, about 100 cases would enable differences between days to be detectable with a 80% power ($P<0.05$). We will ensure that this sample is obtained at the same time as routine bloods taken for clinical purposes to avoid unnecessary venesection and discomfort to the patients.

The estimated time for recruitment is approximately 48 months. Patients will be followed up (from their hospital notes record) after 6 months of recruitment for any hospitalization for Heart Failure, any procedure performed (e.g. revascularization), and for death. The GP surgery will be contacted only for those patients who have died in the community during the follow up time (to find out the cause of death), as normally there is no record of these patient's cause of death in the hospital notes (unless patients died in hospital).

To date no study has addressed the possible correlation in 'Acute Heart Failure' of blood AVP, ADM, ET-1, GN, UGN, MMPs and TIMPs levels with future cardiac events, also its relation to N-BNP (a marker of LV dysfunction and prognosis). Evidence of this correlation may provide evidence for a role of these biomarkers in the pathogenesis of Heart Failure and as a possible therapeutic target in this condition.

Our department has extensive research experience in the area of Heart Failure and Coronary Artery Disease. We are currently carrying out number of research projects in this area. Patients have, and continue to, demonstrate their willingness to participate in this type of project (involving 2 blood & urine samples and Echo scan).

References:

(1) Angerio AD. The role of endothelin in heart failure. Cri Care Nurs Q 2005 Oct-Dec;28(4):355-9.

(2) Chatterjee K. Neurohormonal activation in congestive heart failure and the role of vasopressin. Am J Cardiol 2005 May;95(9A):8B-13B.

(3) Goldsmith SR, Gheorgiade M. Vasopressin antagonism in heart failure. J Am Coll Cardiol 2005 Nov;46(10):1785-91.

(4) Kim JK, Michel JB, Soubrier F, Durr J, Corvol P, Schrier RW. Arginine vasopressin gene expression in chronic heart failure in rats. Kidney Int 1990 Nov;38(5):818-22.

(5) Schrier RW, Martin PY. Recent advances in understanding of water metabolism in heart failure. Adv Exp Med Biol 1998;449:415-26.

(6) Naitoh M, Risvanis J, Balding LC, Johnston CI, Burrell LM. Neurohormonal antagonism in heart failure; beneficial effects of vasopressin V(1a) and V(2) receptor blockade and ACE inhibition. Cardiovasc Res 2002 Apr;54(1):51-5.

(7) Goldsmith SR. Current treatment and novel pharmacologic treatments for hyponatremia in congestive heart failure. Am J Cardiol 2005 May;95(9A):14B-23B.

(8) Rademaker MT, Charles CJ, Lewis LK, Yandle TG, Cooper GJ, Coy DH, Richards AM, Nicholls MG. Beneficial hemodynamic and renal effects of adrenomedullin in an ovine model of heart failure. Circulation 1997 Sep;96(6):1983-90.

(9) Jougasaki M, Stevens TL, Borgeson DD, Luchner A, Redfield MM, Burnett JC Jr. Adrenomedullin in experimental congestive heart failure: cardiorenal activation. Am J Physiol 1997 Oct;273(4 Pt 2):R1392-9.

(10) Nishikimi T, Yoshihara F, Mori Y, Kangawa K, Matsuoka H. Cardioprotective effect of adrenomedullin in heart failure. Hypertens Res 2003 Feb;26 Suppl:S121-7.

(11) Nicholls MG, Charles CJ, Lainchbury JG, Lewis LK, Rademaker MT, Richards AM, Yandle TG. Adrenomedullin in heart failure. Hypertens Res 2003 Feb;26 Suppl:S135-40.

(12) Nishikimi T, Horio T, Sasaki T, Yoshihara F, Takishita S, Miyata A, Matsuo H, Kangawa K. Cardiac production and secretion of adrenomedullin are increased in heart failure. Hypertension 1997 Dec;30(6):1369-75.

(13) Kobayashi K, Kitamura K, Etoh T, Nagatomo Y, Takenaga M, Ishikawa T, Imamura T, Koiwaya Y, Eto T. Increased plasma adrenomedullin levels in chronic congestive heart failure. Am Heart J 1996 May;131(5):994-8.

(14) Rademaker MT, Charles CJ, Espiner EA, Nicholls MG, Richards AM. Long-term adrenomedullin administration in experimental heart failure. Hypertension 2002 Nov;40(5):667-72.

(15) Halawa B, Mazurek W. Levels of adrenomedullin in plasma of patients with chronic congestive heart failure. Pol Arch Med Wewn 1998 Jan;99(1):2-8.

(16) Rademaker MT, Cameron VA, Charles CJ, Lainchbury JG, Nicholls MG, Richards AM. Adrenomedullin and heart failure. Regul Pept 2003 Apr;112(1-3):51-60.

(17) Lainchbury JG. Novel neurohumoral factors in congestive heart failure: adrenomedullin. Curr Cardiol Rep 2001 May;3(3):208-14.

(18) Nishikimi T, Saito Y, Kitamura K, Ishimitsu T, Eto T, Kangawa K, Matsuo H, Omae T, Matsuoka H. Increased plasma levels of adrenomedullin in patients with heart failure. J Am Coll Cardiol 1995 Nov;26(6):1424-31.

(19) Yu CM, Cheung BM, Leung R, Wang Q, Lai WH, Lau CP. Increase in plasma adrenomedullin in patients with heart failure characterized by diastolic dysfunction. Heart 2001 Aug;86(2):155-60.

(20) Sakai S, Miyauchi T, Kobayashi M, Yamaguchi I, Goto K, Sugishita Y. Inhibition of myocardial endothelin pathway improves long-term survival in heart failure. Nature 1996 Nov;384(6607):353-5.

(21) Teerlink JR. Endothelins: pathophysiology and treatment implications in chronic heat failure. Curr Heart Fail Rep 2005 Dec;2(4):191-7.

(22) Spieker LE, Noll G, Luscher TF. Therapeutic potential for endothelin receptor antagonists in cardiovascular disorders. Am J Cardiovasc Drugs 2001;1(4):293-303.

(23) Spieker LE, Noll G, Ruschitzka FT, Luscher TF. Endothelin receptor antagonists in congestive heart failure: a new therapeutic principle for the future? J Am Coll 2001 May;37(6):1493-505.

(24) Ertl G, Bauersachs J. Endothelin receptor antagonists in heart failure: current status and future directions. Drugs 2004;64(10):1029-40.

(25) Krum H, Denver R, Tzanidis A, Martin P. Diagnostic and therapeutic potential of the endothelin system in patients with chronic heart failure. Heart Fail Rev 2001 Dec;6(4):341-52.

(26) Spieker LE, Noll G, Ruschitzka FT, Luscher TF. Endothelin A receptor antagonists in congestive heart failure: blocking the beast while leaving the beauty untouched? Heart Fail Rev 2001 Dec;6(4):301-15.

(27) Hurlimann D, Enseleit F, Noll G, Luscher TF, Ruschitzka F. Endothelin antagonists and heart failure. Curr Hypertens Rep 2002 Feb;4(1):85-92.

(28) Yamamoto K, Masuyama T, Sakata Y, Nishikawa N, Mano T, Hori M. Prevention of diastolic heart failure by endothelin A receptor antagonist through inhibition of ventricular structural remodeling in hypertensive hearts. J Hypertens 2002 Apr;20(4):753-61.

(29) Duchman SM, Thohan V, Kalra D, Torre-Amione G. Endothelin-1: a new target of therapeutic intervention for the treatment of heart failure. Curr Opin Cardiol 2000 May;15(3):136-40.

(30) Mulder P, Richard V, Derumeaux G, Hogie M, Henry JP, Lallemand F, Compagnon P, Mace B, Comoy E, Letac B, Thuillez C. Role of endogenous endothelin in chronic heart failure: effect of long-term treatment with an endothelin antagonist on survival, hemodynamics, and cardiac remodeling. Circulation 1997 Sep;96(6):1976-82.

(31) Love MP, McMurray JJ. Endothelin in heart failure: a promising therapeutic target? Heart 1997 Feb;77(2):93-4.

(32) Sindic A, Schlatter E. Cellular effects of guanylin and uroguanylin. J Am Soc Nephrol. 2006;17(3):607-16.

(33) Carrithers SL, Eber SL, Forte LR, Greenberg RN. Increased urinary excretion of uroguanylin in patients with congestive heart failure. Am J Physiol Heart Circ Physiol. 2000 ;278(2):H538-47.

(34) Kelly D, Cockerill G, Ng LL, Thompson M, Khan S, Samani NJ, Squire IB. Plasma matrix metalloproteinase-9 and left ventricular remodelling after acute myocardial infarction in man: a prospective cohort study. Eur Heart J. 2007 ; 28 : 711-8.

(35) Kelly D, Khan SQ, Thompson M, Cockerill G, Ng LL, Samani N, Squire IB. Plasma tissue inhibitor of metalloproteinase-1 and matrix metalloproteinase-9: novel indicators of left

ventricular remodelling and prognosis after acute myocardial infarction. Eur Heart J. 2008 Jul 8. [Epub ahead of print]
(36) Blankenberg S, Rupprecht HJ, Poirier O, Bickel C, et al. Plasma concentrations and genetic variations of Matrix Metalloproteinase 9 and prognosis of patients with cardiovascular disease. Circulation 2003; 107:1579-1585.
(37) Perttu J et al. Coronary artery complicated lesion area is related to functional polymorphism of the matrix metalloproteinase 9 gene. An autopsy study.Arterioscler Thromb Vasc Biol. 2001;21:1446-1450
(38) Zhang B, Ye S, Herrmann SM, Eriksson P, et al. Functional Polymorphisms in the regulatory region of Gelatinase B gene in relation to severity of coronary atherosclerosis. Circulation. 1999;99:1788-1794.
(39) Morgan AR et al. Haplotypic analysis of the MMP-9 gene in relation to coronary artery disease. J Mol Med. 2003;81(5):321
(40) Vasku A, Goldbergova M, Holla LI, Spinarova L, Spinar J, Vitovec J, Vasha J. Two MMP-2 promotor polymorphisms (-790T/G and –735C/T) in chronic heart failure. Clin Chem Lab Med. 2003;41(10):1299-303
(41) Mizon-Gerard F, de Groote P, Lamblin N, Hermant X, Dallongeville, Amouyel P, Bauters C, helbecque N. Prognostic impact of Matrix metalloproteinase gene polymorphisms in patients with heart failure according to the aetiology of left ventricular dysfunction. Eur Heart J. 2004;25(8):688-93
(42) Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. N Engl J Med. 2007 ; 357 : 443-53.

8.5 Appendix E

List of publications and presentations at scientific meetings

Publications:

Auluck J*, **Hakimi A***, Jones GDD, Ng LL, Jones DJL (**2013**) Assessment of reproducibility in depletion and enrichment workflows from plasma proteomics using label-free quantitative data-independent LC-MS. Manuscript accepted by Proteomics (see attached paper)

Presentations:

Auluck J, NG LL, Jones DJL (**2013**). Searching for the elusive marker of Heart Failure. Poster presentation at the Festival of Postgraduate Research, University of Leicester, UK

Auluck J, NG LL, Jones DJL (**2012**). Proteomic biomarkers for predicting Acute Heart Failure prognosis. Oral presentation at the cardiovascular seminar, University of Leicester, UK

Auluck J, NG LL, Jones DJL (**2012**). $HDMS^E$ for the discovery of prognostic biomarkers. Poster presentation at the 11^{th} East midlands proteomics workshop, Loughborough University, UK

Auluck J, NG LL, Jones DJL (**2012**). Application of $HDMS^E$ for the discovery of prognostic biomarkers in Acute Heart Failure Poster presentation at the 11^{th} annual world congress of HUPO, Hynes convention centre, Boston, MA, USA

Auluck J, NG LL, Jones DJL (**2012**). Apllication of $HDMS^E$ in Heart Failure proteomics. Oral presentation at the 9^{th} Proteomics Methods Forum Meeting, Barts Cancer Institute, London, UK

Auluck J, NG LL, Jones DJL (**2011**). Proteomic biomarkers of Heart Failure. Poster presentation at the 10th East midlands proteomics workshop, Loughborough University, UK

Auluck J, NG LL, Jones DJL (**2011**). Impact of LC-MS^{E}on biomarker discovery in Heart Failure. Poster presentation at the 32nd BMSS annual meeting, Cardiff, UK

Auluck J, NG L.L, Jones D.J.L (**2010**). Searching the human proteome for diagnostic biomarkers of heart failure using a label-free LC-MS^E approach. Poster presentation at the 31st BMSS annual meeting, Cardiff, UK

References

Method of the Year 2012. 2013. **10,** 1-1.

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

Alssema, M., Dekker, J.M., Kuivenhoven, J.A., Nijpels, G., Teerlink, T., Scheffer, P.G., Diamant, M., Stehouwer, C.D., Bouter, L.M., Heine, R.J., 2007. Elevated cholesteryl ester transfer protein concentration is associated with an increased risk for cardiovascular disease in women, but not in men, with Type 2 diabetes: the Hoorn Study. *Diabetic Medicine : A Journal of the British Diabetic Association.* **24,** 117-123.

Alvarez-Diaz, S., Valle, N., Garcia, J.M., Pena, C., Freije, J.M., Quesada, V., Astudillo, A., Bonilla, F., Lopez-Otin, C., Munoz, A., 2009. Cystatin D is a candidate tumor suppressor gene induced by vitamin D in human colon cancer cells. *The Journal of Clinical Investigation.* **119,** 2343-2358.

Anderson, N.L., 2010. The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum. *Clinical Chemistry.* **56,** 177-185.

Anderson, N.L. & Anderson, N.G., 2002. The human plasma proteome - History, character, and diagnostic prospects. *Molecular & Cellular Proteomics.* **1,** 845-867.

Antman, E.M., Tanasijevic, M.J., Thompson, B., Schactman, M., McCabe, C.H., Cannon, C.P., Fischer, G.A., Fung, A.Y., Thompson, C., Wybenga, D., Braunwald, E., 1996. Cardiac-specific troponin I levels to predict the risk of mortality in patients with acute coronary syndromes. *The New England Journal of Medicine.* **335,** 1342-1349.

Araujo, J.P., Lourenco, P., Azevedo, A., Frioes, F., Rocha-Goncalves, F., Ferreira, A., Bettencourt, P., 2009. Prognostic value of high-sensitivity C-reactive protein in heart failure: a systematic review. *Journal of Cardiac Failure.* **15,** 256-266.

Askoxylakis, V., Thieke, C., Pleger, S., Most, P., Tanner, J., Lindel, K., Katus, H., Debus, J., Bischof, M., 2010. Long-term survival of cancer patients compared to heart failure and stroke: A systematic review. *BMC Cancer.* **10,** 105.

Aston, F.W., 1919. LXXIV. A positive ray spectrograph. *Philosophical Magazine Series 6.* **38,** 707-714.

Authors/Task Force Members, McMurray, J.J.V., Adamopoulos, S., Anker, S.D., Auricchio, A., Böhm, M., Dickstein, K., Falk, V., Filippatos, G., Fonseca, C., Gomez-Sanchez, M.A., Jaarsma, T., Køber, L., Lip, G.Y.H., Maggioni, A.P., Parkhomenko, A., Pieske, B.M., Popescu, B.A., Rønnevik, P.K., Rutten, F.H., Schwitter, J., Seferovic, P., Stepinska, J., Trindade, P.T., Voors, A.A., Zannad, F., Zeiher, A., ESC Committee for

Practice Guidelines (CPG), Document Reviewers, 2012. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *European Journal of Heart Failure.* **14,** 803-869.

Barber, M., Bordoli, R.S., Sedgwick, R.D., Tyler, A.N., 1981. Fast atom bombardment of solids as an ion source in mass spectrometry. *Nature.* **293,** 270-275.

Barter, P.J. & Rye, K.A., 2012. Cholesteryl ester transfer protein inhibition as a strategy to reduce cardiovascular risk. *Journal of Lipid Research.* **53,** 1755-1766.

Beaglehole, R. & Bonita, R., 2008. Global public health: a scorecard. *The Lancet.* **372,** 1988-1996.

Bell, A.W., Deutsch, E.W., Au, C.E., Kearney, R.E., Beavis, R., Sechi, S., Nilsson, T., Bergeron, J.J., HUPO Test Sample Working Group, 2009. A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. *Nature Methods.* **6,** 423-430.

Berg, J., Tymoczko, J. and Stryer, L., 2002. Biochemistry. Section 4.3, Immunology Provides Important Techniques with Which to Investigate Proteins. 5th ed. New York: W H Freeman. Section 4.3.

Berhane, B.T., Zong, C., Liem, D.A., Huang, A., Le, S., Edmondson, R.D., Jones, R.C., Qiao, X., Whitelegge, J.P., Ping, P., Vondriska, T.M., 2005. Cardiovascular-related proteins identified in human plasma by the HUPO Plasma Proteome Project pilot phase. *Proteomics.* **5,** 3520-3530.

Berliner, D., Angermann, C.E., Ertl, G., Stoerk, S., 2009. Biomarkers in Heart Failure - Better than History or Echocardiography? *Herz.* **34,** 581-588.

Berrueta, L.A., Gallo, B., Vicente, F., 1995. A Review of Solid-Phase Extraction - Basic Principles and New Developments. *Chromatographia.* **40,** 474-483.

Biorad, . *Proeominer^{TM} Protein enrighment technology: Digging deeper in the proteome.* [online]. Available at: http://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_5635B.pdf [accessed 11/21 2013].

Blanco-Colio, L.M., López, J.A., Martinez-Pinna Albar, R., Egido, J., Martín-Ventura, J.L., 2009. Vascular proteomics, a translational approach: from traditional to novel proteomic techniques. *Expert Review of Proteomics.* **6,** 461-464.

Boekholdt, S.M., Kuivenhoven, J.A., Wareham, N.J., Peters, R.J., Jukema, J.W., Luben, R., Bingham, S.A., Day, N.E., Kastelein, J.J., Khaw, K.T., 2004. Plasma levels of cholesteryl ester transfer protein and the risk of future coronary artery disease in apparently healthy men and women: the prospective EPIC (European Prospective Investigation into Cancer and nutrition)-Norfolk population study. *Circulation.* **110,** 1418-1423.

Boja, E., Hiltke, T., Rivers, R., Kinsinger, C., Rahbar, A., Mesri, M., Rodriguez, H., 2011. Evolution of clinical proteomics and its role in medicine. *Journal of Proteome Research.* **10,** 66-84.

Boschetti, E., Lomas, L., Citterio, A., Righetti, P.G., 2007. Romancing the "hidden proteome", Anno Domini two zero zero seven. *Journal of Chromatography A.* **1153,** 277-290.

Boschetti, E. & Righetti, P.G., 2009. The art of observing rare protein species in proteomes with peptide ligand libraries. *Proteomics.* **9,** 1492-1510.

Braunwald, E., 2008. Biomarkers in Heart Failure. *The New England Journal of Medicine.* **358,** 2148-2159.

Brea, D., Sobrino, T., Blanco, M., Fraga, M., Agulla, J., Rodriguez-Yanez, M., Rodriguez-Gonzalez, R., Perez de la Ossa, N., Leira, R., Forteza, J., Davalos, A., Castillo, J., 2009. Usefulness of haptoglobin and serum amyloid A proteins as biomarkers for atherothrombotic ischemic stroke diagnosis confirmation. *Atherosclerosis.* **205,** 561-567.

Cairns, D.A., 2011. Statistical issues in quality control of proteomic analyses: good experimental design and planning. *Proteomics.* **11,** 1037-1048.

Camm AJ, ed, 2009. *The ESC Textbook of Cardiovascular Medicine.* 2nd ed. Oxford, UK: Oxford University Press.

Capriotti, A.L., Caruso, G., Cavaliere, C., Piovesana, S., Samperi, R., Lagana, A., 2012. Comparison of three different enrichment strategies for serum low molecular weight protein identification using shotgun proteomics approach. *Analytica Chimica Acta.* **740,** 58-65.

Carrasco-Sanchez, F.J., Galisteo-Almeda, L., Paez-Rubio, I., Martinez-Marcos, F.J., Camacho-Vazquez, C., Ruiz-Frutos, C., Pujol-De La Llave, E., 2011. Prognostic value of cystatin C on admission in heart failure with preserved ejection fraction. *Journal of Cardiac Failure.* **17,** 31-38.

Cas, L.D., Metra, M., Nodari, S., Cas, A.D., Gheorghiade, M., 2003. Prevention and management of chronic heart failure in patients at risk. *American Journal of Cardiology.* **91,** 10F-17F.

Casas, J.P., Shah, T., Hingorani, A.D., Danesh, J., Pepys, M.B., 2008. C-reactive protein and coronary heart disease: a critical review. *Journal of Internal Medicine.* **264,** 295-314.

Chelius, D. & Bondarenko, P.V., 2002. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *Journal of Proteome Research.* **1,** 317-323.

Chen, C.B. & Wallis, R., 2004. Two mechanisms for mannose-binding protein modulation of the activity of its associated serine proteases. *The Journal of Biological Chemistry.* **279,** 26058-26065.

Chen, W., Tran, K.D., Maisel, A.S., 2010. Biomarkers in heart failure. *Heart.* **96,** 314-320.

Cheung, M.C., Brown, B.G., Marino Larsen, E.K., Frutkin, A.D., O'Brien, K.D., Albers, J.J., 2006. Phospholipid transfer protein activity is associated with inflammatory markers in patients with cardiovascular disease. *Biochimica Et Biophysica Acta.* **1762,** 131-137.

Cholesterol Treatment Trialists' (CTT) Collaboration, Baigent, C., Blackwell, L., Emberson, J., Holland, L.E., Reith, C., Bhala, N., Peto, R., Barnes, E.H., Keech, A., Simes, J., Collins, R., 2010. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. *Lancet.* **376,** 1670-1681.

Choudhary, R., Iqbal, N., Khusro, F., Higginbotham, E., Green, E., Maisel, A., 2013. Heart Failure Biomarkers. *Journal of Cardiovascular Translational Research.* **6,** 471-484.

Comisarow, M.B. & Marshall, A.G., 1974. Fourier transform ion cyclotron resonance spectroscopy. *Chemical Physics Letters.* **25,** 282-283.

Craddock, R.M., Huang, J.T., Jackson, E., Harris, N., Torrey, E.F., Herberth, M., Bahn, S., 2008. Increased alpha-defensins as a blood marker for schizophrenia susceptibility. *Molecular & Cellular Proteomics : MCP.* **7,** 1204-1213.

Crawford, J.R., Pilling, D., Gomer, R.H., 2012. FcgammaRI mediates serum amyloid P inhibition of fibrocyte differentiation. *Journal of Leukocyte Biology.* **92,** 699-711.

Creaser, C.S., Griffiths, J.R., Bramwell, C.J., Noreen, S., Hill, C.A., Thomas, C.L.P., 2004. Ion mobility spectrometry: a review. Part 1. Structural analysis by mobility measurement. *Analyst.* **129,** 984-994.

Cubedo, J., PadroÌ • , T., GarciÌ • a-Moll, X., PintoÌ • , X., Cinca, J., Badimon, L., 2011. Proteomic Signature of Apolipoprotein J in the Early Phase of New-Onset Myocardial Infarction. *Journal of Proteome Research.* **10,** 211-220.

DardeÌ • , V.M., de, l.C., Gil Dones, F., Alvarez-Llamas, G., Barderas, M.G., Vivanco, F., 2010. Analysis of the Plasma Proteome Associated with Acute Coronary Syndrome: Does a Permanent Protein Signature Exist in the Plasma of ACS Patients? *Journal of Proteome Research.* **9,** 4420-4432.

Davidsson, P., Hulthe, J., Fagerberg, B., Camejo, G., 2010. Proteomics of Apolipoproteins and Associated Proteins From Plasma High-Density Lipoproteins. *Arteriosclerosis, Thrombosis, and Vascular Biology.* **30,** 156-163.

De Beer, F., Soutar, A.K., Baltz, M.L., Trayner, I., Feinstein, A., Pepys, M., 1982. Low density lipoprotein and very low density lipoprotein are selectively bound by aggregated C-reactive protein. *The Journal of Experimental Medicine.* **156,** 230-242.

De Souza, A.I., Cardin, S., Wait, R., Chung, Y., Vijayakumar, M., Maguy, A., Camm, A.J., Nattel, S., 2010. Proteomic and metabolomic analysis of atrial profibrillatory remodelling in congestive heart failure. *Journal of Molecular and Cellular Cardiology.* **49,** 851-863.

Deursen, V., Damman, K., Meer, P., Wijkstra, P., Luijckx, G., Beek, A., Veldhuisen, D., Voors, A., 2012. Co-morbidities in heart failure. *Heart Failure Reviews.* 1-10.

Deutsch, E.W., Lam, H., Aebersold, R., 2008. Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics. *Physiological Genomics.* **33,** 18-25.

Di Palma, S., Hennrich, M.L., Heck, A.J., Mohammed, S., 2012. Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *Journal of Proteomics.* **75,** 3791-3813.

Dickstein, K., Cohen-Solal, A., Filippatos, G., McMurray, J.J.V., Ponikowski, P., Poote-Wilson, P.A., Stromberg, A., van Veldhuisen, D.J., Atar, D., Hoes, A.W., Keren, A., Mebazaa, A., Nieminen, M., Priori, S.G., Swedberg, K., Heart Failure Assoc Esc HFA, ESICM, 2009. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2008 (vol 29, pg 2388, 2008). *European Journal of Heart Failure.* **11,** 110-110.

Ding, Y., Tong, M., Liu, S., Moscow, J.A., Tai, H., 2005. NAD -linked 15-hydroxyprostaglandin dehydrogenase (15-PGDH) behaves as a tumor suppressor in lung cancer. *Carcinogenesis.* **26,** 65-72.

Distelmaier, K., Adlbrecht, C., Jakowitsch, J., Winkler, S., Dunkler, D., Gerner, C., Wagner, O., Lang, I.M., Kubicek, M., 2009. Local complement activation triggers neutrophil recruitment to the site of thrombus formation in acute myocardial infarction. *Thrombosis and Haemostasis.* **102,** 564-572.

Diz, A.P., Truebano, M., Skibinski, D.O., 2009. The consequences of sample pooling in proteomics: an empirical study. *Electrophoresis.* **30,** 2967-2975.

Domon, B. & Aebersold, R., 2006. Challenges and opportunities in proteomics data analysis. *Molecular & Cellular Proteomics : MCP.* **5,** 1921-1926.

Domon, B. & Aebersold, R., 2010. Options and considerations when selecting a quantitative proteomics strategy. **28,** 710-721.

Dwivedi, R.C., Krokhin, O.V., Cortens, J.P., Wilkins, J.A., 2010. Assessment of the Reproducibility of Random Hexapeptide Peptide Library-Based Protein Normalization. *Journal of Proteome Research.* **9,** 1144-1149.

Elliott, M.H., Smith, D.S., Parker, C.E., Borchers, C., 2009. Current trends in quantitative proteomics. *Journal of Mass Spectrometry : JMS.* **44,** 1637-1660.

Elster, S.K., Braunwald, E., Wood, H.F., 1956. A study of C-reactive protein in the serum of patients with congestive heart failure. *American Heart Journal.* **51,** 533-541.

Emdin, M., Vittorini, S., Passino, C., Clerico, A., 2009. Old and new biomarkers of heart failure. *European Journal of Heart Failure.* **11,** 331-335.

Emerging Risk Factors Collaboration, Di Angelantonio, E., Sarwar, N., Perry, P., Kaptoge, S., Ray, K.K., Thompson, A., Wood, A.M., Lewington, S., Sattar, N., Packard, C.J., Collins, R., Thompson, S.G., Danesh, J., 2009. Major lipids, apolipoproteins, and risk of vascular disease. *JAMA : The Journal of the American Medical Association.* **302,** 1993-2000.

Emsley, J., White, H.E., O'Hara, B.P., Oliva, G., Srinivasan, N., Tickle, I.J., Blundell, T.L., Pepys, M.B., Wood, S.P., 1994. Structure of pentameric human serum amyloid P component. *Nature.* **367,** 338-345.

Engvall, E. & Perlmann, P., 1971. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochemistry.* **8,** 871-874.

Ensor, C.M. & Tai, H.H., 1995. 15-Hydroxyprostaglandin dehydrogenase. *Journal of Lipid Mediators and Cell Signalling.* **12,** 313-319.

Ewald, B., Ewald, D., Thakkinstian, A., Attia, J., 2008. Meta-analysis of B type natriuretic peptide and N-terminal pro B natriuretic peptide in the diagnosis of clinical heart failure and population screening for left ventricular systolic dysfunction. *Internal Medicine Journal.* **38,** 101-113.

Fang, X. & Zhang, W.W., 2008. Affinity separation and enrichment methods in proteomic analysis. *Journal of Proteomics.* **71,** 284-303.

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

Florens, L., Carozza, M.J., Swanson, S.K., Fournier, M., Coleman, M.K., Workman, J.L., Washburn, M.P., 2006. Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods (San Diego, Calif.).* **40,** 303-311.

Fonseca, C., 2006. Diagnosis of heart failure in primary care. *Heart Failure Reviews.* **11,** 95-107.

Frauenknecht, V., Thiel, S., Storm, L., Meier, N., Arnold, M., Schmid, J.P., Saner, H., Schroeder, V., 2013. Plasma levels of mannan-binding lectin (MBL)-associated serine proteases (MASPs) and MBL-associated protein in cardio- and cerebrovascular diseases. *Clinical and Experimental Immunology.* **173,** 112-120.

Fung, E.T., Wilson, A.M., Zhang, F., Harris, N., Edwards, K.A., Olin, J.W., Cooke, J.P., 2008. A biomarker panel for peripheral arterial disease. *Vascular Medicine (London, England).* **13,** 217-224.

Fuster, V. & Kelly, B.B., 2010. *Promoting cardiovascular health in the developing world: a critical challenge to achieve global health.* National Academies Press.

Gaggin, H.K. & Januzzi Jr., J.L., Biomarkers and diagnostics in heart failure. *Biochimica Et Biophysica Acta (BBA) - Molecular Basis of Disease.*

Gardner, R.S., Ozalp, F., Murday, A.J., Robb, S.D., McDonagh, T.A., 2003. N-terminal pro-brain natriuretic peptide. A new gold standard in predicting mortality in patients with advanced heart failure. *European Heart Journal.* **24,** 1735-1743.

Gerszten, R.E. & Wang, T.J., 2008. The search for new cardiovascular biomarkers. *Nature.* **451,** 949-952.

Gilar, M., Daly, A.E., Kele, M., Neue, U.D., Gebler, J.C., 2004. Implications of column peak capacity on the separation of complex peptide mixtures in single- and two-dimensional high-performance liquid chromatography. *Journal of Chromatography.A.* **1061,** 183-192.

Gilar, M., Fridrich, J., Schure, M.R., Jaworski, A., 2012. Comparison of orthogonality estimation methods for the two-dimensional separations of peptides. *Analytical Chemistry.* **84,** 8722-8732.

Gilar, M., Olivova, P., Chakraborty, A.B., Jaworski, A., Geromanos, S.J., Gebler, J.C., 2009. Comparison of 1-D and 2-D LC MS/MS methods for proteomic analysis of human serum. *Electrophoresis.* **30,** 1157-1167.

Gilar, M., Olivova, P., Daly, A.E., Gebler, J.C., 2005a. Orthogonality of separation in two-dimensional liquid chromatography. *Analytical Chemistry.* **77,** 6426-6434.

Gilar, M., Olivova, P., Daly, A.E., Gebler, J.C., 2005b. Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *Journal of Separation Science.* **28,** 1694-1703.

Giles, K., Pringle, S.D., Worthington, K.R., Little, D., Wildgoose, J.L., Bateman, R.H., 2004. Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Communications in Mass Spectrometry : RCM.* **18,** 2401-2414.

Giles, K., Williams, J.P., Campuzano, I., 2011. Enhancements in travelling wave ion mobility resolution. *Rapid Communications in Mass Spectrometry : RCM.* **25,** 1559-1566.

Gordon, S.M., Deng, J., Lu, L.J., Davidson, W.S., 2010. Proteomic Characterization of Human Plasma High Density Lipoprotein Fractionated by Gel Filtration Chromatography. *Journal of Proteome Research.* **9,** 5239-5249.

Guerrier, L., Thulasiraman, V., Castagna, A., Fortis, F., Lin, S.H., Lomas, L., Righetti, P.G., Boschetti, E., 2006. Reducing protein concentration range of biological samples using solid-phase ligand libraries. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences.* **833,** 33-40.

Guilhaus, M., Selby, D., Mlynski, V., 2000. Orthogonal acceleration time-of-flight mass spectrometry. *Mass Spectrometry Reviews.* **19,** 65-107.

Guilhaus, M., 2000. Essential elements of time-of-flight mass spectrometry in combination with the inductively coupled plasma ion source. *Spectrochimica Acta Part B: Atomic Spectroscopy.* **55,** 1511-1525.

Hakimi, A., Auluck, J., Jones, G.D., Ng, L.L., Jones, D.J., 2013. Assessment of reproducibility in depletion and enrichment workflows for plasma proteomics using label-free quantitative data-independent LC-MS. *Proteomics.*

Harris, K.R. & Thompson, M.K., 2012. Heart Failure. *InnovAiT: The RCGP Journal for Associates in Training.* **5,** 687-695.

Harvey, S.R., Macphee, C.E., Barran, P.E., 2011. Ion mobility mass spectrometry for peptide analysis. *Methods (San Diego, Calif.).* **54,** 454-461.

Haudek, S.B., Xia, Y., Huebener, P., Lee, J.M., Carlson, S., Crawford, J.R., Pilling, D., Gomer, R.H., Trial, J., Frangogiannis, N.G., 2006. Bone marrow-derived fibroblast precursors mediate ischemic cardiomyopathy in mice. *Proceedings of the National Academy of Sciences.* **103,** 18284-18289.

Heinecke, J.W., 2009. The HDL proteome: a marker–and perhaps mediator–of coronary artery disease. *Journal of Lipid Research.* **50,** S167-S171.

HFM, 2007. *Herat failure matters website.* [online]. Available at: http://heartfailurematters.org [accessed 04/19 2010].

Hoffmann, E.D., Stroobant, V., 2006. *Mass spectrometry priciples and applications.* Third ed. Great Britain: Wiley.

Hortin, G.L. & Sviridov, D., 2010. The dynamic range problem in the analysis of the plasma proteome. *Journal of Proteomics.* **73,** 629-636.

Hortin, G.L., Sviridov, D., Anderson, N.L., 2008. High-abundance polypeptides of the human plasma proteome comprising the top 4 logs of polypeptide abundance. *Clinical Chemistry.* **54,** 1608-1616.

Hu, S., Loo, J.A., Wong, D.T., 2006. Human body fluid proteome analysis. *Proteomics.* **6,** 6326-6353.

Huuskonen, J., Ekstrom, M., Tahvanainen, E., Vainio, A., Metso, J., Pussinen, P., Ehnholm, C., Olkkonen, V.M., Jauhiainen, M., 2000. Quantification of human plasma phospholipid transfer protein (PLTP): relationship between PLTP mass and phospholipid transfer activity. *Atherosclerosis.* **151,** 451-461.

Isaac, D.L., 2008. Biomarkers in heart failure management. *Current Opinion in Cardiology.* **23,** 127-133.

Jabbour, H.N., Sales, K.J., Smith, O.P., Battersby, S., Boddy, S.C., 2006. Prostaglandin receptors are mediators of vascular function in endometrial pathologies. *Molecular and Cellular Endocrinology.* **252,** 191-200.

Jackson, G., Gibbs, C.R., Davies, M.K., Lip, G.Y., 2000. ABC of heart failure. Pathophysiology. *BMJ (Clinical Research Ed.).* **320,** 167-170.

Jenny, N.S., Arnold, A.M., Kuller, L.H., Tracy, R.P., Psaty, B.M., 2007. Serum amyloid P and cardiovascular disease in older men and women: results from the Cardiovascular Health Study. *Arteriosclerosis, Thrombosis, and Vascular Biology.* **27,** 352-358.

Jessup, M. & Brozena, S., 2003. Heart failure. *The New England Journal of Medicine.* **348,** 2007-2018.

Jiang, X.C., Jin, W., Hussain, M.M., 2012. The impact of phospholipid transfer protein (PLTP) on lipoprotein metabolism. *Nutrition & Metabolism.* **9,** 75-7075-9-75.

Jiwan, J.L., Wallemacq, P., Herent, M.F., 2011. HPLC-high resolution mass spectrometry in clinical laboratory? *Clinical Biochemistry.* **44,** 136-147.

Kannel, W.B. & McGee, D.L., 1979. Diabetes and cardiovascular disease. The Framingham study. *JAMA : The Journal of the American Medical Association.* **241,** 2035-2038.

Karas, M. & Hillenkamp, F., 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry.* **60,** 2299-2301.

Karp, N.A. & Lilley, K.S., 2007. Design and analysis issues in quantitative proteomics studies. *Proteomics.* **7 Suppl 1,** 42-50.

Kaufman Jr, R.P., Anner, H., Kobzik, L., Valeri, C.R., Shepro, D., Hechtman, H.B., 1987. Vasodilator prostaglandins (PG) prevent renal damage after ischemia. *Annals of Surgery.* **205,** 195.

Kay, R., Barton, C., Ratcliffe, L., Matharoo-Ball, B., Brown, P., Roberts, J., Teale, P., Creaser, C., 2008. Enrichment of low molecular weight serum proteins using acetonitrile precipitation for mass spectrometry based proteomic analysis. *Rapid Communications in Mass Spectrometry.* **22,** 3255-3260.

Kenchaiah, S., Narula, J., Vasan, R.S., 2004. Risk factors for heart failure. *Medical Clinics of North America.* **88,** 1145-1172.

Ketchum, E.S. & Levy, W.C., 2011. Establishing prognosis in heart failure: a multimarker approach. *Progress in Cardiovascular Diseases.* **54,** 86-96.

Kiernan, U.A., Nedelkov, D., Nelson, R.W., 2006. Multiplexed mass spectrometric immunoassay in biomarker research: a novel approach to the determination of a myocardial infarct. *Journal of Proteome Research.* **5,** 2928-2934.

Kingsmore, S.F., 2006. Multiplexed protein measurement: technologies and applications of protein and antibody arrays. *Nature Reviews.Drug Discovery.* **5,** 310-320.

Kocher, T., Pichler, P., Swart, R., Mechtler, K., 2011. Quality control in LC-MS/MS. *Proteomics.* **11,** 1026-1030.

Koenig, W., 2007. Serum amyloid P component and cardiovascular disease: is there a sensible link? *Arteriosclerosis, Thrombosis, and Vascular Biology.* **27,** 698-700.

Kouris, N.T., Zacharos, I.D., Kontogianni, D.D., Goranitou, G.S., Sifaki, M.D., Grassos, H.E., Kalkandi, E.M., Babalis, D.K., 2005. The significance of CA125 levels in patients with chronic congestive heart failure. Correlation with clinical and echocardiographic parameters. *Eur J Heart Fail.* **7,** 199-203.

Krum, H. & Abraham, W.T., 2009. Heart failure. *Lancet.* **373,** 941-955.

Kubo, S.H., Walter, B.A., John, D.H., Clark, M., Cody, R.J., 1987. Liver function abnormalities in chronic heart failure. Influence of systemic hemodynamics. *Archives of Internal Medicine.* **147,** 1227-1230.

Kuiper, J.L., Lind, J.S., Groen, H.J., Roder, J., Grigorieva, J., Roder, H., Dingemans, A.M., Smit, E.F., 2012. VeriStrat((R)) has prognostic value in advanced stage NSCLC patients treated with erlotinib and sorafenib. *British Journal of Cancer.* **107,** 1820-1825.

Kuriyama, M., Wang, M.C., Lee, C.I., Papsidero, L.D., Killian, C.S., Inaji, H., Slack, N.H., Nishiura, T., Murphy, G.P., Chu, T.M., 1981. Use of human prostate-specific antigen in monitoring prostate cancer. *Cancer Research.* **41,** 3874-3876.

Lassus, J., Harjola, V.P., Sund, R., Siirila-Waris, K., Melin, J., Peuhkurinen, K., Pulkki, K., Nieminen, M.S., FINN-AKVA Study group, 2007. Prognostic value of cystatin C in acute heart failure in relation to other markers of renal function and NT-proBNP. *European Heart Journal.* **28,** 1841-1847.

Latini, R., Masson, S., Anand, I.S., Missov, E., Carlson, M., Vago, T., Angelici, L., Barlera, S., Parrinello, G., Maggioni, A.P., Tognoni, G., Cohn, J.N., 2007. Prognostic value of very low plasma concentrations of troponin T in patients with stable chronic heart failure. *Circulation.* **116,** 1242-1249.

Lee, C.W. & Burnett, J., Jr., 2007. Natriuretic peptides and therapeutic applications. *Heart Failure Reviews.* **12,** 131-142.

Levin, E.R., Gardner, D.G., Samson, W.K., 1998. Natriuretic Peptides. *N Engl J Med.* **339,** 321-328.

Levin, Y., Jaros, J.A.J., Schwarz, E., Bahn, S., 2010. Multidimensional protein fractionation of blood proteins coupled to data-independent nanoLC-MS/MS analysis. *Journal of Proteomics.* **73,** 689-695.

Li, G., Vissers, J.P.C., Silva, J.C., Golick, D., Gorenstein, M.V., Geromanos, S.J., 2009. Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics.* **9,** 1696-1719.

Li, X., Yutani, C., Shimokado, K., 1998. Serum Amyloid P Component Associates with High Density Lipoprotein as well as Very Low Density Lipoprotein but Not with Low Density Lipoprotein. *Biochemical and Biophysical Research Communications.* **244,** 249-252.

Lilley, K.S., Deery, M.J., Gatto, L., 2011. Challenges for proteomics core facilities. *Proteomics.* **11,** 1017-1025.

Liu, H., Sadygov, R.G., Yates, J.R.,3rd, 2004. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Analytical Chemistry.* **76,** 4193-4201.

Lok, D.J., Klip, I.T., Lok, S.I., Bruggink-Andre de la Porte, P.W., Badings, E., van Wijngaarden, J., Voors, A.A., de Boer, R.A., van Veldhuisen, D.J., van der Meer, P., 2013. Incremental prognostic power of novel biomarkers (growth-differentiation factor-15, high-sensitivity C-reactive protein, galectin-3, and high-sensitivity troponin-T) in patients with advanced chronic heart failure. *The American Journal of Cardiology.* **112,** 831-837.

MacIver, D.H., Dayer, M.J., Harrison, A.J.I., 2013. A general theory of acute and chronic heart failure. *International Journal of Cardiology.* **165,** 25-34.

MacIver, D.H., Dayer, M.J., Harrison, A.J.I., A general theory of acute and chronic heart failure. *International Journal of Cardiology.*

Mair, J., Dienstl, F., Puschendorf, B., 1992. Cardiac troponin T in the diagnosis of myocardial injury. *Critical Reviews in Clinical Laboratory Sciences.* **29,** 31-57.

Makarov, A., 2000. Electrostatic Axially Harmonic Orbital Trapping:â€‰ A High-Performance Technique of Mass Analysis. *Analytical Chemistry.* **72,** 1156-1162.

Mallick, P. & Kuster, B., 2010. Proteomics: a pragmatic perspective. *Nature Biotechnology.* **28,** 695-709.

Martens, L., Hermjakob, H., Jones, P., Adamski, M., Taylor, C., States, D., Gevaert, K., Vandekerckhove, J., Apweiler, R., 2005. PRIDE: the proteomics identifications database. *Proteomics.* **5,** 3537-3545.

Martinez-Pinna, R., Barbas, C., Blanco-Colio, L., Tunon, J., Ramos-Mozo, P., Lopez, J., Meilhac, O., Michel, J., Egido, J., Martin-Ventura, J., 2010. Proteomic and Metabolomic Profiles in Atherothrombotic Vascular Disease. *Current Atherosclerosis Reports.* **12,** 202-208.

Martin-Ventura, J.L., Duran, M.C., Blanco-Colio, L.M., Meilhac, O., Leclercq, A., Michel, J., Jensen, O.N., Hernandez-Merida, S., Tuñón, J., Vivanco, F., Egido, J., 2004. Identification by a Differential Proteomic Approach of Heat Shock Protein 27 as a Potential Marker of Atherosclerosis. *Circulation.* **110,** 2216-2219.

McLean, A., Huang, S., Salter, M., 2008. Bench-to-bedside review: The value of cardiac biomarkers in the intensive care patient. *Critical Care.* **12,** 215.

McMurray, J.J. & Stewart, S., 2000. Epidemiology, aetiology, and prognosis of heart failure. *Heart.* **83,** 596-602.

McMurray, J.J. & Pfeffer, M.A., 2005. Heart failure. *The Lancet.* **365,** 1877-1889.

Mebazaa, A., Vanpoucke, G., Thomas, G., Verleysen, K., Cohen-Solal, A., Vanderheyden, M., Bartunek, J., Mueller, C., Launay, J.M., Van Landuyt, N., D'Hondt, F., Verschuere, E., Vanhaute, C., Tuytten, R., Vanneste, L., De Cremer, K., Wuyts, J., Davies, H., Moerman, P., Logeart, D., Collet, C., Lortat-Jacob, B., Tavares, M., Laroy, W., Januzzi, J.L., Samuel, J.L., Kas, K., 2012. Unbiased plasma proteomics for novel diagnostic biomarkers in cardiovascular disease: identification of quiescin Q6 as a candidate biomarker of acutely decompensated heart failure. *European Heart Journal.* **33,** 2317-2324.

Mendis, S., Puska, P. and Norrving, B., 2011. *Global atlas on cardiovascular disease prevention and control.* World Health Organization.

Merrifield, R.B., 1965. Automated Synthesis of Peptides. *Science.* **150,** 178-185.

Michalski, A., Cox, J., Mann, M., 2011. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *Journal of Proteome Research.* **10,** 1785-1793.

Moller-Kristensen, M., Jensenius, J.C., Jensen, L., Thielens, N., Rossi, V., Arlaud, G., Thiel, S., 2003. Levels of mannan-binding lectin-associated serine protease-2 in healthy individuals. *Journal of Immunological Methods.* **282,** 159-167.

Mosterd, A. & Hoes, A.W., 2007. Clinical epidemiology of heart failure. *Heart.* **93,** 1137-1146.

Mueller, C., Laule-Kilian, K., Christ, A., Brunner-La Rocca, H.P., Perruchoud, A.P., 2006. Inflammation and long-term mortality in acute congestive heart failure. *American Heart Journal.* **151,** 845-850.

Mukoyama, M., Nakao, K., Hosoda, K., Suga, S., Saito, Y., Ogawa, Y., Shirakami, G., Jougasaki, M., Obata, K., Yasue, H., 1991. Brain natriuretic peptide as a novel cardiac hormone in humans. Evidence for an exquisite dual natriuretic peptide system, atrial natriuretic peptide and brain natriuretic peptide. *Journal of Clinical Investigation.* **87,** 1402.

Murray, H., Cload, B., Collier, C.P., Sivilotti, M.L., 2006. Potential impact of N-terminal pro-BNP testing on the emergency department evaluation of acute dyspnea. *Cjem.* **8,** 251-258.

Muslin, A.J., 2012. Chapter 37 - The Pathophysiology of Heart Failure. In Joseph Hill, ed, *Muscle.* Boston/Waltham: Academic Press. 523-535.

Nagele, H., Bahlo, M., Klapdor, R., Schaeperkoetter, D., Rodiger, W., 1999. CA 125 and its relation to cardiac function. *Am Heart J.* **137,** 1044-1049.

Napoli, C., Zullo, A., Picascia, A., Infante, T., Mancini, F.P., 2013. Recent advances in proteomic technologies applied to cardiovascular disease. *Journal of Cellular Biochemistry.* **114,** 7-20.

National Institute for Health and Clinical Excellence, 2010. *Chronic Heart Failure.* [online]. Available at: http://guidance.nice.org.uk/Topic/Cardiovascular/HeartFailure [accessed October/26 2010].

Niebauer, J., Pflaum, C., Clark, A.L., Strasburger, C.J., Hooper, J., Poole-Wilson, P.A., Coats, A.J., Anker, S.D., 1998. Deficient insulin-like growth factor I in chronic heart failure predicts altered body composition, anabolic deficiency, cytokine and neurohormonal activation. *Journal of the American College of Cardiology.* **32,** 393-397.

Nielsen, F.C., 1992. The molecular and cellular biology of insulin-like growth factor II. *Progress in Growth Factor Research.* **4,** 257-290.

Nilsson, T., Mann, M., Aebersold, R., Yates, J.R.,3rd, Bairoch, A., Bergeron, J.J., 2010. Mass spectrometry in high-throughput proteomics: ready for the big time. *Nature Methods.* **7,** 681-685.

Omenn, G.S., States, D.J., Adamski, M., Blackwell, T.W., Menon, R., Hermjakob, H., Apweiler, R., Haab, B.B., Simpson, R.J., Eddes, J.S., Kapp, E.A., Moritz, R.L., Chan, D.W., Rai, A.J., Admon, A., Aebersold, R., Eng, J., Hancock, W.S., Hefta, S.A., Meyer, H., Paik, Y.K., Yoo, J.S., Ping, P., Pounds, J., Adkins, J., Qian, X., Wang, R., Wasinger, V., Wu, C.Y., Zhao, X., Zeng, R., Archakov, A., Tsugita, A., Beer, I., Pandey, A., Pisano, M., Andrews, P., Tammen, H., Speicher, D.W., Hanash, S.M., 2005. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics.* **5,** 3226-3245.

Pan, S., Aebersold, R., Chen, R., Rush, J., Goodlett, D.R., McIntosh, M.W., Zhang, J., Brentnall, T.A., 2009. Mass spectrometry based targeted protein quantification: methods and applications. *Journal of Proteome Research.* **8,** 787-797.

Patel, N.A., Crombie, A., Slade, S.E., Thalassinos, K., Hughes, C., Connolly, J.B., Langridge, J., Murrell, J.C., Scrivens, J.H., 2012. Comparison of one- and two-dimensional liquid chromatography approaches in the label-free quantitative analysis of Methylocella silvestris. *Journal of Proteome Research.* **11,** 4755-4763.

Paul, W. & Steinwedel, H., 1953. Ein neues Massenspektrometer ohne Magnetfeld. *Zeitschrift Naturforschung Teil A.* **8,** 448.

Pepys, M.B., Booth, D., Hutchinson, W., Gallimore, J., Collins, I., Hohenester, E., 1997. Amyloid P component. A critical review. *Amyloid.* **4,** 274-295.

Pepys, M., Baltz, M., Gomer, K., Davies, A., Doenhoff, M., 1979. Serum amyloid P-component is an acute-phase reactant in the mouse.

Petricoin, E.F., Ardekani, A.M., Hitt, B.A., Levine, P.J., Fusaro, V.A., Steinberg, S.M., Mills, G.B., Simone, C., Fishman, D.A., Kohn, E.C., Liotta, L.A., 2002. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet.* **359,** 572-577.

Pfeffer, M.A. & Braunwald, E., 1990. Ventricular remodeling after myocardial infarction. Experimental observations and clinical implications. *Circulation.* **81,** 1161-1172.

Pogue-Geile, K.L., Chen, R., Bronner, M.P., Crnogorac-Jurcevic, T., Moyes, K.W., Dowen, S., Otey, C.A., Crispin, D.A., George, R.D., Whitcomb, D.C., Brentnall, T.A., 2006. Palladin mutation causes familial pancreatic cancer and suggests a new cancer mechanism. *PLoS Medicine.* **3,** e516.

Polanski, M. & Anderson, N.L., 2007. A list of candidate cancer biomarkers for targeted proteomics. *Biomarker Insights.* **1,** 1-48.

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-Analytical Technologies in the Biomedical and Life Sciences.* **785,** 263-275.

Purvine, S., Eppel, J.T., Yi, E.C., Goodlett, D.R., 2003. Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer. *Proteomics.* **3,** 847-850.

Rathi, S. & Deedwania, P.C., 2012. The Epidemiology and Pathophysiology of Heart Failure. *Medical Clinics of North America.* **96,** 881-890.

Remme, W.J. & Swedberg, K., 2001. Guidelines for the diagnosis and treatment of chronic heart failure. *Eur Heart J.* **22,** 1527-1560.

Ricciotti, E. & FitzGerald, G.A., 2011. Prostaglandins and inflammation. *Arteriosclerosis, Thrombosis, and Vascular Biology.* **31,** 986-1000.

Rifai, N., Gillette, M.A., Carr, S.A., 2006. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature Biotechnology.* **24,** 971-983.

Righetti, P.G. & Boschetti, E., 2007. Sherlock Holmes and the proteome - a detective story. *Febs Journal.* **274,** 897-905.

Righetti, P.G., Boschetti, E., Lomas, L., Citterio, A., 2006. Protein Equalizer (TM) Technology: The quest for a "democratic proteome". *Proteomics.* **6,** 3980-3992.

Righetti, P.G., Boschetti, E., Zanella, A., Fasoli, E., Citterio, A., 2010. Plucking, pillaging and plundering proteomes with combinatorial peptide ligand libraries. *Journal of Chromatography A.* **1217,** 893-900.

Righetti, P.G., Campostrini, N., Pascali, J., Hamdan, M., Astner, H., 2004. Quantitative proteomics: a review of different methodologies. *European Journal of Mass Spectrometry.* **10,** 335-348.

Righetti, P.G., Castagna, A., Boschetti, E., Lomas, L., 2005. Equalizer beads; the quest for a "Democratic Proteome". *Molecular & Cellular Proteomics.* **4,** S12-S12.

Righetti, P.G., Boschetti, E., Kravchuk, A.V., Fasoli, E., 2010. The proteome buccaneers: how to unearth your treasure chest via combinatorial peptide ligand libraries. *Expert Review of Proteomics.* **7,** 373-385.

Robins, S.J., Lyass, A., Brocia, R.W., Massaro, J.M., Vasan, R.S., 2013. Plasma lipid transfer proteins and cardiovascular disease. The Framingham Heart Study. *Atherosclerosis.* **228,** 230-236.

Rocchiccioli, J.P., McMurray, J.J.V., Dominiczak, A.F., 2010. Biomarkers in heart failure: a clinical review. *Heart Failure Reviews.* **15,** 251-273.

Roepstorff, P. & Fohlman, J., 1984. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical Mass Spectrometry.* **11,** 601.

Rundlett, K.L. & Armstrong, D.W., 1996. Mechanism of signal suppression by anionic surfactants in capillary electrophoresis-electrospray ionization mass spectrometry. *Analytical Chemistry.* **68,** 3493-3497.

Schmidt, A., Karas, M., Dulcks, T., 2003. Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI? *Journal of the American Society for Mass Spectrometry.* **14,** 492-500.

Schunkert, H., Konig, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., Preuss, M., Stewart, A.F.R., Barbalic, M., Gieger, C., Absher, D., Aherrahrou, Z., Allayee, H., Altshuler, D., Anand, S.S., Andersen, K., Anderson, J.L., Ardissino, D., Ball, S.G., Balmforth, A.J., Barnes, T.A., Becker, D.M., Becker, L.C., Berger, K., Bis, J.C., Boekholdt, S.M., Boerwinkle, E., Braund, P.S., Brown, M.J., Burnett, M.S., Buysschaert, I., Carlquist, J.F., Chen, L., Cichon, S., Codd, V., Davies, R.W., Dedoussis, G., Dehghan, A., Demissie, S., Devaney, J.M., Diemert, P., Do, R., Doering, A., Eifert, S., Mokhtari, N.E.E., Ellis, S.G., Elosua, R., Engert, J.C., Epstein, S.E., de Faire, U., Fischer, M., Folsom, A.R., Freyer, J., Gigante, B., Girelli, D., Gretarsdottir, S., Gudnason, V., Gulcher, J.R., Halperin, E., Hammond, N., Hazen, S.L., Hofman, A., Horne, B.D., Illig, T., Iribarren, C., Jones, G.T., Jukema, J.W., Kaiser, M.A., Kaplan, L.M., Kastelein, J.J.P., Khaw, K., Knowles, J.W., Kolovou, G., Kong, A., Laaksonen, R., Lambrechts, D., Leander, K., Lettre, G., Li, M., Lieb, W., Loley, C., Lotery, A.J., Mannucci, P.M., Maouche, S., Martinelli, N., McKeown, P.P., Meisinger, C., Meitinger, T., Melander, O., Merlini, P.A., Mooser, V., Morgan, T., Muhleisen, T.W., Muhlestein, J.B., Munzel, T., Musunuru, K., Nahrstaedt, J., Nelson, C.P., Nothen, M.M., Olivieri, O., Patel, R.S., Patterson, C.C., Peters, A., Peyvandi, F., Qu, L., Quyyumi, A.A., Rader, D.J., Rallidis, L.S., Rice, C., Rosendaal, F.R., Rubin, D., Salomaa, V., Sampietro, M.L., Sandhu, M.S., Schadt, E., Schafer, A., Schillert, A., Schreiber, S., Schrezenmeir, J., Schwartz, S.M., Siscovick, D.S., Sivananthan, M.,

Sivapalaratnam, S., Smith, A., Smith, T.B., Snoep, J.D., Soranzo, N., Spertus, J.A., Stark, K., Stirrups, K., Stoll, M., Tang, W.H.W., Tennstedt, S., Thorgeirsson, G., Thorleifsson, G., Tomaszewski, M., Uitterlinden, A.G., van Rij, A.,M., Voight, B.F., Wareham, N.J., Wells, G.A., Wichmann, H., Wild, P.S., Willenborg, C., Witteman, J.C.M., Wright, B.J., Ye, S., Zeller, T., Ziegler, A., Cambien, F., Goodall, A.H., Cupples, L.A., Quertermous, T., Marz, W., Hengstenberg, C., Blankenberg, S., Ouwehand, W.H., Hall, A.S., Deloukas, P., Thompson, J.R., Stefansson, K., Roberts, R., Thorsteinsdottir, U., O'Donnell, C.,J., McPherson, R., Erdmann, J., 2011. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genetics.* **43,** 333-338.

Schwaeble, W.J., Lynch, N.J., Clark, J.E., Marber, M., Samani, N.J., Ali, Y.M., Dudler, T., Parent, B., Lhotta, K., Wallis, R., Farrar, C.A., Sacks, S., Lee, H., Zhang, M., Iwaki, D., Takahashi, M., Fujita, T., Tedford, C.E., Stover, C.M., 2011. Targeting of mannan-binding lectin-associated serine protease-2 confers protection from myocardial and gastrointestinal ischemia/reperfusion injury. *Proceedings of the National Academy of Sciences of the United States of America.* **108,** 7523-7528.

Shah, A. & Bano, B., 2009. Cystatins in health and diseases. *International Journal of Peptide Research and Therapeutics.* **15,** 43-48.

Shi, Y., Xiang, R., Horvath, C., Wilkins, J.A., 2004. The role of liquid chromatography in proteomics. *Journal of Chromatography.A.* **1053,** 27-36.

Shliaha, P.V., Bond, N.J., Gatto, L., Lilley, K.S., 2013. Effects of Traveling Wave Ion Mobility Separation on Data Independent Acquisition in Proteomics Studies. *Journal of Proteome Research.* **12,** 2323-2339.

Siirilä-Waris, K., Lassus, J., Melin, J., Peuhkurinen, K., Nieminen, M.S., Harjola, V., 2006. Characteristics, outcomes, and predictors of 1-year mortality in patients hospitalized for acute heart failure. *European Heart Journal.* **27,** 3011-3017.

Silbiger, V.N., Luchessi, A.D., Hirata, R.D.C., Neto, L.G.L., Pastorelli, C.P., Ueda, E.K.M., Santos, E.S.d., Pereira, M.P., Ramos, R., Sampaio, M.F., Armaganijan, D., Paik, S.H., Murata, Y., Ooi, G.T., Ferguson, E.W., Hirata, M.H., 2011. Time course proteomic profiling of human myocardial infarction plasma samples: An approach to new biomarker discovery. *Clinica Chimica Acta.* **412,** 1086-1093.

Silva, J.C., Denny, R., Dorschel, C.A., Gorenstein, M., Kass, I.J., Li, G.Z., McKenna, T., Nold, M.J., Richardson, K., Young, P., Geromanos, S., 2005. Quantitative

proteomic analysis by accurate mass retention time pairs. *Analytical Chemistry.* **77,** 2187-2200.

Silva, J.C., Gorenstein, M.V., Li, G.Z., Vissers, J.P.C., Geromanos, S.J., 2006. Absolute quantification of proteins by LCMSE - A virtue of parallel MS acquisition. *Molecular & Cellular Proteomics.* **5,** 144-156.

Simi, M., Leardi, S., Tebano, M.T., Castelli, M., Costantini, F.M., Speranza, V., 1987. Raised plasma concentrations of platelet factor 4 (PF4) in Crohn's disease. *Gut.* **28,** 336-338.

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., Klenk, D.C., 1985. Measurement of protein using bicinchoninic acid. *Analytical Biochemistry.* **150,** 76-85.

Sorensen, R., Thiel, S., Jensenius, J.C., 2005. Mannan-binding-lectin-associated serine proteases, characteristics and disease associations. *Springer Seminars in Immunopathology.* **27,** 299-319.

Stahl-Zeng, J., Lange, V., Ossola, R., Eckhardt, K., Krek, W., Aebersold, R., Domon, B., 2007. High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Molecular & Cellular Proteomics : MCP.* **6,** 1809-1817.

Steen, H. & Mann, M., 2004. The ABC's (and XYZ's) of peptide sequencing. *Nature Reviews.Molecular Cell Biology.* **5,** 699-711.

Stewart, C.R., Tseng, A.A., Mok, Y.F., Staples, M.K., Schiesser, C.H., Lawrence, L.J., Varghese, J.N., Moore, K.J., Howlett, G.J., 2005. Oxidation of low-density lipoproteins induces amyloid-like structures that are recognized by macrophages. *Biochemistry.* **44,** 9108-9116.

Street, J.M. & Dear, J.W., 2010. The application of mass-spectrometry-based protein biomarker discovery to theragnostics. *British Journal of Clinical Pharmacology.* **69,** 367-378.

Surinova, S., Schiess, R., Huttenhain, R., Cerciello, F., Wollscheid, B., Aebersold, R., 2011. On the development of plasma protein biomarkers. *Journal of Proteome Research.* **10,** 5-16.

Tabb, D.L., Vega-Montoto, L., Rudnick, P.A., Variyath, A.M., Ham, A.J., Bunk, D.M., Kilpatrick, L.E., Billheimer, D.D., Blackman, R.K., Cardasis, H.L., Carr, S.A., Clauser, K.R., Jaffe, J.D., Kowalski, K.A., Neubert, T.A., Regnier, F.E., Schilling, B., Tegeler, T.J., Wang, M., Wang, P., Whiteaker, J.R., Zimmerman, L.J., Fisher, S.J., Gibson,

B.W., Kinsinger, C.R., Mesri, M., Rodriguez, H., Stein, S.E., Tempst, P., Paulovich, A.G., Liebler, D.C., Spiegelman, C., 2010. Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *Journal of Proteome Research.* **9,** 761-776.

Taguchi, F., Solomon, B., Gregorc, V., Roder, H., Gray, R., Kasahara, K., Nishio, M., Brahmer, J., Spreafico, A., Ludovini, V., Massion, P.P., Dziadziuszko, R., Schiller, J., Grigorieva, J., Tsypin, M., Hunsucker, S.W., Caprioli, R., Duncan, M.W., Hirsch, F.R., Bunn, P.A.,Jr, Carbone, D.P., 2007. Mass spectrometry to classify non-small-cell lung cancer patients for clinical outcome after treatment with epidermal growth factor receptor tyrosine kinase inhibitors: a multicohort cross-institutional study. *Journal of the National Cancer Institute.* **99,** 838-846.

Tambor, V., Fucikova, A., Lenco, J., Kacerovsky, M., Rehacek, V., Stulik, J., Pudil, R., 2010. Application of proteomics in biomarker discovery: a primer for the clinician. *Physiological Research / Academia Scientiarum Bohemoslovaca.* **59,** 471-497.

Thomson, J., 1899. On the Masses of the Ions in Gases at Low Pressures. *The Chemical Educator*. 547-567.

Thulasiraman, V., Lin, S.H., Gheorghiu, L., Lathrop, J., Lomas, L., Hammond, D., Boschetti, E., 2005. Reduction of the concentration difference of proteins in biological liquids using a library of combinatorial ligands. *Electrophoresis.* **26,** 3561-3571.

Townsend N, Wickramasinghe K, Bhatnagar P, Smolina K, Nichols M, Leal J, Luengo-Fernandez R and Rayner M, 2012. *Coronary heart disease statistics 2012 edition.* London: British Heart Foundation.

Tu, C., Rudnick, P.A., Martinez, M.Y., Cheek, K.L., Stein, S.E., Slebos, R.J., Liebler, D.C., 2010. Depletion of abundant plasma proteins and limitations of plasma proteomics. *Journal of Proteome Research.* **9,** 4982-4991.

Turner, M.W., 2003. The role of mannose-binding lectin in health and disease. *Molecular Immunology.* **40,** 423-429.

Urbonaviciene, G., Martin-Ventura, J.L., Lindholt, J.S., Urbonavicius, S., Moreno, J.A., Egido, J., Blanco-Colio, L.M., 2011. Impact of soluble TWEAK and CD163/TWEAK ratio on long-term cardiovascular mortality in patients with peripheral arterial disease. *Atherosclerosis.* **219,** 892-899.

Vaisar, T., Mayer, P., Nilsson, E., Zhao, X., Knopp, R., Prazen, B.J., 2010. HDL in humans with cardiovascular disease exhibits a proteomic signature. *Clinica Chimica Acta.* **411,** 972-979.

Valentine, S.J., Liu, X., Plasencia, M.D., Hilderbrand, A.E., Kurulugama, R.T., Koeniger, S.L., Clemmer, D.E., 2005. Developing liquid chromatography ion mobility mass spectometry techniques.

van Kimmenade, R.R.J. & Januzzi, J.L., 2012. Emerging Biomarkers in Heart Failure. *Clinical Chemistry.* **58,** 127-138.

Vanloo, B., Peelman, F., Deschuymere, K., Taveirne, J., Verhee, A., Gouyette, C., Labeur, C., Vandekerckhove, J., Tavernier, J., Rosseneu, M., 2000. Relationship between structure and biochemical phenotype of lecithin:cholesterol acyltransferase (LCAT) mutants causing fish-eye disease. *Journal of Lipid Research.* **41,** 752-761.

Vitek, O., 2009. Getting started in computational mass spectrometry-based proteomics. *PLoS Computational Biology.* **5,** e1000366.

Voller, A., Bartlett, A., Bidwell, D.E., 1978. Enzyme immunoassays with special reference to ELISA techniques. *Journal of Clinical Pathology.* **31,** 507-520.

Voyksner, R.D. & Lee, H., 1999. Investigating the use of an octupole ion guide for ion storage and high-pass mass filtering to improve the quantitative performance of electrospray ion trap mass spectrometry. *Rapid Communications in Mass Spectrometry : RCM.* **13,** 1427-1437.

Wang, M.C., Papsidero, L.D., Kuriyama, M., Valenzuela, L.A., Murphy, G.P., Chu, T.M., 1981. Prostate antigen: a new potential marker for prostatic cancer. *The Prostate.* **2,** 89-96.

Wang, Y., Yang, F., Gritsenko, M.A., Wang, Y., Clauss, T., Liu, T., Shen, Y., Monroe, M.E., Lopez-Ferrer, D., Reno, T., Moore, R.J., Klemke, R.L., Camp, D.G.,2nd, Smith, R.D., 2011. Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics.* **11,** 2019-2026.

Watanabe, S., Tamura, T., Ono, K., Horiuchi, H., Kimura, T., Kita, T., Furukawa, Y., 2010. Insulin-like growth factor axis (insulin-like growth factor-I/insulin-like growth factor-binding protein-3) as a prognostic predictor of heart failure: association with adiponectin. *European Journal of Heart Failure.* **12,** 1214-1222.

Watson, C.J., Ledwidge, M.T., Phelan, D., Collier, P., Byrne, J.C., Dunn, M.J., McDonald, K.M., Baugh, J.A., 2011. Proteomic Analysis of Coronary Sinus Serum

Reveals Leucine-Rich alpha 2-Glycoprotein as a Novel Biomarker of Ventricular Dysfunction and Heart Failure. *Circulation-Heart Failure.* **4,** 188-U126.

Weber, O., Bischoff, H., Schmeck, C., Bottcher, M.F., 2010. Cholesteryl ester transfer protein and its inhibition. *Cellular and Molecular Life Sciences : CMLS.* **67,** 3139-3149.

White, M., Ducharme, A., Ibrahim, R., Whittom, L., Lavoie, J., Guertin, M.C., Racine, N., He, Y., Yao, G., Rouleau, J.L., Schiffrin, E.L., Touyz, R.M., 2006. Increased systemic inflammation and oxidative stress in patients with worsening congestive heart failure: improvement after short-term inotropic support. *Clin Sci (Lond).* **110,** 483-489.

Wiley, W. & McLaren, I.H., 1955. Time-of-flight mass spectrometer with improved resolution. *Review of Scientific Instruments.* **26,** 1150.

Wilkins, M.R., Sanchez, J.C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F., Williams, K.L., 1996. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology & Genetic Engineering Reviews.* **13,** 19-50.

Wilm, M., 2009. Quantitative proteomics in biological research. *Proteomics.* **9,** 4590-4605.

Wilm, M. & Mann, M., 1996. Analytical properties of the nanoelectrospray ion source. *Analytical Chemistry.* **68,** 1-8.

Wirtz, K.W., 2006. Phospholipid transfer proteins in perspective. *FEBS Letters.* **580,** 5436-5441.

Wolfe, M.L. & Rader, D.J., 2004. Cholesteryl Ester Transfer Protein and Coronary Artery Disease: An Observation With Therapeutic Implications. *Circulation.* **110,** 1338-1340.

Wolff, M., M. & Stephens, W., E., 1953. A Pulsed Mass Spectrometer with Time Dispersion. **24,** 616.

Yamashita, M. & Fenn, J.B., 1984. Electrospray ion source. Another variation on the free-jet theme. *The Journal of Physical Chemistry.* **88,** 4451-4459.

Yang, F., Shen, Y., Camp, D.G.,2nd, Smith, R.D., 2012. High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert Review of Proteomics.* **9,** 129-134.

Yates, J.R.,3rd, 1998. Mass spectrometry and the age of the proteome. *Journal of Mass Spectrometry : JMS.* **33,** 1-19.

Yates, J.R., Ruse, C.I., Nakorchevsky, A., 2009. Proteomics by mass spectrometry: approaches, advances, and applications. *Annual Review of Biomedical Engineering.* **11,** 49-79.

Ying, S., Gewurz, A.T., Jiang, H., Gewurz, H., 1993. Human serum amyloid P component oligomers bind and activate the classical complement pathway via residues 14-26 and 76-92 of the A chain collagen-like region of C1q. *The Journal of Immunology.* **150,** 169-176.

Yuhki, K., Kojima, F., Kashiwagi, H., Kawabe, J., Fujino, T., Narumiya, S., Ushikubi, F., 2011. Roles of prostanoids in the pathogenesis of cardiovascular diseases: Novel insights from knockout mouse studies. *Pharmacology & Therapeutics.* **129,** 195-205.

Zhang, Z., 2012. An in vitro diagnostic multivariate index assay (IVDMIA) for ovarian cancer: harvesting the power of multiple biomarkers. *Reviews in Obstetrics and Gynecology.* **5,** 35.

Zhang, H. & Ge, Y., 2011. Comprehensive analysis of protein modifications by top-down mass spectrometry. *Circulation.Cardiovascular Genetics.* **4,** 711.

Zhang, X., Fang, A., Riley, C.P., Wang, M., Regnier, F.E., Buck, C., 2010. Multi-dimensional liquid chromatography in proteomics--a review. *Analytica Chimica Acta.* **664,** 101-113.

Zhang, Z., Bast, R.C.,Jr, Yu, Y., Li, J., Sokoll, L.J., Rai, A.J., Rosenzweig, J.M., Cameron, B., Wang, Y.Y., Meng, X.Y., Berchuck, A., Van Haaften-Day, C., Hacker, N.F., de Bruijn, H.W., van der Zee, A.G., Jacobs, I.J., Fung, E.T., Chan, D.W., 2004. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Research.* **64,** 5882-5890.

Zhang, Z. & Chan, D.W., 2010. The road from discovery to clinical diagnostics: lessons learned from the first FDA-cleared in vitro diagnostic multivariate index assay of proteomic biomarkers. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology.* **19,** 2995-2999.

Zhou, F., Cardoza, J.D., Ficarro, S.B., Adelmant, G.O., Lazaro, J.B., Marto, J.A., 2010. Online nanoflow RP-RP-MS reveals dynamics of multicomponent Ku complex in response to DNA damage. *Journal of Proteome Research.* **9,** 6242-6255.

Zhou, J., Zhou, T., Cao, R., Liu, Z., Shen, J., Chen, P., Wang, X., Liang, S., 2006. Evaluation of the application of sodium deoxycholate to proteomic analysis of rat hippocampal plasma membrane. *Journal of Proteome Research.* **5,** 2547-2553.

Zhu, W., Smith, J.W., Huang, C.M., 2010. Mass spectrometry-based label-free quantitative proteomics. *Journal of Biomedicine & Biotechnology.* **2010,** 840518.