Ovarian cancer serum biomarker discovery using proteomics

Musarat Kabir

Department of Gynaecological Oncology,
University College London (UCL),
Gower Street, London

THIS THESIS IS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY FROM UNIVERSITY COLLEGE LONDON

I [Musarat Kabir] confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Ovarian cancer is a lethal gynaecological malignancy which is known as the silent killer. It has a poor prognosis due to the lack of major symptoms in early stage disease and hence its late detection. Cancer antigen-125, the most widely used biomarker for ovarian cancer detection, lacks appropriate sensitivity and specificity. Thus, early biomarkers of the disease are urgently required. Proteomic analysis of human serum promises to be a valuable approach for the discovery of putative biomarkers for human malignancies like ovarian cancer, which could be developed into non-invasive blood tests. In this study, serum samples from a pilot study for ovarian cancer screening which were collected prior to diagnosis were processed at Memorial Sloan Kettering Cancer Research Centre, in collaboration with Prof. Tempst's group, who had developed a novel mass spectrometry (MS)-based technology platform for the high-throughput extraction and measurement of serum peptides. Several marker peaks were identified, which when used in combination with the ovarian cancer biomarker CA-125, assisted in the discrimination of case versus healthy samples at an earlier point prior to diagnosis. Work then involved the establishment and optimisation of a similar serum profiling platform at UCL. This involved the optimisation of a liquid-handling robot to provide semi-automated high-throughput sample purification and spotting, and optimisation of spectral acquisition and processing. The reproducibility of the platform was tested and the effects of different sample handling conditions on peptide profiles examined. The method was then used to search for putative markers of ovarian cancer, using identically processed samples from women diagnosed with malignant or benign ovarian cancer and healthy controls. Finally, as a complementary approach to discover protein biomarkers, the same samples were profiled using 2D Difference Gel Electrophoresis, employing different fractionation strategies to overcome the very large dynamic range of protein expression in serum. Mass spectrometry was used to identify several previously reported and some novel putative biomarkers of ovarian cancer, which warrant further validation.

Acknowledgements

The work presented in this thesis would not have been possible without the help of many people who have been tortured by me. This section is therefore an attempt to thank all of those who have played a crucial role. I would first and foremost like to thank Prof. Ian J. Jacobs and Prof. Mike D. Waterfield for providing me with the opportunity to carry out my research within the EGA Institute for Women's health at UCL and making the whole ordeal possible. I would like to thank Prof. Jacobs for the funding of my PhD via the Helen Feather Trust. And also Prof. Waterfield for organising the collaboration with Prof. Paul Tempst's group at the Memorial Sloan Kettering Cancer Research Centre, New York, through the Ludwig Institute for Cancer Research UCL branch. The experience and friends I gained there are irreplaceable and unforgettable.

However, the solid base of my PhD work was carried out under the supervision of Dr John Timms in the Cancer Proteomics Laboratory at UCL. I would like to thank him for his support, advice and patience during the experimental work. I am grateful for his guidance on the writing of this report.

All the samples used were provided by Dr Usha Menon's team from the UKCTOCS and UKOPS collections. For this I am grateful to many people and would like to thank everyone from the group including Alecs Gentry-Maharaj, Jeremy Ford, Rachel Hallett and all the MLA's.

As part of the MRC collaboration I am grateful to many people involved in the project including; Prof. Rainer Cramer, Ali Tiss, Celia Smith, Prof. Alex Gammerman, Ilia Nouretdinov, Zhiyuan Luo, Brian Burford, and Stephane Camuzeaux, many thanks to you all.

I would like to thank Dr Mark Weeks and Dr M^{a}Carmen Duran-Ruiz for useful discussions, proof reading of the work presented in this thesis and for their patience. I am sorry for torturing you both! I would also like to thank John Sinclair (soon to be Dr) for imparting his 2D-DIGE expertise and Dr Pedro Cutillas from the Centre for

Cell Signalling, Institute of Cancer, Queen Mary, University of London, for his help with the protein identification.

In addition, I would like to thank all my wonderful colleagues and friends from LICR, the Translation Lab and UCL for their support and advice throughout my PhD. In no particular order I would like to thank; Akunna Akpan (Oshe pupo), Ayo Afonja (Oshe pupo), Eli Jovceva, Severine Garbi, Hong-Lin Chan, Mariana Bertani, Bertran Gerrits, Richard Jacob, Ken Choi, Lydia Quaye, Edmund George and Indira Thandi (my yoga instructor). Thank you all for everything you have taught me!

I would also like to thank Dr Joseph Villanueva from MKSCC for his help and support during my time in New York. I can never forget the amazing lunches and political conversations with Iana Aranda and John Philip.

The most important people to thank are my lovely parents, my sister and brothers. Thank you, all. Three small words, so much to add. For all your love, prayers and your support a million words would be too short. Al-hamdu-illal wa Shukruall-illah.

Allahumma infa’nii bimaa ‘allamtanii wa’allimnii maa yanfa’ unii.

Allahumma ijal leesanee ‘amiran bi thikrika wa qalbi bi khashyatika. Ameen.

مسرت کبیر

Abbreviations

The abbreviations that are used throughout this thesis are listed below. The abbreviations are subdivided into four categories: General; Chemicals: Mass spectrometry and separation techniques and Proteins. The meaning of the abbreviation is also given the first time used.

General

AJCC	American Joint Committee on Cancer
CA-125	Cancer antigen 125
CEA	Carcinoembryonic antigen
Da	Dalton
FIGO	International Federation of Gynaecology and Obstetrics
HMR	High Mass Range 4–15 kDa m/z
HNPCC	Hereditary non-polyposis colorectal cancer syndrome
k-NN	K-nearest neighbours classification algorithm
Laser	Light Amplification by Stimulated Emission of Radiation
LMR	Low Mass Range 700-4000Da m/z
MARS	Multiple Affinity Removal System
Mw	Molecular Weight
mmu	Millimass units
pI	Isoelectric point
ppm	parts per million
PSA	Prostate-specific antigen
RPM	Revolution Per Minute
RT	Room Temperature
SVM	Support Vector Machine
TNM	Classification of Malignant Tumours cancer staging system
UKCTOCS	United Kingdom Collaborative Trial of Ovarian Cancer Screening
UKOPS	United Kingdom Ovarian cancer Population Study
UV	Ultra Violet

Chemicals

α-CCA	Alpha cyano 4-hydroxy-cinnamic acid
AmBic	Ammonium bicarbonate (NH_4HCO_3)
ACN	Acetonitrile
APS	Ammonium Persulphate
BPB	Bromophenol Blue
C8	Octyl silane material
C18	Octadecyl silane material
CCB	Colloidal Coomassie Blue
CHAPS	(3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate)
Cy2	3-(4-carboxymethyl)phenylmethyl)-3'-ethyloxacarbocyanine halide N-hydroxy-succinimidyl ester
Cy3	1-(5-carboxypentyl)-1'-propylindocarbocyanine halide N-hydroxy-succinimidyl ester

Cy5	1-(5-carboxypentyl)-1'-methylindodicarbocyanine halide N-hydroxy-succinimidyl
DTT	dithiothreitol
IAM	Iodoacetamide
SDS	Sodium Dodecyl Sulphate
TEMED	N,N,N',N'-Tetramethylethylenediamine
TFA	Trifluoroacetic Acid
Tris	Tris-(hydroxymethyl) aminomethane
NHS-ester	N-hydroxy-succinimidyl
PBS	Phosphate-buffered saline

Mass spectrometry and separation techniques

1DE	One-Dimensional Gel Electrophoresis
2DE	Two-Dimensional Gel Electrophoresis
2D-LC	Two-Dimensional Liquid Chromatography
BVA	Biological Variation Analysis
DE	Delayed Extraction
DIA	Differential In-gel Analysis
DIGE	Difference gel electrophoresis
ESI	Electrospray Ionisation
HPLC	High Performance Liquid Chromatography
IEF	Isoelectrical focusing
IPG	Immobilised pH Gradient
LC	Liquid Chromatography
LC-MS	Liquid Chromatography coupled online with Mass Spectrometry
MALDI	Matrix-Assisted-Laser-Desorption/Ionisation
m/z	mass-to-charge ratio
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
TOF	Time of Flight
PAGE	Polyacrylamide gel electrophoresis
RP	Reverse phase

Proteins

ALB	Albumin
APO	Apolipoproteins
HP	Haptoglobin
IgA	Immunoglobulin A
IgG	Immunoglobulin G
SERPIN	Serine protease inhibitor (Antitrypsin)
TF	Transferrin

Figures

Tables

Chapter 1 Introduction

1.1 Overview of cancer

Cancer is a disease which can affect people of all nationalities and age groups. A number of cancer types are sex specific, such as cervical, uterine sarcoma and ovarian cancer in females and prostate cancer in males. More than 100 distinct types of human cancer have been described, and subtypes of tumours can be found within specific organs [Grizzi and Chiriva-Internati, 2006]. It is theorised that all cancers start with a mutation in a single cell in the body. Although in rare cases a single mutation may be enough, cancer is typically an accumulation of mutations that irreversibly transforms a normal cell into a cancerous one over a prolonged period of time.

Human cancer is predominately a disease of the various cell surface and glandular epithelia [Cairns, 1975; Grizzi and Chiriva-Internati, 2006; Grizzi et al., 2006]. It is a multistage process involving dynamic changes in the genome with alterations in different families of the cell cycle regulatory mechanisms. For example, the production of oncogenes with dominant gain of function and tumour suppressor genes with recessive loss of function as shown in Figure 1.1 [Hanahan and Weinberg, 2000]. Essentially, cancers arise through clonal selection, an evolutionary process promoting proliferation of mutated cells as a result of gene expression changes which also confer a survival advantage. These gene expression changes may manifest as the appearance of new proteins, differences in the amount of expressed proteins, and/or changes in post-translational modifications [Bell, 2005; Hanahan and Weinberg, 2000].

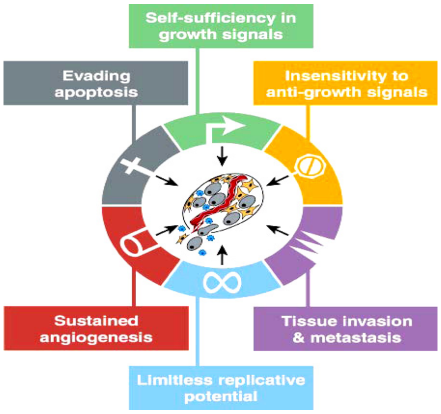

Self-sufficiency in growth signals
Evading apoptosis
Insensitivity to anti-growth signals
Sustained angiogenesis
Tissue invasion & metastasis
Limitless replicative potential

Figure 1.1 The acquired capabilities of cancer cells. Adapted from Hanahan & Weinberg (2000).

The majority of cancer cell genotypes are believed to be a manifestation of six essential alterations in the cell physiology that collectively dictate malignant growth. These alterations include evading apoptosis, self-sufficiency in growth signals, insensitivity to anti-growth signals, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis [Hanahan and Weinberg, 2000]. Only 10% of cancers are the result of inherited genetic susceptibility. Thus, a 90% risk of developing cancer is attributed to a combination of environment, diet, cultural and lifestyle factors [Bell, 2005; Hanahan and Weinberg, 2000].

A diagnosis of cancer can be a very stressful event for the patients and their families. Patients, partners and other family members can suffer from clinical levels of depression and severe levels of anxiety and stress reactions [Steck et al., 2007]. However, cancer does not have to be terminal and treatment can be very successful for example treatment of prostate and some breast cancers. Successful treatment is invariably dependent on early detection. In the case of late detection where advanced metastatic disease has already developed, there is an extremely poor prognosis. For example, pancreatic and ovarian cancers, while relatively uncommon, have extremely poor prognoses which are directly attributable to their late diagnosis. Thus, early stage disease biomarkers are urgently needed.

1.1.1 Overview of ovarian cancer

Ovarian cancers are one of the most lethal gynaecological malignancies worldwide. The tumours range from benign to aggressive malignant including an intermediate class referred to as borderline carcinomas. Ovarian cancers arise in the ovaries which are responsible for hormone and egg production. As Figure 1.2 shows they lie in the pelvis either side of the uterus. Like other cancers, ovarian tumours are believed to arise through clonal selection of a mutated cell.

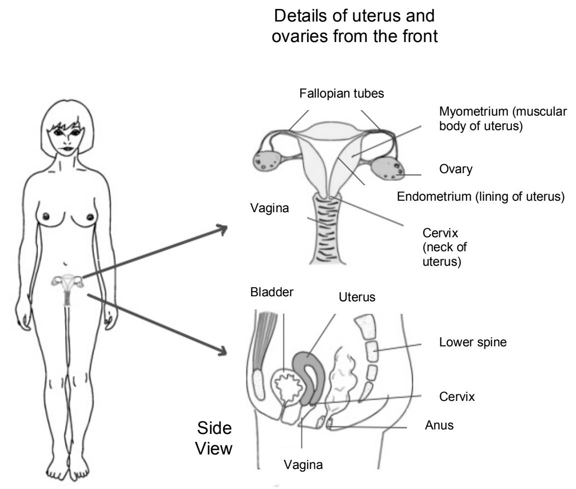

Details of uterus and ovaries from the front
Fallopian tubes
Myometrium (muscular body of uterus)
Ovary
Endometrium (lining of uterus)
Vagina
Cervix (neck of uterus)
Bladder
Uterus
Lower spine
Cervix
Anus
Side View
Vagina

Figure 1.2 Schematic illustration of the female reproductive system and position of the ovaries

(Copyright EMIS and PiP 2008, as distributed on www.patient.co.uk, reproduced with permission.)

There are three types of ovarian cancer named according to the cell type from which the tumour originates. The most common are epithelial cancers (carcinomas), which arise in the ovarian surface epithelium (OSE) and account for 90% of diagnosed cases. Epithelial carcinomas can be histologically classified depending on their microscopic appearance and site; namely, endometrioid, mucinous, serous, clear cell, and undifferentiated carcinomas, [Rosenthal et al., 2006].

The second type is malignant germ cell tumours that form in the part of the ovary responsible for egg production; overall they account for approximately 5% of ovarian cancers. The third type comprises malignant sex-cord/stromal tumours. These arise in the connective tissue and hormone producing parts of the ovary and also account for 5% of malignant ovarian tumours.

1.1.2 Ovarian cancer aetiology and risk factors

In around 90% of ovarian cancer cases, there is no identifiable cause. However, family history plays an important role in ovarian cancer. Ovarian cancer has been linked with several hereditary syndromes including: breast-ovarian cancer syndrome, hereditary non-polyposis colorectal cancer syndrome, site-specific ovarian cancer syndrome, Li-Fraumeni syndrome and Cowden’s syndrome. All of these are within the 10% of cases that have known identifiable hereditary causes of ovarian cancer.

Breast-ovarian cancer syndrome is characterised by a hereditary mutation in the BRCA1 gene, which has been linked to increased risk of both breast and ovarian cancer. About 30-40% of women who have this mutation develop ovarian cancer. A mutation of the BRCA2 gene also increases the risk of ovarian cancer, but to a lesser degree. Clues that may indicate the presence of these mutations include family members who have ovarian cancer or breast cancer or both, especially those who are diagnosed with breast cancer when younger than 50 years. Ashkenazi Jews demonstrate increased frequency of BRCA1 and BRCA2 gene mutations [Ramus et al., 2007; King et al., 2003].

Hereditary non-polyposis colorectal cancer (HNPCC) syndrome (also known as Lynch syndrome II) is a genetic syndrome that is also caused by inherited gene mutations that reduce the body's ability to repair damage to its DNA. This results in a greatly increased risk for colorectal, endometrial and ovarian cancer. HNPCC is predominately associated with colon cancer developing in people younger than 50 years. The risk for ovarian cancer with HNPCC syndrome is much less than that associated with BRCA1 or BRCA2 and probably causes about 1% of all ovarian epithelial cancers. Other organs that can be involved include the breast, stomach, and pancreas. Germline mutations in one of five mismatch repair genes are responsible for this syndrome. These are MSH2 (chromosome 2q), MLH1 (chromosome 3p), PMS1 (chromosome 2q) PMS2 (chromosome7p) and MSH6 (chromosome 2p) [Zweemer, 2002].

Li-Fraumeni syndrome is a rare, hereditary cancer syndrome that is linked to an inherited mutation in the p53 gene, which normally prevents cells with DNA damage from replicating. Cancers in Li-Fraumeni families typically occur between the ages of 15 and 44 [Fallows et al., 2001]. Cowden’s syndrome or multiple hamartoma syndrome is an inherited genetic disease caused again by mutations in the PTEN tumour suppressor gene. These mutations prevent the PTEN protein from effectively regulating cell survival and division, which can lead to the formation of tumours.

Cowden syndrome is one of several inherited diseases caused by mutations in the PTEN gene. This syndrome primarily affects women, causing skin rashes, tiny wart-like bumps, thyroid disease, and severe benign fibrocystic disease. By age 40, 50-75% of women with Cowden's syndrome develop breast or ovarian cancer. Cowden's syndrome can also be associated with kidney, Merkel cell skin and thyroid cancers [Wright and Whitney, 2006; 2005].

Other factors that increase ovarian cancer risk include age greater than 50 years; 50% of all ovarian cancers are found in women over the age of 63. Studies have indicated a relationship between the number of menstrual cycles in a woman's lifetime and her risk of developing ovarian cancer. Nulliparity (no pregnancies) can also contribute to the risk of developing ovarian cancer. Some studies have shown that the use of fertility drugs also increases the risk of ovarian cancer, such as the prolonged use of the fertility drug clomiphene citrate, especially without achieving pregnancy, may increase the risk for developing ovarian tumours, particularly a type known as low malignant potential or borderline tumours. However, results have not been consistent [Ayhan et al., 2004]. Furthermore, Ashkenazi Jewish heritage (by virtue of an increased frequency of BRCA1 and BRCA2 gene mutations) or European heritage also increases ovarian cancer risk. As some studies have shown that white women are much more likely to have ovarian cancer than African American women, although this may due to better detection in more developed countries [Ness et al., 2000]. In addition, some studies suggest that the use of estrogen replacement therapy may promote ovarian cancer in women who have been through menopause [Moorman et al., 2005]. The risk among women who used ERT for longer than 10 years was almost double that of women who had never used it and the risk among those who used it for 20 years or more was tripled. For normal healthy women the average lifetime risk for developing ovarian cancer is about 2% [Moorman et al., 2005].

In contrast, factors that inhibit ovulation seem to protect against the development of ovarian cancer. This may be because ovulation disrupts the epithelial layer of the ovary. As cells divide to repair the damage, mutations may occur, increasing the risk of developing cancer (the "incessant ovulation theory") [Purdie et al., 2003; Holschneider and Berek, 2000]. For example, higher numbers of full-term pregnancies (>37 weeks gestation) also significantly reduces the risk of ovarian cancer as does the use of oral contraceptives. This may be through a combination of ovulation suppression and induction of apoptosis by progestagens [Holschneider and Berek, 2000]. Furthermore, surgery involving tubal ligation to prevent pregnancy, or having a hysterectomy, also lowers the risk of ovarian cancer as does breastfeeding [Tung et al., 2005]. Finally, the removal of the ovaries before cancer occurs reduces the risk to zero. This may be an option for women with a strong family history of the disease, a known mutation in BRCA1 or 2, or in women over 45 years old undergoing abdominal surgery for other reasons.

1.1.3 Staging of ovarian cancers

All ovarian cancers are classified according to the terms of the staging scheme developed by the International Federation of Gynecology and Obstetrics (FIGO system) and the classification system developed by the American Joint Committee on Cancer (AJCC, TNM system), which indicate likely prognosis and help to define treatment (Table 1.1). Once an ovarian cancer is assigned a stage, the classification does not change, even if the cancer recurs or metastasizes to other sites within the body. Ovarian cancer treatment ultimately depends upon such staging. In general, the lower the stage, the more favourable is the prognosis.

Stage	Criteria
I	Confined to one (IA) or both (IB) ovaries. The tumour may be on the surface of the ovaries, the tumour may have ruptured, or malignant cells are found in peritoneal fluid or washings (IC).
II	Found outside the ovary and has spread to the uterus or fallopian tubes (IIA) or other areas in the pelvis (IIB). A stage II tumour may involve the capsule of the ovary, or peritoneal fluid or washings contain malignant cells (IIC).
III	Spread to abdominal organs and/or lymph nodes. Microscopic deposits of tumour are on abdominal peritoneal surfaces (IIIA), or small (<2cm) implants of tumour on abdominal peritoneal surfaces (IIIB). Abdominal implants may be larger (>2cm) or pelvic or retroperitoneal abdominal lymph nodes may be involved (IIIC).
IV	Spread outside the abdominal cavity (e.g. malignant cells are found in the fluid surrounding the lungs), or cancer has spread within the intra-abdominal organs (e.g. liver, spleen)

Table 1.1 Staging in ovarian cancer according to the International Federation of Gynaecology and Obstetrics (FIGO) staging system.

Ovarian cancers are typically diagnosed at a late stage. This probably reflects the absence of major symptoms in early stage disease, due to the anatomic position of the ovaries, which results in minimal interference with surrounding structures until the ovarian enlargement is considerable, or metastatic disease supervenes. When symptoms do occur, they are frequently nonspecific, often requiring multiple consultations with a primary care physician before further investigation is prompted. Thus, the prognosis is typically poor [Rosenthal et al., 2006].

There are three main forms of treatment for ovarian cancer which include surgery to remove cancerous tissue, chemotherapy to destroy cancer cells using strong anti-cancer drugs and radiotherapy to destroy cancer cells by high-energy radiation exposure. There are also many combinations of these treatment methods. The success of the treatment depends upon a number of factors (e.g., stage and grade of the disease, the histopathologic type, and the patient's age and overall health). Surgery usually is required to treat ovarian cancer. Most patients undergo surgery in addition to another form of treatment (e.g., chemotherapy and/or radiotherapy). Surgery helps the physician to accurately stage the tumour, make a diagnosis, and perform debulking (removal of as much tumour mass as possible). Debulking surgery is especially important in ovarian cancer because aggressive removal of cancerous tissue is associated with improved survival. Once ovarian cancer is confirmed, a total hysterectomy (removal of the uterus), bilateral salpingo-oophorectomy (removal of the fallopian tubes and ovaries on both sides), omentectomy (removal of the fatty tissue that covers the bowels), lymphadenectomy (removal of one or more lymph nodes) may be performed. Tissue removed during debulking is sent for histopathological examination. Patients with no residual tumour mass or tumour masses that measure less than 1 cm have the best survival rate. Modified ("conservative") surgery that leaves tumour-free reproductive organs intact may be conducted in women who still wish to still have children if (a) the tumour is confined (usually not serous or endometriotic in type, which tend to be bilateral tumours), and (b) wedge biopsy of the opposite ovary shows no evidence for disease involvement. Such a procedure carries an increased risk of relapse, therefore, total hysterectomy and salpingo-oophorectomy is recommended immediately after childbearing is complete.

1.1.4 The rationale for ovarian cancer screening & novel biomarker discovery

In developed countries, ovarian cancer remains a highly lethal disease. The American Cancer Society estimates that about 22,430 new cases of ovarian cancer were diagnosed in the United States during 2007. In the UK nearly 7,000 cases of ovarian cancer are diagnosed resulting in more than 4,400 deaths each year. Ovarian cancer accounts for about 3% of all cancers in women. The ovarian cancer incidence rate has increased by about 0.5% per year since 1975 (Figure 1.3). Ovarian cancer is predominantly a disease of older, post-menopausal women with almost 85% of cases being diagnosed in women over 50 years. There is a steep increase in incidence after the usual age of the menopause [Breedlove and Busenhart, 2005]. As a result of the advances in surgical management and chemotherapeutic options over the last few decades, the medium term survival for ovarian cancer patients has improved. However, overall long-term survival has not been significantly improved. Poor survival rates are mainly attributable to late diagnosis of the disease as most suffers of ovarian cancer do not show specific symptoms until the later stages of disease [Rosenthal et al., 2006].

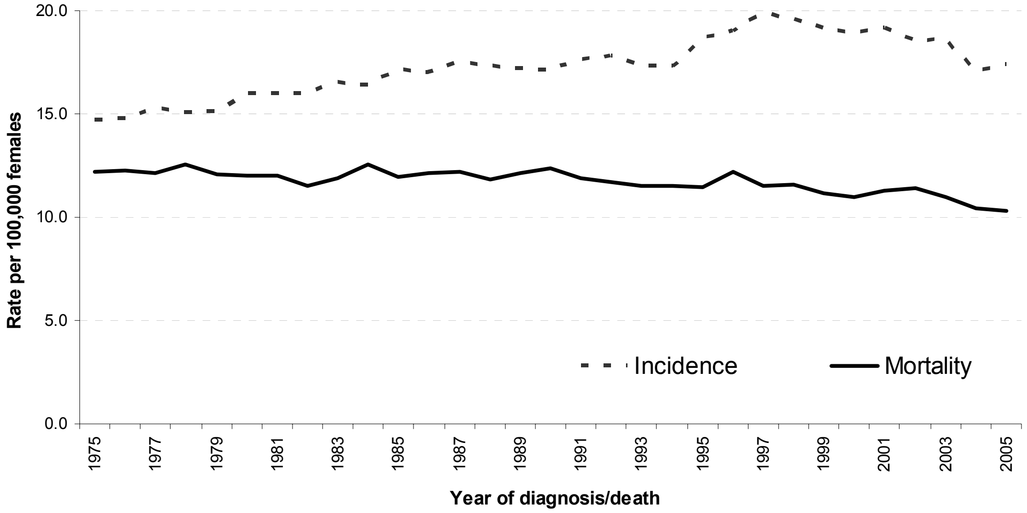


Figure 1.3 Age-standardised (to only include women over the age of 50) ovarian cancer incidence and mortality rates, Great Britain (1975-2005).

(Reproduced with permission from Cancer Research UK, May 2008
http://info.cancerresearchuk.org/cancerstats/types/ovary/incidence/)

The overall 5-year survival rate for ovarian cancer is 15% to 30%, whereas the 5-year survival rate of women with stage I at the time of diagnosis can be as high as 95%. However, there is no recommended screening method for women considered to be at low risk of ovarian cancer due to a lack of evidence of a long-term survival benefit, the lack of accurate and sensitive markers and the risks of false positive screening results i.e. unnecessary anxiety and surgery [Breedlove and Busenhart, 2005].

Current screening techniques such as transvaginal sonography and the serum cancer antigen 125 (CA-125) assay are only recommended for women with known strong risk factors. However, both transvaginal sonography and serum CA-125 are currently of unproven diagnostic use. CA-125, the most widely used biomarker for ovarian cancer detection, is a celomic epithelium–related glycoprotein protein that is secreted into the bloodstream by ovarian cells. A CA-125 test result of greater than 35 U/ml is generally accepted as being elevated. The CA-125 test has an 80% chance of returning true positive results from stage II, III, and IV ovarian cancer patients. The other 20% of ovarian cancer patients do not show any increase in CA-125 concentrations. The CA-125 test only returns true positive results for about 50% of stage I ovarian cancer patients. Thus, the CA-125 test is not an adequate early detection tool when used alone [Bosse et al., 2006]. CA-125 is also produced by other mesothelium-derived tissues (e.g. the peritoneum) and consequently may be elevated in many benign gynaecologic diseases and other types of cancer, leading to false positive results. For example, 70% of people with cirrhosis, 60% of people with pancreatic cancer and 20%-25% of people with other malignancies have elevated levels of CA-125. The CA-125 test also has a lower specificity in pre-menopausal women than post-menopausal women [Bosse et al., 2006]. Furthermore, because ultrasound cannot determine the histology of any mass detected, an ovary that looks suspicious on ultrasound may need to be removed surgically in order to exclude the diagnosis of cancer, creating a burden on the health service.

Although a large randomised trial of ovarian cancer screening in the general population is already underway to evaluate the use of an algorithm incorporating rate of change of CA-125 over time, to increase sensitivity and specificity, there is still only preliminary evidence that such a screening technique will reduce the mortality from ovarian cancer [Jacobs et al., 1999; Rosenthal et al., 2006].

There is thus an urgent need to find new biomarkers of ovarian cancer amenable to mass screening with high sensitivity and specificity for early-stage ovarian cancer detection and diagnosis that would enable early diagnosis, so that surgical therapy can be offered to all patients rather than a select few. This would ultimately decrease the morbidity and mortality rates from this disease [Bast, Jr. et al., 1998; 2002]. Ideally, for large-scale screening biomarkers would be detectable in the blood, facilitating the development of relatively non-invasive collection and assays.

1.2 Human Serum

Human serum is the clear yellowish fluid obtained upon separating whole blood into its solid and liquid components after it has been allowed to clot. Serum is a complex bodily fluid that contains approximately 60 to 80 mg of protein per mL in addition to various small molecules including salts, lipids, amino acids and sugars. The major protein constituents of serum include albumin (ALB), immunoglobulins (IgG, IgA), transferrin (TF), antitrypsin (SERPIN), haptoglobin (HP), complement proteins and lipoproteins (APO). Twenty-two of these abundant proteins make up 99% of the total protein content (Figure 1.4). It is estimated that 1000s of relatively low abundances proteins and peptides may be commonly present in serum [Fusaro and Stone, 2003]. In addition to the major protein constituents, serum contains any other proteins that are actively synthesized and secreted, or shed from cells and tissues throughout the body. As serum constantly perfuses tissues in their microenvironment, it potentially holds an archive of histological information. Therefore, the background matrix of serum represents a complex milieu in which unique disease-specific biomarkers may be found in extremely low abundance.

However, while the easily obtainable nature and the high protein content of serum deem it a valuable specimen for biomarker determination, human serum is one of the most complex proteomes known and there are still numerous hurdles to overcome when analysing it. For example, albumin is the most abundant protein in serum and may be 10 or more orders of magnitude higher in concentration than the scarcest of proteins [Fusaro and Stone, 2003]. In addition, many serum proteins have similar molecular weight and overall charge, making protein separation difficult. Therefore, biomarkers for disease at low concentrations in serum may be hidden by more abundant proteins with similar biophysical characteristics. As such, the reliable proteomic characterisation of serum and identification of biomarkers could be dramatically improved by reducing the complexity of the serum proteome through additional fractionation.

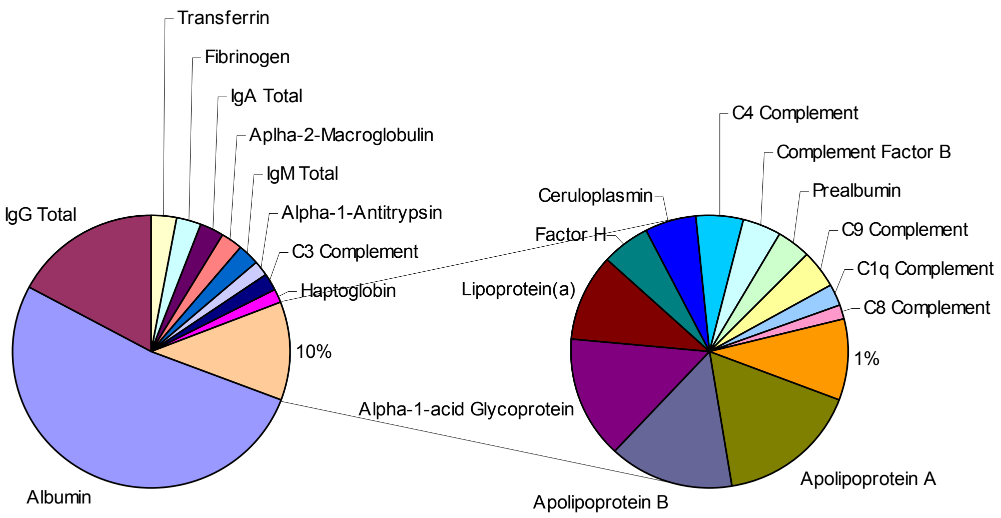


Figure 1.4 The relative contribution of proteins within serum. It is hypothesized that putative biomarkers will be found in the 1% portion of serum proteins [Fusaro and Stone, 2003].

1.2.1 Serum protein biomarkers

A biomarker can be broadly defined as any characteristic that can be objectively measured and evaluated as an indicator of normal biological or pathological processes. Serum protein biomarkers are produced by tissues or tumours. When detected in higher or lower amounts in blood, they can be suggestive of the presence of a tumour.

Tests based on biomarkers have been around for more than half a century, but interest in their application for diagnostics and for clinical screening has increased remarkably since the beginning of the 21st century [Baker, 2005]. Biomarkers have the potential to have a tremendous impact in clinical oncology by facilitating the identification of individuals at risk for developing cancer, assisting in the preclinical detection of cancer and ultimately allowing real-time monitoring of therapeutic responses. Several serological markers are already routinely used for a number of cancers (Table 1.2).

One example of a serum biomarker is cancer antigen 125 (CA-125). As previously mentioned, CA-125 is used as a biomarker for measuring the risk of ovarian cancer or as an indicator of malignancy. However, evidence suggests CA-125 lacks the sensitivity and specificity for general screening as it can be elevated in other malignant cancers, including those originating in the endometrium, fallopian tubes, lungs, breast and gastrointestinal tract. CA-125 may also be elevated in a number of relatively benign conditions, such as endometriosis, several diseases of the ovary, and pregnancy. Hence, there is a need to find putative markers for ovarian cancer which can be used in combination with CA-125 levels to offer non-invasive screening which is robust, highly sensitive and disease-specific. Examples of existing serum markers for ovarian cancer include carcinoembryonic antigen, ovarian cystadenocarcinoma antigen, lipid-associated sialic acid, NB/70K, TAG 72.3 as well as CA-125.

Biomarker	Cancer type	Specificity	Example of non-cancer pathology	Primary clinical use
α-fetoprotein	Hepatocellular, non-seminomatous testicular	Moderate	Prostatitis	Staging
Human chorionic gonadotropin-β	Testicular, ovarian	Low	Pregnancy	Staging
CA-15-3	Breast	Poor	Cirrhosis, benign diseases of ovaries and breast	Disease monitoring
CA19-9	Gastro, pancreatic, stomach	Poor	Gastritis	Disease monitoring
CA-125	Ovarian, cervical, uterine, fallopian tube	Moderate	Pancreatitis, kidney or liver disease	Disease monitoring
CA27-29	Breast	Not known	Not known	Disease monitoring
CEA	Colorectal, pancreas, lung, breast, medullary thyroid	Low	Non-malignant disorders	Disease monitoring
Epidermal growth factor receptor	Colon, non-small cell lung cancer	Low	Non-malignant disorders e.g. benign prostatic hyperplasia	Selection of therapy
Her2/Neu	Breast, ovarian	Moderate	Benign breast disease	Disease monitoring; selection of therapy
PSA	Prostate	High	Benign prostatic hyperplasia	Screening; disease monitoring
Thyroglobulin	Thyroid	Poor	Grave's disease thyroiditis	Disease monitoring

CA: Cancer anitgen; CEA Carcinoembryonic antigen; PSA: Prostate-specific antigen

Table 1.2 Common serum cancer markers used in primary care. Adapted from aoui-Jamali & Xu (2006).

1.2.2 Serum proteomics in cancer biomarker discovery

In recent years there has been immense interest in applying proteomics to the development of new serological biomarker platforms for early diagnosis of chronic diseases such as cancer. Proteomics can be defined as the qualitative and quantitative comparison of proteomes (PROTEin complement of a genOME) under different conditions to further unravel biological processes. Since at least the 1950s there has been support for the idea that plasma/serum protein patterns might provide important insight into the presence and activity of disease. Undeniably, extracellular fluids such as serum represent a major link among all cells, tissues and organs of an organism and contain a complex collection of peptides, proteins and protein fragments that are produced in the entire body. Thus, the analysis of human serum has great potential for novel putative biomarker discovery. Indeed assays to measure > 100 different proteins in blood have been developed and are in routine use in clinical chemistry laboratories today. The fibrinogen functional turbidimetric assay is an example of a blood-based assay which has a diagnostic value in pathology-disseminated intravascular coagulation and in assessing risk for atherothrombosis [Stief, 2008].

Conventionally, differential two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) was the mainstay of proteomic biomarker discovery [Anderson and Anderson, 2002; Anderson et al., 2004]. However, as a result of improvements in the sensitivity and accuracy of mass spectrometry (MS), proteomics has become increasingly popular for the analysis of complex protein samples such as human plasma and serum. This has led to the identification of more than 1500 different gene products in the serum/plasma of healthy donors [Villanueva et al., 2006]. In addition to proteomic studies looking for biomarkers of cancer in serum [Villanueva et al., 2006; Zimmerman et al., 2005; Zhang et al., 2004], several cancer specific proteins have been identified in the urine of patients with visceral cancers, such as lung [Tantipaiboonwong et al., 2005], ovarian cancer [Chambers and Vanderhyden, 2006] and breast cancer [Roy et al., 2004]. These studies support the view that bodily fluids like plasma, serum and urine have the potential to be a valuable source of diagnostic and prognostic markers of disease [Mor et al., 2005; Liotta and Petricoin, 2006].

It has been hypothesised that proteolysis activity within the tissue microenvironment generates protein fragments that passively diffuse into the circulation. Diagnostic peptides can also be generated ex vivo by circulating enzymes derived from the diseased tissue microenvironment acting on exodogenously derived protein fragments produced by during the clotting process as shown in Figure 1.5 [Villanueva et al., 2006].

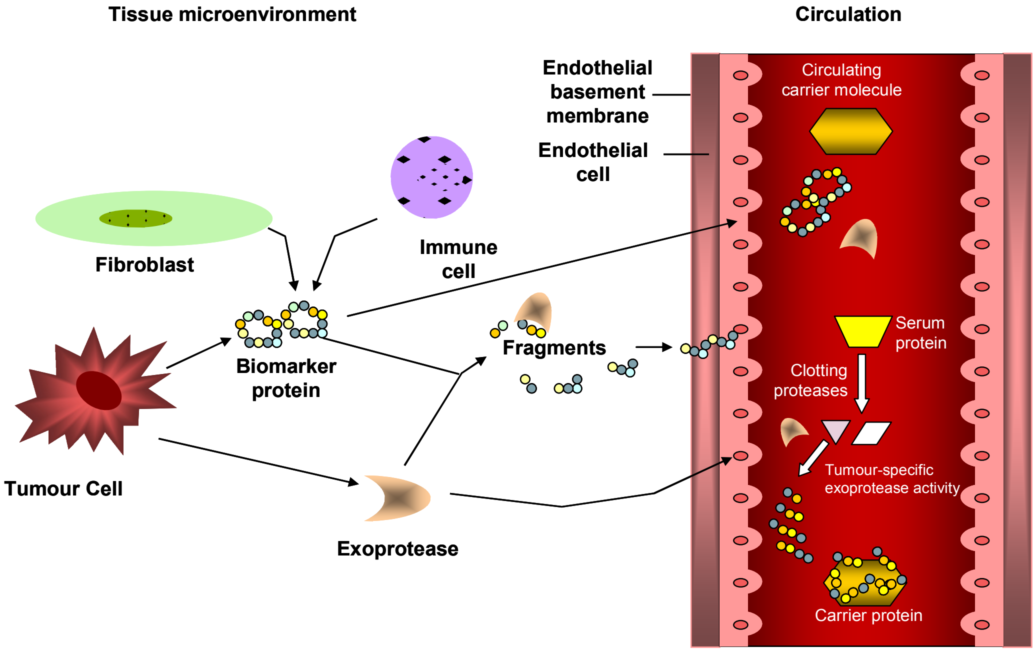


Figure 1.5 Proteases generate surrogate biomarker fragments. Circulating proteins generated in the diseased tissue microenvironment may serve as diagnostic protein markers. [Villanueva et al., 2006].

Numerous groups have identified a number of putative cancer biomarkers using mass spectrometry-based proteomics tools which could be potentially useful for diagnosis (Table 1.3) [aoui-Jamali and Xu, 2006]. This may have the added benefit of increasing our understanding of the pathways involved in the initiation and progression of cancer as well as for identifying key cancer biomarkers [Ransohoff, 2005; Robbins et al., 2005].

Biomarker	Cancer type	References
Apolipoprotein A1	Ovarian, pancreatic	Zhang et al., 2004; Kozak et al., 2005
Haptoglobin α-subunit	Ovarian, pancreatic, lung	Ye et al., 2003
Transthyretin fragment	Ovarian	Kozak et al., 2005
Inter-alpha-trypsin inhibitor fragment	Ovarian, pancreatic	Zhang et al., 2004
Vitamin D-binding protein	Prostate, breast	Corder et al.,1993; Pawlik et al., 2006
Serum amyloid A	Nasopharyngeal, pancreatic, ovarian	Orchekowski et al., 2005; Moshkovskii et al.,2005; Helleman et al., 2007
α1-antitrypsin and α1-antichymotrypsin	Pancreatic	Orchekowski et al., 2005: Yu et al., 2005
Haemoglobin-alpha & -beta subunits	Ovarian	Woong-Shick et al., 2005
EPCA-2	Prostate	Leman et al,. 2007
Afamin	Ovarian	Jackson et al., 2007

Table 1.3 Examples of putative serum biomarkers. Adapted from aoui-Jamali et al. (2006).

The use of mass spectrometry for the direct analysis of proteins and peptides from biological fluids, i.e. human serum, for putative disease biomarker discovery was first reported in 2002. Using peak pattern discrimination several groups reported on the correct classification of (a) ovarian cancer [Petricoin et al., 2002] (b) prostate cancer [Adam et al., 2002; Qu et al., 2002] and (c) breast cancer [Li et al., 2002]. As such, proposals for a blood test-based on MS pattern-recognition of human serum proteins for detecting cancer were put to the U.S. Food and Drug Administration. Subsequent to this commercial laboratories planned to market a blood test for ovarian cancer in late 2003 or early 2004 [Pan et al., 2005; Petricoin and Liotta, 2002; Petricoin et al., 2002; Petricoin, III et al., 2002; Rodland, 2004; Villanueva et al., 2006; Yu et al., 2003; Zhang et al., 2004; Zimmerman et al., 2005]. However, questions were raised about whether the technology's results were reproducible and reliable enough for application in practice [Baggerly et al., 2004; Baggerly et al., 2005]. Important limitations were found in the design of serum proteomics analysis by MS. Bias and chance (or overfitting of data) were considered as probable explanations of the misinterpretation of data. Hence, plans for the blood test were delayed by the U.S. Food and Drug Administration.

While serum proteomics may offer a non-invasive method for population-based screening programmes, the technology has several draw backs and is highly sensitive to pre- and post-analytical variations. Bias is encountered from inherent properties of the samples studied. Many intrinsic specimen characteristics may have no relationship with the disease in question, but are introduced by any number of factors dependent on logistics of sample collection or selection of subjects included in the study [Drake et al., 2004; Timms et al., 2007; Villanueva et al., 2005]. Strict adherence to standard operating protocols of specimen collection and appropriate matching of case versus control subjects are essential to minimise such sources of error. Collection of case and control specimens by the same site(s) is preferred as each collection site will unavoidably introduce its own bias into how samples are collected [Villanueva et al., 2005; Rai et al., 2005; Drake et al., 2004; Timms et al., 2007; West-Nielsen et al., 2005]. The controversy sparked by the reports in 2002 stimulated improvements in many areas of serum proteomics, from sample collection and serum preparation to the development of new bioinformatics tools to analyse and compare large numbers of data points that are typical of mass spectrometry.

However, another major drawback to MS-based serum profiling is the limitation of this technology to effectively analyse highly complex protein mixtures. Detection of ionized molecules using time-of-flight platforms is inversely related to molecular size such that peptides and small proteins are more readily detected by MS. This places a significant limitation on MS proteomic profiling. Blood protein levels range from nearly millimolar down to femtomolar concentrations. Typical cancer biomarkers are found in the pico- to subnanomolar range. Although MS is highly sensitive, its application to serum profiling has demonstrated the ability to identify proteins at low to submicromolar concentrations. As a result of the complexity of the serum proteome and the limited detection range of MS-based platforms several pre-fractionation techniques have become popular. Another drawback to using MS profiling of the serum proteome is the difficulty of standardising and calibrating instrumentation across multiple sites in order to directly compare findings from different laboratories.

Furthermore, most proteomic studies published to date have identified relatively abundant host response (acute-phase) proteins as candidate biomarkers [Fung et al., 2005]. Host response proteins, e.g. haptoglobin, serum amyloid A, α-1-antitrypsin, α-1-antichymotrypsin, inter-α-trypsin inhibitor and the apolipoproteins, are often dismissed because of an apparent lack of specificity. However, it is hypothesized that the peptide/protein changes comprising the diagnostic patterns in MS based analysis are derived (directly or indirectly) from the molecular state of the tumour–host microenvironment. The proteomic pattern that originates from this microenvironment may signal the presence of an early-stage lesion. Under this hypothesis, the discriminatory markers are likely to be metabolic products, enzymatic fragments, modified proteins, peptides, or cytokines that could be highly specific for the microenvironment of the lesion [Liotta and Petricoin, 2006].

There are several examples of host response proteins which have been identified as markers for ovarian cancer. Recently, three 'host-response' proteins, apolipoprotein A1 (down-regulated in cancer), a truncated form of transthyretin (down-regulated) and a cleavage fragment of inter-α-trypsin inhibitor heavy chain H4 (up-regulated) were also identified as putative markers for ovarian cancer using a proteomics based approach [Zhang et al., 2004]. Furthermore, apolipoprotein A1 (Apo A1), transferrin (TF) and transthyretin (TTR) are also reported as a panel of markers for ovarian cancer [Nossov et al., 2008]. However, attempts to independently validate these proteins were unsuccessful. Furthermore, these proteins may not be specific for ovarian cancer.

1.3 Sample fractionation and protein separation methods

To overcome the problems posed by the large dynamic range of the serum proteome and the interference of the abundant proteins, sample fractionation has become a prerequisite for MS-based serum profiling. Extensive fractionation is thought to improve protein coverage, but adds to the cost of throughput and affects method reproducibility.

1.3.1 Magnetic bead-based peptide extraction

As a result of the complexity of serum and the presence of salts a polypeptide extraction and desalting step is almost always necessary prior to MS-based protein profiling. Methods such as magnetic bead extraction have been combined with MALDI-TOF MS for mass spectral profiling of serum peptides and proteins, often using automated extraction to improve sample throughput and reproducibility [Villanueva et al., 2004; Villanueva, 2006]. Fractionation simplifies complex samples and separates peptides and proteins from non-protein species hence removing contaminants and improving the detection limits for serum peptide ions. The high sensitivity of modern mass spectrometers, combined with advanced bioinformatics makes this technique ideally suited for proteome profiling and protein identification. Other available techniques for serum profiling include Surface-enhanced laser desorption/ionization (SELDI) MS. SELDI is a variation of matrix-assisted laser desorption/ionization (MALDI) MS that uses a target modified to achieve biochemical affinity with the analyte compound. In MALDI-MS, a protein or peptide sample is mixed with the matrix molecule in solution and small amounts of the mixture are deposited on a surface and allowed to co-crystallize as the solvent evaporates. While in SELDI-MS the protein mixture is spotted on a surface modified with a chemical functionality. Binding to the SELDI surface acts as a separation step and the subset of proteins that bind to the surface are easier to analyse. Common surfaces include CM10 (weak-positive ion exchange), H50 (hydrophobic surface, similar to C6-C12 reverse phase columns), IMAC30 (metal-binding surface), and Q10 (strong anion exchanger). Surfaces can also be functionalized with antibodies, other proteins, or DNA.

Within the published literature there is an overall lack of consensus over the optimal method for serum peptide extraction. Different papers detail the use of different types of stationary phases and co-ordinating ligands and employ different conditions for polypeptide binding, washing and elution. The lack of agreement between published protocols, combined with the widespread reports of low inter-laboratory reproducibility highlights the need for method development of a generalised protocol targeted at mass spectrometry [Baumann et al., 2005; de Noo et al., 2005; Martorella and Robbins, 2007; Villanueva et al., 2004].

The most commonly used resins for serum extraction are reverse phase (RP) beads that have modified alkyl groups and can be made of a polymer shell with an iron core. These beads are superparamagnetic, which means that the beads exhibit magnetic properties in a magnetic field, with no residual magnetism once removed (Figure 1.6). The beads separate gently and no columns or centrifugation steps are necessary. They are spherical in shape and have defined surface chemistry minimising chemical agglutination and non-specific binding. This allows uniformity (co-efficient of variance (CV) <3%) of size, shape and surface area provides optimal accessibility and reaction kinetics, for rapid and efficient binding, batch-to-batch consistency (typically within 5%) improving the reproducibility across different runs. A typical MS-based serum profiling workflow is shown in Figure 1.7.

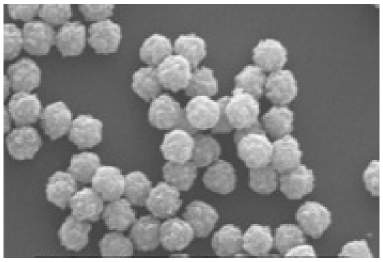


Figure 1.6 Superparamagnetic beads used for serum extraction

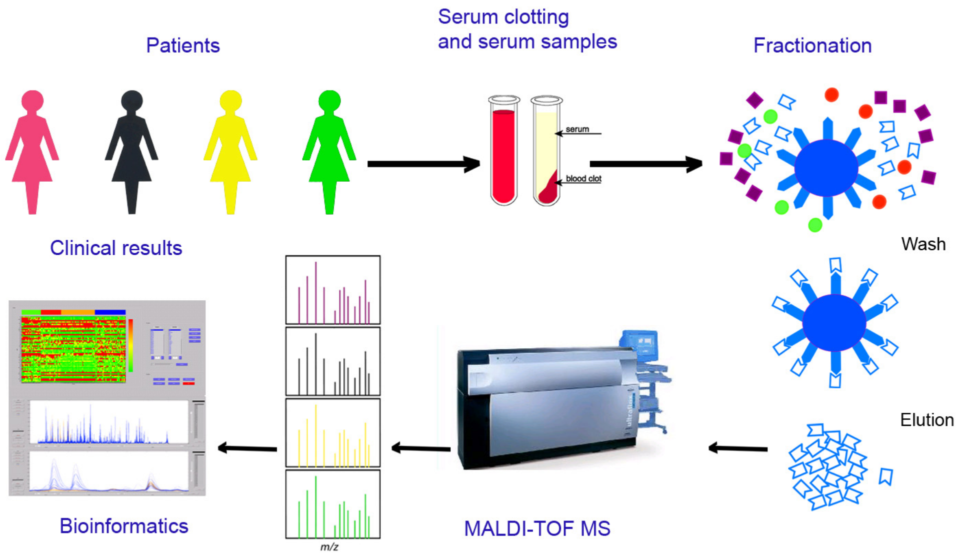


Figure 1.7 Basic principles of proteomics-based serum profiling. Blood samples from volunteers are collected and processed using a standardised protocol. Samples are then fractionated to extract peptide and proteins for analysis by mass spectrometry coupled with bioinformatics tools to mine for differentially expressed peaks between cases and controls.

1.3.2 Serum Depletion

Albumin constitutes anywhere from 55% to 75% of the total protein content of human serum and consequently, is an overwhelming signal in separation and detection assays. Even following albumin removal, serum still contains other high-abundance proteins, the most abundant being IgG, IgA, transferrin, haptoglobin, fibrinogen and antitrypsin. Collectively, these seven proteins constitute ~90% of the total protein in serum. Therefore, their removal represents a fundamental improvement toward characterisation of the lower abundant serum proteins. Classically, Cibacron Blue and protein A/G chromatography methods have been used to deplete serum of albumin and the immunoglobulins. However, an increasing number of methods for the removal of high-abundance proteins from serum are becoming commercially available, making serum analysis a more routine laboratory procedure. The work presented in this thesis has involved the use of two recently commercialised enrichment strategies including the Multiple Affinity Removal System (MARS, Agilent) and the ProteoMiner protein enrichment kit (BioRad).

1.3.3 The Multiple Affinity Removal System (MARS)

MARS consists of a reusable high-capacity affinity liquid chromatography column containing polyclonal antibodies for the removal of the top seven abundant proteins from human serum. It is designed to bind and remove 85-90% of albumin, IgG, transferrin, haptoglobin, IgA, antitrypsin & fibrinogen which constitutes 90% of the total protein amount in serum. Thus, it facilitates the downstream expression profiling of lower abundant protein species in the flow through (Figure 1.8).

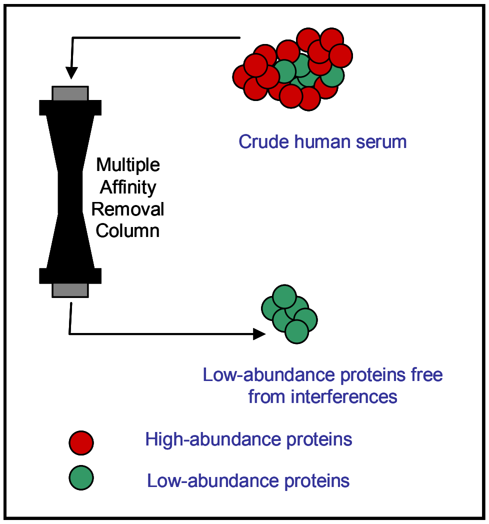

Crude human serum
Multiple Affinity Removal Column
Low-abundance proteins free from interferences
High-abundance proteins
Low-abundance proteins

Figure 1.8 Schematic illustration of the MARS column. The “top-seven” abundant proteins are captured by affinity binding to antibodies and removed from serum. This facilitates the analysis of the lower abundant protein fraction for putative biomarker discovery.

1.3.4 ProteoMiner technology

The ProteoMiner technology is a sample preparation tool used for the compression of the dynamic range of protein concentrations. It is based on treatment of complex protein samples with a large, highly diverse library of hexa-peptides bound to chromatographic supports in a spin column. In theory, the library contains binding sites for all protein sequences in the sample. Since the bead capacity limits binding capacity, high-abundance proteins quickly saturate their binding sites and excess protein is washed out during the procedure. In contrast, low-abundance proteins are concentrated on their specific ligands, thereby decreasing the dynamic range of protein expression in the sample. When analysed in downstream applications, the number of proteins detected dramatically increases (Figure 1.9) [Guerrier et al. 2006; Guerrier et al. 2008; Boschetti et al. 2008].

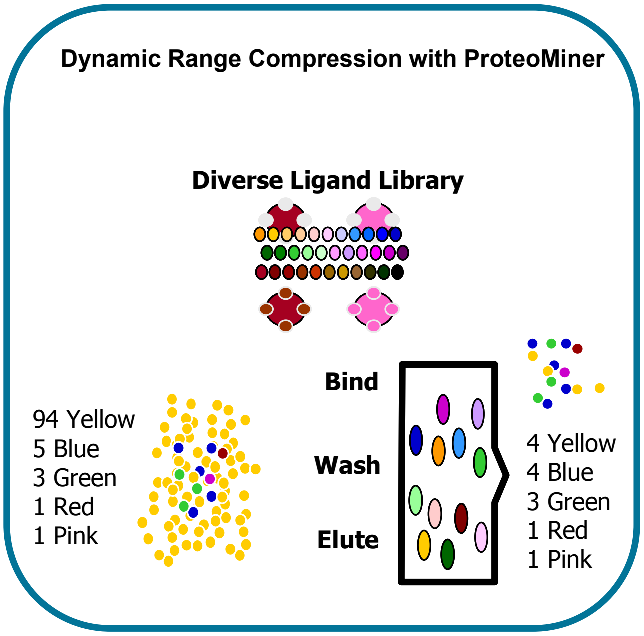


Figure 1.9 Schematic illustration of the ProteoMiner protein enrichment strategy. An 'equalised' amount of all serum proteins are captured by the combinatorial peptide ligands.

1.3.5 High performance two-dimensional gel electrophoresis

In 1975 O'Farrell, Klose, and Scheele almost simultaneously published methods based on isoelectrical focusing (IEF) of proteins in the first dimension and SDS-polyacrylamide gel electrophoresis (SDS-PAGE) in the second. This marked the introduction of two-dimensional electrophoresis (2-DE) for proteins separation and the beginning of the proteomics era [Scheele, 1975; O'Farrell, 1975; Klose, 1975].

In general, 2-DE sorts proteins in two dimensions based on protein charge and molecular weight. In the first dimension, proteins are separated by IEF in a pH gradient, where proteins become focused at their isoelectric points (pI) when they reach zero net charge [Righetti, 1989; Righetti et al., 1988]. The three dimensional configuration of the proteins does not play a role as the protein is assumed to be completely denatured because of the chaotropic chemicals used in the solubilisation buffer. Post-translational modifications (PTMs), such as phosphorylation or glycosylation may influence the net charge of a protein, and can be visualised as spot trains on the gel. Isoelectric focusing is in principle an end point method.

Furthermore, a major development to overcome the problems of pH gradient instability and irreproducibility was the introduction of immobilized pH gradients (IPG) for IEF [Bjellqvist et al., 1982]. IPGs are based on the principle that the pH gradient is generated by a limited number (6-8) of well-defined chemicals (the ‘Immobilines’) which are co-polymerized with the acrylamide matrix. Thus cathodic drift is eliminated, reproducibility enhanced and pattern matching and inter-laboratory comparisons were simplified. IPGs allow the generation of pH gradients of any desired range (broad, narrow or ultra-narrow) between pH 3 and 12. Since the sample loading capacity of IPG-IEF is also higher than with CA-IEF, especially in combination with narrow (1 pH unit) or ultra-narrow (0.1 pH unit) IPGs, 2D-PAGE with IPGs is the method of choice for micropreparative separation and spot identification.

Following IEF, IPG strips are equilibrated for the second dimension. This treatment has three functions; reduction and alkylation of disulfide bonds, acetylation and SDS treatment. In order to maintain solubilisation in the second dimension, disulphide bonds are once more reduced with dithiothreitol (DTT) according to reaction (1) and alkylated with iodoacetamide according to reaction (2). Alkylation prevents the formation of new disulfide bonds.

$$R\text{-}CH_2\text{-}S\text{-}S\text{-}CH_2\text{-}R + C_4\text{-}H_{10}O_2S_2 \rightarrow 2\ R\text{-}CH_2\text{-}SH + C_4H_8O_2S_2\ (1)$$

$$R\text{-}CH_2\text{-}SH + ICH_2CONH_2 \rightarrow R\text{-}CH_2\text{-}S\text{-}CH_2\text{-}CO\text{-}NH_2 + HI\ (2)$$

In the second dimension, proteins are separated according to their relative molecular weight (Mw) by conventional SDS-PAGE. The detergent SDS binds to the proteins at a ratio of about one SDS molecule per two amino acid residues in such a way that all proteins have the same net negative charge density and thus migrate in an electrical field according to their relative molecular mass. The strips are transferred to second dimension polyacrylamide gels. An electric field is applied and the proteins migrate towards the anode due to their negative charge and the sieving effects of cross-linked gels.

The 2D electrophoretic mobility of a protein is reasonably specific thus allowing accurate comparison of protein amounts in different samples analysed on distinct gels. Usually each spot on the resolving 2D gel corresponds to a single protein species of the sample, however, in certain cases more than one protein can be found in a single spot on a gel. This happens when proteins present in the same spot have the same pI and Mw. 2-DE allows separation of thousands of different proteins as well as providing protein information such as the protein pI, Mw, PTM and the amount of each protein. At present, there are no other techniques that are capable of simultaneously resolving thousands of proteins in one separation procedure. However, there are some drawbacks of 2-DE (i.e. poor resolution of high and low molecular weight proteins and hydrophobic and basic proteins, low gel-to-gel reproducibility in different experimental runs; it is also a labour intensive and expensive technique) that have limited its application in some proteomics studies.

An important step made in the 2-DE strategy for protein study was the introduction of protein labelling and detection in order to further define quantitative and qualitative profiles of complex protein samples. The sensitivity and specificity of protein stains has always been the major factor influencing the amount of information that can be extracted from 2D gels. In addition, the most significant breakthrough in proteomics has been the mass spectrometric identification of gel-separated proteins, which has extended analyses beyond the mere display of proteins. Mass spectrometry is very sensitive, can deal with a mixture of proteins and is amenable to high-throughput operation. In the last decade, the sensitivity of analysis and accuracy of results for protein identification by MS have increased by several orders of magnitude, and nowadays it is estimated that proteins in the femtomolar range can be identified from complex samples if appropriate fractionation strategies are applied.

Several reviews have outlined the different methods of choice for detection of gel separated proteins [Patton, 2000; Rabilloud, 2002]. Most strategies for protein detection in 2D gels use post-electrophoretic protein staining and a multitude of different methods have been described which differ in their sensitivity, specificity, linear dynamic range and compatibility with downstream identification mainly by mass spectrometry. Common post-electrophoretic protein stains employed for protein detection before MS include: i) Colloidal Coomassie Blue G-250, which has a linear detection range of 100ng-10μg, is simple to use and is compatible with MS [Neuhoff et al., 1988]; ii) Silver stains have a higher sensitive range of detection, typically 2-4 ng of protein per spot. Some silver-staining methods are incompatible with MS, because the aldehyde-based cross-linkers used in the sensitisation steps can cross-link proteins, and because the silver ions can interfere with mass spectrometric data analysis. Silver staining is only linear over a small dynamic range and can also stain differently based on the protein post-translational modifications and amino acid composition. This makes silver staining a poor choice for quantitation of protein expression. iii) fluorescent stains, such as the ruthenium-based fluorescent dyes SYPRO Ruby SYPRO Orange and Deep purple (Molecular Probes). SYPRO Ruby is a transition metal organic complex that binds directly to proteins by electrostatic interactions. It provides sensitivity similar to that of classical silver staining (1-2 ng of protein/spot) but without the complex methodology, limitation on linear dynamic range, or the problems with MS compatibility. However, fluorescent dyes are expensive and require fluorescent scanners for analysis. v) Phosphoprotein stains are becoming increasingly important because of the growing interest in studying phosphorylation, an important post-translational modification that influences and determines the function of proteins. Pro-Q Diamond (Molecular Probes, Eugene, OR, USA) has been introduced as a fluorescence detection method for gel separated phosphoproteins with a detection limit of 1-2 ng. However, it was shown that this dye is not highly specific and it can label non phosphorylated proteins. Several reviews have outlined the different methods of choice for detection of gel separated proteins [Schulenberg et al., 2003; Steinberg et al., 2003].

1.3.6 Two-dimensional Difference Gel Electrophoresis (2D-DIGE)

Fluorescence two-dimensional difference gel electrophoresis (2D-DIGE) is a 2-DE gel-based proteomics technique that provides a sensitive, rapid and quantitative analysis of differential protein expression between two or more biological samples. Developed by Unlu et al. in 1997, the technique utilizes charge- and mass- matched chemical derivatives of spectrally distinct fluorescent cyanine dyes which are used to covalently label lysine residues in different samples prior to mixing and separating on the same 2-DE gel [Unlu et al., 1997]. In this way, the labelled samples would be subjected to identical electrophoretic conditions to generate directly superimposable images for relative quantification. Cyanine dyes were first described by Mujumdar et al. [Mujumdar et al., 1989; Mujumdar et al., 1993]. These fluorophors have a structure which can be modified to create a panel of reactive fluorescent tags. Unlu et al., (1997) developed the N-hydroxy-succinimidyl (NHS)-ester derivatives of the fluorescent cyanines 3 and 5 (NHS-propyl-Cy3 and NHS-methyl-Cy5). These dyes and a third cyanine dye, NHS-Cy2, are now commercially available from GE Healthcare (Figure 1.10). All 3 dyes possess a single net charge of +1, thus maintaining the charge of the lysine residue which they modify.

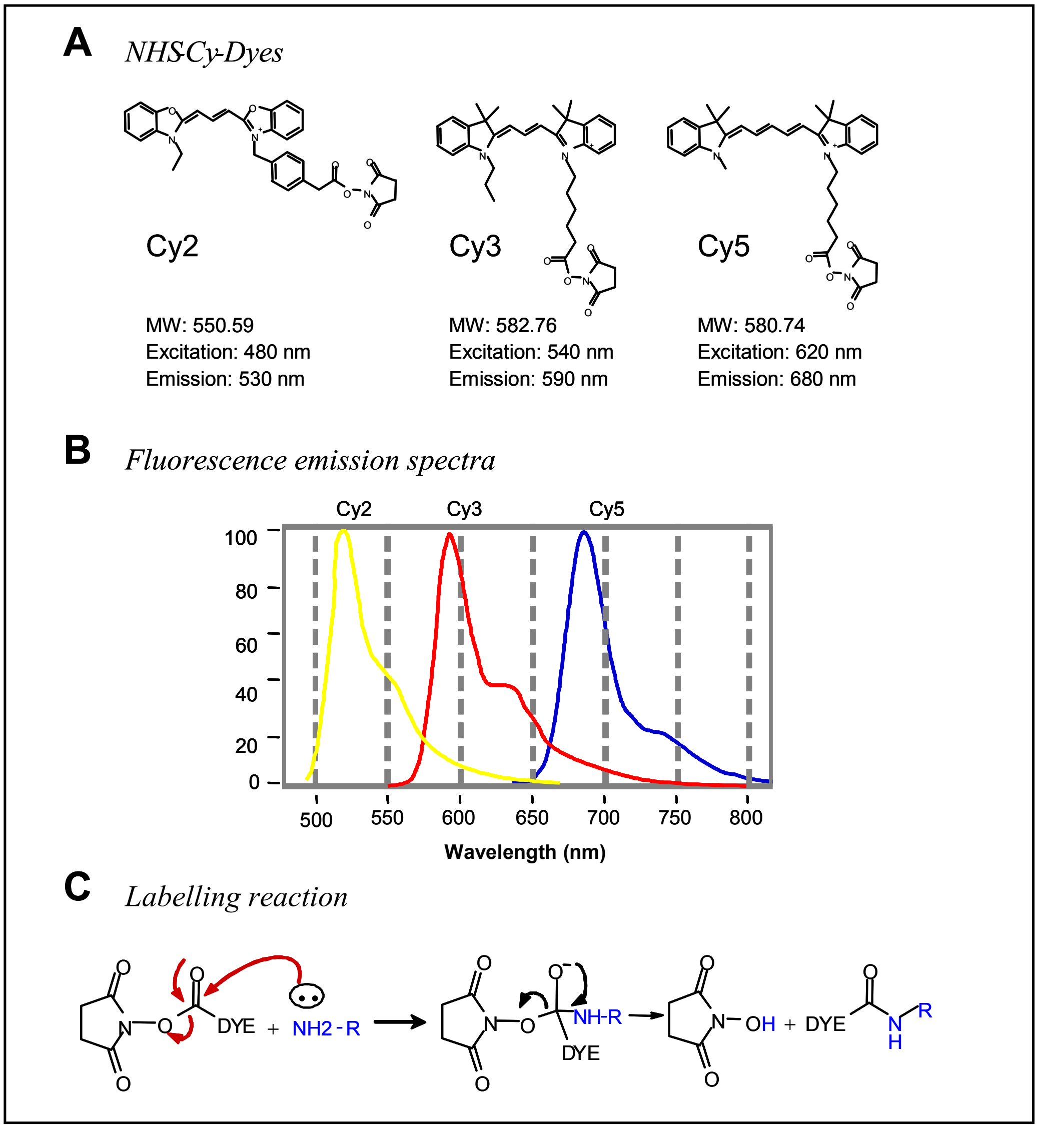


Figure 1.10 Characteristics of the NHS-Cy-dyes. A) Structure of the NHS-Cyanine dyes. Cy2, 3-(4-carboxymethyl-phenylmethyl)-3'-ethyloxacarbocyanine halide N-hydroxy-succinimidyl ester; Cy3, 1-(5-carboxypentyl)-1'-propylindocarbocyanine halide N-hydroxysuccinimidyl ester; Cy5; 1-(5-carboxypentyl)-1'-methylindodicarbocyanine halide N-hydroxysuccinimidyl. Each dye has a similar molecular weight and single positive charge matching the charge of the modified primary amino group. B) Each dye displays distinct emission spectra enabling the individual detection of differentially labelled proteins at the appropriate wavelength without overlap of signals. C) The dyes have an N-hydroxysuccinimidyl ester reactive group triggering covalent interaction with the primary amine groups of lysine residues or the N-terminus.

Initially, the NHS Cy3 and Cy5 dyes were used to label different protein samples prior to mixing and running them on the same 2-DE [Unlu et al., 1997]. This allowed the samples to run under identical electrophoretic conditions in a type of differential display format. Theoretically, to compare the same proteins derived from two differently labelled samples, the dyes should be mass and charge matched and the dye modifications should not perturb the electrophoretic mobility of labelled proteins. For this reason, the size of the aliphatic chain (Figure 1.10A) was originally modulated to maintain a similar molecular weight between each dye and the dyes possess a positive charge which matches the positively charged amino groups they modify.

The advantage of using lysine labelling is that almost all proteins contain at least one lysine residue and it contains a reactive amino group. For expression profiling the dyes are typically used under conditions of minimal stoichiometrical labelling. Ideally, just a single lysine residue is labelled in around 5% of the molecules of a particular protein. This helps to keep protein soluble and limits the shift in Mw of the labelled versus unlabelled population of proteins during SDS-PAGE run, whilst keeping the sensitivity of detection high. The reported sensitivity of DIGE labelling is ~1ng protein per spot. The method of using the three Cy dyes was originally evaluated and applied by Tonge et al. and Gharbi et al. [Tonge et al., 2001; Gharbi et al., 2002] and further optimised and commercialised by GE Healthcare.

For quantification, one of the dyes (usually Cy2) is used to label an internal standard sample which is run on all gels and usually comprises an equal pool of proteins from all samples under investigation [Gharbi et al., 2002]. Thus, the Cy2 labelled pool is used for normalisation of data across gels, thereby reducing experimental variation and increasing the accuracy of quantitation and statistical confidence of protein expression differences. Since fluorescence detection also provides a superior linear dynamic range of detection and sensitivity compared to visible staining methods [Patton, 2000], this technology is suited to the analysis of biological samples with their large dynamic ranges of protein abundance. As shown in Figure 1.11 this labelling strategy is also compatible with downstream identification of gel spots by mass spectrometry (MS) [Tonge et al., 2001; Gharbi et al., 2002].

For protein expression profiling, gels are converted to digital images using scanning devices and these are processed to detect the protein features. Spot volumes are quantified and spot patterns matched across different gels. Statistical methods are then employed to detect protein spots with statistically significant changes in expression. This kind of image analysis is usually performed with dedicated software programmes.

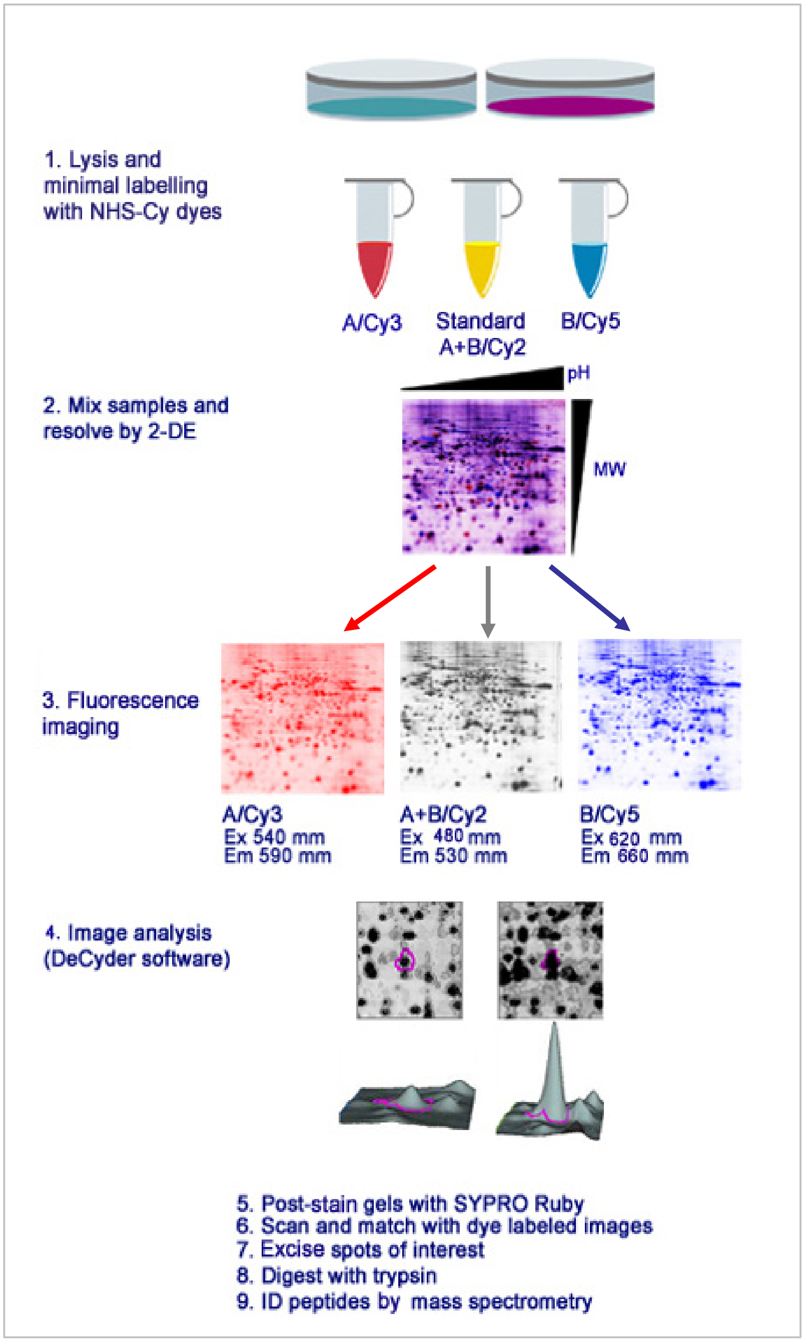

1. Lysis and minimal labelling with NHS-Cy dyes
A/Cy3
Standard A+B/Cy2
B/Cy5
pH
2. Mix samples and resolve by 2-DE
MW
3. Fluorescence imaging
A/Cy3
Ex 540 mm
Em 590 mm
A+B/Cy2
Ex 480 mm
Em 530 mm
B/Cy5
Ex 620 mm
Em 660 mm
4. Image analysis (DeCyder software)
5. Post-stain gels with SYPRO Ruby
6. Scan and match with dye labeled images
7. Excise spots of interest
8. Digest with trypsin
9. ID peptides by mass spectrometry

Figure 1.11 Schematic representation of 2D-DIGE protocol for minimal lysine labelling and using an internal standard for normalization.

1.4 Mass Spectrometry

1.4.1 Biological Mass Spectrometry

Mass spectrometry (MS) is a powerful analytical technique used for the accurate measurement of the mass-to-charge ratio (m/z) of molecules. The development of the first mass spectrometer is attributed to J.J. Thomson, who at the beginning of the 20th century measured the m/z ratios of several atoms and small molecules. In the first half of the 20th century, developments in ionisation methods and analysers occurred with the parallel application of mass spectrometry in the field of organic chemistry for the elucidation of chemical structures. It was not until the beginning of the 1990's however, that the field of biological mass spectrometry became significant. This was due to the introduction of soft ionisation methods, e.g. Matrix-Assisted-Laser-Desorption/Ionisation (MALDI) by Tanaka, and Karas and Hillenkamp, and Electron Spray Ionisation (ESI) by Fenn that allowed for the ionisation of macromolecules such as proteins and peptides [Lin et al., 2003; Zhang et al., 2004]. In recognition of the development of ESI and for the development of soft laser desorption (SLD) Fenn and Tanaka received the Nobel Prize for Chemistry in 2002.

A mass spectrometer can be defined as an instrument capable of measuring the mass-to-charge ratio of molecules. Mass spectrometers generally couple three devices, namely i) an ionisation device, ii) a mass analyser and iii) a detector. In addition, sample inlet and data output recorders are needed, but they are not part of the mass spectrometer as such. There are many different kinds of mass spectrometers described generally by the types of ionisation sources, mass analysers, and detectors that are used. In all MS methods, analyte molecules must be converted into gaseous ions using an ionisation source. The most commonly used ionisation sources for biological molecules are MALDI. The other most used ionisation source is ESI. Other ionisation methods include Fast Atom Bombardment (FAB), Chemical Ionisation (CI), Thermal Ionisation (TIMS), Secondary Ionisation (SIMS) and Plasma Desorption (PD).

An electric or magnetic field can deflect charged particles, and since the kinetic energy imparted by motion through an electric field gives the particles an inertia dependent on the particle's mass, the mass analyser can use this to steer certain ions to a detector based on their m/z ratio by varying the electrical or magnetic field. It can be used to select a narrow mass range (i.e. to select peptides of interest for tandem mass spectrometry (MS/MS)) or to scan through a range of masses to catalogue the ions present (survey scan). Examples of mass analysers are quadrupole mass analysers, time-of-flight (TOF), ion trap (IT), ion cyclotron resonance (ICR), orbitrap and magnetic sector instruments. There are numerous combinations of mass analysers in so called hybrid instruments. The first three are the most commonly used analysers in biological mass spectrometry.

There are several ways to detect ions. Routinely these are recorded when an ion hits a detector plate such as Multi Channel Plates, or MCP. As ions hit the plate a cascade of electrons is released, amplifying the single ion detection. This flow is called image current and can be detected and amplified. When a scan is conducted in the mass analyser, the charge induced in the detector during the course of the scan will produce a mass spectrum, a record of the m/z values at which ions are present and their intensities.

1.4.2 Ionisation methods

Biological mass spectrometry has been and is being developed at a rapid pace since the development of the soft ionisation techniques MALDI and ESI.

1.4.2.1 Matrix Assisted Laser Desorption Ionisation (MALDI)

MALDI was first introduced by Karas & Hillenkamp and Tanaka in 1988 as a 'soft' ionisation method with which relatively large macromolecules could be ionised and analysed in the gaseous phase [Karas and Hillenkamp, 1988; Hoffmann, 2002]. In this ionisation method the sample is mixed with an excess of matrix molecules and allowed to crystallise and then a laser is used to excite and ionize analytes from the solid to the gas phase. The matrices used in MALDI are typically acidic compounds (e.g. carboxylic acids) with an absorption in the region of the laser wavelength. The most commonly used matrices for protein/peptide analysis are 2,5-dihydroxybenzoic acid (DHB) and α-cyano-4-hydroxycinnamic acid (α-CCA) (Figure 1.12).

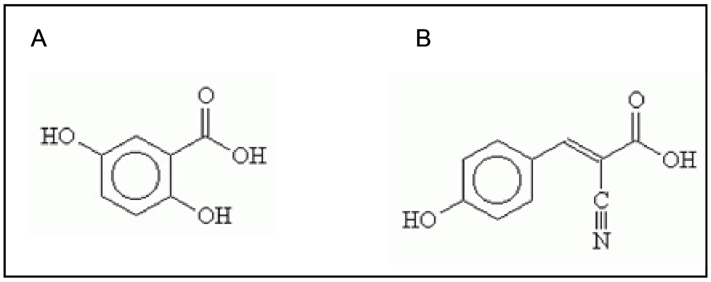


Figure 1.12 Examples of commonly used MALDI matrices. Two commonly used matrices used for protein/peptide analysis are A) 2,5-dihydroxybenzoic acid (DHB) and B) α-cyano-4-hydroxycinnamic acid (α-CCA). DHB forms a crystal rim and 'hot spots' of crystallised peptides, and is relatively salt tolerant, α-CCA forms semi-homogenous spots which makes it more amenable for automated spectral acquisition.

1.4.2.1.1 Principles of the MALDI process

(i) The Formation of a 'Solid Sample'

The analyte sample is mixed with a suitable matrix compound at a 1-10 times molar excess and allowed to co-crystallise with the evaporation of the solvent. The number of matrix molecules exceeds those of the analyte, separating its molecules and thereby preventing the formation of sample clusters, which inhibit the appearance of molecular ions. The incorporation of the sample molecules into the lattice structure of the matrix is a pre-condition of the functioning of the laser desorption/ionisation process. The matrix serves to minimise sample damage from the laser pulse by absorbing most of the incident energy and increases the efficiency of energy transfer from the laser to the analyte. As such, the sensitivity is increased.

(ii) Matrix Excitation

This step involves ablation of portions of the solid solution by pulses of laser energy for a short duration. Some of the laser energy incident on the co-crystallised sample is absorbed by the matrix, causing rapid vibrational excitation, bringing about localised disintegration of the solid solution forming clusters made up of a single analyte molecule surrounded by neutral and excited matrix molecules. The matrix molecules evaporate away from these clusters to leave the excited analyte molecule.

(iii) Analyte Ionisation

The analyte molecules become ionised by simple protonation by the photo-excited matrix, leading to the formation of the typical $[M+X]^+$ type species (where X= H, Li, Na, K, etc.). Some multiply charged species, dimers and trimers can also be formed. Negative ions are formed from reactions involving deprotonation of the analyte by the matrix to form $[M-H]^-$ and from interactions with photoelectrons to form the $[M]-^{\cdot}$ radical molecular ions [Dreisewerd, 2003]. These ionisation reactions occur in the first tens of nanoseconds after irradiation, and within the initial desorbing matrix/analyte plume. These ions are then accelerate through an electrical field toward a mass analyser. It is important to note that these principles are hypothetical since some aspects of the MALDI process are not yet fully understood (Figure 1.13).

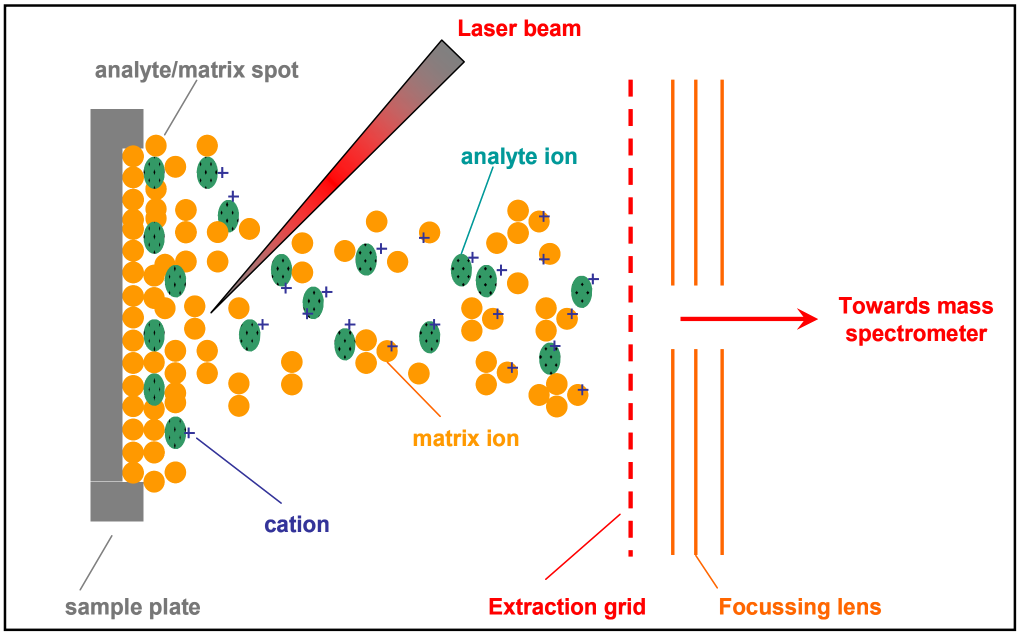


Figure 1.13 The MALDI process. Analyte molecules are co-crystallised with an excess of matrix molecules. A hypothesis to account for ion formation by MALDI is that irradiation of these crystals with a laser beam desorbs matrix-analyte ion clusters, which undergo gas phase reactions. As a result such clusters dissociate to leave free analyte ions and matrix ions. These ions are then accelerated through an electrical field toward a mass analyser.

1.4.2.2 Electrospray Ionisation

In contrast to MALDI, ESI ionises analytes from a liquid phase. The analyte is dissolved in an organic solvent mixture, typically methanol or acetonitrile, containing a small concentration of acid (e.g. formic acid 0.1-1%). The introduction of the sample into the mass spectrometer can be carried out by a number of methods. In the simplest case, sample is directly infused through a syringe and a narrow transfer capillary. Another example is the so-called nano-spray sample delivery method. In this system, a small amount of sample is placed into a needle, which has a very small tapered opening on one side. Sample is forced out because of capillary forces and high voltage is applied to the needle. The most commonly used method is the coupling of the electrospray directly with reverse phase chromatography. In this setup the capillary end of the chromatographic system is connected to the needle to which the voltage is applied. Typical flow rates of 200 to 500 nL min^{-1} (nanospray) are used for the chromatography. A fine spray of charged droplets emerges from the capillary and is directed into the vacuum chamber of the mass spectrometer through a small orifice. An electrostatic field is formed between the capillary and the walls of the mass spectrometer, and as the droplets travel they evaporate resulting in the formation of gas-phase ions. The magnitude of the charge-repulsion effect becomes more significant, and at a certain charge/solvent composition (termed the Rayleigh limit), Coulomb explosion of the analyte-solvent clusters occurs. The clusters become smaller and more highly charged within the skimmer region until single molecular ions are formed either by further explosion of clusters or by desorption of molecular ions from the clusters. The charged ions are accelerated through the analyser towards the detector. These ions can then be analysed according to their mass-to-charge ratio (Figure1.14) [Mann and Wilm, 1995].

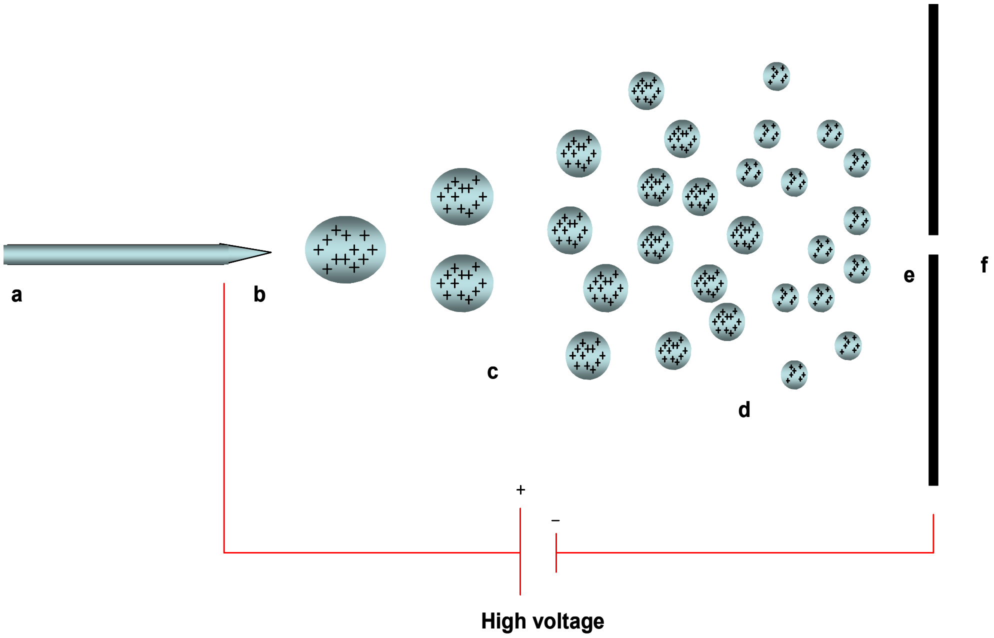


Figure 1.14 Schematic representation of electrospray ionisation. This figure shows the schematic representation of the ESI process. Sample is delivered through a capillary (a) and a tapered needle (b). Through a high voltage, droplets are extracted (c). Due to the evaporation of solvent, the charge repulsion reaches a critical value (d The Rayleigh limit), when droplets explode (Coulomb explosion), creating multiply charged ions. These enter the mass spectrometer (e) under the applied electrical field (f).

1.4.2 Mass analysers

1.4.2.1 Time-of-Flight (TOF)

Due to the pulsed nature of MALDI ionisation, it is most commonly used in combination with a time-of-flight (TOF) analyser. The TOF mass spectrometer was introduced commercially more than 40 years ago, yet only recently its high mass range and high sensitivity multichannel recording capabilities have been realised, making this type of spectrometer an attractive instrument in biological research. The TOF mass spectrometer is the simplest type of mass analyser and has a very high sensitivity over a virtually unlimited mass range (Figure 1.15). The sample ions are generated in the source zone (s), and are expelled in bundles that are produced by the laser desorption on the source-focusing lenses. A potential, (Vs - the source extraction) is applied across the source to extract ions which accelerate from the source into the field-free 'drift' zone of the instrument (d).

TOF is a measure of the duration of time required for ionised proteins and peptides to travel through the MS chamber to the detector plate. As Figure 1.15 illustrates, the fundamental principle that permits MS to separate analytes is the fact that small ions fly faster than larger ones. The ions' m/z ratios may be calculated from the time that each one requires to reach the detector plate. The flight time is proportional to the square root of the mass to charge ratio. Knowing the acceleration voltage and the length of the drift region the m/z ratio can be determined by measuring the flight time. The range of typical MALDI flight times is between a few µs and some 100µs. The drift regions are typically 1-4m long. Differences in TOF and thus m/z ratios allow the distinction and, in many cases, the identification of different proteins and peptides. Calibration of the instrument is performed with known reference masses.

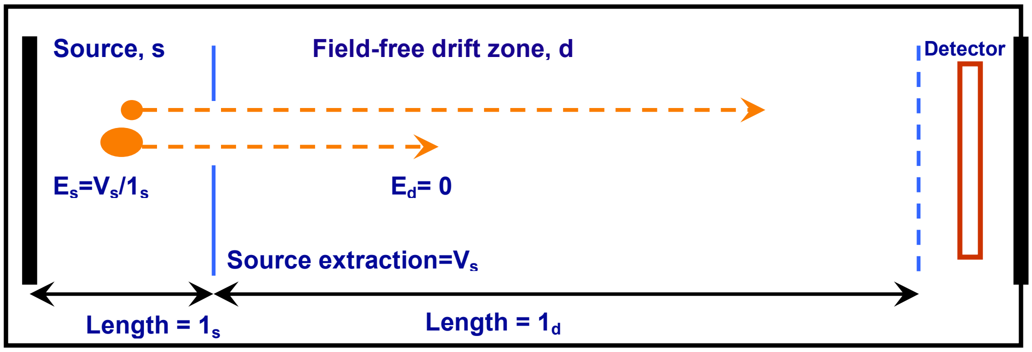


Figure 1.15 Schematic diagram of the process of time-of-flight mass spectrometry

Once the ions produced in the ion source have been separated by the mass analyser they are commonly detected using an electron multiplier detection device. The signal is sent to a computer, which records incoming signals and displays them graphically in a chromatograph or mass spectrum. Thus, the final product of this kind of analysis is a list of m/z ratios that represent a peptide mass map, also called a peptide mass fingerprint [Mann et al., 1993; Pappin et al., 1993].

1.4.2.2 Quadrupole analysers

ESI sources are typically coupled with ion traps and quadrupole analysers or hybrid instruments combining different analysers in tandem such as Q-TOF, triple Q, IT-Orbitrap and IT-FT-ICR. Protein identification of samples presented in this thesis were analysed with a Q-TOF instrument. The quadrupole analysers consist of four parallel metal rods. Each opposing rod pair is connected together and a radio frequency (RF) voltage is applied between each pair. A current voltage is then directly superimposed on the RF voltage. Ions travel down the quadrupole in between the rods. Only ions of a certain m/z are able to reach the detector for a given ratio of voltages, while other ions have unstable trajectories and will collide with the rods. This allows selection of particular ion, or scanning by varying the voltages.

A triple quadrupole mass spectrometer has a linear series of three quadrupoles. The first (Q1) and the third (Q3) quadrupoles act as mass filters, while the middle (Q2) quadrupole is employed as a collision cell. The collision cell is an RF only quadrupole and uses Ar or He gas to induce collisional dissociation of selected parent ions from Q1. Subsequent fragments (daughter or product ions) are passed through to Q3 or a TOF where they may be filtered or scanned fully, generating a collision-induced dissociation (CID) spectrum. Peptide fragmentation caused by collision mainly occurs at the lowest energy amide bonds of peptides. When the charge is retained by the amino terminal fragment, a, b, and c type ions are formed, while x, y and z type ions are formed when the charge is retained by the carboxy-terminal fragments (Figure 1.16). The mass difference between sequential b- and y- ions thus corresponds to the mass of the amino acids in the sequence. The nomenclature was proposed by Roepstroff et al. in 1984 (Figure 1.16; Roepstorff and Fohlman, 1984).

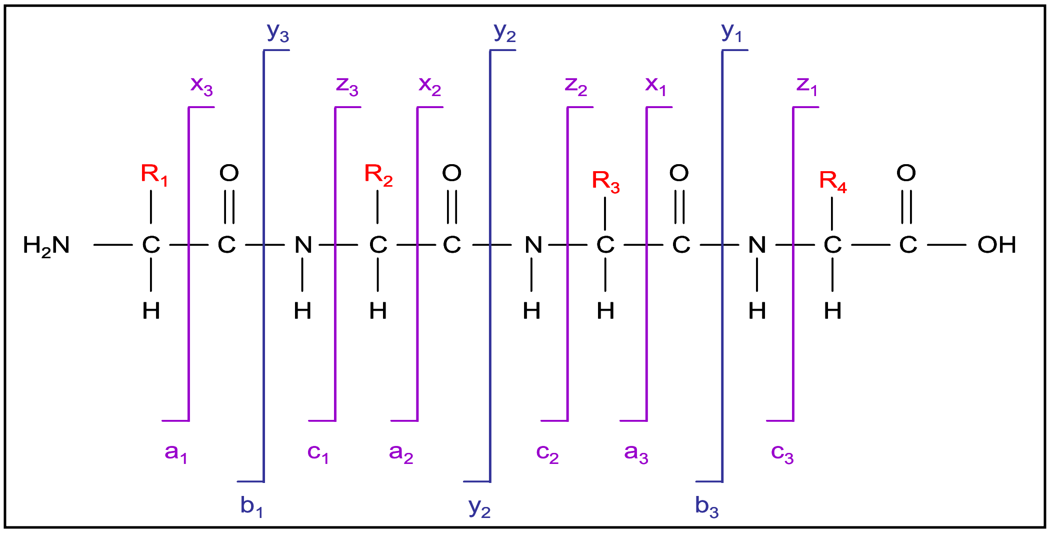


Figure 1.16 Schematic representation of peptide fragment ions nomenclature. The overview of the possible fragmentation on the peptide backbone and the nomenclature of the resulting fragment ions are shown. Different amino acids are distinguished by the side group displayed in red (R). The nomenclature was proposed by Roepstroff et al. in 1984 [Roepstorff and Fohlman, 1984]. When the charge is retained on the C-terminal side of the product ions, fragments are named x, y and z, while when the charge is retained on the N-terminal side a, b and c type ions are formed. The resulting ions are dependent on many factors, such as the type of fragmentation method used, analyser principle and primary sequence dependency. Also fragmentation of the side chains is possible. These ions are indicated as v and w ions (not shown in this figure).

1.5 Aims and scope

The primary aim of this study was to set up a separation and analysis scheme to rigorously investigate the serum proteome using mass spectrometry as a central technique. As discussed, the serum proteome contains an archive of histological information. Ovarian cancer is a lethal gynaecological malignancy that has an urgent need for new biomarkers. The serum of ovarian cancer patients has been investigated for several decades, yet, only a handful of putative markers have been identified. In this work several strategies have been employed to address the complexity of the human serum proteome and to identify putative biomarkers of ovarian cancer. Chapter three is concerned with the analysis of serum samples pre-dating diagnosis of ovarian cancer using a previously established magnetic bead-based and MALDI-TOF MS technology platform focussing on the serum peptidome. The hypothesis here is that tumour-specific exopeptidase activities generate surrogate markers during the blood clotting process.

Chapter four describes the establishment and optimisation of this technology platform in the host laboratory at UCL. The main focus of this chapter is the reproducibility of the technology platform. Chapter five describes the analysis of serum samples from the UKOPS collection to discovery statistically significant MS peaks which could discriminate case versus control samples. As a complementary approach Chapter six shows the analysis of these samples using MARS and ProteoMiner protein enrichment strategies coupled with 2D-DIGE based separation and identification of differentially expressed protein features followed by MALDI-TOF PMF and LC-MS/MS protein identification.

Chapter 2: Materials and Methods

2.1 Serum samples and sample collections

In this study serial serum samples collected from women later diagnosed with ovarian cancer and matched healthy controls from a UKCTOCS (United Kingdom Collaborative Trial for Ovarian Cancer Screening) pilot study were also used for proteomic analysis using the platform developed in the laboratory of Professor Paul Tempst at the Memorial Sloan Kettering Cancer Centre (MSKCC) in New York USA (Chapter 3) [Villanueva et al., 2004]. Volunteers were selected using criteria which ensured volunteers had no previous history of ovarian cancer upon recruitment and in total 22,000 women were recruited.

Serum samples were taken every year from 1996/1997 up to and including 2001. All cases and controls satisfied the following criteria: (a) age ≥ 50 years, and (b) > 12 months amenorrhoea following a natural or surgical menopause or > 12 months of hormone replacement therapy commenced for menopausal symptoms. The exclusion criteria were: (a) a history of bilateral oophorectomy, (b) active malignancy (women with a past history of non-ovarian cancer malignancy were eligible if they had no documented persistent or recurrent disease), (c) increased risk of ovarian cancer due to familial predisposition – exclusion criteria were entry criteria for the UK Familial Ovarian Cancer Screening Study and (d) a previous history of ovarian cancer. Information on reproductive history, family history of cancer and hormone replacement therapy use was obtained from all women prior to the start of screening. These volunteers underwent annual screening for 2-6 years.

From this sample collection, a subset consisting of 92 samples from 19 women, which pre-dated a diagnosis of ovarian or fallopian tube cancer from < 1 to 6 years was analysed at the MSKCC (Figure 2.1). The matched controls for this set of samples came from 183 healthy volunteers (from the general population) with no previous history of cancer and no incidence of cancer as far as known since recruitment onto the pilot study. The general procedure for selecting controls was based on using sera that reached the laboratory, and were processed on the same day and placed in the same sample freezer rack as the case sample. If it was not possible to find a match using this method then preceding samples were chosen. Matching was also done based on age (within 5 yrs) and HRT use. Each case sample had two matched controls except for one sample which only had one.

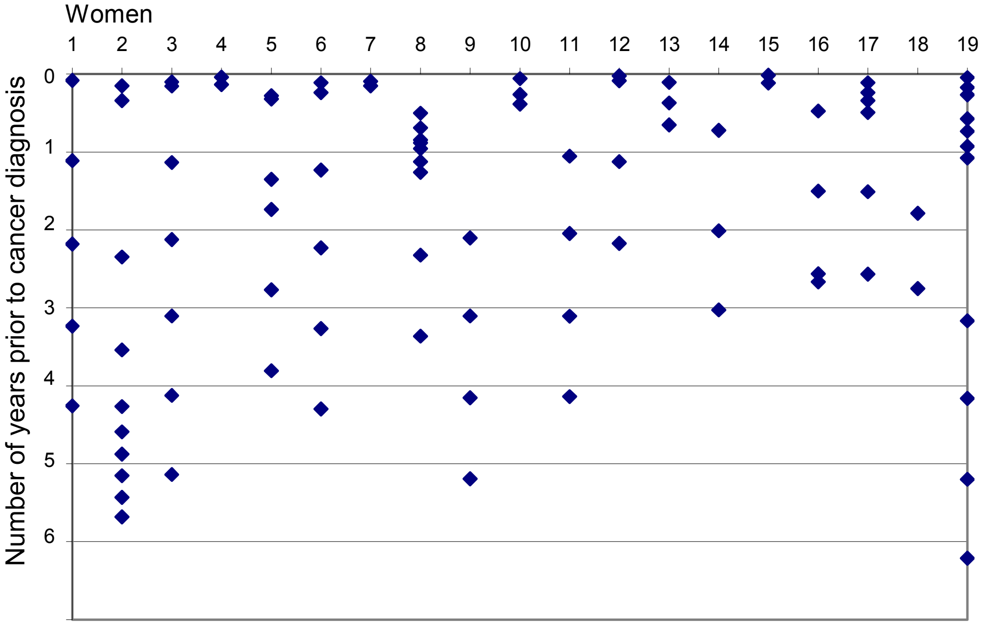


Figure 2.1 Scatter plot illustrating the timescale of serial serum samples collected from women predating a diagnosis of ovarian cancer. A total of 92 serial samples from 19 volunteers were collected and analysed using a MALDI-TOF MS-based proteomic profiling platform. '0' on the x-axis represents time of diagnosis.

Serum samples from the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) study were used for the protocol comparison study (Chapter 4). The UKCTOCS trial involves just over 200,000 apparently healthy postmenopausal women aged 50-74 years of age. All participants provide a serum sample at registration and 50,000 participants provide additional samples annually for 6 years. All participants are followed up using data from the Office of National Statistics which provides information on cancer diagnoses and cause of death in this cohort. In addition, all the participants are sent health questionnaires 3 and 7 years after recruitment to the study.

Serum samples from the United Kingdom Ovarian Cancer Population Study (UKOPS) collection were also used (Chapters 5 & 6). UKOPS is a multi-centre study set up by the Institute for Women's Health at University College London (UCL) that aims to predict which women in the population are at greatest risk of getting ovarian cancer. The study aims to recruit 1,000 women from the UK who have been diagnosed with ovarian cancer. This includes women undergoing surgery for possible ovarian cancer or benign tumours or who have had a previous diagnosis, and 2000 healthy women who act as controls for the study.

Finally, commercial serum (Human Sera S7023-1 Sigma-Aldrich) was used for the platform optimisation and as a quality control (QC) in all profiling experiments.

2.1.1 Sample collection and handling procedures

All UKOPS serum samples were collected according to a previously established protocol outlined by Villanueva et al. who used an optimised magnetic bead peptide extraction method coupled with MALDI-TOF MS to profile low mass serum peptides and proteins from healthy and diseased serum samples [Villanueva et al., 2004]. Venous blood was collected into BD Vacutainer SST tubes, (Becton Dickinson # 367988). The tube was gently inverted 5 times to mix clot activator with blood. The blood was then allowed to clot for 1 hour at room temp (RT) in a vertical position. The SST tubes were placed on wet ice in a vertical position prior to centrifugation at 2,000 x g for 10 min at RT. The serum (upper phase) was then aliquoted into Falcon tubes and frozen at -80 °C. Samples were then shipped on dry ice to the collection laboratory where they were thawed, aliquoted (500 µL) into bar-coded straws, heat sealed and stored at -80 °C.

All UKCTOCS samples used in this study were collected in Greiner gel tubes. These samples had been allowed to clot, centrifuged at room temperature (RT) then divided into aliquots in straws that were heat sealed and stored at -80°C. The time from venipuncture to centrifugation was 30 hours for each sample (protocol 1; Green; GN). Additionally, for the handling protocol comparison study, samples from the same 25 UKCTOCS volunteers were collected in Becton Dickinson red-top tubes, allowed to clot at RT for 60 minutes. These samples were then placed on wet ice for 2 hours before centrifugation. Following this samples were transferred to straws and stored at -80°C (protocol 2; Yellow; YE). A third protocol used a 5 minute clotting time at RT, followed by incubation on wet ice for 3 hours before centrifugation. This set of samples was also transferred to straws for storage at -80°C (protocol 3; Gray; GY). Three variants of protocol 3 were also prepared where samples were stored in cryovials at -80° instead of straws (protocol 4; Cryovial; CR); were placed on wet ice for 6 hours instead of 3 hours (protocol 5; Orange; OR); were incubated for 3 hours at RT instead of on wet ice (protocol 6; White; WH). These protocols were chosen to assess the effects of different transfer times, temperatures, clotting times and storage tubes on the serum proteome in the context of clinically feasible collection protocols.

From 1995-2001 the pre-UKCTOCS samples (Pilot study) were collected in Becton Dickinson red-top tubes. All samples were transported to the central laboratory by courier at ambient temperature. All samples had recorded transit times of less than 48 hours. Upon reaching the central laboratory, samples were immediately centrifuged and aliquoted into storage tubes with plastic push-on caps, which were stored at -20°C. In 2004 all samples were moved to a -80°C freezer. For the analysis performed at the MSKCC, a 100 µL aliquot of each of these samples was shipped on dry ice. Upon arrival at the MSKCC samples were thawed and further aliquoted (50 µL) and stored at -80°C. Samples were thawed before processing on the MSKCC automated platform and analysed by MALDI-TOF MS.

2.2 Automated magnetic bead-based serum peptide extraction

2.2.1 Preparation of calibrant mixture for automated runs

For instrument and spectral calibration, a calibrant mixture was spotted beside sample spots on the MALDI targets. The calibrant mixture was prepared fresh before each run. A standard peptide mixture was purchased from Bruker and resuspended in 125 µL of 0.1% TFA as per the manufacturer's instructions. 20 µL of this mixture was then diluted 1:3 with 50% (v/v) acetonitrile (ACN). Synthetic peptide 782 was prepared at a concentration of 2 µg/µL, then diluted 1:25 with 0.1% TFA followed by a 1:10 dilution with 50% (v/v) ACN. 20 µL of both the diluted standard peptide mixture and the synthetic peptide 782 were combined and diluted a further 1:5 in 50% (v/v) ACN and labelled 'peptide mix' (Table 2.1). Protein mixture 1 was also purchased from Bruker and resuspended as per the manufacturer's instructions. 10 µL of the resuspended protein mixture was diluted 1:1 in 50% (v/v) ACN and labelled 'protein mix' (Table 2.1). To prepare a master mix for 96 calibrant spots, 102 µL of peptide mix was mixed with 17 µL of protein mix and 17 µL of 50% (v/v) ACN was added for a final volume of 136 µL. All the preparations were kept on ice until they were placed in a cooled rack in the robot for mixing with pre-made α-cyano-4-hydroxycinnamic acid (α-CCA) MALDI matrix and automated spotting. The final amounts on target were 30 fmol of each peptide and 500 fmol of each protein.

Calibrant	m/z
Peptide mix	
Peptide 782	782.04
Angiotensin II	1,047.20
Angiotensin I	1,297.51
Substance P	1,348.66
Bombesin	1,620.88
ACTH fragment 1-17	2,094.46
ACTH fragment 18-39	2,466.73
Protein mix	
Insulin	5,734.56
Ubiquitin	8,565.89
Cytochrome C	12,361.09
Myoglobin	8,476.77

Table 2.1 Calibrants used for MALDI-TOF spectral calibration. The final amounts on target were 30 fmol for each peptide and 500 fmol for each protein. All m/z values are calculated for singly charged ions except myoglobin which was doubly charged.

2.2.2 Automated liquid handling protocol for bead-based serum peptide extraction

A Tecan Genesis Freedom liquid handling workstation with 1 mL syringes was used for all steps of automated serum peptide bead-based extraction and MALDI target spotting. The robot was set up during the morning of the day of the runs by cleaning the liquid handler, degassing the water used for the runs, ensuring the waste container had enough room to receive discarded liquids and ensuring that the fixed robot tips were clean and accurately aligned. The system was also flushed twice with degassed water to ensure there were no air bubbles in the system.

Serum sample plates were removed from the -80°C freezer and allowed to thaw at room temperature for approximately 10 min before each run. Samples were assigned randomised positions in the 96-well plates. C18 Dynabeads at 2 µg/µL stock were prepared by washing 320 µL of Dynabead slurry twice in 200 µL of deionised water and then resuspending in 320 µL of deionised water. This was done to remove ethanol from the stock solution.

The Dynabeads were gently mixed by aspiration and dispersal for 5 min until completely re-suspended and 40 µL was added to each tube of a strip of eight 0.2 mL thin-wall tubes and placed in a holder on the robot deck. Fresh 0.1% TFA was added to the TFA trough on the robot before each run. Next, a 96-well skirted microtiter plate was prepared with 75 µL of 50% (v/v) acetonitrile in each well of the first column and 95 µL of pre-prepared α-CCA matrix solution in each well of the second column. The third column contained 65 µL of calibrant mix in row A and 70 µL of matrix solution in row B. The wells of the plate were tightly then sealed with self-adhesive foil using a rubber roller. The plate was then attached to a cooler rack with a piece of Parafilm wrapped around the microtiter plate and the inner part of the cooler rack to ensure that the plate would remain in place when the robot tips pierced the foil. Adhesive tape was used to secure the cooling rack to the robot deck. The cooler rack had been chilled at -20 °C for several hours before use. A cleaned MALDI target plate was placed in its position on the robot deck.

The magnetic bead-based reversed-phase extraction protocol adopted for this project was previously outlined by Villanueva et al. [Villanueva et al., 2004]. However, some modifications of the published protocol were made before it was adopted at the host laboratory. These were mostly used to minimise bead loss and carry-over during the processing steps. These modifications are highlighted below.

First the magnetic beads in the thin-wall tubes were re-suspended by pipetting up and down 10 times. A measured volume of bead suspension (12.5 µL) was transferred to the well of a 96-well microtiter plate (Starlab 1402-9700) containing an aliquot of serum (50 µL). Magnetic beads and serum were then mixed by aspirating and dispensing 10 times, incubated for 2 minute to allow sufficient binding of serum polypeptides to the beads. Next, the beads were pulled to the side of each well using magnetic force (Jancox Metal Products Inc. REL033-01) and the supernatant was removed and discarded. Then 200 µL of washing solution (0.1 % trifluoroacetic acid, TFA) was added, the beads were pulled five times from left to right and back on the side magnets and then were allowed to settle on one side of the tube wall for 30 seconds. The washing solution (200 µL) was removed. This washing step was repeated once more, however during the second wash only 120 µL of washing solution was initially removed. At the conclusion of the washing step, beads were further re-suspended in the remaining 80 μL of washing solution and then pulled to the tip of the wells by magnets positioned beneath the plate. A further 60 μL of supernatant was removed, the sample plate was then moved to the side magnet and the remaining 20 μL of washing solution was removed. In the original protocol the removal of the remaining washing solution was performed while the beads were at the tip of the wells. However, in our laboratory, bead loss was noted at this step and so beads were therefore moved to the side magnet to prevent contact of the fixed tips with the magnetic beads.

For elution, 10 μL of elution solvent 50% (v/v) acetonitrile was added to the bead pellet. Beads were mixed with elution solution by moving the plate back and forth on the side magnets 5 times. The beads were then allowed to settle on one side of the tube and a 5 μL fraction of the eluate transferred to another well. In the original protocol elution solution was added while beads were at the tip of the wells and mixing was done with the fixed tips. However, bead carry-over and losses were noted. Thus, modifications were made to enable elution and mixing on the side magnet. Finally, 5 μL of pre-made α-CCA matrix solution (Agilent Technologies, UK) was added to the eluate and mixed, and 1 μL aliquots were deposited in replicates onto a stainless steel MALDI target in every other column of the 384-spot layout. Spots were allowed to dry at room temperature prior to MS profiling (Figure 2.2)

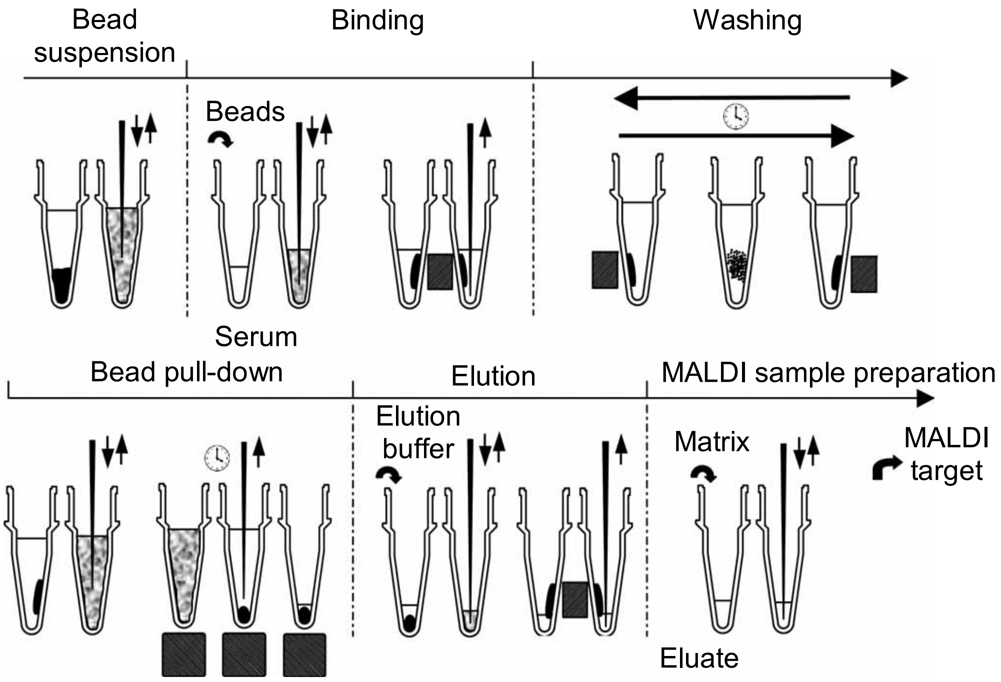


Figure 2.2 Schematic illustration of the magnetic bead-based extraction of serum peptides. Briefly, magnetic bead pellets are re-suspended, a measured volume of bead suspension is transferred to a tube containing an aliquot of serum and the magnetic beads and serum are mixed. Beads are pulled to the side by magnetic force and supernatant is removed and discarded. Washing solution is added, beads are pulled five times from left to right and back, beads are pulled to the side and washing solution removed. At the conclusion of the washing step, beads are further re-suspended, pulled to the tip of the tube by magnets positioned underneath, and the supernatant removed. An elution solvent is mixed with the bead pellet and beads are pulled to the side and a fraction of the eluate transferred to another tube. Matrix solution is added to the eluate and mixed and 1 µL is spotted onto a MALDI target.

2.2.3 MALDI-TOF MS serum profiling

2.2.3.1 Autoflex MALDI-TOF MS at MSKCC

Serum polypeptide profiles were generated in two mass ranges; low mass range (LMR) 700-4000Da m/z segment and high mass range (HMR) 4–15 kDa m/z segment using the 'AutoXecute' function of the software on an Autoflex MALDI-TOF/TOF mass spectrometer (Bruker Daltonics, Bremen, Germany). For spectral acquisition a sum of 400 laser shots, delivered in four sets of 100 shots (at 50 Hz) to each of four different locations on the surface of the matrix spot were acquired in linear mode geometry under 20 kV (18.6 kV during delayed extraction) of ion acceleration and -1.3 kV multiplier potentials, and with suppression of mass ions set to m/z <400. Delayed extraction was maintained for 80 ns (≤ 4 kDa) or 50 ns for (≥ 4 kDa) to give appropriate time-lag focusing after each laser shot.

2.2.3.2 Ultraflex MALDI-TOF MS at UCL

MALDI-TOF MS at UCL was performed using an Ultraflex MALDI-TOF/TOF instrument (Bruker Daltonics, Bremen, Germany). In the LMR a sum of 400 laser shots, delivered in eight sets of 50 shots (at 10 Hz) to each of eight different locations on the surface of the matrix spot were acquired in linear mode geometry under 20 kV (18.6 kV during delayed extraction) of ion acceleration and -1.3 kV multiplier potential, and with suppression of mass ions set to m/z <400. Delayed extraction was maintained for 80 ns to give appropriate time-lag focusing after each laser shot.

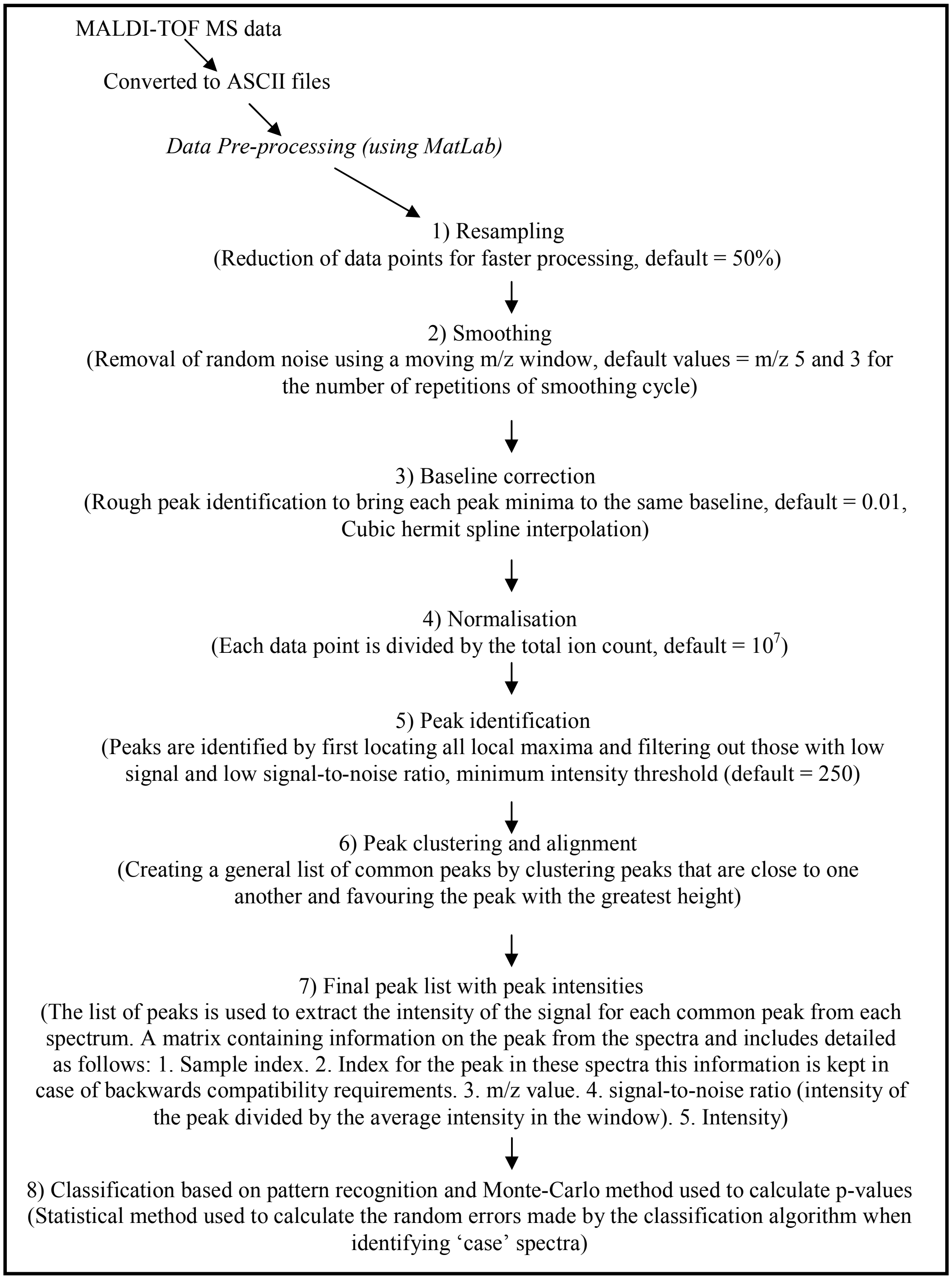
For acquisition of the HMR 4–15 kDa m/z segment a sum of 500 laser shots were acquired in linear mode with suppression of mass ions set to m/z <3,000. Delayed extraction was maintained for 50 ns to give appropriate time-lag focusing after each laser shot. The initial 100 shots for the HMR acquisition were delivered at the same location as the first 50 shots used for the LMR since the AutoXecute function requires a home position for each individual MALDI spot. Therefore, an extra 100 shots were required for the HMR owing to the ablation of sample at the home position.

2.2.4 Data processing and analysis

The MALDI-TOF MS data presented in this study were analysed with two different systems.

2.2.4.1 Data processing and analysis 1

Data collected at the MSKCC for the UKCTOCS pilot study (Chapter 3) was analysed in collaboration with the Computer Learning Centre at Royal Holloway, University of London, UK. The workflow used for spectral processing is shown in Figure 2.3 (MatLab script for each function can be found in Appendix 1). Spectra were externally calibrated using 13 calibrant peaks from the calibrant spots associated with each sample (Table 2.1). Smoothing was done by averaging the intensities within a moving window and baseline subtraction involved estimation of the baseline from the mass spectrum. Spectra were normalised by dividing each intensity value by the total ion count. The obtained values were then multiplied with a constant *C* (C = $2*10^5$). Peak definition/detection involved finding local peak maxima in the mass spectra with a signal-to-noise ratio exceeding an optimised threshold (set to 4 in this study). The noise level was defined as the average of the intensities at the m/z ratio within a moving window with a fixed size (e.g. 500 Da). Local maxima were located by finding the m/z ratios with the highest intensities among their neighbours. The peaks identified were quantified as the intensity at the local maximum. Peaks were then internally aligned across all spectra. At the peak alignment step, the peaks of multiple mass spectra within the mass error rate (100 ppm) were grouped together as a “peak group”. Since not all peaks occur in all spectra, a given number of peak points acted as unique anchors for alignment and every other sample was aligned with this ‘superset’. To this end, the superset was split into clusters which were defined in two steps. Firstly, all the intervals between neighbouring peak positions in the superset exceeding a mass resolution of 1500 ppm, were found. These intervals split the m/z into clusters of order 1. This was then checked to see that each sample had no more than 1 peak in a cluster. When more than one peak was found, the cluster was divided into smaller clusters to ensure only 1 peak per cluster. Thus, all peaks were aligned to certain clusters. Once peaks had been aligned spectra were then labelled ‘case’ or ‘control’. Two pattern recognition algorithms (the nearest neighbour algorithm and the support vector machine) were used to classify the samples in their respective groups. Pattern recognition algorithms construct decision rules on the basis of the training set of spectra used. Due to the limited number of ‘case’ samples the ‘training’ and ‘test’ spectra were the same. The Monte-Carlo method was used to calculate the p-values for the errors made by the classification algorithm [Gammerman et al., 2008].



MALDI-TOF MS data

Converted to ASCII files

Data Pre-processing (using MatLab)

1) Resampling
(Reduction of data points for faster processing, default = 50%)

2) Smoothing
(Removal of random noise using a moving m/z window, default values = m/z 5 and 3 for the number of repetitions of smoothing cycle)

3) Baseline correction
(Rough peak identification to bring each peak minima to the same baseline, default = 0.01, Cubic hermit spline interpolation)

4) Normalisation
(Each data point is divided by the total ion count, default = 10^7)

5) Peak identification
(Peaks are identified by first locating all local maxima and filtering out those with low signal and low signal-to-noise ratio, minimum intensity threshold (default = 250)

6) Peak clustering and alignment
(Creating a general list of common peaks by clustering peaks that are close to one another and favouring the peak with the greatest height)

7) Final peak list with peak intensities
(The list of peaks is used to extract the intensity of the signal for each common peak from each spectrum. A matrix containing information on the peak from the spectra and includes detailed as follows: 1. Sample index. 2. Index for the peak in these spectra this information is kept in case of backwards compatibility requirements. 3. m/z value. 4. signal-to-noise ratio (intensity of the peak divided by the average intensity in the window). 5. Intensity)

8) Classification based on pattern recognition and Monte-Carlo method used to calculate p-values
(Statistical method used to calculate the random errors made by the classification algorithm when identifying ‘case’ spectra)

Figure 2.3 Workflow used to process MS spectra. The MatLab scripts for each of these steps can be found in Appendix 1.

2.2.4.2 Data processing and analysis 2

All data collected at UCL was processed and analysed using ClinProTools™ software (V2.0 & 2.2, Bruker Daltonics, Germany). Spectra were first subjected to a 0.80 level convex hull baseline subtraction. Following this the detection of peaks was based on the analysis of a smoothed first derivative where the smoothing was determined by the "resolution" parameter. Spectral settings were optimised using '400 and 200' for the resolution parameter for the LMR and HMR respectively. 1 cycle of smoothing was done using the Savitsiky Golay algorithm with an m/z width of 1 for the LMR and 5 for the HMR. Once the peaks were detected, peak areas were calculated by integrating the intensities over the region of the peak (between the start and end positions) using the zero level integration function. Peak areas were then normalised to make the total signal equal in all spectra. These peak areas were given as arbitrary units (arb.u.). Spectral recalibration was done with a maximum peak shift of 0.3% and a 15% match of automatically selected internal calibrant peaks. Peak selection was done using a signal-to-noise ratio of 3 for the LMR and 5 for the HMR regions. When different classes of data e.g. groups of spectra generated on different days were loaded into the software, as part of the spectral pre-processing step p-values are automatically generated for all the identified peaks. It is important to note that ClinProTools requires at least 2 spectra in each class. In addition, a 2D peak distribution view is generated to show the distribution of the average peak area of two selected peaks which may be the top discriminatory peaks between two classes of samples (user defined). The peak statistics calculated by ClinProTools were used to calculate the co-efficient of variance.

For classification of clinical samples, each condition was loaded as a separate class and the Support Vector Machine (SVM) algorithm available in the ClinProTools software was used. The optimised model used 1-25 peaks selected automatically by ClinProTools to include 3 k-nearest neighbours in both mass ranges.

2.3 Sample preparation for 1D and 2D SDS-PAGE analysis

2.3.1 Unfractionated pooled serum

Prior to 2D-DIGE analysis, unfractionated serum samples were pooled according to volume into healthy, benign or malignant groups. Pooled samples were then diluted 1:100 for an accurate measure of protein concentration using the Pierce BCA protein assay, using BSA to generate a standard curve. Equal amounts of protein were labelled with each of the three Cy dyes for 2D-DIGE analysis (see section 2.4).

2.3.2 Protocol for HPLC immunoaffinity depletion using the Multiple Affinity Removal System (MARS).

As a complementary approach, the Multiple Affinity Removal System was used to facilitate the differential analysis of the pooled serum samples by allowing the removal of the 7 most abundant proteins in serum. The Multiple Affinity Removal System (MARS) is comprised of an affinity HPLC column (size: 4.6 x 50 mm, Agilent part number 5185-5984) packed with immobilised affinity-purified polyclonal antibodies for removal of albumin, transferrin, IgG, IgA, haptoglobin, antitrypsin and fibrinogen with high specificity and optimised mobile phases. All chromatographic steps were performed at 20°C on an Agilent 1100 HPLC system.

From each pool, 30 µL of serum was diluted five times with MARS Buffer A containing protease inhibitors (COMPLETE™, Roche) and centrifuged at 16,000 x g at room temperature for 5 minutes to remove particulates. Automated sample injection was set up for 30 µL of diluted serum sample per injection in Buffer A at a flow rate of 0.25 mL/min for 9 min. Flow-through fractions, ~0.75 mL per injection, containing the lower abundant protein species were collected from each injection manually at 2-4 min into 0.5 mL Eppendorf tubes and stored at -20°C until further analysis. The bound fractions were eluted with 100% Buffer B at a flow rate of 1.0 mL/min for 3.5 min. The column was regenerated by equilibrating with Buffer A for 10 min.

2.3.3 Protein desalting and concentration

In order to resolve proteins from depleted serum samples on 2D gels, flow-through fractions from five injections were pooled into a 5 mL Zeba™ desalting spin column (Pierce, Rockford, IL) and desalted. Desalted samples were concentrated to 0.5 mL using a spin concentrator with a 5 kDa Molecular Weight Cut-Off (MWCO) membrane; samples were spun at 4000 x g for approximately 1 hour at 10 °C. Concentrated retentates were transferred into fresh Eppendorf tubes and speed vacuumed to dryness. Samples were resuspended in 2D lysis buffer (8 M urea, 2 M thiourea, 4% CHAPS, 0.5% NP40 and 10 mM Tris pH 8.3). Protein content was estimated using the Pierce BCA protein assay using BSA to generate a standard curve.

2.3.4 ProteoMiner Protein Enrichment Kit

Pooled samples were applied to spin columns from the ProteoMiner Protein Enrichment Kit (BioRad Catalogue # 163-3000). The ProteoMiner Protein Enrichment Kit is comprised of several spin columns packed with a large, highly diverse bead-based library of combinatorial peptide ligands. It is considered to be a novel sample preparation tool used for the compression of the dynamic range of the protein concentration in complex biological samples. Best results are obtained with protein concentrations greater than 50 mg/mL. When complex biological samples (e.g. human serum) are applied to the beads, the high abundance proteins saturate their high affinity ligands and the excess protein is washed away. In contrast, the medium and low abundance proteins are concentrated on their specific affinity ligands. This reduces the dynamic range of protein concentrations, while maintaining representatives of all proteins within the original sample.

First the spin columns (20% beads, 20% v/v aqueous EtOH, 0.5% v/v acetonitrile) were prepared by centrifugation at 1,000 x g for 2 min to remove the storage solution. The collected material was discarded. The columns were then washed by adding 1 mL deionised water and rotating them end-to-end over a 5 min period. Again the columns were centrifuged at 1,000 g for 2 min to remove the water and the collected material was discarded. The wash steps were repeated twice using 1 mL wash buffer (PBS; 150 mM NaCl, 10 mM NaH_2PO_4, pH 7.4). At the final wash step columns were centrifuged again for an additional 1 min at 1,000g to remove any remaining buffer.

Pooled serum samples were centrifuged at 10,000 x g for 10 min to remove particulates. 1 ml of serum (>50mg/mL), normalised for protein concentration was applied to the spin columns and incubated with the beads by rotation for 2 hr at room temperature. Columns were centrifuged at 1,000 x g for 2 min and the collected material was retained for analysis. The columns were centrifuged again at 1,000g to remove residual material. Next, 1 ml of wash buffer (PBS) was added to each column and the columns were rotated over a 5 min period. The columns were centrifuged at 1,000 x g for 2 min and the collected material was discarded. Again the columns were centrifuged at 1,000 x g for an additional 1 min to remove any remaining material. This wash step was repeated twice and then a final wash step using deionised H_2O was carried out.

Bound proteins were eluted with 100 µL of 2D lysis buffer (8 M urea, 2 M thiourea, 4% CHAPS, 0.5% NP40 and 10 mM Tris pH 8.3) by gentle vortexing over a 15 min period at ambient temperature and the eluate collected by centrifugation at 1,000 x g for 2 min to elute bound proteins. The elution step was repeated twice to ensure all bound material was collected. Protein concentration of the eluates was estimated using the Pierce BCA protein assay using BSA to generate a standard curve (632.4 µg recovered from healthy pool, 687.4 µg from benign and 786.8 µg from malignant pool). The eluted samples were stored at -20°C prior to downstream 2D-DIGE analysis.

2.4 Two-Dimensional Difference Gel Electrophoresis (2D-DIGE)

Two-dimensional electrophoresis is a powerful and widely used method for the analysis of complex protein mixtures extracted from biological samples. This technique sorts proteins according to two independent properties in two discrete steps: the first dimension step, isoelectric focusing (IEF), separates proteins according to their isoelectric points (pI) and the second-dimension step, SDS-polyacrylamide gel electrophoresis (SDS-PAGE), separates proteins according to their molecular weights (M_r relative molecular weight). A few thousand different proteins can thus be separated, and information such as the protein pI, the apparent molecular weight, and the amount of each protein provided. Fluorescence two-dimensional difference gel electrophoresis (2D-DIGE) is a more recently developed 2D gel-based proteomics technique that provides a sensitive, rapid and quantitative analysis of differential protein expression between two or more biological samples (see below).

2.4.1 Protein labelling with NHS-cyanine dyes (DIGE-labelling)

The NHS-cyanine dye Cy2 was purchased from GE Healthcare, whilst NHS-Cy3 and NHS-Cy5 were synthesised "in-house" by Dr P. Gaffney [Chan et al., 2005]. Protein labelling and 2D-DIGE were performed according to Gharbi et al. [Gharbi et al., 2002]. For this study, protein pools and fractions were labelled in triplicate with NHS-Cy3 or NHS-Cy5 at 4 pmol dye/µg protein on ice in the dark for 30 min. Equal amounts of protein from each clinical condition were also pooled together and labelled with NHS-Cy2 to create an internal standard which was run on all the gels against the Cy3- and Cy5-labelled samples to aid in spot matching and quantitation. Labelling reactions were quenched with a 20-fold molar excess of free L-lysine to dye and left on ice for 10 min. Equal amounts of proteins labelled with Cy3 and Cy5 were mixed appropriately and the same amount of Cy2-labelled pool was added to each mixture.

Samples were reduced by adding 1.3 M dithiothreitol (DTT) to a final concentration of 65 mM. Ampholine/Pharmalyte carriers (1:1 mix, pH 3-10), were added to a final concentration of 2% and bromophenol blue was added to each sample. The final volume of each sample was adjusted to 450 µL with 2D-DIGE lysis buffer plus DTT. For isoelectric focusing (IEF), 24 cm, non-linear pH 3-10 IPG strips (GE Healthcare) were rehydrated with Cy-dye labelled samples in a re-swelling tray overnight in the dark at RT, according to the manufacturer's guidelines. The separation of the Cy-dye labelled proteins in the first dimension by IEF was carried out on a Multiphor II apparatus (GE Healthcare) for a total of 80 kVh at 18°C.

For protein separation in the second dimension, 1.5 mm 12% SDS-PAGE gels were cast between 24 cm low-fluorescence glass plates. The inner surface of one plate of each set was coated with Bind Silane solution (PlusOne, GE HealthCare) to bond the gels. This allowed easier handling of gels during scanning and protein post-staining, storage and spot excision. Fluorescent reference markers were placed at the edges of these bonded plates to facilitate the generation of coordinates for each protein feature in the final pick lists. The inner surface of the other plate was treated with Repel Silane (PlusOne, GE Healthcare) to ensure easy separation of plates after running. After IEF, IPG strips were equilibrated in equilibration buffer (6 M urea, 30 % (v/v) glycerol, 50 mM Tris-HCL pH 6.8 and 2% (w/v) SDS) in two steps for 15 minutes each with gentle rocking. In the first step, the equilibration buffer was supplemented with 65 mM DTT to reduce disulphide bonds, while in the second step 240 mM iodoacetamide (IAM) was added to the equilibration buffer to alkylate reduced thiol groups. IPG strips were then rinsed with Tris-Glycine-SDS electrophoresis buffer (Severn Biotech) and transferred onto the second dimension gels. Strips were overlaid with 0.5% (w/v) low-melting point agarose in Tris-Glycine-SDS electrophoresis buffer with bromphenol blue. Gels were run in an Ettan 12 apparatus (GE Healthcare) at 2 W per gel at 8°C until the dye front had run off, thereby avoiding the fluorescence signal from bromophenol blue and free dye. All steps were carried out in a dedicated clean room.

2.4.2 Detection of Cy-Dye labelled proteins

Gel images were obtained by scanning the gels between plates on a Typhoon™ 9400 multiwavelength fluorescence scanner using ImageQuant software (both from GE Healthcare). Excitation and emission wavelengths for each dye used in this study are shown in Table 2.2. The photomultiplier tube voltage of the Typhoon scanner was adjusted for each channel (Cy2, Cy3, and Cy5) in preliminary low-resolution scans (1000 μm) to give maximum pixel values within 10% for each channel, but below the saturation level. These setting were then used for high-resolution (100 μm) scanning. Images were generated as .gel/TIFF files and exported to image analysis software for further analysis.

Dye	Excitation (nm)	Emission (nm)
Cy2	480	530
Cy3	540	590
Cy5	620	680
CCB	620	/

Table 2.2 Excitation and emission wavelengths used to detect each of the Cy-dyes and Colloidal Coomasie Blue (CCB) post-staining.

2.4.3 Image analysis

Gel images were analysed using DeCyder™ image analysis software V5.0 (GE Healthcare). Firstly, images were analysed using the Differential In-Gel Analysis (DIA) module. DeCyder processes the three images derived for the three Cy dyes (Cy2, Cy3 and Cy5) representing profiles of each of the three samples run on a single gel. DIA performs automatic normalisation, spot detection, filtering and background subtraction and also quantifies protein spot abundance or volume on each image and expresses these values as ratios, indicating changes in expression levels by direct comparison of the corresponding spots on each gel. This ratio can be used to directly evaluate changes between two labelled protein samples run on a single gel and between the test samples and the same spot in the internal standard to give a standard spot volume that allows accurate inter-gel protein spot comparisons. Features resulting from non-protein sources (e.g. scratches on glass plates and dust particles) were filtered out.

Subsequently, the Biological Variance Analysis (BVA) module of DeCyder was used for matching protein spots from different conditions across gels by matching to a user defined master gel image, which identified common protein spots across the sets of gels. User intervention was required at this stage to set landmarks on gels for accurate cross-gel matching. Standardised spot volumes were then averaged across replicate samples for each experimental condition and data plotted graphically within BVA. Statistical analysis was performed and spots displaying a ≥ 1.5 average-fold increase or ≤ 1.5 average-fold decrease in abundance between clinical conditions with *P* values <0.05 or <0.01 from a Student t-test were selected for spot picking and MS-based identification.

2.4.4 Protein post-staining and spot excision

Bonded 2D gels were post-electrophoretically stained with colloidal Coomassie Blue G-250 (CCB) to visualise proteins for accurate spot picking. Gels were stained according to a modified protocol by Neuhoff et al. 1988 [Neuhoff et al., 1988]. Briefly, bonded gels were fixed in 35% (v/v) ethanol with 2% (v/v) phosphoric acid for more than three hours on a shaking platform and then washed three times for 30 min each in ddH_2O. Gels were then incubated in 34% (v/v) methanol, 17% ammonium sulphate and 3% (v/v) phosphoric acid for one hour prior to the addition of 0.5 g/L Coomassie Blue G-250 (Merck Biosciences) and left to stain for two to three days. De-staining was not required. Post-stained gels were scanned on the Typhoon™ 9400 scanner using the red laser with no emission filter (Table 2.2). Post-stained images were imported into the BVA module of DeCyder and matched with the processed Cy-Dye images. Using the reference markers fixed onto the glass plates during gel casting, a pick list of coordinates (.txt file) for protein features that were differentially expressed was created for automated spot picking. An Ettan automated spot picker (GE Healthcare) was used with a 2 mm picking head, which excised protein features from gels submerged under 1-2 mm of ddH_2O. Spots were collected in 96-well plates, drained and stored at -20°C prior to MS analysis.

2.4.5 Protein in-gel digestion

For protein sequence analysis by mass spectrometry, protein spots were subjected to trypsin digestion. Gel pieces were washed three times with 50% (v/v) acetonitrile, dried in a SpeedVac for 10 min, reduced with 10 mM DTT in 5 mM ammonium bicarbonate pH 8.0 (AmBic) for 45 min at 50°C and then alkylated with 50 mM iodoacetamide (IAM) in 5 mM AmBic for one hour in the dark at RT. Gel pieces were then washed three times in 50% (v/v) acetonitrile and vacuum-dried prior to re-swelling with 50 ng of modified trypsin (Promega) in 5 mM AmBic pH 8.0. The gel pieces were then overlaid with 10 µL of 5 mM AmBic and digested for 16 hours at 37°C. Supernatants were collected and trypsin digests were further extracted by washing the gel pieces twice with 5 % (v/v) trifluoroacetic acid in 50 % acetonitrile.

Peptide extracts from each gel piece were pooled, vacuum-dried and resuspended in 5 μL of 0.1 % formic acid and stored at -20°C prior to MS analysis.

2.4.6 Protein identification

Protein identification was carried out using Matrix-Assisted Laser Desorption/Ionisation Time-of-Flight (MALDI-TOF) MS by peptide mass fingerprinting. For this, 0.75 μL of the trypsin digest was mixed with 0.5 μL of matrix solution (saturated aqueous 2,5-dihydroxybenzoic acid, DHB), and applied to a sample target plate and air dried. MALDI-TOF mass spectra were acquired using an externally calibrated UltraFlex mass spectrometer (Bruker Daltonics). Firstly, the mass spectrometer was calibrated using a standard mixture of peptides (calibration mixture 2 from the Sequazyme™ kit, Applied Biosystems). Then a sum of 200 laser shots, delivered in sets of 30 shots (at 6.7 Hz) to several locations on the surface of the matrix spot were acquired in the reflector, positive ion geometry under 25 kV for ion source 1, 21.2 kV for ion source 2, 73 kV on the lens, 26.1 kV on the first reflector and 14.9 kV on the second reflector. The reflector detector was set at 5.75 V, delayed extraction of ion acceleration was maintained for 150 ns to give appropriate time-lag focusing after ach laser shot and -1.3 kV multiplier potential, and with suppression of mass ions set to m/z <400. Internal calibration of each mass spectrum was performed using reference trypsin autolysis peaks 842.51 m/z and 2211.10 m/z. Prominent peaks in the mass range m/z 700-5000 were then used to generated a peptide mass fingerprint, which was searched against updated NCBI and IPI-Human databases using the Mascot search engine, version 2.0.02 (Matrix Sciences Ltd.). For the search criteria, carbamidomethylation of cysteines was selected as a fixed modification, while oxidation on methionine, N-acetylation and pyro-glutamate were selected as variable modifications. A positive identification was accepted when a minimum of 6 peptide masses matched a particular protein (mass error of ± 50 or 100 ppm, allowing 1 missed cleavage), sequence coverage was >25%, MOWSE scores were higher than a threshold value where $p=0.05$ and the predicted protein mass agreed with the gel-based mass. When a protein 'hit' fulfilled the specified thresholds for identification, unmatched peptides were systematically re-submitted to the database in a search for possible multiple proteins per gel piece and potential sites of post-translational modifications (e.g. phosphorylation or glycosylation).

In addition, some identifications were made using nano-LC-electrospray ionization collision-induced dissociation tandem MS (LC-MS/MS). This was preformed on an ACQUITY Ultra performance LC system (Waters) with a PepMap C18 100-µm inner diameter column (LC Packings) at a flow rate of 400 nL/min, coupled to a Quadrupole Time-Of-Flight (QTOF Premier) mass spectrometer (Waters/Micromass, Manchester, UK). Spectra were processed using MassLynx software (Waters) and submitted to Mascot database search routines including +2 and +3 peptide charges, and a mass tolerance of ± 100 ppm. Positive identifications were accepted when at least two peptide sequences matched an entry with MOWSE scores above the $p=0.05$ threshold value.

Chapter 3: MALDI-TOF MS analysis of serum samples pre-dating diagnosis of ovarian cancer

3.1 Introduction

One of the first reports on the use of magnetic beads for serum peptide and protein extraction coupled with MALDI-TOF MS profiling was published by Villanueva et al. [Villanueva et al., 2004]. The group described an automated technology platform for the simultaneous measurement of serum peptides that was simple, scalable and generated reproducible patterns. Peptides and proteins were captured and concentrated using reversed-phase (RP) batch processing on C8-coated magnetic beads in an automated format on a liquid handling robot followed by a MALDI-TOF mass spectrometric analysis. The optimised protocol was based on a detailed investigation of serum handling conditions, RP ligands, eluant selection, small-volume robotics design and spectral acquisition across a study set. The improved sensitivity and resolution allowed detection of approximately 400 polypeptides (0.7-15 kDa range) from a single droplet (50 µL) of serum and almost 2,000 unique peptides in larger sample set [Villanueva et al., 2004]. The group also described a pilot study which indicated that sera from brain tumour patients could be distinguished from healthy controls based on a pattern of 274 peptide masses. This in turn allowed the generation of a learning algorithm that correctly predicted 96.4% of the samples as either healthy or diseased [Villanueva et al., 2004].

In summary, the system that they described and further optimised satisfied all criteria of MALDI-TOF compatibility, high resolution, reproducibility and throughput, and the limited application provided a proof-of-concept that sera from cancer patients with solid tumours contain peptides detectable by MALDI-TOF MS that reflect the activity of the cancer. They further identified these peptide peaks by tandem MS sequencing and database interrogation. In their study they found signature peptides that fell into several tight clusters or ‘ladders’ of serum protein–derived peptides such as C3F and FPA generated by a 2-step process involving endoproteases such as kallikerins, thrombin and factor I as well as unknown exopeptidase activities that produce cancer type–specific differences which were superimposed on the proteolytic events of the ex vivo coagulation and complement degradation pathways. Thus, they went on to hypothesize that tumour-specific exopeptidase activities produce ‘surrogate’ markers from abundant polypeptides generated during the clotting stage, facilitating correct classification of diseased samples [Villanueva et al., 2004; Villanueva et al., 2005]. Through collaboration, this platform was used at the MSKCC to analyse a set of samples from a pilot study designed to explore the possibility of using MS-based pattern recognition to detect onset of ovarian cancer at an early stage. To this end, during a visit to the Memorial Sloan Kettering Cancer Centre, (MSKCC, New York, USA) in 2004/5, training was provided and reproducibility of the platform assessed, prior to analysis of the pilot study samples.

The first aim of the study presented in this chapter was to assess the reproducibility of the automated bead-based serum peptide extraction protocol and MS-based profiling platform developed at the MSKCC. The second aim was to use this automated platform for the analysis of serum samples that pre-dated diagnosis of ovarian cancer and to determine whether the MALDI-TOF MS profiles obtained could be used to predict ovarian cancer and to compare this with the performance of serum cancer antigen CA-125.

3.2 Reproducibility of the automated platform at MSKCC

As mentioned previously, critics have argued that the published results of serum profiling studies looking for cancer biomarkers do not demonstrate biomarker reproducibility. Reproducibility is a measure of the robustness of any technology and is vital for providing support for new and emerging platforms. Hence, the reproducibility of the platform developed by Villanueva et al. [Villanueva et al., 2004] was evaluated using commercial serum from Sigma (Cat. No. S-7023 Lot. 034K8937) as a quality control at the MSKCC. Four replicates of 10 µL of serum were each mixed with 5 µL of C8-coated magnetic beads (Chemicell) for peptide extraction following the protocol outlined in Chapter 2 (Materials and Methods section 2.2) and peptide profiles in the 0.7 to 15 kDa range were generated by MALDI-TOF-MS. The coefficient of variance for all peaks detected in the low mass range (LMR; 700-4000 Da) (average standard deviation divided by the average peak area) was calculated between the 4 samples as a measure of intra-assay reproducibility. The preparation and analysis was then repeated once a day over 5 days to yield inter-assay reproducibility values. All data processing was performed using ClinProTools software (V2.0). The intra-assay reproducibility for all detected peaks varied across the runs from 12 ± 1.2% to 33.9 ± 6.9%, although this was < 20.11% when run 1 was excluded. Taking the average peak areas for all 5 runs (i.e. 20 samples processed together) the overall inter-assay CV value was 32.8 ± 6.4% (Table 3.1).

Run Number	Ave No. of peaks	Ave peak area (arb. u.)	peak area StdDev	peak area CV (%)
1	78	28.81	6.9	33.9
2	70	30.06	4.0	18.3
3	117	22.01	1.2	12.0
4	92	26.73	2.7	17.9
5	116	20.71	2.5	20.1
Overall inter-assay	92	25.42	6.4	32.8

Table 3.1 Intra/inter-assay reproducibility of MSKCC platform. The average standard deviation and coefficient of variance (Av. SD / Av. peak area) were calculated for all detected peaks using ClinProTools software (V2.0). Intra-assay reproducibility was calculated by running 4 samples per run, while inter-assay reproducibility was calculated by assessing the variation between 5 runs.

Spectral heat maps created in ClinProTools software (V2.0) showed some obvious peak area differences in the intra-assay comparisons (Figure 3.1). These differences, which could be the result of inconsistencies in the complex MALDI ionisation process, are likely to account for the seemingly high CV values. Furthermore, the average number of peaks obtained are comparable to the relative number of peaks previously reported [Villanueva et al., 2004]. The assay CVs are also considered acceptable.

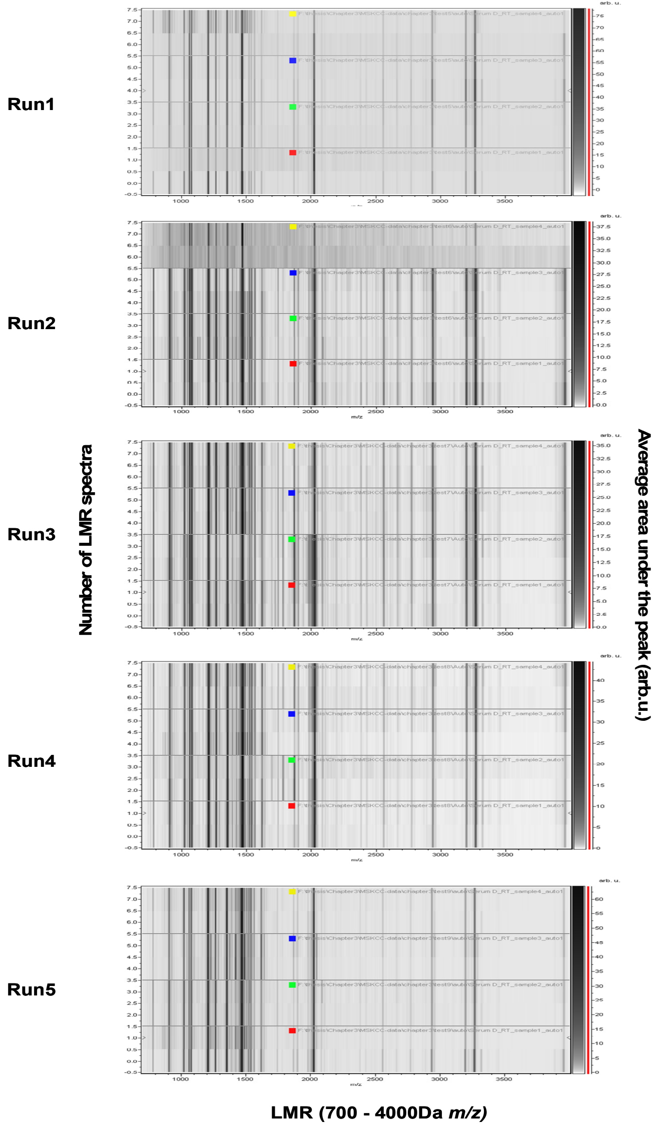


Figure 3.1 Spectral comparisons across runs. Four replicates of 10 μL of serum each were mixed with 5 μL of C8 beads (Chemicell) for peptide extraction following the protocol outlined in Chapter 2 (Materials and Methods section) prior to MALDI-TOF MS analysis. Five replicate runs were then preformed on different days. Spectra were processed and displayed as heat maps using ClinProTools software (V2.0).

3.3 Analysis of serial samples pre-dating diagnosis of ovarian cancer

Following the evaluation of the reproducibility of the platform, 92 serial serum samples from 19 women which pre-dated a diagnosis of ovarian cancer were analysed. The samples came from a pilot study of ~22,000 women which involved regular serum sample collection for cancer antigen CA-125 assay and following up volunteers for up to 7 years (1995-2001) prior to actually diagnosis. Each of the 19 volunteers who developed ovarian cancer (in one case, non-epithelial) had between 2 and 11 serial samples. Each case sample had 2 matched healthy controls which were taken and handled at approximately the same time as the case sample. In addition samples were matched on the use of hormone replacement therapy. Other information such as date of birth, CA125, the date samples were taken, the date samples were received at the laboratory and tube type used for serum collection was also available. It is hypothesised that during the clotting process tumour-specific proteases generate fragment 'ladders' of marker peptides in serum producing subtle changes in the MS profiles. These changes have been used to discriminate diseased sera from the sera of matched healthy controls which had been collected at the same centre at approximately the same time and thus handled and sorted in the same manner [Villanueva et al., 2004; Villanueva et al., 2005]. The main aim of the work presented in this chapter was to demonstrate that the information contained in mass spectra, in combination with the level of CA-125, is useful for early detection of ovarian cancer.

The data presented in this chapter came from 80 samples from 18 volunteers as the non-epithelial ovarian cancer case, and samples which did not have both of the matched control samples were subsequently removed. The 240 samples were processed, extracted and spotted in duplicate for MALDI-TOF analysis. One spectrum was chosen from each pair at the data processing and analysis stage. Example spectra from serial samples of a healthy control and cancer case are shown in Figure 3.2. These show generally that many of the peaks detected were common across the sample sets, but that there was considerable variation in peak areas. A total of 240 spectra were analysed in collaboration with Professor Alex Gammerman's group at the Computer Learning Research Centre, Royal

Holloway, University of London, who used this data to build class prediction algorithms [Gammerman et al., 2008].

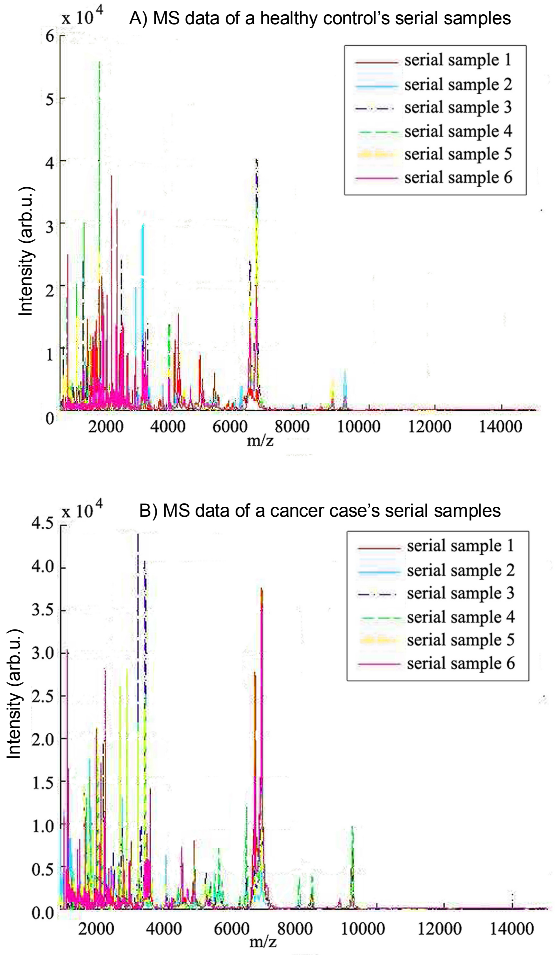


Figure 3.2 Spectral comparisons across serial samples. MS profiles of six serial samples from (A) a healthy control and (B) a cancer case are overlaid for comparison. The data from the low and high mass range have been combined.

After spectral pre-processing, (described in Chapter 2 section 2.2.4), 7216 ‘non-aligned peaks’ were identified in the 265 spectra, (including the non-epithelial cancer case plus controls). After peak alignment these peaks were clustered into 402 peak groups. The peaks were ordered according to their frequency (i.e. the percentage of samples having the same peak). Overall, 20 peaks were found to exceed 40% frequency (Table 3.2 and Figure 3.3). The peak intensities of the 20 most frequent peaks in each sample and the corresponding CA-125 measurement were then used to create a vector consisting of 21 numbers for a classification algorithm.

Peak number	mean *m/z*	*m/z* range	Number of samples having the peak	% of total
1	3188.9	3185.3-3191.0	246	92.8
2	6646.1	6636.5-6652.8	245	92.5
3	3330.5	3325.2-3333.4	198	74.7
4	2004.1	2001.2-2005.3	190	71.7
5	1764.6	1762.0-1766.6	184	69.4
6	818.5	817.4-818.9	165	62.3
7	9307	9294.7-9319.7	154	58.1
8	2982.3	2978.8-2985.5	146	55.1
9	2020.9	2019.7-2021.9	144	54.3
10	4292.5	4288.0-4300.8	140	52.8
11	3280	32.76.2-3282.1	139	52.5
12	2548.2	2543.5-2550.2	137	51.7
13	2562.8	2561.4-2564.2	123	46.4
14	8942.2	8930.8-8955.3	118	44.5
15	3296.9	3294.3-3299.1	115	43.4
16	1888.8	1887.7-1889.7	114	43
17	899.9	898.6-901.0	113	42.6
18	3172.1	3169.1-3176.4	109	41.1
19	3229.8	3226.7-3234.2	109	41.1
20	5010.4	5004.0-5017.0	109	41.1
39	2016.8	2013.6-2019.6	56	21.1

Table 3.2 Top 20 peaks and peak 39. The top 20 peaks are those present in more than 40% of the samples. Peak number 39 occurred more rarely but was found to be useful for case versus control discrimination.

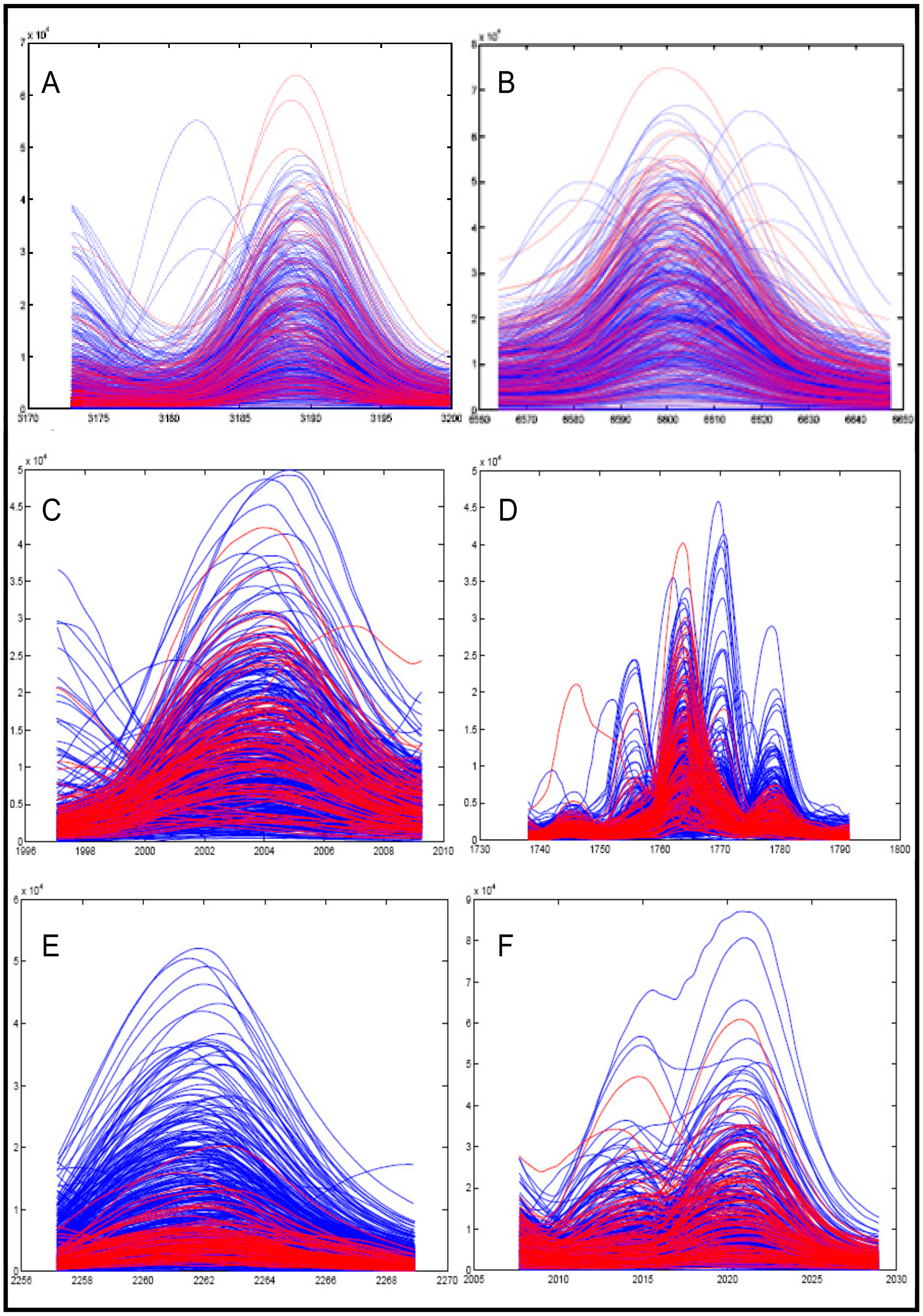


Figure 3.3 Spectral views of peaks. After spectral pre-processing (described in Chapter 2 section 2.2.4) 20 peaks were found to be common in over 40% of the samples. Representative peak profiles for peaks (A) 3188.9 m/z, (B) 6646.1 m/z, (C) 2004.1 m/z, (D) 1764.6 m/z, (E) 2562.8 m/z and (F) 2016.8 m/z, are shown here in blue for healthy and red for cancer case.

For classification, for each t (time) = 1-18, the null hypothesis was that the assignment of the label "case" within each triplet in St (start time) was random. For example, the correct identification of the 'case' labelled samples, the 240 samples were divided into 80 triplets, each consisting of a case sample and the two matched control samples. The 80 triplets were further divided into 18 triplet groups corresponding to the 18 cases (with the size of the group varying between 2 and 10 samples). Each triplet was assigned a non-negative value *t*, the time to diagnosis (months) for the case measurement in each triplet. For each *t* = 0, 1, 2,…St (start time) was assigned to be the set of triplets taken *t* months before the diagnosis. As such the largest St (for *t* = 0, 1) contained 18 triplets, whereas the smallest St contained 14 triplets. The classification algorithm used a rather limited set of rules for the identification of the cancer 'case' labelled sample within each triplet. Each classification rule is described by three numbers, (p; w1; w2), which are a peak number p (1 – 20) and weights w1 (0; 1) and w2 (-1; 1). For each triplet, the classification rule was used to predict the sample with the largest (w1 log C+w2 log P), where C was the CA-125 level and P is the area of peak p, labelled as "case". The logarithms were taken to remove the arbitrary units of measurement of CA-125 and the peak area.

The baseline rule in the classification algorithm used CA-125 measurements alone to classify the triplet samples. The output for the algorithms reported on the number of errors (E1), made on identifying the triplets in each St. Then another dimension was added using the top 20 peaks (E2) and finally the top 100 peaks (E3). The number of errors in E1 increased as the time to diagnosis increased, demonstrating that CA-125 measurements alone are insufficient for predicting ovarian cancer and for early diagnosis (Table 3.3). Using the top 20 peaks (E2) the number of errors also increased with time from diagnosis, but not to the same degree as E1, suggesting that early diagnosis information was present in the spectra.

Results showed that the classification algorithm made the smallest numbers of errors when the top 100 peaks (E3) were used from each of the triplets sets (St) (Table 3.3). For example, in Table 3.3 the entry 2 corresponding to *t* = 6 means that out of 15 cases with samples taken at least 6 months before diagnosis the best classification algorithm made 2 errors on the most recent of those samples. This is a very small number of errors. The Monte-Carlo method was used to calculate valid p-values and involved randomly reassigning “case” labels and counting the number of errors which were as good as or better than the number of errors obtained for the true labels (Q). The p-value was then estimated as the ratio of Q/N where N = 10^6. It should be noted that these results are heavily biased since the training set St included the test samples. However, the p-value of 0.012% for 7 months prior to diagnosis is highly significant. Furthermore, peak number 10 (4292.5 m/z) was the most frequent feature used in the classification algorithm.

t	|St|	E 1	p-value 1	E 2	p 2	w 2	p-value 2	E 3	p 3	w 3	p-value 3
0	18	2	0.000001	1	-9	1	0.000001	1	-9	1	0.00006
1	18	2	0.000001	2	3	1	0.00002	1	-14	2	0.00001
2	15	4	0.0018	3	8	1	0.0094	2	18	2	0.0052
3	15	5	0.0086	3	8	1	0.0095	2	18	2	0.0056
4	15	6	0.031	3	8	1	0.0097	2	18	2	0.0056
5	15	6	0.031	3	-14	1	0.0017	2	-14	1	0.0047
6	15	7	0.088	2	-10	1	0.011	2	-10	0.5	0.0037
7	15	7	0.088	3	-10	1	0.011	2	-10	0.5	0.0037
8	14	8	0.31	3	-10	1	0.025	2	-39	1	0.021
9	14	7	0.15	3	-10	1	0.025	2	18	2	0.012
10	14	7	0.15	3	-10	1	0.025	2	18	2	0.012
11	14	7	0.15	3	-14	1	0.032	2	18	2	0.012
12	14	7	0.15	4	-10	1	0.13	2	18	2	0.012
13	14	8	0.31	3	-10	1	0.024	3	-10	1	0.063
14	14	9	0.52	4	-10	0	0.12	4	-10	0	0.29
15	14	9	0.52	5	-10	0	0.39	4	-60	0	0.51
18	14	10	0.74	4	-10	0	0.12	3	-60	0	0.15

Table 3.3 Classification results. The columns are *t*, the time to diagnosis in months; |St|, the number of cases with measurements taken ≥ *t* months before diagnosis; E1, the number of errors when classifying the triplets in St with CA-125 alone; p-value 1, the corresponding p-value; E2, the minimal number of errors when classifying with CA-125 plus one of the peaks 1-20; p2, the peak number with the best discriminative power, (the minus sign indicates an assigned weight of -1, otherwise +1); w2, is the value of weight assigned to CA-125 (0 or 1); p-value 2, the corresponding p-value; E3, the minimal number of errors when classifying with CA-125 and one of the peaks 1-100 (weight -1 or 1); p3 and w3, the peak number and the value of the weights for CA125, respectively, attaining E3 errors; p-value 3, the corresponding p-values.

These results suggest that using a combined classification approach (CA-125 plus MS peaks) ovarian cancer can be detected up to 12 months prior to diagnosis at a significance level of <0.05. Furthermore, when CA-125 and the 20 most frequent peaks were used in equal weight (E2), peak 10 (4292.5 m/z) was found to be the most discriminating feature. Peak 18 (3171.1 m/z) was also a significant discriminator when the 100 most frequent peaks were used. Interestingly, the changes in the area of peak 10 were complementary to changes in CA-125 levels in samples prior to the time of diagnosis (Figure 3.4).

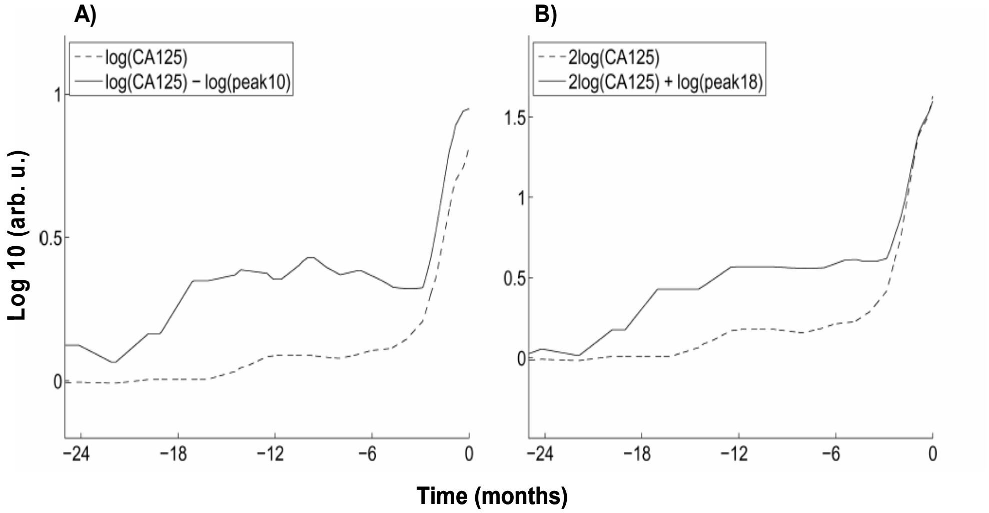


Figure 3.4 Dynamics of peak 10 combined with CA-125. Results suggest that using a combined classification approach (CA-125 plus MS peaks) ovarian cancer can be detected up to 12 months prior to diagnosis at a significance level of < 0.05. Here the dynamics of A) peaks 10 in combination with CA-125 (solid line) and CA-125 alone (dashed line). B) Peaks 18 in combination with CA-125 (solid line) and CA-125 alone (dashed line) are shown on a logarithmic scale. Each plot essentially shows the deviation of the value taken by the corresponding classification rule on the cases from the values taken on the controls. The horizontal axis shows the time to diagnosis in months.

3.4 Discussion

It is universally agreed that any technology platform for biomarker discovery needs to demonstrate high reproducibility and robustness. Analytical reproducibility is a significant challenge in protein profiling. A number of MALDI-TOF MS protein-profiling strategies have been developed for improved analytical performance. In particular, MALDI-TOF-MS protein profiling has been combined with advanced biostatistics to identify proteomic biomarker patterns for human diseases and improved reproducibility of the spectral output has been critical for avoiding false discoveries using this approach.

The use of mass spectrometry (MS) for the direct analysis of proteins and peptides from human serum for disease biomarker discovery was first reported in 2002 [Petricoin et al., 2002; Petricoin et al., 2002]. However, critics argued that the published results of serum profiling for diagnosis of ovarian cancer did not demonstrate reproducibility in independent subjects [Baggerly et al., 2004; Baggerly et al., 2005]. The apparent discrimination reported could be explained through the over-fitting of the data that occurs when multivariable models are used to fit a large number of possible predictors (such as mass spectrometry peaks) or by differences in sample handling of diseased and control samples to discriminate among a group of subjects with or without cancer [Petricoin and Liotta, 2004; Petricoin, III et al., 2002; Ransohoff, 2005].

Previously reported intra experiment CVs of peak intensities vary greatly between individual peaks, and reported mean CVs of the peak area varies across studies from 4% to 26% [de Noo et al., 2005; West-Norager et al., 2007]. However, often, only the most abundant and stable peaks are considered. In biomarker research utilizing MALDI-TOF MS profiling, the aim is to identify peaks that show consistent differences in intensities (or peak areas) between case and control samples, and thus the reproducibility of peak areas is of highest importance.

The data presented in this chapter was obtained through collaboration with a group at the MSKCC, who have previously established an automated magnetic beads-based strategy for serum peptide and protein extraction coupled MALDI-TOF MS-based profiling platform [Villanueva et al., 2004; Villanueva, 2006]. The work presented here assessed the reproducibility of the existing platform and found intra-assay reproducibility for the peaks to be below 20% (except run 1). However, significant inter-experiment variation was found (~33%) which is comparable to the CVs reported by others and thus considered acceptable. Run-to-run variation is a well-known problem with these types of high-throughput profiling strategies. Inter-assay variation is often high because of differences in the numbers of peaks obtained and possibly due to other confounding factors such as temperature and humidity variation from day-to-day affecting matrix-analyte co-crystallisation prior to MS analysis. Results presented in this chapter demonstrate how the analysis of the same sample over several runs under identical conditions can yield different peak numbers.

The matrix (co)crystallization and desorption/ionization steps in MALDI-TOF MS have been derived empirically, and the thermodynamic and physicochemical processes of phase transition and ionisation are poorly understood [Cohen and Chait, 1996]. Matrix molecules crystallize in different shapes and dimensions, proteins tend to accumulate at the droplet periphery, and the composition of the matrix solution and the rate of crystal growth influence the spectral output. These phenomena produce shot-to-shot variations, and are related to sampling different parts of the target surface and progressive sample ablation with repeated laser shots [Cohen and Chait, 1996]. Studies have also demonstrated ion suppression effects in MALDI-TOF MS. Ion suppression occurs when an ion suppresses the peak signal of other ions in the sample, and peptides with greater hydrophobicity show the greatest suppression effects. Peak area is thus associated with the concentration of the individual protein, to its primary structure, and to the complexity of the sample [Cohen and Chait, 1996]. In summary, peak area in MALDI-TOF MS profiling has significant analytical variation and is poorly understood.

Following the initial assessment of platform reproducibility an analysis of serum samples that pre-dated diagnosis of ovarian cancer was performed. The results presented here demonstrate that the predictive power of CA-125 alone was limited for early diagnosis. However, using a combined classification approach (CA-125 plus MS peaks) ovarian cancer can be detected up to 12 months prior to diagnosis at a significance level of <0.05. Of the 20 most frequently occurring peaks, statistical analysis showed that the information provided by peak 10 (4292.5 m/z) was complementary to the information provided by serum CA-125 measurements and the majority of OC cases show either a tendency of serum CA-125 growth or a tendency of peak 10 to decrease, or both. Furthermore, peak 18 (3171.1 m/z) was also a significant discriminator when the 100 most frequent peaks were used. The identification of these two peaks would facilitate immune-based assay development for detailed validation of these findings.

It is important to note that these conclusions are based on exploratory results (namely, the p-values obtained) and further work in needed to validate these findings. It can be speculated that these peaks maybe a fragments of host immune/acute phase proteins e.g. inter-α-trypsin inhibitor heavy chain (ITIH4) or Complement C3F, which are abundant serum proteins. These proteins have been shown to form peptide 'ladders' through tumour specific exopeptidase activity during the clotting process. Although, none of the peak masses matched those reported by Villanueva et al. it is important to note that the study involved an analysis of ovarian cancer samples, whereas previously, breast, bladder and prostate cancer samples were analysed [Villanueva et al., 2004; Villanueva, 2006]. In addition, the sample collection which was not standardized, handling and data analysis procedures were also different to those used by the group at MSKCC.

Finally, the technology platform was adapted in the host lab at UCL for the analysis of other cohorts of serum samples (following chapters).

Chapter 4: Optimisation of an automated magnetic bead-based extraction protocol for mass spectral profiling of human serum

4.1 Introduction

High-throughput sample preparation and protein profiling with MALDI-TOF MS analysis is a relatively new tool for diagnosis of human diseases. Indeed several groups have reported the use of peak pattern discrimination for the correct classification of ovarian cancer [Petricoin et al., 2002], prostate cancer [Adam et al., 2002; Qu et al., 2002], breast cancer [Li et al., 2002] breast, prostate and bladder cancer [Villanueva et al., 2006], although, at present none of these are in clinical use. It is universally agreed that any new technology platform for biomarker discovery needs to demonstrate high reproducibility and robustness [Srinivas et al., 2002]. However, a major obstacle to reliably determining quantitative changes in protein expression is to overcome errors imposed by technical and biological variation [Molloy et al., 2003].

A general mass spectrometry-compatible bead-based protocol for the extraction of serum peptides and proteins requires the optimisation of numerous experimental conditions. Factors to be considered include bead type used for peptide extraction, bead-to-serum ratio, pH, loading, washing, and elution conditions. Furthermore, the effects of pre-analytical variation from sample handling and storage need to be considered. Indeed the humidity, temperature, storage, and time for preparation of sera have all been shown to induce spectral changes [Timms et al., 2007; Villanueva et al., 2005; West-Norager et al., 2007].

The aims of the studies presented in this chapter were to adapt and establish a previously reported bead-based peptide and protein extraction and MALDI-TOF MS serum profiling platform in the host laboratory at UCL. The original protocol was optimised and adapted for MALDI-TOF MS analysis of human serum. The instrument used in the earlier studies was a Bruker Daltonics Autoflex TOF-TOF. However, in the host lab at UCL the available instrument was a Bruker Daltonics

Ultraflex TOF-TOF, thus mass spectral acquisition parameters were also optimised to establish a robust protocol at UCL. With respect to data processing and analysis ClinProTools software (Bruker) was used to evaluate the effects of modifications on the methods, since the 'bespoke' analysis methods developed at MSKCC were unavailable. In addition, the effects of different sample handling conditions on MALDI-TOF MS serum peptide profiles were also explored to determine a clinically feasible handling method for serum samples.

4.2 Optimisation of MALDI-TOF MS spectral acquisition parameters

Initial experiments used a cocktail of commercially available peptides and proteins (Table 4.1) in combination with manually prepared commercial serum (Sigma-Aldrich) to establish optimised spectral acquisition methods for the low (700-4000 Da) and high mass ranges (4-15 kDa). Spectral acquisition was split into two mass ranges to minimise errors associated with data processing i.e. baseline subtraction and smoothing. The purpose of baseline subtraction is to remove the broad structures of a spectrum and to create a baseline for the accurate selection of peaks based on signal-to-noise and intensity thresholds. Notably, noise levels increase in the high mass range and requires additional smoothing compared to the low mass range. Thus, it is advantageous to acquire and process spectral data in two separate mass ranges.

Calibrant	m/z
Peptide mix	
Peptide 782	782.04
Angiotensin II	1,047.20
Angiotensin I	1,297.51
Substance P	1,348.66
Bombesin	1,620.88
ACTH fragment 1-17	2,094.46
ACTH fragment 18-39	2,466.73
Protein mix	
Insulin	5,734.56
Ubiquitin	8,565.89
Cytochrome C	12,361.09
Myoglobin	8,476.77

Table 4.1 Calibrants used for MALDI-TOF MS spectral acquisition optimisation and calibration. Calibrant mixture was freshly prepared on the day of each experiment according to the method outlined in Chapter 2. Briefly, peptide and proteins were diluted and mixed to a final concentration of 30 femtomoles per peptide and 500 fmol per protein. All m/z values are calculated for single-charged ions except for myoglobin which generated a doubly-charged ion.

For MALDI-TOF MS profiling of human serum a number of MS spectral acquisition parameters were tested and optimised. These included the voltages applied to the lens and ion source, laser energy being delivered to sample spots and the raster file associated with the laser firing pattern. This was necessary to facilitate optimised and accurate automated detection of the peptides and proteins in the sample eluates using the instruments 'AutoXecute' function. The mass-to-charge ratios of the reference calibrant peptides and proteins were used to make the necessary adjustments to the MS acquisition protocol to ensure the observed masses matched expected masses within a mass error of ± 10 ppm (Figure 4.1).

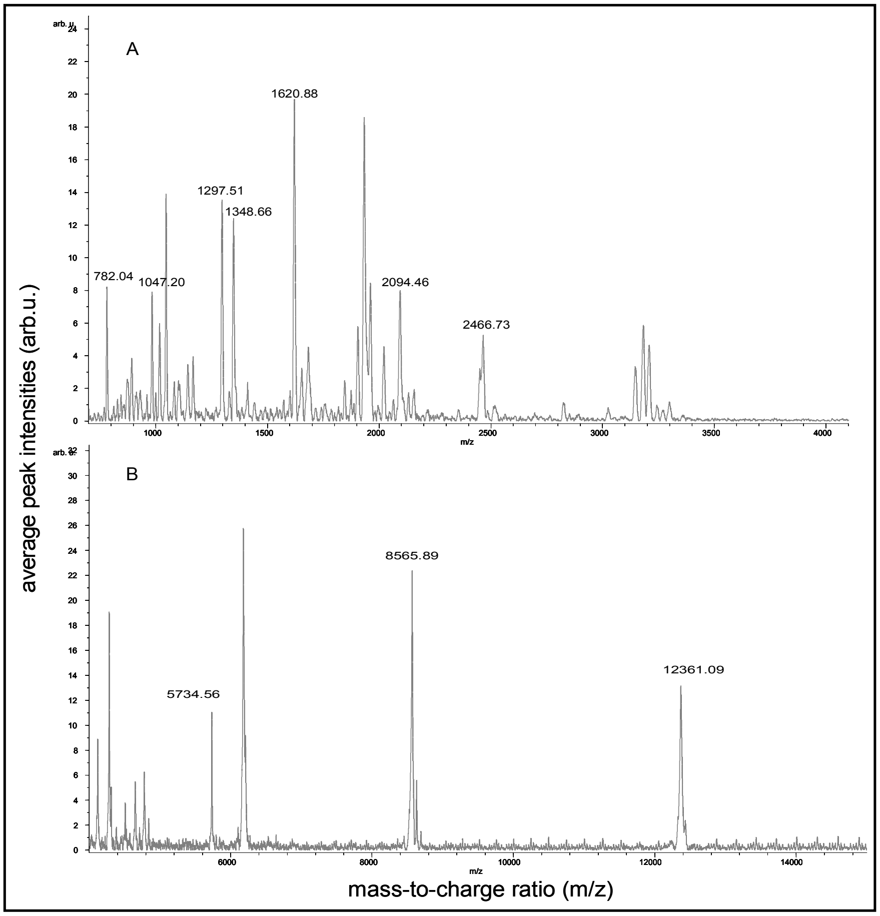


Figure 4.1 Mass spectral chromatographs illustrating acquired peaks from the calibrant mixture. A) Low mass range (700-4000Da) showing peptide calibrants and B) High mass range (4-15kDa) showing protein calibrants. Calibrants were freshly prepared according to the protocol described in Chapter 2 (Materials and Methods section) and spotted 1:1 (0.5 µL calibrant mix plus 0.5 µL α-CCA matrix) before MALDI-TOF MS analysis. MS parameters were adjusted to ensure accurate measurement of the peptide and protein m/z ratios with high intensity and low noise.

In parallel, to ensure efficient detection of serum peptides and proteins these adjustments were tested on serum eluates prepared using a manual bead extraction procedure. For this part of the study, C8 porous magnetic beads (Chemicell) used in the original study were used and were obtained from the MSKCC. This reagent is now no longer commercially available. MS spectra of two replicate samples were analysed in ClinProTools software to determine peak numbers. At this stage, an average of 143 and 109 serum peaks were detected in the LMR and HMR respectively (Figure 4.2). Overall average CVs were 18.7 % in the LMR and 19.4 % in the HMR. After the optimisation of the MS spectral acquisition parameters, the bead-based extraction protocol was optimised on a liquid handling robot for high-throughput sample analysis (Chapter 2 section 2.2.2).

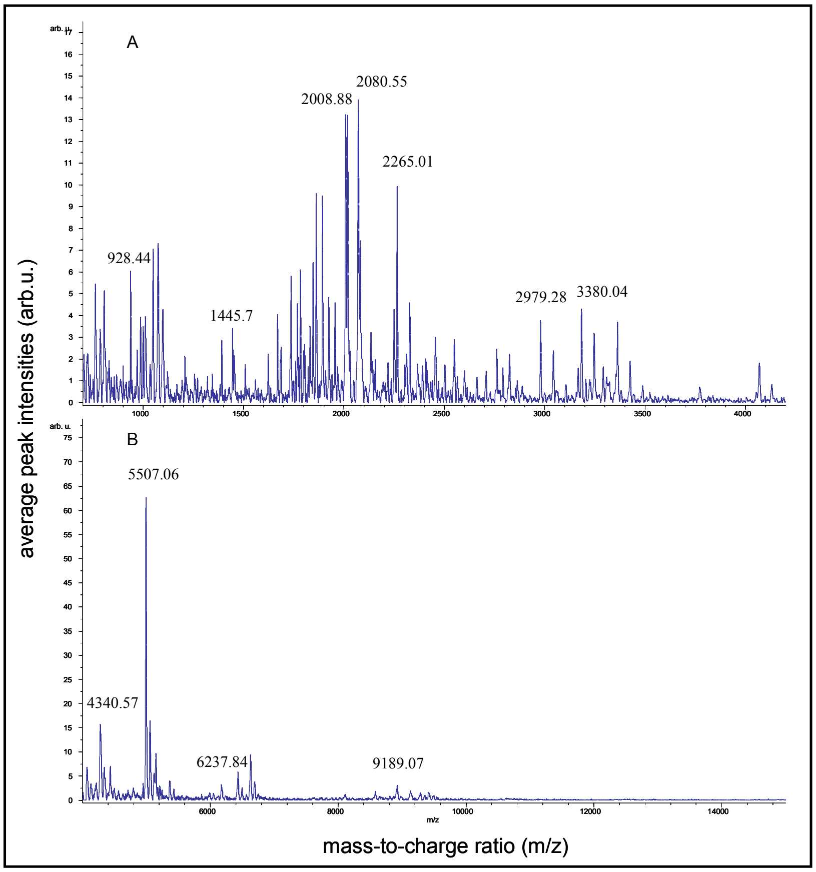


Figure 4.2 Mass spectral chromatographs of commercial serum. Spectral profiles generated during MS spectral acquisition optimisation are shown. Serum samples were manually processed as described in Chapter 2 and spotted 1:1 (0.5 µL sample eluate plus 0.5 µL α-CCA matrix) before MALDI-TOF MS analysis. Data was analysed in ClinProTools software A) Low mass range (700-4000 Da) and B) High mass range (4-15 kDa).

4.3 Optimisation of magnetic bead-based extraction protocol at host institute.

One of the methods recently adopted for the extraction of peptides and proteins from human serum is reversed-phase (RP) capture on magnetic beads derivatised with a variety of binding chemistries. By using magnetic beads coated with these chemistries it is possible to automate the capture of peptides and proteins using liquid handling robotics thereby providing high throughput and reproducibility in sample processing. Indeed, the capture of peptides and proteins using RP batch processing in a magnetic bead-based format was previously automated at the MSKCC on a liquid handling robot allowing simultaneous processing of 100s of serum samples at a time [Villanueva et al., 2004; Villanueva et al., 2006]. The RP reagents utilised in the original study (C8-coated magnetic beads) were no longer commercially available therefore the performance of beads from alternative sources was investigated for the establishment of an automated serum profiling platform at UCL.

The nature of the stationary phase is a vital factor in any chromatographic separation, since it determines the retention of specific ligands. For example, weak cation exchange beads carry negative surface charges which reversibly adsorb oppositely charged proteins and large peptides. Bound samples can be sequentially removed using a step-wise elution with increasing concentrations of salts, decreasing the pH or a combination of both. Reverse phase (RP) beads are usually coated with carbon chains for peptide and protein binding through hydrophobic interactions. Sequential removal is then achieved by increasing the concentrations of organic solvents such as acetonitrile. The length and deposition of carbon chains onto the bead surfaces affect the selectivity for peptide binding. Larger polypeptides are preferentially captured on less hydrophobic surfaces (C1-3), while smaller peptides are captured on more hydrophobic surfaces (C8-18). These chemistries have been applied to magnetic beads to facilitate batch-wise binding and elutions.

4.3.1 Optimisation of bead type and bead-to-serum ratio

Initial experiments were performed to compare the number of peptide species captured by different commercially available magnetic beads. This was done to determine the most appropriate stationary phase and ‘bead-type’ for serum peptide and protein extraction. Three main types of RP magnetic beads are commercially available from a number of manufactures including BioClone, Bruker Daltonics and Invitrogen (Dynal beads). The polypeptide affinity of C4, C8 and C18 RP beads, as well as a combination of C4 and C18 beads from BioClone, Bruker Daltonics and Invitrogen (Dynal) were compared. In addition, differing bead slurry-to-serum volume ratios (1:1, 1:2, 1:4 and 1:8) were tested after equalisation of bead concentration (to 2µg/µL) to determine the optimal ratio for capturing and recovery of peptides and proteins. Since batch preparation methods were the most effective way to carry out cross-comparisons of multiple conditions concurrently, sample conditions were standardised by using 50 µL of Sigma serum for all comparisons. Each condition was run in triplicate on the robot and spotted in duplicate onto a MALDI target and the 6 spectra per condition were acquired using the MS parameters optimised in the previous section. The MALDI-TOF MS spectra were processed using ClinProTools software (V2.0) as described in Chapter 2 section 2.2.4, and the average spectra and peak statistics were compared. Overall, a similar number of peaks were detected in each bead type in both the LMR and HMR (Figures 4.3 and 4.4). However, C8 beads alone performed marginally better (more consistently) then C18 beads alone in the high mass range-as expected.

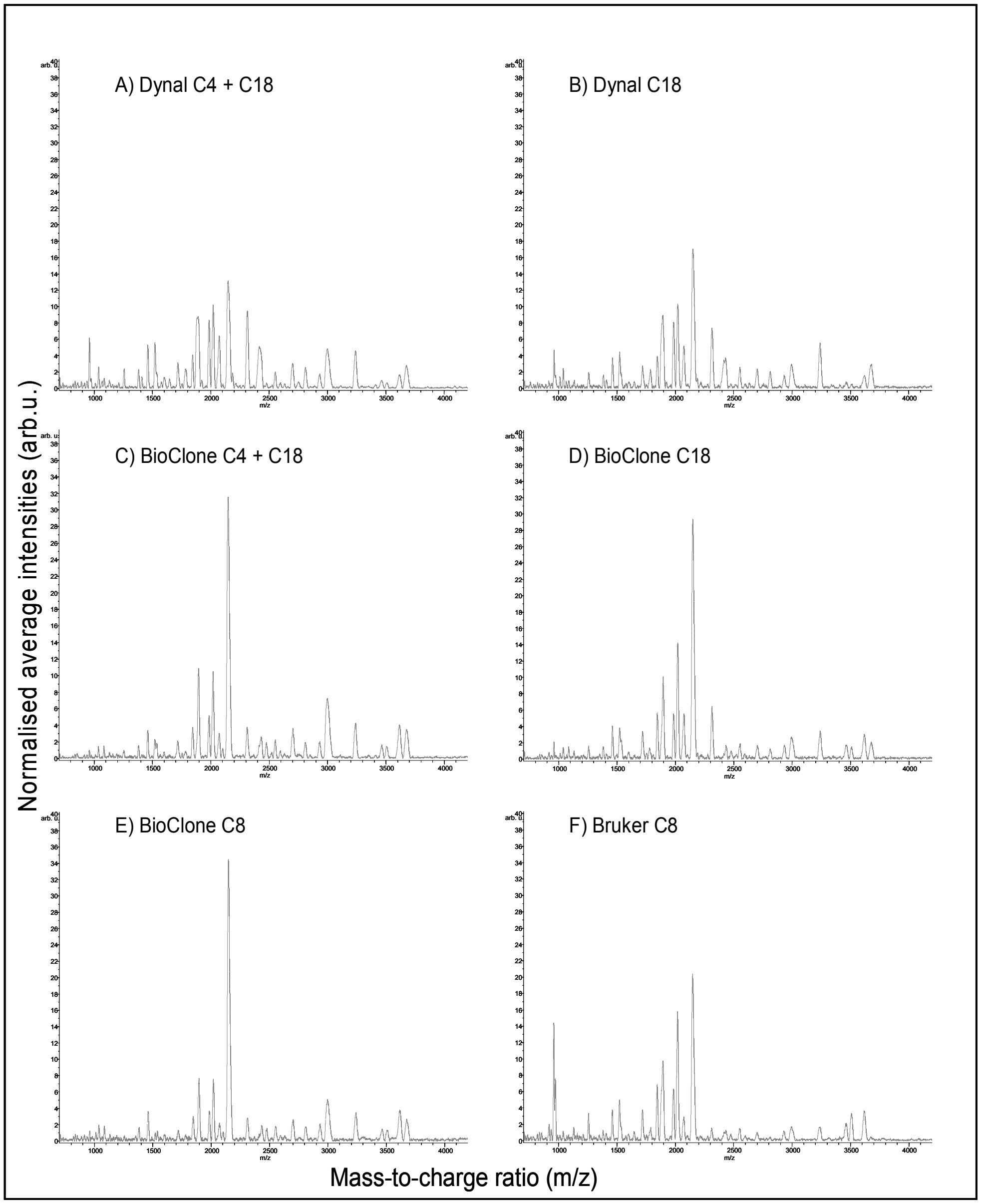


Figure 4.3 MALDI-TOF spectra illustrating the average LMR (700-4000Da) serum peptide profiles used in the comparison of magnetic beads. Average spectra are shown for Sigma serum (n=6) extracted with equal amounts of (A) Dynal C4 and C18 RP bead slurry. (B) Dynal C18 beads. (C) Equal amounts of BioClone C4 and C18 RP bead slurry; (D) BioClone C18 beads; (E) BioClone C8 beads; (F) Bruker C8 beads. 12.5 µL of bead slurry was used for extraction of Sigma serum essentially as described in Chapter 2 section 2.2.2, eluate was mixed 1:1 with α-HCCA, and 1 µL was spotted in duplicate for MALDI-TOF MS analysis. Samples were run in triplicate and the average spectra generated by ClinProTools (V2.0) are presented here.

Generally for all bead types, a 25 µL bead slurry volume was found to capture the highest number of peptide peaks above an S/N ratio of 3 in the LMR and 5 in the HMR (Figure 4.4), although there were some fluctuations. This was perhaps surprising since one may expect more peptides to be detected at higher bead-to-serum ratios. This hints at ion suppression effects.

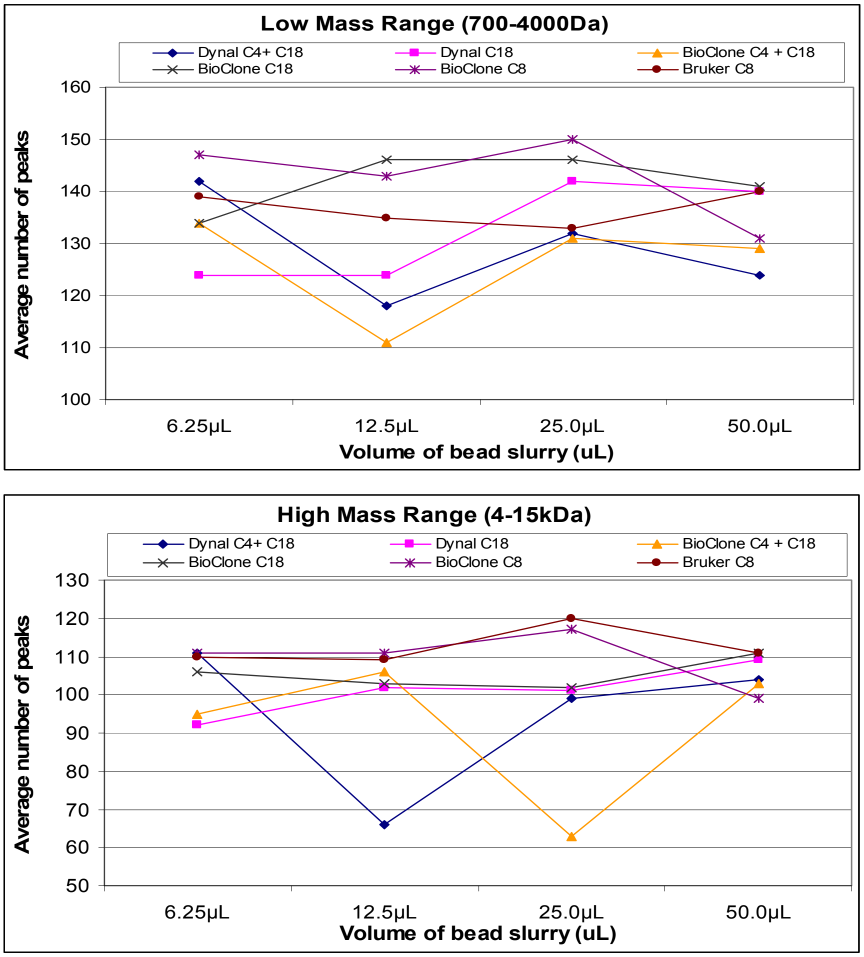


Figure 4.4 MALDI-TOF comparison of the extraction of serum polypeptides using different coated magnetic beads. A volume of 50 µL of neat serum was mixed with 6.25 µL, 12.5 µL, 25 µL and 50 µL of bead slurry (all equalised to 2 µg/µL). This mixture was incubated at room temperature for 2 minutes. Beads were then pulled side-to-side 10 times with a magnet and allowed to settle on one side of a PCR tube before supernatant was removed and discarded. Beads were then washed twice with 0.1% TFA. Bound peptides were eluted with 5 µL of 50% ACN. Eluate was mixed 1:1 in MALDI matrix α-HCCA and 1 µL was deposited on a normal stainless steel MALDI target for MS analysis. The number of peaks obtained for bead conditions were calculated for (A) LMR and (B) HMR using ClinProTools software (V2.0) and S/N cutoffs of 3 & 5, respectively.

Indeed, analysis of the changes in the spectra showed that increasing the bead amount led to a reduction in the peak area of lower mass range peaks (Figure 4.5). The high mass range peaks were relatively unaffected. Furthermore, there was some fluctuation in the number of peptide peaks detected using 12.5 µL versus 25 µL in the LMR (Table 4.2). This suggests that concentration of the eluate affects the low mass range signal possibly as a result of poor sample crystallisation and/or ion suppression which maybe due to 'competitive' binding of peptide and protein species.

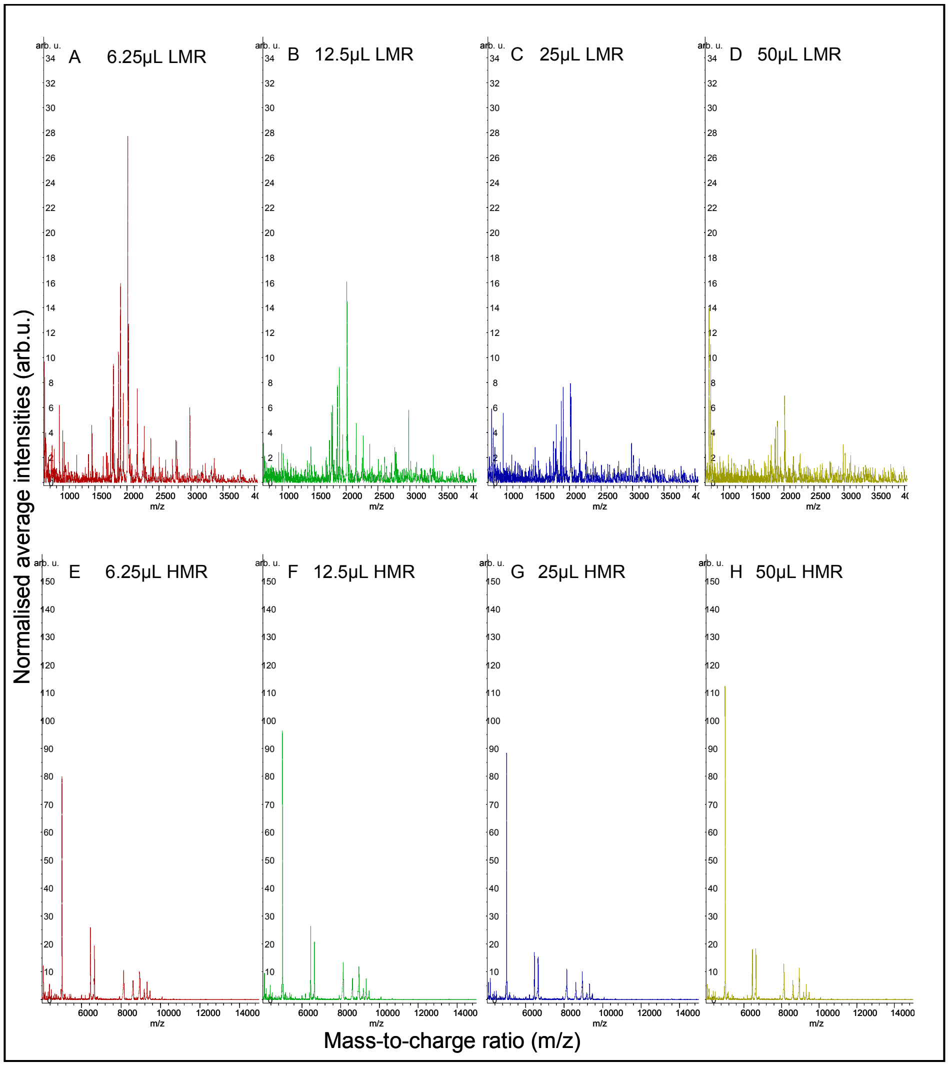


Figure 4.5 MALDI-TOF spectral comparisons of extraction using different serum-to-bead ratios. A-D show Low Mass Range (700-4000 Da) peptide profiling while E-H show High Mass Range (4-15 kDa). Results for Dynal C18 beads and representatives spectra of all bead slurry volumes tested are shown here. Data were processed using ClinProTools software (V2.0).

The reproducibility of each bead type was also investigated (Table 4.2). The overall coefficient of variance (CV) for all detected peaks in the LMR ranged varied from 19.4% to 45.4% and 9.9% to 15.8% in the HMR. These CVs were subsequently improved through modifications of the robotic volume aspirations and dispension speeds which minimised bead loss during the washing steps. Apart from a few outliers reproducibility across the beads and dilutions was found to be similar. On comparison of the quality of spectra, Dynal C18 beads were found to be marginally better overall at capturing serum polypeptides. Dynal C18 beads also proved to be the most cost effective option.

Bead type	Slurry Vol. µL	Ave number of LMR peaks	Ave LMR peak area (arb. u.)	CV (%)	Ave number of HMR peaks	Ave LMR peak area (arb. u.)	CV (%)
Dynal C4+ C18	6.25	142	21.30	30.59	111	75.02	12.95
	12.50	118	28.76	20.43	66	98.60	12.50
	25.00	132	24.49	20.14	99	92.31	11.73
	50.00	124	27.17	20.27	104	56.71	12.06
Dynal C18	6.25	124	25.73	23.37	92	85.99	11.52
	12.50	124	26.81	20.53	102	58.55	11.56
	25.00	142	20.15	22.29	101	61.06	10.65
	50.00	140	22.19	19.67	109	73.47	10.53
BioClone C4 + C18	6.25	134	25.37	20.06	95	82.11	9.87
	12.50	111	30.92	19.37	106	70.71	11.63
	25.00	131	17.78	20.98	63	103.39	12.54
	50.00	129	26.13	20.21	103	79.93	11.50
BioClone C18	6.25	134	24.19	23.46	106	78.85	12.01
	12.50	146	13.83	27.45	103	77.19	14.21
	25.00	146	19.88	25.59	103	76.56	11.56
	50.00	141	23.51	21.58	111	48.14	12.48
BioClone C8	6.25	147	22.32	30.15	111	75.95	15.80
	12.50	143	23.80	22.06	111	73.86	13.83
	25.00	150	20.95	20.57	117	58.05	10.37
	50.00	131	25.39	23.76	99	82.53	12.20
Bruker C8	6.25	139	23.26	20.30	110	73.51	11.10
	12.50	135	23.95	29.19	109	75.08	14.73
	25.00	133	23.58	33.55	120	49.90	12.66
	50.00	140	21.45	45.40	111	71.18	14.19

Table 4.2 Intra-assay reproducibility for each bead type and condition. Spectra were analysed using ClinProTools software (V2.0). Resulting average peak areas were used to calculate co-efficients of variance. The overall average peak number, average peak area (arb. u.) and CV in the LMR (S/N > 3) and HMR (S/N > 5) are shown.

Dynal C18 beads are reported to have a binding capacity of 14 µg peptide per mg of beads. The concentration of protein, determined by Bradford assay in Sigma serum was 90 mg/mL and was therefore 3000 fold in excess of the bead binding capacity. It can be speculated that perhaps more non-specific binding occurs with an increase in the volume of the bead slurry. Furthermore, the presence of high abundant protein species may lead to suppression from signal of low abundant polypeptides. Indeed Figure 4.6 shows several high-abundant serum proteins can bind to C18 magnetic beads. Using larger bead volumes may also lead to poorer elution of bound peptides, as the large bead volume makes the removal of the entire elution fraction more difficult. Perhaps repeat elutions could overcome this problem.

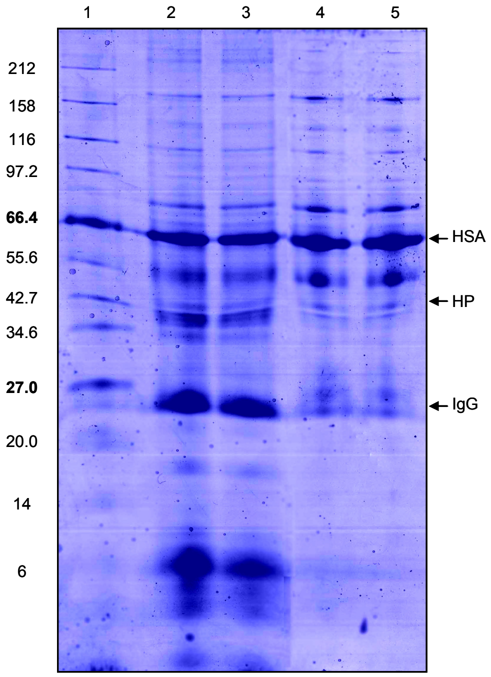


Figure 4.6 1D-gel view of C18 magnetic bead-extracted serum proteins. Lane 1 shows molecular weight markers, lane 2 and 3 10µg of Dynal C18 bead-extracted serum eluate and lanes 4 and 5 show 10µg of unfractionated serum. It is evident that C18 beads capture larger proteins such as human serum albumin (HSA), haptoglobin (HP) and immunoglobulin G (IgG).

Subsequent to this work, experiments showed that diluting eluates 1:5 with 50% ACN increased the signal and peptide peak numbers. This provided evidence for ion suppression with the dilution reducing the concentration of sample applied to the MALDI target, allowing for more effective detection of sample peptides by mass spectrometry. Furthermore, the MALDI-TOF analysis of eluates showed higher signal-to-noise ratios for peptide species detected from Dynal C18 beads. This increased ion signal may have been partially a result of stronger hydrophobic interactions and better compatibility with elution solvents.

These initial experiments demonstrated that effective peptide capture could be achieved using all bead types, however, because of the quality of spectra and cost per assay, Dynal C18 were selected for further use. Although not optimal, for economical reasons 12.5 µL bead slurry volume was selected as being most reproducible during automated pipetting on the liquid handling robot.

4.3.2 Further investigation on the effects of altering serum volume

In further experiments to define the optimal serum-to-bead volume ratio, 12.5 µL of 50% Dynal C18 bead slurry (2 µg/µL) was manually mixed and incubated with 10, 20 and 50 µL of commercial serum for 2 minutes (incubation time recommended in manufacturer's instructions). The supernatants were removed and the beads were washed twice. After washing the bound polypeptides were eluted with 5 µL 50% ACN and mixed with matrix solution. The eluate/matrix mixture was applied directly to a MALDI target and left to air dry before MALDI-TOF MS analysis using the optimised acquisition parameters described in section 4.1. Results showed that similar numbers of peptides could be captured across the serum volume range, although overall the 50 µL volume gave the highest total number of peptides (Figures 4.7 & 4.8).

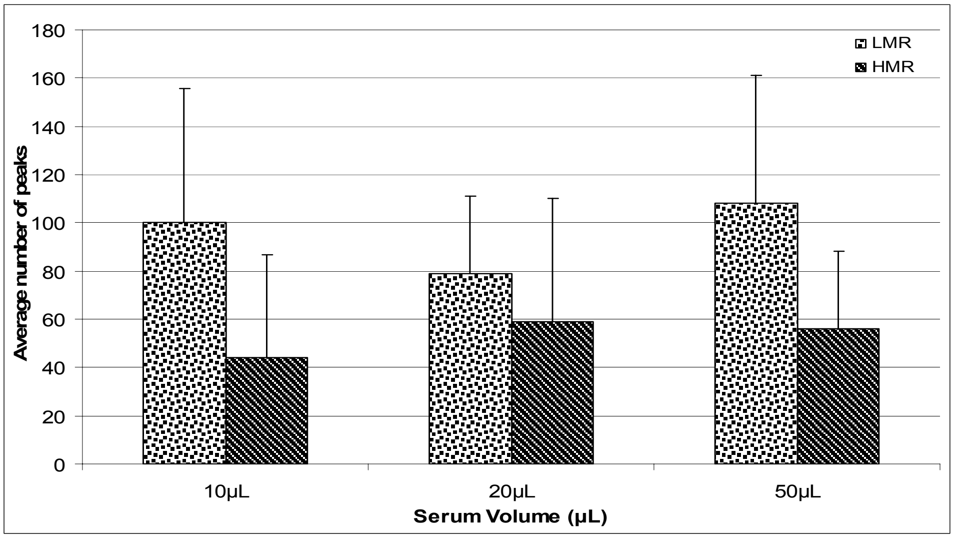


Figure 4.7 Optimisation of serum volume. Serum peptides were extracted as described in the text above. Data was analysed using ClinProTools software (V2.0) to calculate the number of peaks detected in the LMR (S/N >3) and HMR (S/N >5).

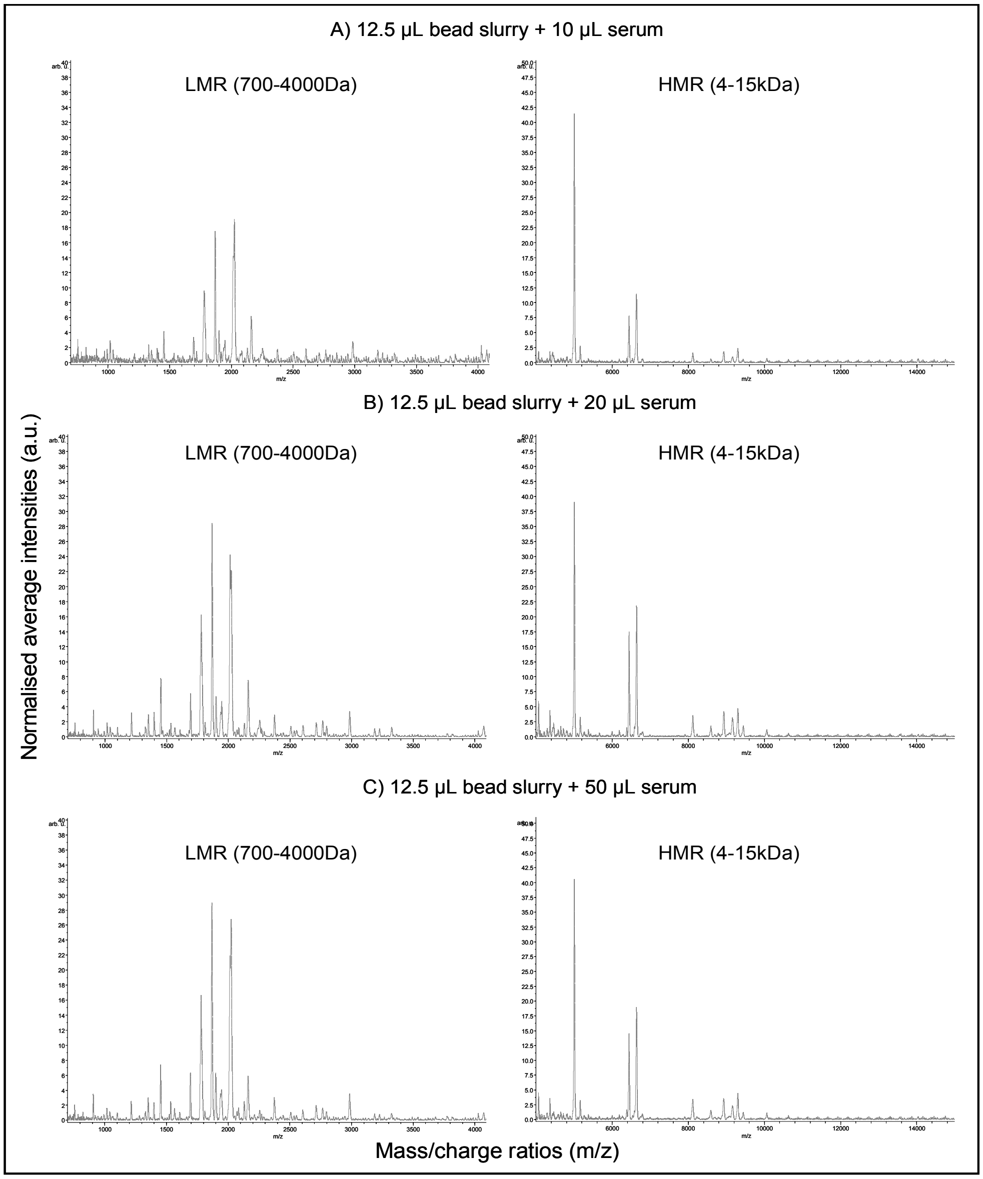


Figure 4.8 MALDI-TOF spectral comparisons following extraction of peptides and proteins with varying serum volume. Serum samples were manually processed using 12.5 µL of 50 % Dynal C18 bead slurry (2 µg/µL) using (A) 10 µL, (B) 20 µL and (C) 50 µL of commercial serum. Five µL elution solution (50% ACN) was added to each bead pellet, incubated, and the eluate transferred to a fresh tube. Eluate was mixed 1:1 with α-CCA matrix solution, 1 µL was spotted onto a MALDI plate and allowed to air dry before MALDI-TOF MS analysis. Data was analysed in ClinProTools software (V2.0) and the average spectra are shown here on the same scale.

To assess the reproducibility, the average CVs for the peak areas from all matched peaks across 8 spectra within each condition were compared (Table 4.3). A serum volume of 50 µL gave an average peak area CV of 11.1% ± 1.5 for the LMR and 10.8 % ± 0.9 for the HMR. Results suggest that changing serum volume made little difference to the number of peaks detected and the peak CVs.

Intra sample peak variability

	Ave peak CV (%)	
Serum Volume	LMR	HMR
10µL	11.03 (+/-1.5)	9.54 (+/- 1.8)
20µL	10.14 (+/- 1.0)	11.24 (+/-2.4)
50µL	11.1 (+/- 1.5)	10.82 (+/- 0.9)

Table 4.3 Intra-sample average peak area variability. Varying volumes of sigma serum samples were processed manually in quadruplicate and spotted in duplicate on a MALDI target before MALDI-TOF MS analysis. Data was analysed in ClinProTools software to calculate the average CVs for all peaks.

4.4 Intra- and inter-assay reproducibility

A major criticism of early MS-based serum profiling was the lack of experimental reproducibility of datasets within and between studies [Baggerly et al., 2004]. Many studies only report reproducibility based on a few of the most intense peaks. Following the platform optimisation studies discussed so far, C18 Dynabeads were selected to undergo a more thorough investigation of the robustness of the automated technology platform. Inter- and intra-assay reproducibility was evaluated using 50 µL commercial serum with 12.5 µL of C18 Dynabeads (2 µg/µL). Six aliquots of standard serum were processed in the same automated run, and the CVs were calculated for all the peak areas between the samples as a measure of the intra-assay reproducibility. The preparation and analysis was then repeated once a day over 7 days to yield inter-assay reproducibility values (Figure 4.9).

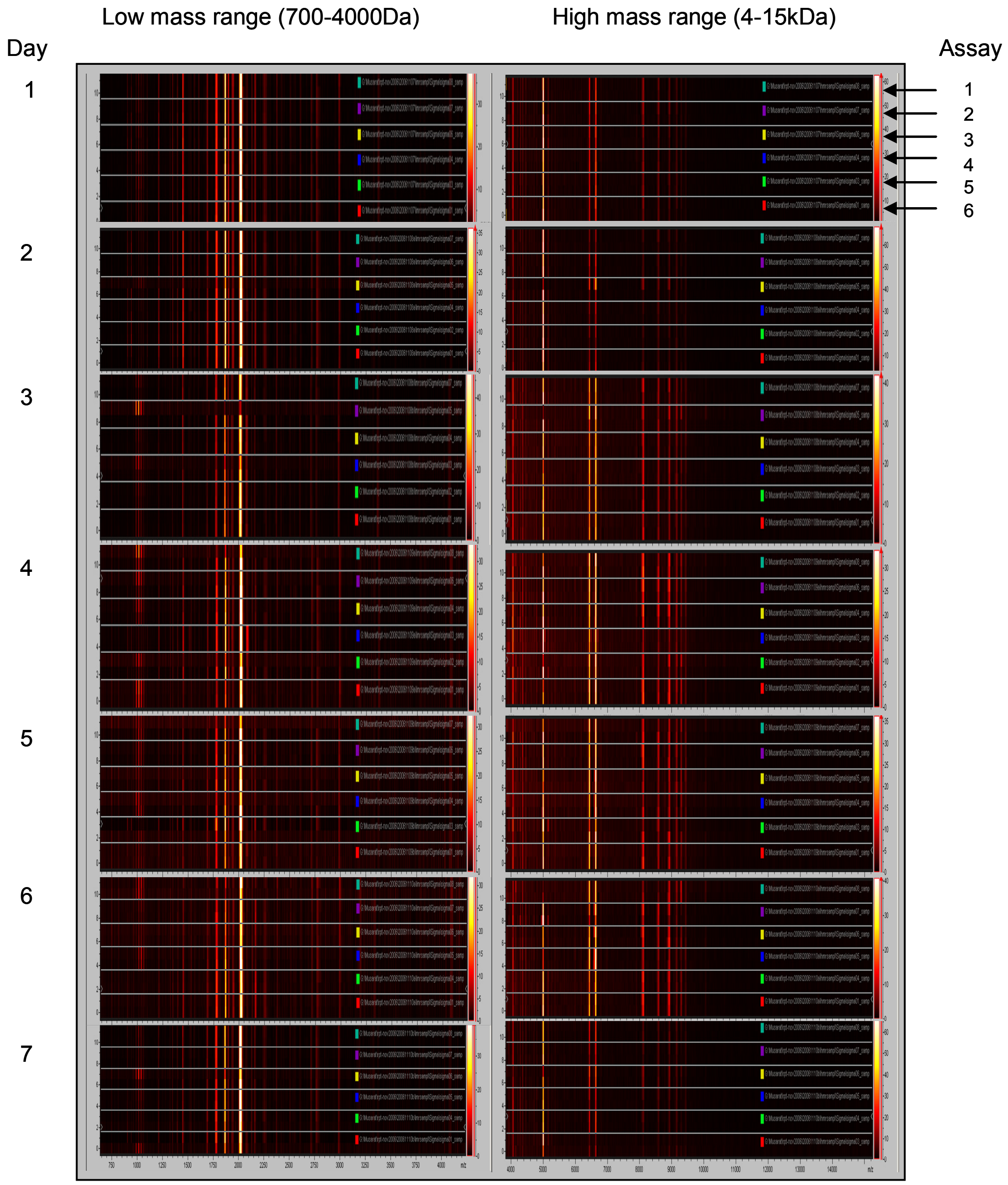


Figure 4.9 Heat map representations of spectra from intra- and inter-assay reproducibility test. Over the course of 7 days, six aliquots of commercial serum were processed each day on the automated platform. Data was analysed by ClinProTools software (V2.0) to calculate intra-and inter assay reproducibility based on peak area.

The intra-assay reproducibility for all the peaks detected varied significantly across the runs. In the LMR an average number of 69 peaks with an average CV of all peak areas of 10.3% ± 1.6 were found in Run 1 compared to an average of 133 peaks with an average CV of 20.9% ± 17.3 in Run 4. In the HMR an average of 34 peaks with an average CV of 14.3% ± 3 was found in Run 1 compared to an average of 105 peaks with average CV of 11.4% ± 1.6 in Run 7. Taking the average peak statistics for all 7 runs the overall inter-assay CV values were 28.4% ± 5.4 and 17.68% ± 3 for the LMR and HMR respectively (Table 4.4). The CV values are comparable to those previously reported in literature and in chapter 3.

	Ave number of peaks		Overall Ave CV	
Run Number	LMR	HMR	LMR	HMR
1	69	34	10.3 ± 1.6	14.3 ± 3
2	75	89	12.8 ± 12.4	14.4 ± 5.4
3	93	79	20.9 ± 17.3	12.6 ± 3.6
4	133	74	24.1 ± 15	14.5 ± 3.4
5	73	59	24.4 ± 11.7	15.4 ± 5.4
6	79	42	12.5 ± 7.5	14.6 ± 3.3
7	87	105	16.1 ± 9.0	11.4 ± 1.6
Inter-assay	118	123	28.4 ± 5.4	17.7 ± 3

Table 4.4 Intra/inter-assay reproducibility. The intra-assay reproducibility was calculated by running 6 samples per run, while inter-assay reproducibility was measured by assessing the variation between 7 runs. Average peak areas for the matched peaks were calculated using ClinProTools software and average co-efficient of variance (SD/peak area) were calculated for the areas of all peaks detected.

4.5 Quantifying variations in serum polypeptide profiles under different sample handling conditions.

Several factors termed pre-analytical variables have been reported to influence the observed serum proteome by affecting the stability of proteins [Rai et al., 2005]. These include the potential variations introduced during the actual blood collection and handling procedure. For example, variation in sample storage tube type, variable blood clotting times, transport conditions, time to centrifugation, storage conditions and freeze-thaw cycles may all affect the resulting profiles. Evidence also suggests disease specific exo-proteases are active during the clotting process which produce changes in protein degradation that are detectable by mass spectrometry [Villanueva et al., 2006]. Prolonged clotting times and repeated freeze thaw cycles could accelerate these complex break-down processes. As a result, peptides and proteins shed from tumours, typically low in abundance, could be lost in the background variance produced by inconsistencies in sample handling [Timms et al., 2007; Villanueva et al., 2005; West-Nielsen et al., 2005].

The effect of six different processing protocols on MALDI-TOF MS profiles was thus assessed using serum samples from 25 healthy volunteers from the UKCTOCS study. These women gave six blood samples that were processed as indicated in Table 4.5. Changes in the protocol were made to test the effects of storage tube type, varied clotting time, transport conditions, time to centrifugation, storage conditions and freeze thaw cycles. These protocols were chosen to be clinically feasible for the collection and processing of multiple samples, possibly from different centres, and included the standard UKCTOCS protocol (Protocol 1, Green) and that used by the Tempst group at the MSKCC (Protocol 2, Yellow).

Protocol No.	Colour code	Tube type	Mixing	Clotting	Storage	Time to centrifugation	Aliqouting & Stroage at -80°C
1	GN	Greiner gel tubes	Slowly inverted 5x	RT	RT	30 hrs from collection	Straws at RT
2	YE	BD tiger top tubes	Slowly inverted 5x	RT-60'-vertical position	Wet ice-vertical	3 hrs from collection	Straws at RT
3	GY	BD tiger top tubes	Slowly inverted 5x	RT-5'-vertical position	Wet ice-vertical	3 hrs from collection	Straws at RT
4	CR	BD tiger top tubes	Slowly inverted 5x	RT-5'-vertical position	Wet ice-vertical	3 hrs from collection	Cryovials one freeze thaw
5	OR	BD tiger top tubes	Slowly inverted 5x	RT-5'-vertical position	Wet ice-vertical	6 hrs from collection	Straws at RT
6	WH	BD tiger top tubes	Slowly inverted 5x	RT-5'-vertical position	RT	3 hrs from collection	Straws at RT

Table 4.5 Protocol comparison using 25 healthy volunteer samples. The shaded boxes indicated where changes were introduced in each protocol.

For variance analysis each volunteer sample was run in triplicate using 50 µL of serum, and 12.5 µL of Dynal C18 beads. The average number of peaks from each triplicate was first compared across the different handling methods using ClinProTools software V2.0, (optimised parameters for peak detection are outlined in Chapter 2 section 2.2.4.2). Briefly, spectra were first subjected to a 0.80 level baseline subtraction. Following this the detection of peaks was based on the analysis of a smoothed first derivative where the smoothing was determined by a given resolution parameter. Once the peaks were detected peaks were aligned and peak areas were calculated by integrating the intensities over the region of the peak. It was found that the number of peaks for each triplicate was protocol dependent (Table 4.6 and Figure 4.10). Protocol 2 (YE) gave the highest average number of peaks in the low mass range, (LMR 700-4000 Da), closely followed by the protocol 1 (GN), while protocol 3 (GY) gave the highest number of peaks in the high mass range (HMR 4-15 kDa).

Volunteer Ref	GN-LMR	GN-HMR	YE-LMR	YE-HMR	GY-LMR	GY-HMR	CR-LMR	CR-HMR	OR-LMR	OR-HMR	WH-LMR	WH-HMR
11005404	95	34	41	27	39	44	26	38	71	28	71	30
11006217	146	58	97	NA	96	26	118	35	97	42	61	34
11006440	65	30	40	38	36	40	102	26	74	76	68	41
11006863	70	30	169	33	63	36	59	30	62	55	90	32
11010114	98	28	167	35	88	57	92	54	41	31	57	46
11023647	134	44	61	49	33	19	88	49	31	26	73	56
11026386	64	45	82	30	62	36	23	40	92	30	80	32
11034276	142	35	80	30	47	44	131	38	121	40	38	31
11056019	50	29	50	33	104	27	41	NA	98	29	127	35
11068507	40	21	166	34	107	43	23	26	76	19	71	31
11074916	136	21	83	43	48	27	69	33	56	34	36	39
11084213	101	23	120	33	152	55	69	39	59	21	167	53
11093385	42	31	79	53	122	47	55	28	108	39	54	33
11099224	101	34	42	39	59	50	114	31	64	45	73	36
11099568	56	38	119	44	72	33	113	69	59	37	82	39
11101746	36	33	73	40	83	39	41	25	58	39	36	32
11101936	126	33	33	NA	47	47	60	37	76	36	66	33
11102206	54	37	105	42	76	44	83	37	34	24	53	26
11102296	109	46	164	29	40	37	41	28	96	33	116	35
11102557	66	29	67	37	79	27	42	35	106	26	71	61
11118143	84	47	52	58	68	31	112	33	63	33	52	38
11122713	36	42	148	32	154	39	71	38	47	35	72	39
11122791	95	36	53	44	107	55	62	42	93	41	168	48
11126095	96	28	14	53	61	44	118	35	73	45	135	34
11126328	42	35	101	43	134	45	130	37	106	36	52	16
Overall Average	83	35	88	39	79	40	75	37	74	36	79	37
SD	35	9	46	8	35	10	35	10	24	12	37	10

Table 4.6 Average number of spectral peaks. The overall average peak numbers and SD were calculated by aligning the spectra from all 25 volunteers in ClinProTools. Average peak numbers by collection method and volunteer are shown.

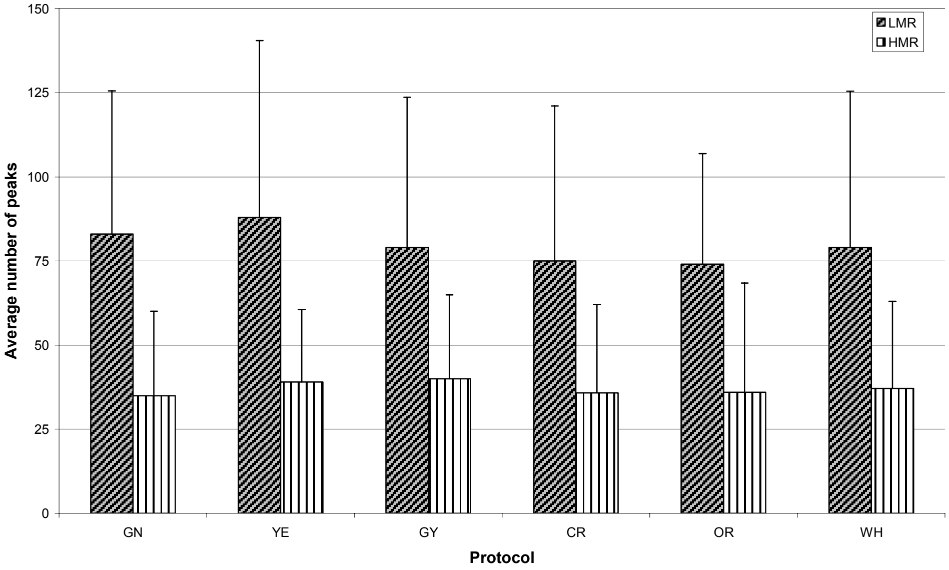


Figure 4.10 Overall average numbers of peaks across all protocols. The average numbers of peaks from triplicate samples of 25 volunteers are shown for each protocol in each mass range; LMR (700-4000 Da) and HMR (4-15 kDa). Protocol 2 (YE) showed the highest number of peaks in the LMR, while protocol 3 (GY) gave a slightly higher number of peaks in the HMR.

Despite the high variance in the peak numbers, certain peaks were common to one or more of the protocols. The average variance of all the peaks detected was compared across the protocols. Overall average variance was ~ 12% in the LMR and ~ 11% in the HMR. No one protocol produced higher variance in all volunteers in either mass range (Figure 4.11).

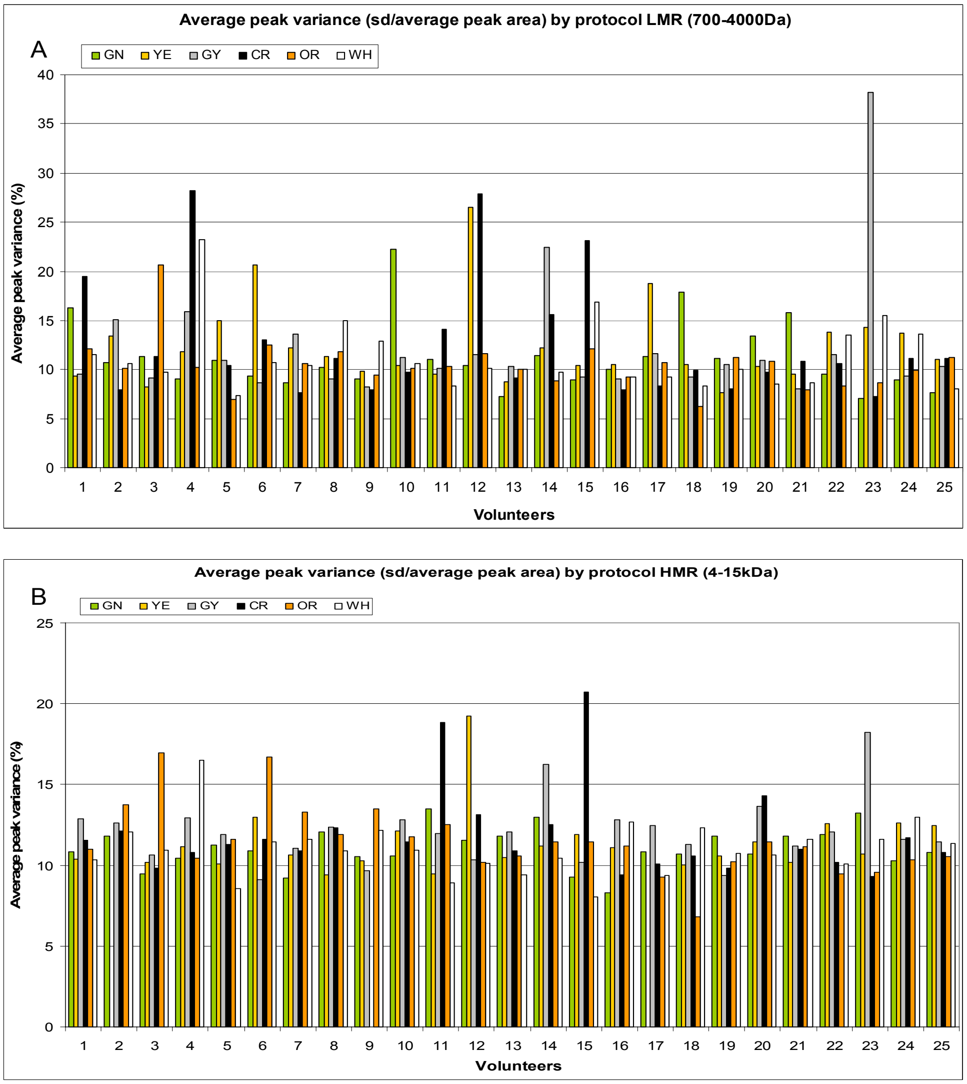


Figure 4.11 Average peak area variance analyses. (A) Low mass range (700-4000Da), overall average CV (standard deviation/average peak area) of all peaks from 3 replicates of the same sample by protocol and volunteer, where GN is protocol 1, YE ; 2, GY ; 3, CR ; 4 OR ; 5 & WH ; 6. (B) High mass range (4-15 kDa), average CV of all peaks by protocol and volunteer.

Heat maps were generated in ClinProTools software for peak profile analysis. The average peak areas from each individual, for each protocol were compared. This was done to assess which of the peaks could be directly attributed to differential sample preparation. Discriminatory peaks in several m/z regions from both mass ranges were found to discern protocol 1 (GN) from the other 5 protocols. In the LMR three peaks (1060, 1465 and 3198 m/z) were found to have lower peak areas in protocol 1 compared to the other 5 protocols (Figure 4.12). In the HMR two peaks (5895 and 12589 m/z) consistently had lower peak areas in protocol 1 and two peaks (4048 and 8122 m/z) had higher peak areas (Figure 4.12).

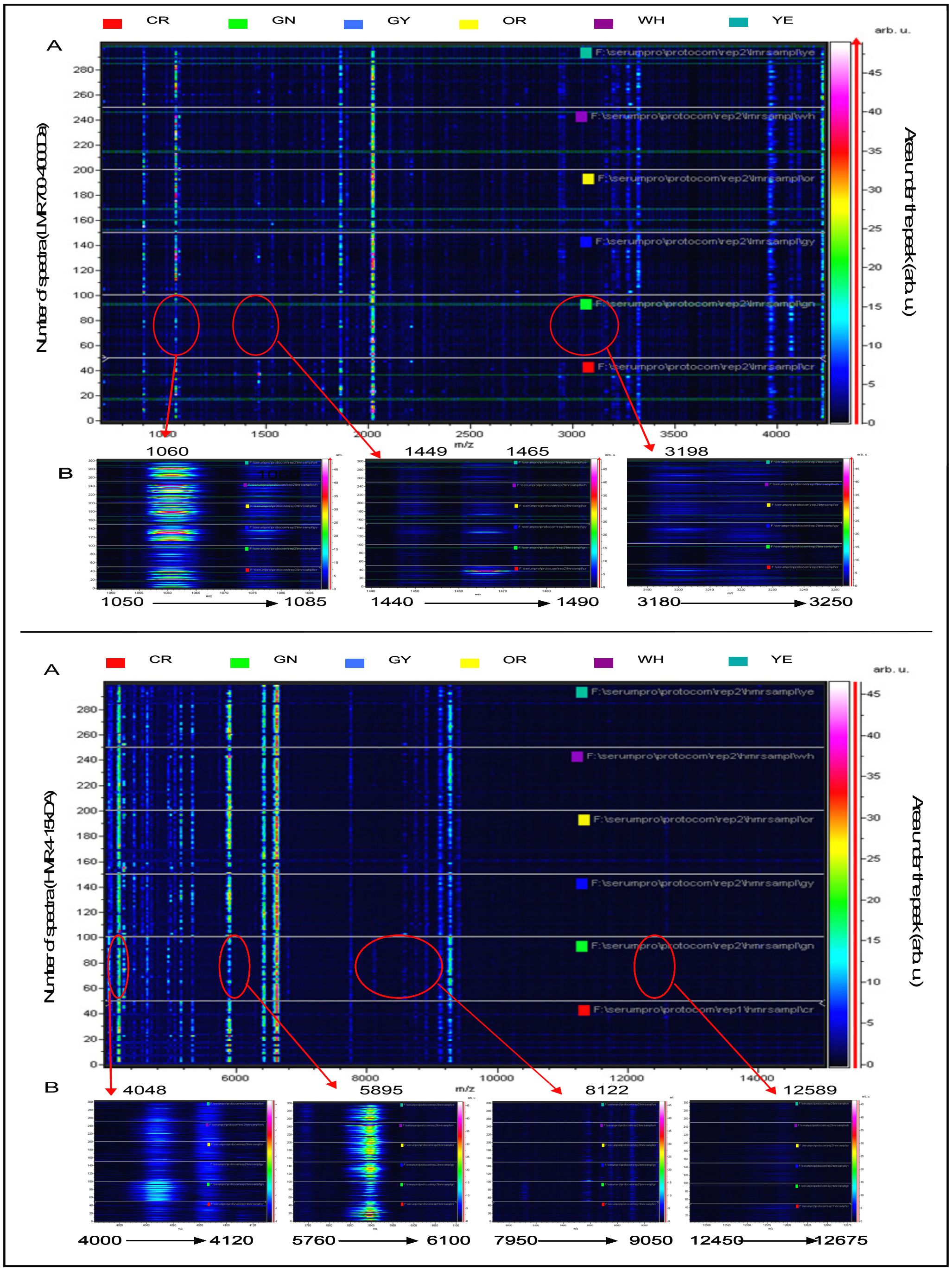
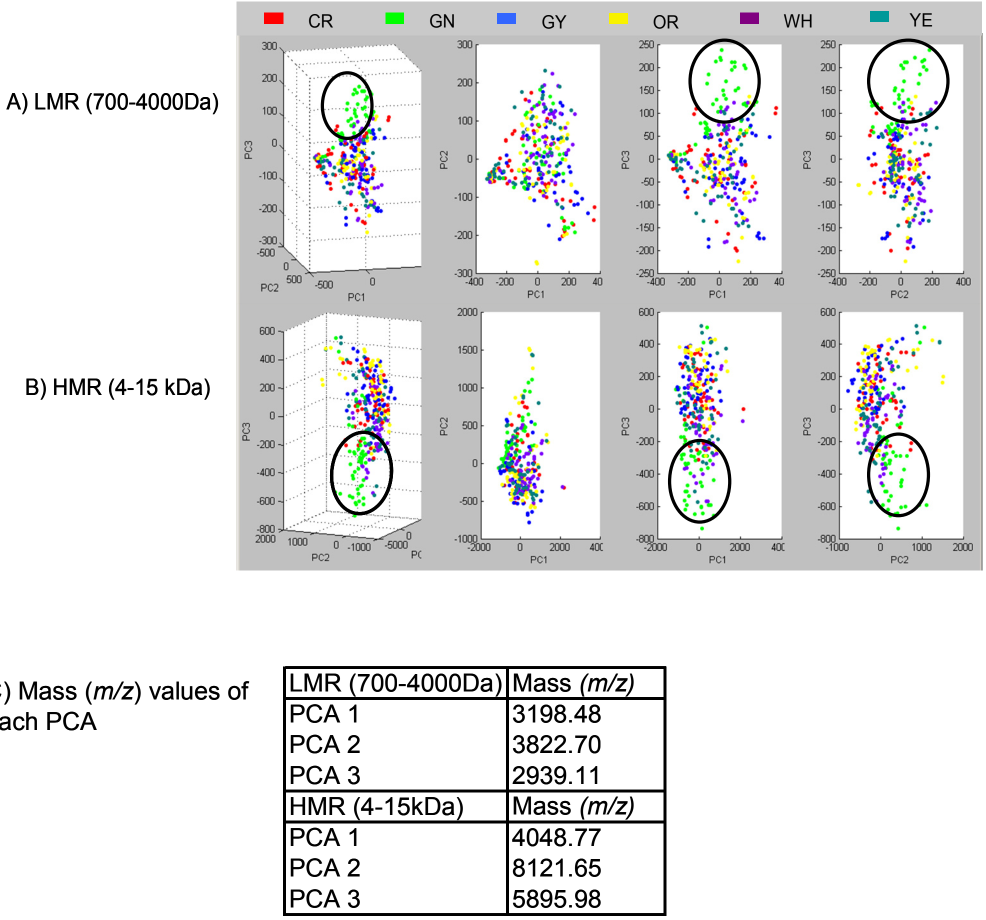


Figure 4.12 Spectral comparisons across the protocols. Top panel (A) 300 low mass range (LMR) spectra from experimental replica 2 (25 volunteers had 2 spotting replicates for each of the 6 protocols) and (B) Several LMR m/z areas where clear differences between protocol 1 (GN) and the other 5 protocols are shown. Bottom panel (A) 300 high mass range (HMR) spectra (B) Several HMR m/z areas where clear difference between protocol 1 (GN) and the other 5 protocols are shown.

A principal component analysis (PCA) was performed to determine how samples would group together according to the preparation method. Protocol 1 (GN) was the most distinctive method, with most volunteer samples grouping together and away from samples handled using the other methods (Figure 4.13). This was true for both mass ranges. The most likely explanation for this separation trend is the extended transport/storage time used in this protocol.



C) Mass (*m/z*) values of each PCA

LMR (700-4000Da)	Mass *(m/z)*
PCA 1	3198.48
PCA 2	3822.70
PCA 3	2939.11
HMR (4-15kDa)	Mass *(m/z)*
PCA 1	4048.77
PCA 2	8121.65
PCA 3	5895.98

Figure 4.13 Protocol comparisons by principal component analysis. A PCA was performed using ClinProTools software (V2.2) to compare samples handled using the different protocols. In the example shown, peaks from replicate number 2 for all samples were used for analysis. The circles denote the clustering of most protocol 1 (GN) samples in (A) the low mass range samples and (B) high mass range and (C) the m/z values of the mass used for clustering. Please note that the colours in this figure differ from those used for protocol labelling.

A pairwise comparison of protocols by PCA more clearly showed this distinction between protocol 1 (GN), and the others (Figure 4.14).

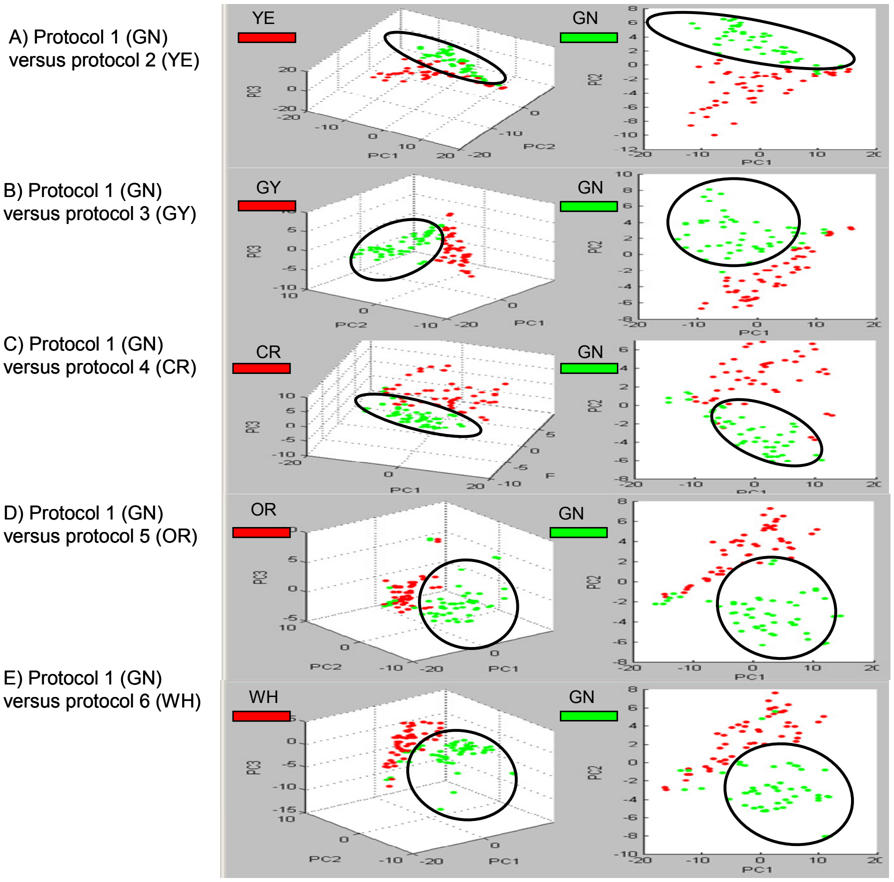


Figure 4.14 Pair-wise protocol comparisons by principal component analysis. A PCA was performed using ClinProTools software (V2.2) to compare samples handled using pairs of protocols. A, B, C and D are the clustering results between pairs of protocols. Protocol 1 (GN) is the most discriminatory protocol.

Overall, 180 peaks (80 on average per protocol) were detected in the LMR (Table 4.6). The peak which discriminated most between the protocols (3198.5 m/z) had an average peak area 3 times lower in protocol 1 (GN) compared to the other protocols. The ANOVA p-value was highly significant (p = 6.45E-32). The second most discriminatory peak (3822.7 m/z) showed a 1.5 times greater peak area in comparison to the other protocols (p = 3.85E-25, Table 4.7).

In the HMR, 91 peaks (35 on average per protocol) were detected in total (Table 4.6). The most discriminatory peak (4048.76 m/z) had an average peak area 2.5 times greater in protocol 1 (GN) compared to the other protocols. The ANOVA p-value was highly significant (p = 2.17E-24). This peak was also elevated in protocol 6 (WH). The second most discriminatory peak in the HMR (8122.05 m/z) showed a 2 times greater peak area in comparison to the other protocols (p = 3.85E-25, Table 4.7).

A) LMR Peak number	Mass *m/z*	Diff in Ave peak area	ANOVA TEST	CR Ave peak area	GN Ave peak area	GY Ave peak area	OR Ave peak area	WH Ave peak area	YE Ave peak area	CR CV	GN CV	GY CV	OR CV	WH CV	YE CV
1	3198.54	31.2	6.45E-32	40.76	15.43	42.55	46.63	38.9	42.4	70	34	58	29	40	58
2	3822.66	14.89	3.85E-25	22.54	35.49	21.41	22.21	20.61	21.34	31	16	32	27	28	27
3	2939.13	23.81	1.46E-23	37.4	18.83	36.66	42.64	35.97	35.52	61	24	52	34	43	52
4	3033.47	3.26	8.64E-23	5.43	8.45	5.32	5.24	5.19	5.29	31	17	28	24	25	27
5	2580.49	4.51	1.09E-22	10.1	13.75	9.26	9.57	9.24	9.82	28	13	24	24	22	29
6	1793.73	2.54	4.11E-20	3.58	5.55	3.29	3.21	3.01	3.37	28	21	33	31	35	24
7	2956.98	21.57	4.93E-20	32.66	18.55	40.12	36.47	40.07	36.86	46	30	67	46	40	53
8	3274.89	40.84	6.68E-20	78.61	43.12	82.98	83.97	76.49	83.62	55	30	57	29	42	61
9	2770.88	10.93	1.32E-19	22.69	12.97	21.6	23.9	22.26	21.2	43	19	43	41	38	59
10	1030.62	2.79	5.06E-16	3.99	5.79	3.01	3	3.2	3.31	44	30	24	23	41	28
11	1954.19	3.71	1.54E-15	8.87	11.57	7.86	8.26	7.92	9.18	28	15	38	31	33	30
12	1741.47	11.53	1.56E-15	22.7	11.17	21.91	22.38	15.74	18.1	59	28	58	48	37	45
13	1172.34	2.65	1.80E-15	5.08	6.3	4.3	3.64	3.79	4.34	42	24	38	26	38	38
14	1704.79	2.38	1.83E-15	2.86	5.12	2.92	2.9	2.74	3	30	27	35	35	39	30
15	4071.12	114.41	2.94E-15	55.46	157.69	46.71	43.28	78.76	52.81	55	50	39	54	51	46
16	1281.44	1.71	6.71E-15	3.26	4.05	2.47	2.33	2.76	3.15	42	23	43	30	54	61
17	1060.7	89.04	1.35E-14	85.06	49.5	119.13	109.34	138.54	105.1	76	42	73	64	69	65
18	4050.61	6.02	1.48E-14	9.27	14.66	8.64	9.21	9.57	10.09	35	22	39	48	27	45
19	2492	2.78	1.74E-14	5.39	7.36	4.58	4.96	5.04	5.41	29	20	38	28	24	33
20	1076.28	11.47	1.75E-14	11.49	6.37	14.63	14.11	17.84	12.01	68	31	74	60	87	65
B) HMR Peak number	Mass *m/z*	Diff in Ave peak area	ANOVA TEST	CR Ave peak area	GN Ave peak area	GY Ave peak area	OR Ave peak area	WH Ave peak area	YE Ave peak area	CR CV	GN CV	GY CV	OR CV	WH CV	YE CV
1	4048.76	180.87	2.17E-24	102.76	283.63	108.73	108.94	155.04	117.97	35	30	41	39	38	46
2	8122.05	68.13	1.38E-22	62.5	115.16	47.03	49.95	49.52	54.1	41	30	19	36	29	33
3	5330.18	129.09	6.20E-20	237.3	115.89	212.27	244.98	204.48	181.69	55	24	51	45	45	50
4	12589.6	50.56	1.25E-19	97.77	47.21	75.69	83.1	49.13	57.03	49	25	30	37	22	52
5	5895.89	593.99	6.76E-19	717.23	338.36	797.95	864.4	932.34	713.6	67	59	56	49	45	60
6	9276.3	382.25	5.76E-18	375.96	738.72	362.54	356.46	622.91	522.86	53	33	35	57	38	47
7	7754.93	49.91	8.29E-16	139.01	148.02	98.1	98.94	138.67	122	23	27	23	29	36	26
8	5061.09	57.12	2.81E-15	105.79	84.19	103.73	117.9	136.79	141.31	41	29	36	46	35	25
9	4957.46	88.21	5.08E-14	67.93	146.62	58.4	65.24	125.01	65.23	26	56	27	29	51	42
10	5744.65	36.31	1.59E-11	78.32	72.22	65.67	70.3	84.27	101.98	28	19	27	31	26	25
11	6790.14	32.38	8.79E-11	65.12	88.2	56.02	55.83	61.53	59.66	25	29	21	25	24	20
12	4086.28	34.9	4.14E-10	99.35	75.7	110.6	106.89	105.5	98.55	29	29	26	25	22	25
13	4999	61.4	6.81E-10	65.38	38.56	94.78	99.96	52.38	56.86	64	33	91	82	76	64
14	4637.14	80.8	4.85E-09	152.04	225.11	151.04	144.31	197.93	176.56	41	32	36	46	24	44
15	5244.9	15.31	6.71E-08	52.48	54.16	38.85	41.59	43.25	47.73	45	20	31	50	33	41
16	7221.34	9.87	6.71E-08	44.87	43.82	35	38.12	36.28	42.71	36	18	19	34	23	29
17	5586.37	11.73	4.79E-07	42.31	38.61	30.58	33.46	35.91	39.48	45	18	23	32	25	31
18	7913.92	13.87	7.34E-07	50.15	44.17	36.27	39.39	38.89	44.16	45	16	18	41	23	38
19	12431.06	16.17	1.14E-06	47.18	42.19	33.39	36.81	31.01	35.25	55	21	26	43	39	49
20	7003.11	9.5	2.64E-06	51.54	51.33	42.04	43.74	44.14	48.06	34	16	19	30	23	26

Table 4.7 Comparisons of average peak areas by protocol. Peak areas from A) LMR and B) HMR were compared by loading all protocols against each other in ClinProTools software (V2.0). Peaks were sorted by ANOVA test p-values. The most discriminatory peak is listed first (only the top 20 are shown here). To avoid exceeding the software's memory capacity only samples from experimental replicate 2 were used.

A Wilcoxon test was also used to find the most discriminatory peaks between pairs of protocols and an overlap between the m/z values of the most discriminating peaks between protocol 1 (GN) and protocol 2 (YE), protocol 3 (GY) and protocol 4 (CY) (3034 m/z p=0.000001) was found (Table 4.8). Furthermore, a peak at 4071 m/z was also found to discriminate protocols 6 (WH) from 2 (YE p=0.023), 3 (GY p=0.000701), 4 (CR p=0.000701) and 5 (OR p=0.000252). These peaks had a higher average peak area in the samples from protocol 1 (GN) suggesting they are likely to be protein degradation products which become more abundant with prolonged clotting times at room temperature.

Moreover, several peaks in the LMR had lower peak areas in protocol 1 (GN) compared with the other 5 protocols. For example, in table 4.7 peak number 1 3198.58 m/z had an average peak area almost 3 times lower in GN compared with the other protocols. Peaks 2939.13 m/z and 2956.98 m/z also had average peak areas 2 times lower in the GN protocol versus the others. This supports the hypothesis that prolonged clotting times could also lead to protein degradation and loss of signal, and highlights the importance of consistent sample handling.

Protocol	GN	YE	GY	CR	OR	WH
GN		m/z 3034.12 p=0.000001	m/z 3034.23 p=0.000001	m/z 3034.26 p=0.000001	m/z 2273.59 p=0.00713	m/z 3823.1 p=0.000001
YE	m/z 3034.12 p=0.000001		m/z 2023.78 p=0.00785	m/z 3435.71 p=0.0122	m/z 2273.59 p=0.00713	m/z 4071.06 p=0.023
GY	m/z 3034.23 p=0.000001	m/z 2023.78 p=0.00785		m/z 1030.61 p=0.0397	m/z 3322.77 p=0.0069	m/z 4071 p=0.023
CR	m/z 3034.26 p=0.000001	m/z 3435.71 p=0.0122	m/z 1030.61 p=0.0397		m/z 1015.6 p=0.00989	m/z 4071.04 p=0.000701
OR	m/z 3197.33 p=0.000001	m/z 2273.59 p=0.00713	m/z 3322.77 p=0.0069	m/z 1015.6 p=0.00989		m/z 4070.99 p=0.000252
WH	m/z 3823.1 p=0.000001	m/z 4071.06 p=0.023	m/z 4071.04 p=0.000701	m/z 4071.04 p=0.000701	m/z 4070.99 p=0.000252	

Table 4.8 Protocol comparisons based on Wilcoxon test p-values for the most discriminating peaks. The Wilcoxon test p-values were calculated for pairs of protocols using average peak area values.

4.6 Discussion

Ideally, for comprehensive serum proteome analysis, universal pre-analytical processing and data handling standards are essential for the discovery of novel protein biomarkers. Furthermore, appropriate biomarker-based tests should be minimally invasive and reproducible. In cancer a simple blood or urine test that detects molecules specific to tumour tissues would be ideal. In addition, the screening technology must be sufficiently sensitive to detect early cancers, but specific enough to classify individuals without cancer as being free of malignancies. The potential use of mass spectrometry (MS) based methods for analysing proteins and peptides from biological fluids for potential biomarker discovery was demonstrated in 2002 [Pearl, 2002; Petricoin et al., 2002]. However, these initial studies were severally criticized on the experimental design and method-induced variability [Baggerly et al., 2004]. Subsequent to this, improvements in many areas of blood-based proteomics studies have been made. As a result, a pool of knowledge about the effect of different variables on MS-based proteome analyses has been generated. Pre-analytical variation from sample handling and storage leading to artefacts and the influences of humidity, temperature, and time for preparation of sera on spectral changes has been evaluated by several groups [Timms et al., 2007; Villanueva et al., 2005; West-Norager et al., 2007].

The aim of the work presented in this chapter was to establish an optimal workflow for the high-throughput analysis of low abundant serum polypeptides. This high-throughput screening platform coupling magnetic bead-based serum polypeptide extraction with mass spectrometry and bioinformatics tools for data analysis would facilitate the discovery of putative ovarian cancer serum markers. This aim was tackled by adapting and modifying a previously established bead-based semi-automated serum peptide extraction protocol and optimising it in the host laboratory at UCL.

Firstly, the automated liquid handling robot was programmed for optimal performance. Various other factors were investigated including the specific beads to be used for peptide extraction, bead-to-serum volume ratio, and intra/inter-assay reproducibility. Finally different sample collection protocols were examined to aid in determining the most reproducible and clinically feasible method for sample handling.

In order to adapt the automated platform established by Villanueva et al. at the MSKCC the initial experiments involved the use of a cocktail of commercially available peptides and proteins in combination with manually extracted commercial serum to establish optimised spectral acquisition methods in the low mass range (LMR 700–4000 Da) and high mass range (HMR 4-15 kDa). Results from the optimised MS method showed an average of 143 and 109 serum peaks were detected in the LMR and HMR respectively with an overall average CVs were 18.7% and 19.4%.

As previously mentioned, the magnetic beads used in the original study by Villanueva et al. (2004) were no longer commercially available, thus, a comparison of different coated magnetic beads from alternative vendors was performed. Results showed that although spectral profiles of commercially available serum did not differ greatly when peptides and proteins were extracted with different beads, Dynal C18 beads did on average capture more peptide and protein peaks in both mass ranges and were cheaper. Moreover, the efficiency of C18 Dynal beads has been reported and they have been recommended for use by the MSKCC group [Villanueva, 2006].

Investigations into the bead-to-serum volume ratio using Dynal C18 beads demonstrated that a lower volume of bead slurry was optimal for the capture and analysis of serum peptides and low mass proteins. However, to avoid bead loss during the automated peptide extraction process, the next higher bead slurry volume (12.5 μL prepared at 2 μg/μL) was selected for use in subsequent experiments. It can be hypothesised that high bead slurry volumes lead to the non-specific capture of larger molecular weight serum peptides and proteins leading to crystallisation and ionisation competition during the MALDI process. As a result of the inconsistencies in the initial MALDI process (i.e. discrepancies in ionisation of peptides and proteins caused by temporal and spatial variations), not all ions receive the same charge and thus, do not reach the linear detector at the same time. Therefore, not all peptides and proteins are detectable to the same degree. Moreover, the resolution of ions is also limited in the linear mode as charged states cannot be determined. This may also explain the high CVs found during reproducibility analysis of each bead type.

Intra-and inter-assay reproducibility experiments were preformed using Dynal C18 beads and commercial serum. They showed that the automated technology platform was robust and reproducible. Average inter-assay CV values for all peaks were 28.4% and 17.7% for the LMR and HMR, respectively. This supports our view that the optimised protocol is applicable to serum screening for putative biomarker discovery. Inconsistencies in automated sample spotting and crystallisation are likely to be accountable for the inherent variability of the assay platform. Intra-assay variances of 14-23% have been reported by others using a similar profiling strategy [de Noo et al., 2005]. one report used 10 major peaks to report CV values of 18% from MALDI-TOF MS measurement, with variance of 14% from sample preparation and 26% variance in inter-day runs [Zhang et al., 2004]. Other studies reported accumulated CVs of 30% in a time/temperature study and 15-36% for normalised intensity of 3 serum peaks in an intra-and inter laboratory reproducibility tests, respectively [West-Norager et al., 2007; Semmes et al., 2005]. It is important to note these studies only report CV values for a few common and relatively higher abundance peaks, without reporting peaks that may show higher variability. Presenting the CVs for all peaks detected in each data set provides a better overview of the robustness of the technology platform.

There are several factors here that can affect the assay reproducibility, including liquid handling errors especially when using organic solvents. Evaporation and crystallisation can be affected by fluctuations in the ambient temperature and relative humidity of the area where the samples are processed and left to air dry before MS analysis. Indeed several groups have reported the influences of humidity on the reproducibility of serum profiles [de Noo et al., 2005; Villanueva et al., 2005; West-Norager et al., 2007]. It is therefore important to establish the assay reproducibility and identify the experimental factors which affect it before trying to distinguish between these and disease-specific peaks in the pursuit of putative biomarkers.

A protocol comparison study using different sample collection and handling methods compatible with the optimised workflow was used for selecting a method for a large scale case-control study which would be practical for both clinical collection and protein profiling. Variance analysis showed that protocol number 2 (YE) gave the lowest overall variance when all peaks were considered and the highest average number of peaks, and therefore this method was selected as the preferred serum collection/handling protocol for the UKOPS serum collection. This method, whilst clinically feasible, does however require rapid processing of samples on ice in the collection centres. Therefore, dedicated staff are required for collection and processing. This method is also recommended by Villanueva et al. (2006) for assessing the effects of tumour specific protease activities in serum during the clotting process. No consistent effects on variance were evident between the various protocols where samples were placed on ice (GY, YE, CR, OR). This suggests that as long as samples are kept on ice, with a transport time less than 6 hours, then there is little effect on serum proteins profiles. Also, it appears that an extra freeze thaw cycle or storage tube has little effect on variance (CR protocol).

Profiles from protocol number 1 (GN) were clearly different from the other protocols. In both mass ranges, several peaks were found to distinguish protocol 1 from the other protocols. PCA also found the greatest separation for protocol 1 over the other methods. Separation was representative for both mass ranges. Thus, although strict control on sample handling protocols has been proposed by many groups to reduce variability, the data shows that sample collection protocols can be more flexible. The critical issue is that all individual samples must be treated in exactly the same manner. This way protease activity can be better controlled and assessed for tumour-specific activity.

The results presented here also show that irrespective of transit/storage at ambient temperature for up to 30 hours, protocol 1 (GN) gave a relatively low overall variance. This finding is important for samples collected in older studies where longer transit times at ambient temperature have been used. Although again, one must ensure that case control samples for comparison have been handled identically. The finding that numerous peaks from both mass ranges had altered peak areas in protocol 1 compared to the other protocols may well be the result of proteolysis products appearing or substrates disappearing over time. Importantly this could make a difference if hypothesised tumour-specific exoprotease substrates have already been degraded [Villanueva et al., 2005]. For this reason, the more stringent protocol 2 (YE) was chosen for further biomarker discovery experiments.

Having adapted and optimised a semi-automated bead-based extraction and MALDI-TOF MS profiling platform and selected a clinically feasible protocol for sample collection and handling, case and healthy control serum samples from the UKOPS collection were analysed to look for potential markers of ovarian cancer and those which can discriminate malignant from benign samples. The results of this are presented in the next chapter.

Chapter 5: MALDI-TOF MS profiling of UKOPS ovarian cancer, benign cases and healthy control samples.

5.1 Introduction

The protocols set out and optimised in chapter 4 demonstrated that the technology platform developed herein for serum profiling is sufficiently reproducible. A natural progression from this point was to apply this platform to well-characterised clinical sample sets. The UKOPS sample set used contains 60 healthy volunteers, 43 and 22 cases of benign and malignant ovarian cancer samples, respectively (Table 5.1).

Group	Number	OC stage	Ave CA125 (U/mL)
H	60		N/A
B	43		50.16
Me	6	I/II	232.06
ML	16	III/IV	1895.3

Table 5.1 Details for the clinical samples from the UKOPS collection. Sample details of each cohort including the number of volunteers, ovarian cancer stage (OC stage) and average CA-125 levels are shown.

The collection of these samples was rigorously controlled with adherence to standard operation procedures to limit introduced variation and to maintain, as much as possible, the biological integrity of every sample. Importantly, all samples were identically collected and treated following a previously optimised protocol for MALDI-TOF MS-based profiling [Villanueva et al., 2006]. Briefly, samples were collected in Becton Dickinson tiger-top tubes and allowed to clot at RT for 60 mins. Samples were then stored on wet ice for 2 hours before centrifugation and then were transferred to straws for storage at -80°C until they were analysed. The sample handling was controlled to facilitate identification of tumour-specific exopeptidase-generated peptides which could serve as putative markers for ovarian cancer. Thus, this collection represents a valuable resource ideally suited to high-throughput discovery of cancer biomarkers.

It is important to note that although standard procedures for sample collection and handling now exist, there are still no standards for donors. Biological variation caused by diet i.e. products of digestion from the intake of food before a sample is taken, and cyclic variations due to time of day could lead to variability in serum protein levels. The psychological stress suffered by donors especially those who have been diagnosed with cancer may also contribute to biological variation [Morita et al., 2005; Van et al., 1998]. For example, psychosocial factors have been shown to affect serum interleukin-6 levels among women with advanced ovarian cancer [Costanzo et al., 2005]. Interleukin-6 is the chief stimulator of the production of most acute-phase proteins [Gabay and Kushner, 1999]. Moreover, biological variation in specific cytokine levels could also affect the rate of clotting and hence the degree of protein degradation. Indeed, many of the peptides previously identified as ‘surrogate’ cancer markers using the techniques described in this thesis, are derived from acute-phase proteins which maybe involved in the complex coagulation and complement pathways [Villanueva et al., 2006].

Commercial serum was used as a quality control sample (QC) to establish baseline parameters for assessing the reproducibility of the technology platform and for optimising the Support Vector Machine (SVM) algorithm used to classify clinical samples. The aims of this chapter were to define experimental variation within the protocol using a previously analysed commercial serum sample (Sigma serum) and to establish a cut-off threshold for technical variance. In addition, to define potential biomarkers of ovarian cancer by analysis of healthy, benign and malignant ovarian cancer cases from the UKOPS collection, which is a tightly controlled case control serum sample collection. The sample processing and data analysis conditions, optimised as described in the previous chapter, were utilised to detect and statistically evaluate differentially expressed peaks generated using MALDI-TOF MS profiling and to determine if these peaks could be used for the discrimination of ovarian cancer samples and for stage classification.

5.2 Spectral analysis of QC samples

To assess the technical variability caused by automated polypeptide spotting, spot-to-spot inconsistencies in crystallisation and the inherent variability in the MALDI-TOF ionisation process, (discussed in the previous chapter), the intra-assay reproducibility was assessed using Sigma serum samples (QC). The QC samples (n=13) were randomly assigned positions in the assay plates and processed in parallel to the clinical samples.

Following the automated extraction on Dynal C18 magnetic beads, the 13 QC samples were spotted in duplicate onto a MALDI-target plate and a total of 26 MS spectra were acquired in each mass range. Data from all QC samples were analysed in ClinProTools software using the parameters defined in Chapter 2 (section 2.2.4.2) to generate a set of normalised peak areas for each spectrum. The spotted QC replicates were loaded into the analysis software in 13 pairs, the software then generated an average spectrum for each pair. The averaged spectra were aligned and average peak areas for matching peaks were calculated. Thus, for each averaged spectrum the same number of peak areas was obtained. This then allowed the calculation of an overall average spectrum. From this a total of 101 common peaks were detected in the LMR and 59 peaks were detected in the HMR (Figure 5.1).

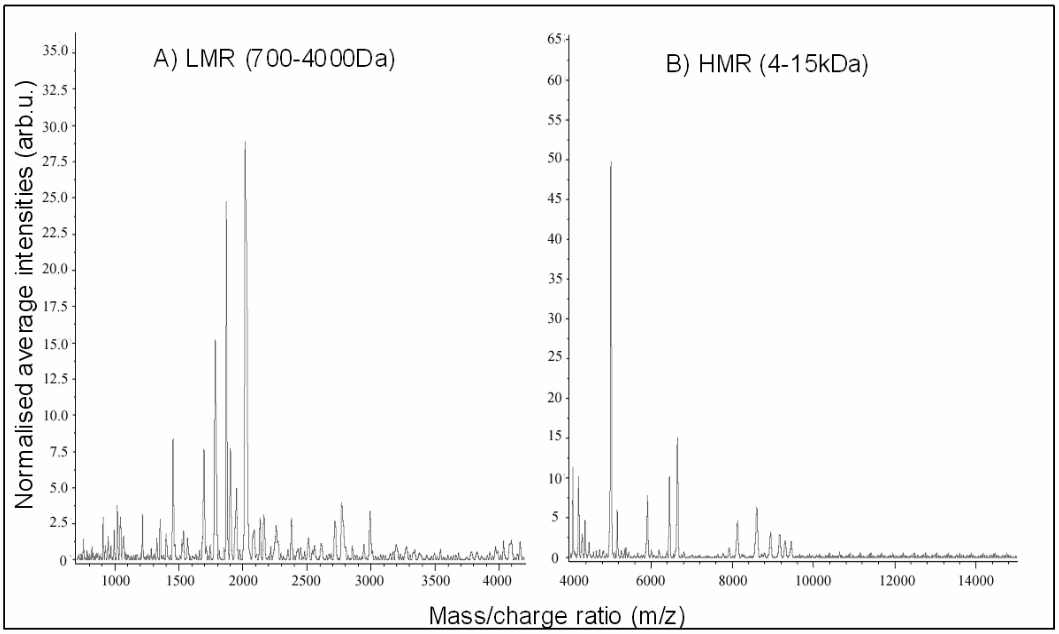


Figure 5.1 MALDI-TOF profiles of QC serum. The averaged spectral profiles from 13 Sigma serum samples and 26 individual spectra in the A) LMR and B) HMR are shown.

Heat maps were used to visualise peak profiles from each QC sample pair to quantify possible spot-to-spot variation. The spectral heat maps showed that several regions in the peak profiles did indeed indicate spot-to-spot variation for the same sample preparations, as well as across the 13 QC samples (Figure 5.2).

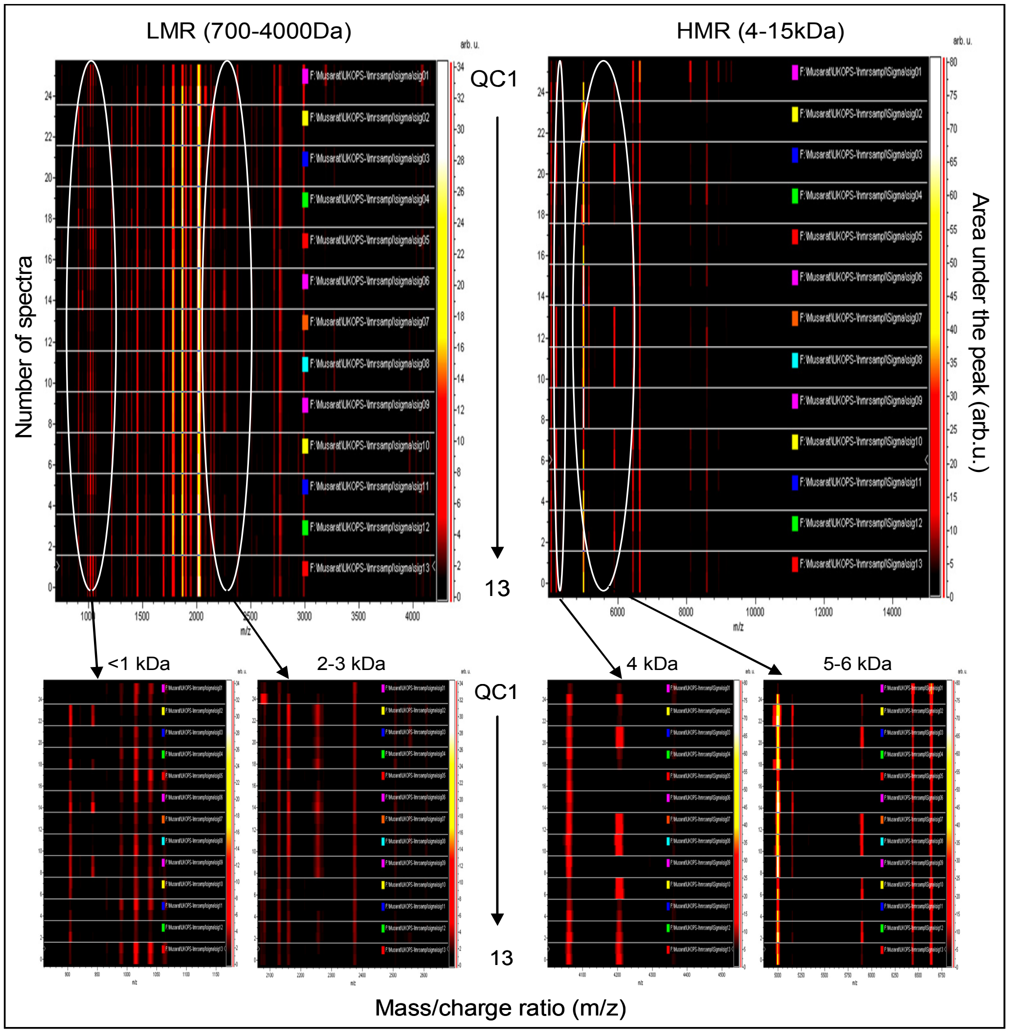


Figure 5.2 Heap map representations of spectra from QC serum samples. Each of the 13 QC samples processed was spotted in duplicate. Spot-to-spot variation is apparent in both mass ranges with several peaks showing discrepancies in peak area between spotting duplicates. The top panel shows the heat map peak profiles from the LMR and HMR and the bottom panel shows expanded sections of several areas where difference were obvious.

The intra-assay reproducibility was assessed by measuring the co-efficient of variance for averaged peak areas across the 13 QC samples (from spotting duplicates). The co-efficient of variance for all peaks ranged from 8.1% in QC sample 10 to 23.9% in QC sample 1 in the LMR and from 8.6% in QC sample 9 to 19.0% in QC sample 10 in the HMR (Figures 5.3). In the LMR, 85% of all peaks were found have CVs below 15% and in the HMR 75% of all peaks fell below this level of variance. The average co-efficient of variance (CV) of all the peaks ranged from 3.7-21.4% and 4.5-23.9% in the LMR and HMR, respectively. Thus, any changes in peak area in the clinical samples would need to exceed a 1.24 fold change to be considered as true biological differences.

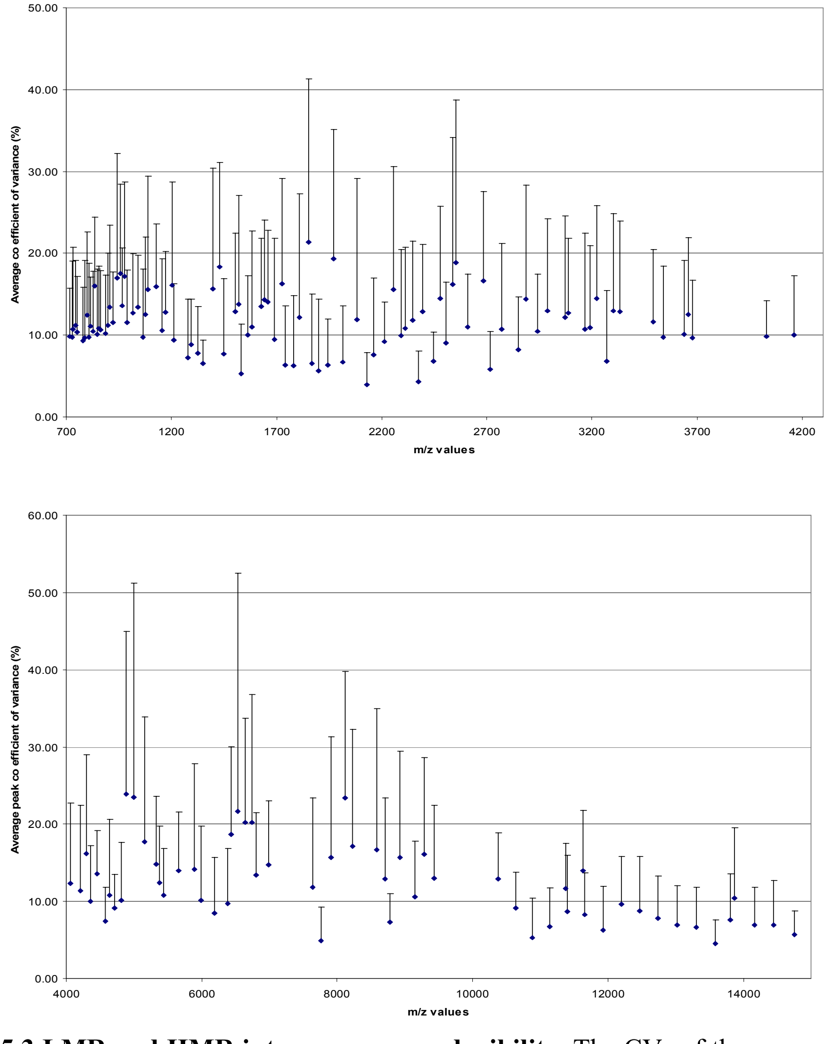


Figure 5.3 LMR and HMR intra-assay reproducibility. The CVs of the average peak area of all peaks from the 13 randomly assayed QC serum samples in the LMR (700-4000 Da top panel) and HMR (4-15kDa bottom panel). The error bars indicate the standard deviation of the average CV values.

The QC samples were distributed across two MALDI plates. The inter-plate variation was found to be the similar across the two plates. In the LMR, the overall average peak area of QC samples was 30 (arb. u.) ± 2.42 (CV=8.1%) and 28 (arb. u.) ± 2.42 (CV=8.6%) on plate 1 and plate 2, respectively. In the HMR, the overall average peak area of was 122.28 ± 18.9 (CV 15.4%) and 116.17 ± 16.93 (CV 14.57%) on plate 1 and plate 2, respectively. Together these results demonstrate that the platform reproducibility was close to 80% with 20% technical variance. As discussed in the previous chapter, several factors can affect the assay reproducibility including fluctuation in the ambient temperature and relative humidity of the area where the samples are processed and left to air dry which can result in changes in the evaporation rate of the organic solvents used. This in turn affects the co-crystallisation and hence the ionisation of peptides during MALDI. Subsequent studies in the host laboratory have shown that diluting the sample eluates 1:5 significantly improves the intra-assay variability. It is fair to say that highly variable peaks are likely to be poor biomarkers.

5.3 Spectral analysis of clinical samples

Clinical samples from healthy controls, benign and ovarian cancer cases from the UKOPS collection were processed on the automated bead-based polypeptide extraction protocol followed by MALDI-TOF MS analysis (detailed in Chapter 2). As with the QC samples, each clinical sample was spotted in duplicate on to one of the two MALDI targets and spectra (average of 400 laser shots over 8 locations around the spot) for each spot were acquired in each mass range.

5.3.1 Intra-condition variation

To assess the variation within each clinical group, spectra from the duplicate spots were analysed in ClinProTools. For both mass ranges each pair belonging to the same sample were loaded into the analysis software as a separate condition. These were then averaged and peak statistics were generated. In the LMR, a total of 174 peaks were detected in the aligned and averaged healthy (H) spectrum (n = 60), 166 in the benign (B) spectrum (n = 43), 171 in the malignant early stage (Me) spectrum (n = 6) and 183 in the malignant late stage (Ml) spectrum (n = 16). In the HMR, a total of 99 peaks were detected in the averaged healthy spectrum, 108 in the benign, 98 in the malignant early stage and 73 in the malignant late stage (Figure 5.4A & B). The difference in peak numbers is likely to be the result of better peak recognition in one condition compared with another.

The co-efficient of variance for the average peak area across the samples in the healthy condition ranged from 7.5% to 29.7% in the LMR and 9.0% to 45.1% in the HMR. While in the benign condition, the variance in average peak area ranged from 7.6% to 40.4% and 9.0% to 29.5% in the LMR and HMR, respectively. In the malignant early stage condition, the variance in average peak area ranged from 8.2% to 30.1% and 11.8% to 28.8% in the LMR and HMR, respectively. Finally, in the malignant late stage condition, the variance in average peak area ranged from 8.2% to 18.4% and 9.5% to 20.2% in the LMR and HMR, respectively. The high CVs are likely to be reflective of intra-condition biological variation.

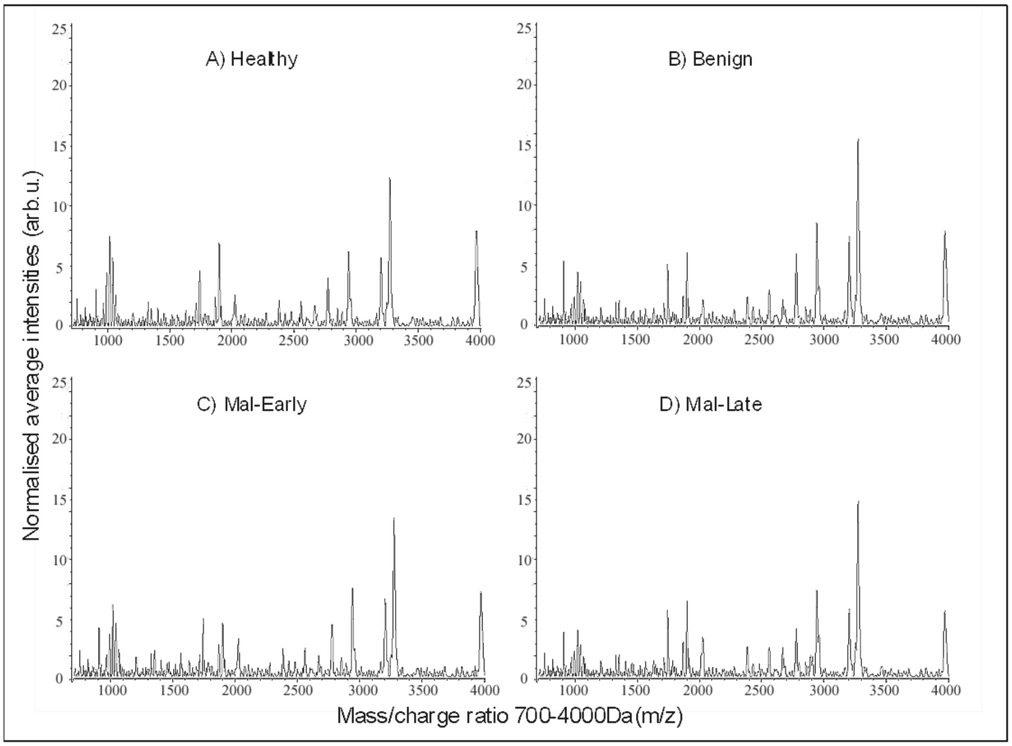


Figure 5.4A Average LMR spectrum of the clinical samples from the UKOPS collection. Serum peptides were extracted using the automated bead-based extraction protocol, subjected to MALDI-TOF MS profiling and data processed using ClinProTools. The average spectrum for each clinical condition is shown A) Healthy, B) Benign, C) Malignant early stage and D) Malignant late stage. A total of 174 peaks were detected in the average healthy spectra, 166 in the benign, 183 in the malignant early stage and 171 in the malignant late stage.

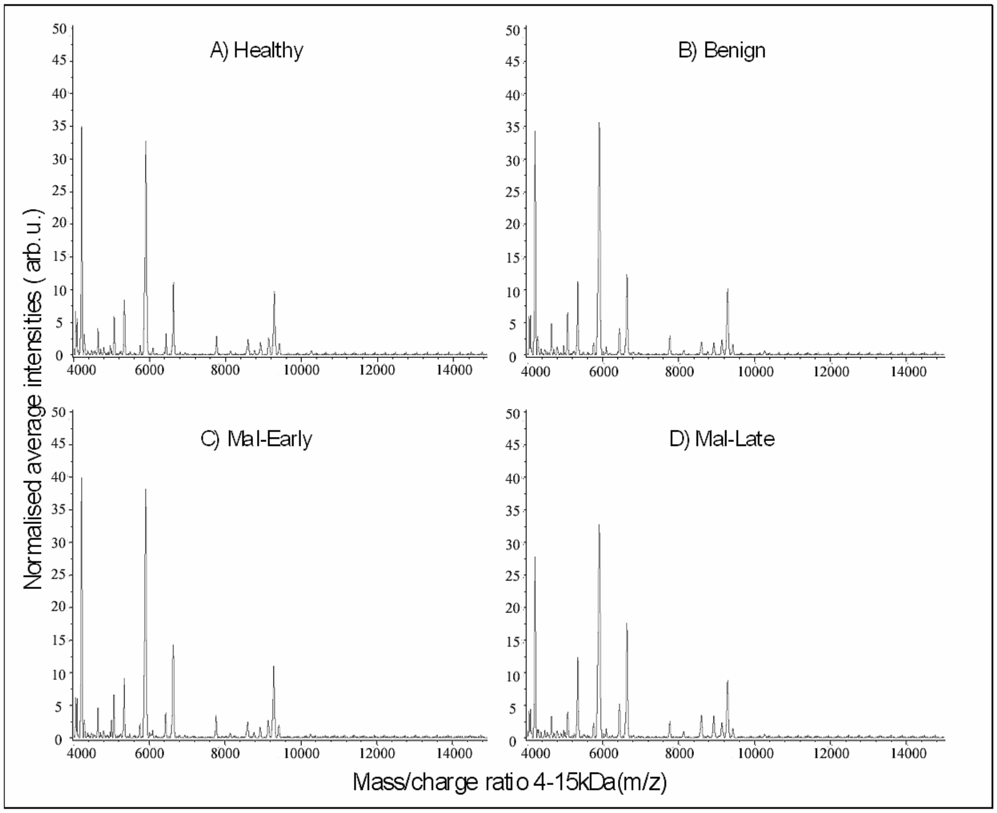
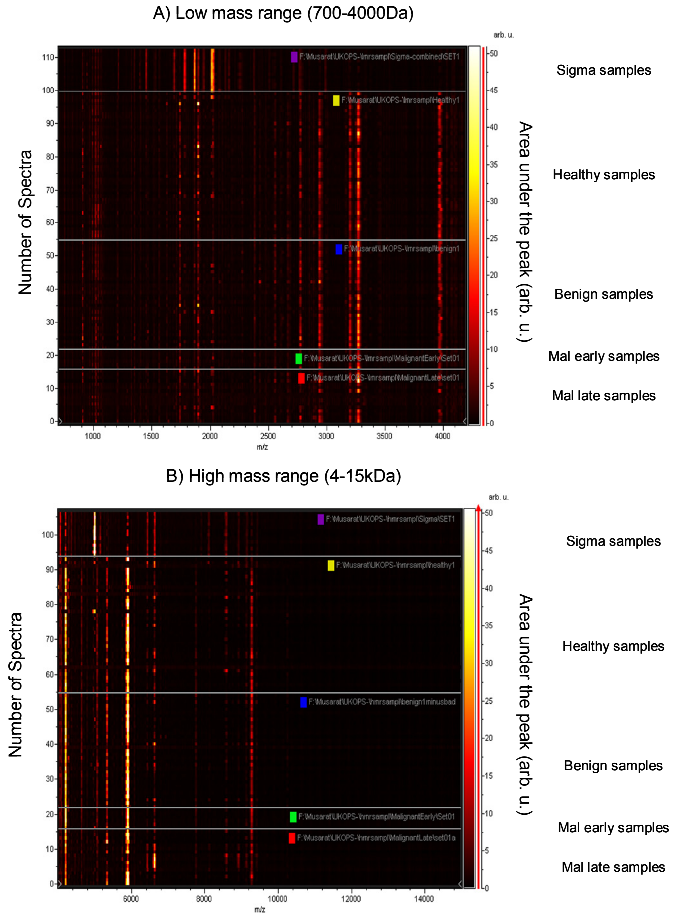


Figure 5.4B Average HMR spectrum of the clinical samples from the UKOPS collection. The average spectrum for each clinical condition is shown A) Healthy, B) Benign, C) Malignant early stage and D) Malignant late stage. A total of 99 peaks were detected in the average healthy spectra, 108 in the benign, 73 in the malignant early stage and 98 in the malignant late stage.

5.3.2 Comparison of QC and clinical samples

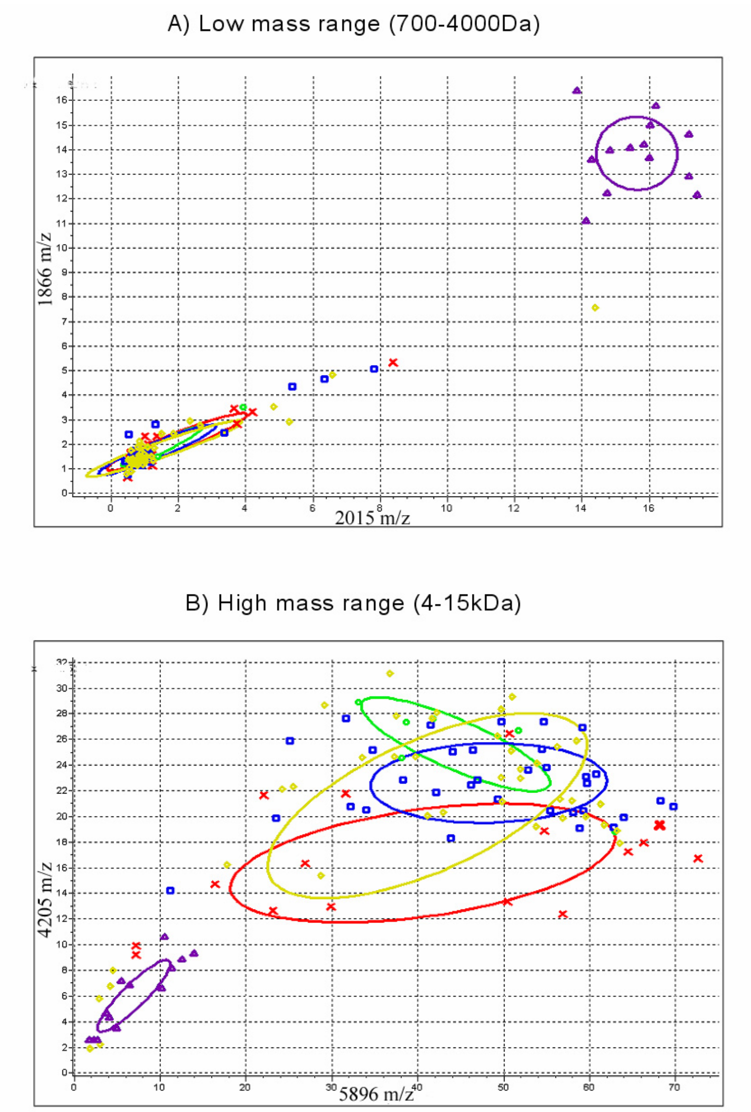
Heat maps were generated to compare spectral peak profiles from QC samples with UKOPS clinical samples. The peak profiles of QC samples were clearly different from UKOPS clinical samples in both mass ranges. For example, prominent peaks at 1779.7 m/z, 2021.8 m/z and 5060 m/z in the QC samples were less apparent in the clinical samples, whilst prominent peaks at 3250 m/z, 4205 m/z and 5950 m/z in the clinical samples were of much reduced intensity in the QC samples (Figure 5.5). The QC sample is a pool of 10 donors' serum and no other details of handling were available from the suppliers. Thus, the differences are likely to be caused by differences in sample handling leading to differential proteolysis rather than differences in the proteomes of the donors.


A) Low mass range (700-4000Da)

B) High mass range (4-15kDa)

Figure 5.5 Heat maps of peak profiles for all samples. The peak profiles of QC samples were compared with UKOPS clinical samples using ClinProTools software in A) LMR and B) HMR.

Peak distribution analysis of the QC and clinical samples confirmed these differences. The ellipses represent the standard deviation of the average peak area. A cluster representing the QC samples grouped away from the clinical samples in both mass ranges (Figure 5.6).


A) Low mass range (700-4000Da)

B) High mass range (4-15kDa)

Figure 5.6 2D peak distributions of the top two discriminatory peaks. The distribution of peak areas from the top two discriminatory peaks in the A) LMR and B) HMR from the healthy (gold), benign (blue), malignant early (green), malignant late (red) and QC samples (purple) are shown.

5.3.3 Discriminatory peak analysis for clinical samples

For discriminatory peak analysis, spectra from the first duplicate spot on the MALDI target were assigned to a training set ('set01') which was used to determine statistically significant discriminatory peaks and to create classification models. Spectra from the second duplicate spot were assigned to a test set ('set02'), which were later used to test the classification algorithm's ability to correctly assign spectra to each of the clinical conditions. Comparison of all four conditions was performed using 179 common peaks (S/N > 3) in the LMR and peak data was sorted according to Wilcoxon T-test p-values (Table 5.2). Results demonstrated that seven peaks had a statistically significant difference in the average peak area between the clinical conditions using a cut-off of p=0.05. The peak at 2905 m/z had an average peak area of 9.31 (arb. u.) in the healthy condition, 10.43 in benign, 9.04 in malignant early and 25.42 in malignant late stage condition. The greatest difference in peak area was found between the malignant early and malignant late at conditions with ~2.5-fold increase in the average peak area (p = 0.0003). The second highest scoring peak at 989.5 m/z (p = 0.0006) had a higher average peak area in the healthy condition and was similar in the benign, malignant early stage and malignant late stage samples. However, it is important to note that despite these peaks being statistically significant the CVs were rather high, a likely consequence of biological variation across the samples in each condition. Average spectra detailing the top 4 peaks are shown in Figure 5.7.

Mass	Greatest Difference in Ave Peak Area	Wilcoxon T-test	Healthy	Benign	Mal Early	Mal Late	StdDev Healthy	StdDev Benign	StdDev Mal Early	StdDev Mal Late	CV Healthy	CV Benign	CV Mal Early	CV Mal Late
2905.44	16.38	0.000298	9.31	10.43	9.04	25.42	3.63	5.43	2.98	16.31	38.99	52.06	32.96	64.16
989.53	17.11	0.000628	39.27	22.82	25.06	22.15	26.46	11.9	9.32	16.03	67.38	52.15	37.19	72.37
1039.69	26.98	0.00118	60.51	39.61	41.21	33.53	28.06	14.85	8.91	16.15	46.37	37.49	21.62	48.17
2796.19	8.87	0.0114	12.03	13.9	9.85	18.72	3.79	5.77	1.96	7.42	31.50	41.51	19.90	39.64
1561.25	11.24	0.0114	15.9	19.44	27.14	19.95	5.98	6.26	8.87	4.97	37.61	32.20	32.68	24.91
1391.55	2.04	0.0147	7.07	7.69	8.68	9.11	1.45	1.7	1.98	1.76	20.51	22.11	22.81	19.32
1015.66	22.82	0.0148	75.47	56.19	58.4	52.65	28.84	19.75	10.59	14.98	38.21	35.15	18.13	28.45

Table 5.2 Comparison of average peak areas in the LMR. Peak areas from the LMR were compared by loading all the conditions against each other in ClinProTools software. Peaks were sorted by the Wilcoxon T-test p-values. The most discriminatory peak is listed first (significant peaks (p<0.05) are shown).

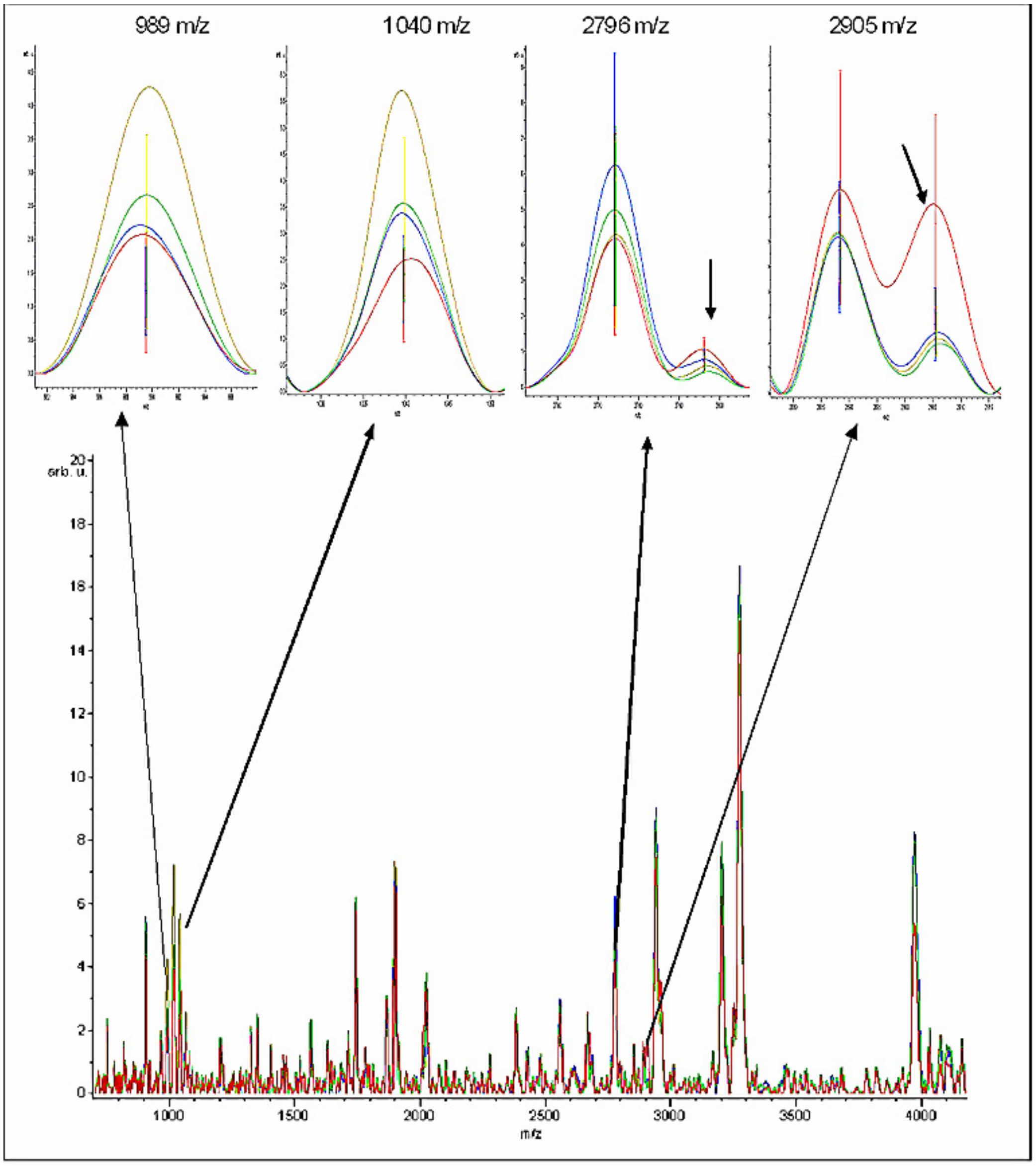


Figure 5.7 MALDI-TOF spectral overlays illustrating the average LMR (700-4000 Da) peptide profiles. Clinical samples were processed using the automated bead-based extraction protocol followed by MALDI-TOF MS analysis. Data was analysed with ClinProTools software (V 2.0). Average spectra of healthy (gold), benign (blue), malignant early (green) and malignant late stage (red) samples are shown and representative discriminatory peaks have been enlarged to show the differences in peak area with error bars as standard deviation.

In the HMR, 94 common peaks (SN > 5) were compared between the four conditions (Table 5.3). The most discriminatory peak at 4049.54 m/z had an average peak area 157.78 (arb. u.) in the healthy condition, 134.8 in the benign, 159.98 in the malignant early and 91.38 in the malignant late stage condition. The greatest difference was therefore found between malignant early stage and malignant late stage conditions where there was a difference of ~1.75 fold which was statistically significant at p = 0.02. Using a cut-off of p=0.05 four discriminatory peaks were found in the HMR. The peak CVs were also high in the HMR, but lower than the LMR (Table 5.3 & Figure 5.8).

Mass	Difference in Ave Peak Area	Wilcoxon T-test	Healthy	Benign	Mal Early	Mal Late	StdDev Healthy	StdDev Benign	StdDev Mal Early	StdDev Mal Late	CV Healthy	CV Benign	CV Mal Early	CV Mal Late
4049.54	68.6	0.0206	157.78	134.8	159.98	91.38	65.51	40.62	45.17	41.56	41.52	30.13	28.23	45.48
4204.56	451.51	0.0228	1015.34	1052.82	1250.23	798.72	338.93	170.68	178.33	226.22	33.38	16.21	14.26	28.32
6241.38	15.67	0.0228	45.75	40.25	39.07	54.73	13.67	10.88	7.6	20.02	29.88	27.03	19.45	36.58
4636.66	42.92	0.0245	135.78	157.15	152.2	114.24	40.77	41.86	37.5	33.2	30.03	26.64	24.64	29.06

Table 5.3 Comparison of average peak areas in the HMR. Peak areas from the HMR were compared by loading all the conditions against each other in ClinProTools software. Peaks were sorted by Wilcoxon T-test p-values. The most discriminatory peak is listed first (only significant peaks (p <0.05) are shown here).

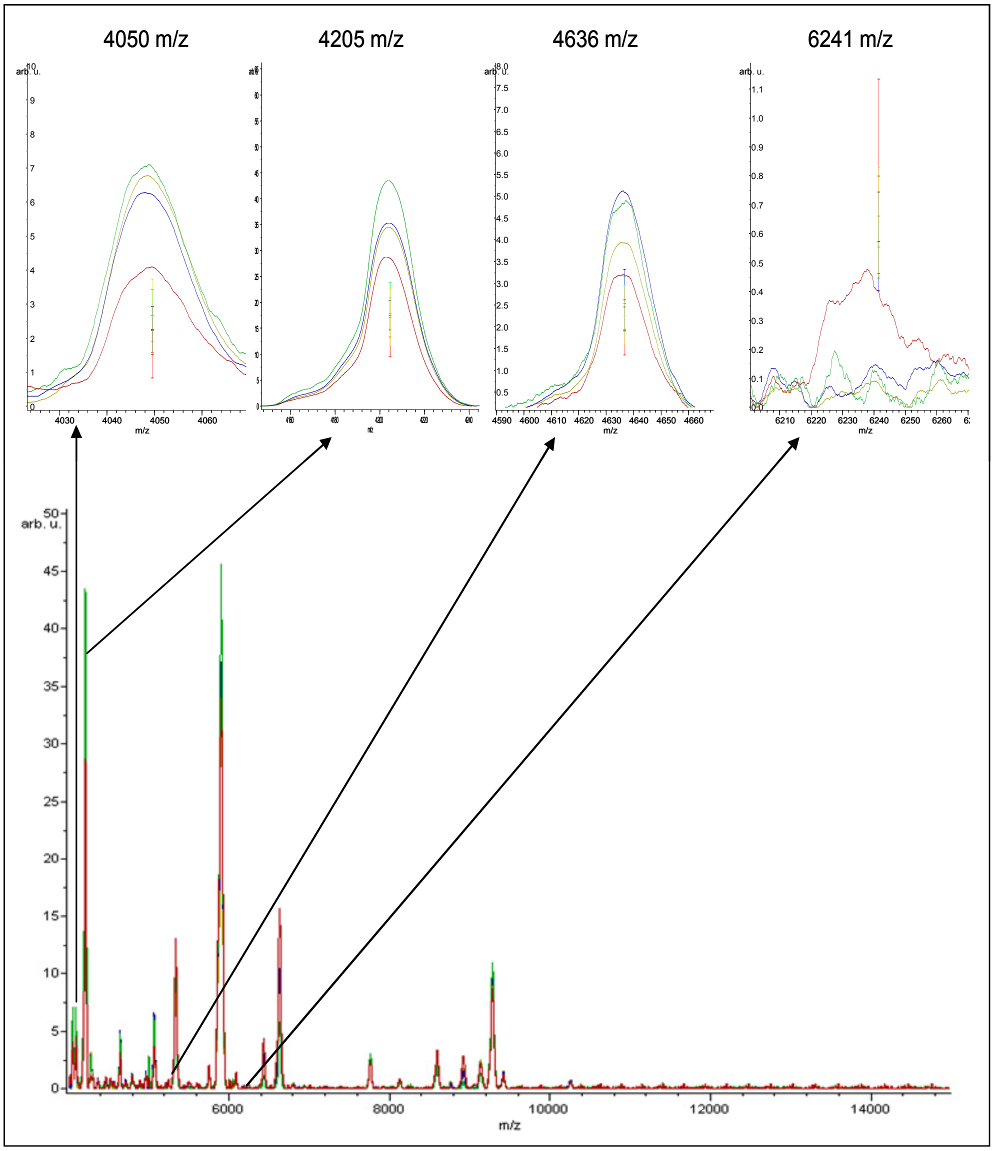


Figure 5.8 MALDI-TOF spectral overlays illustrating the average HMR (4-15 kDa) peptide profiles. Clinical samples were processed using the automated bead-based extraction protocol followed by MALDI-TOF MS analysis. Data was analysed with ClinProTools software (version 2.0). Average spectra of healthy (gold), benign (blue), malignant early (green) and malignant late stage (red) samples are shown and representative discriminatory peaks have been enlarged to show the differences in peak area with error bars representing the standard deviation.

Further analysis of the statistically significant discriminatory peaks suggested that fluctuations in the amount of these peptide masses could be related to proteolysis. For example, in the LMR peaks at 2905 m/z and 2796 m/z which showed a greater average peak area in the malignant late stage condition could be fragments of the same polypeptide. Peaks at 989.5 m/z, 1015.3 m/z and 1039.5 m/z which all showed a greater peak area in the healthy condition are approximately 25 Da apart and maybe modified fragments of the same peptide (Figure 5.9). In the HMR peaks at 4049.5 m/z and 4204.5 m/z which showed a greater average peak area in the malignant early stage condition and are approximately 155 Da apart (the amino acid Arginine has a mass of 156 Da). These peaks may also be fragments of the same polypeptide (Figure 5.10).

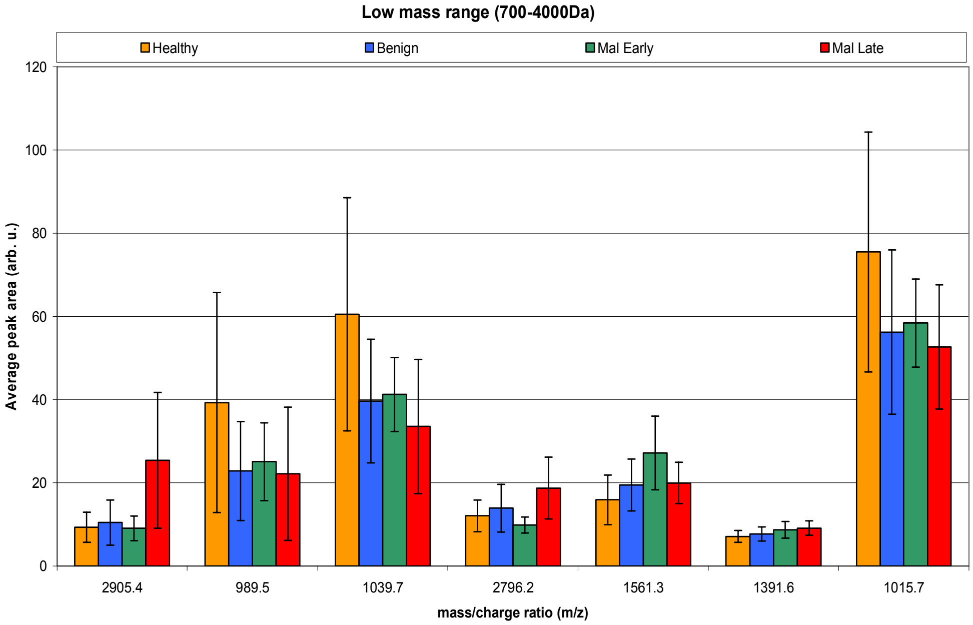


Figure 5.9 Comparison of discriminatory peaks in the LMR. Data analysis showed several peaks with differences in average peak area which discriminate clinical samples. Peaks with statistically significant difference (p<0.05) are shown here.

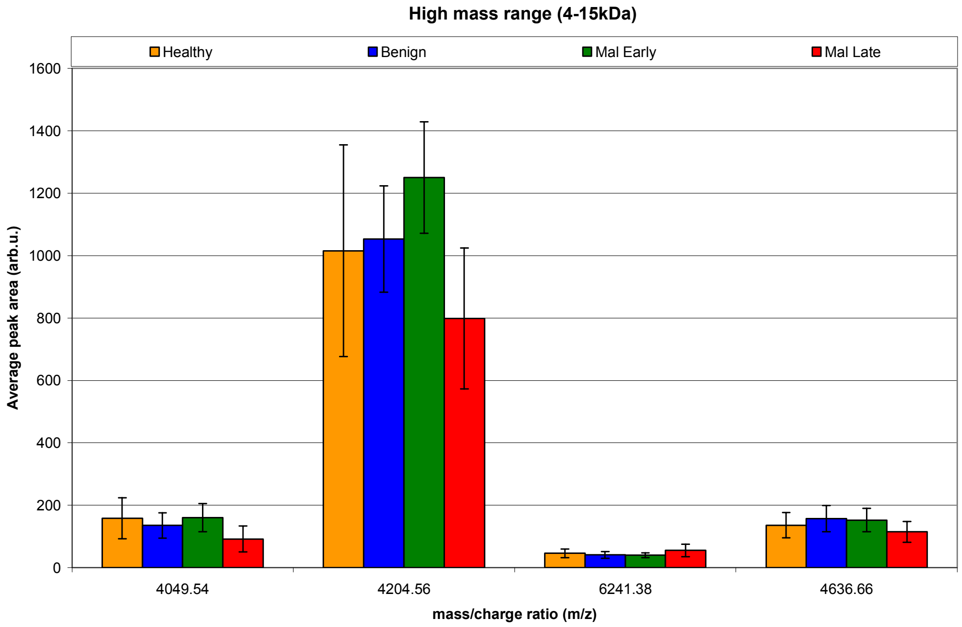
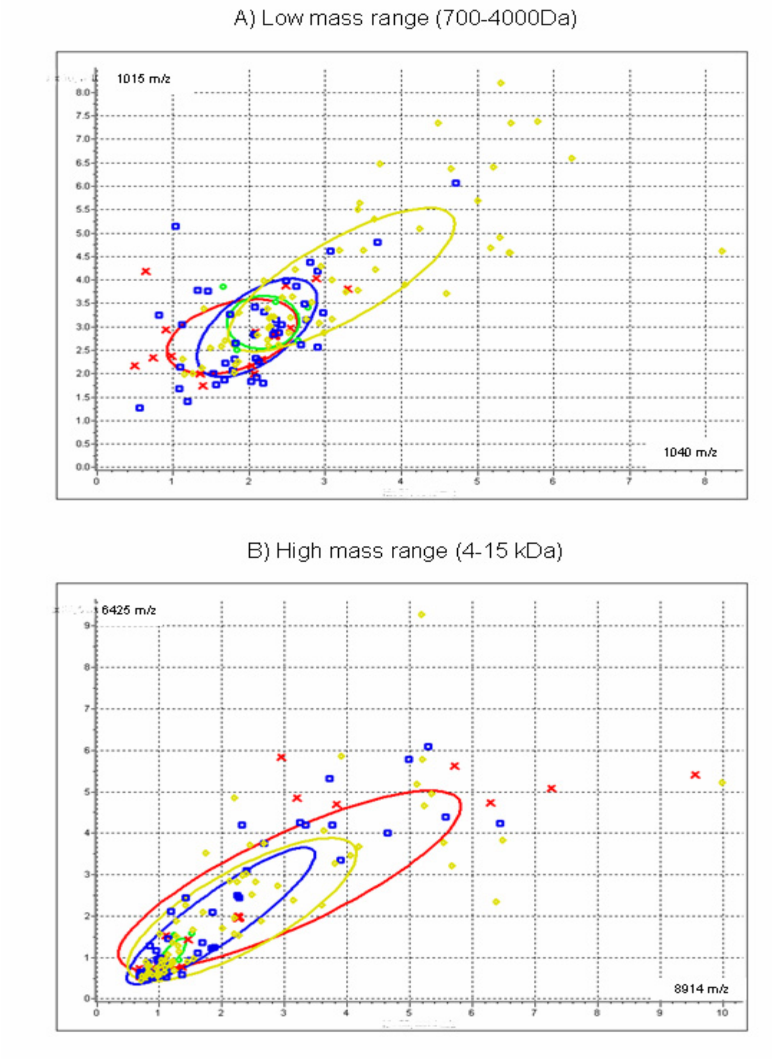


Figure 5.10 Comparison of discriminatory peaks in the HMR. Data analysis showed several peaks with differences in average peak area which discriminate clinical samples. Peaks with statistically significant difference ($p<0.05$) are shown here.

Despite this, a peak distribution analysis for pairs of discriminatory peaks revealed poor separation of the healthy, benign, malignant early stage and malignant late stage conditions. The ellipses in Figure 5.11 represent the standard deviation of the average peak area.


A) Low mass range (700-4000Da)

B) High mass range (4-15 kDa)

Figure 5.11 2D peak distributions of the top two discriminatory peaks. The distribution of peak areas from two discriminatory peaks in the A) LMR and B) HMR from the healthy (gold), benign (blue), malignant early (green), malignant late (red) are shown.

Pair-wise comparisons using the Wilcoxon test between clinical conditions revealed a number of additional discriminatory peaks in both mass ranges. In the LMR, a peak at 904.8 m/z had a 1.5-fold difference in average peak area between the healthy and benign conditions. This peak was significantly up-regulated in the benign condition (p=0.04). Peaks at 1203 m/z and 1064 m/z discriminated between the healthy and malignant late stage conditions. The peak at 1203 m/z was up-regulated by 1.4 fold in the malignant late stage condition (p=0.03), while the peak at 1064 m/z was up-regulated by 1.5 fold in the healthy condition. In the LMR, the greatest number of differentially expressed peaks were found between the healthy and the malignant late stage conditions.

In the HMR, a greater number of differences were found between the benign and the malignant late stage conditions. In addition to the 4 statistically significant peaks found in the four conditions comparison, peaks at 4086.6 m/z and 5062.6 m/z were found to discriminate between the benign and the malignant late stage conditions. Both peaks were up-regulated in the benign condition. The peak at 4086.6 m/z was up-regulated by approximately 1.3-fold (p=0.05), and the peak at 5062.6 m/z was up-regulated by approximately 1.4-fold (p=0.01).

Peak distribution analysis was performed using pairs of discriminatory peaks (Figure 5.12). In the LMR the ellipses representing the standard deviation for the average peak area for the two most discriminatory peaks showed the large spread of the data.

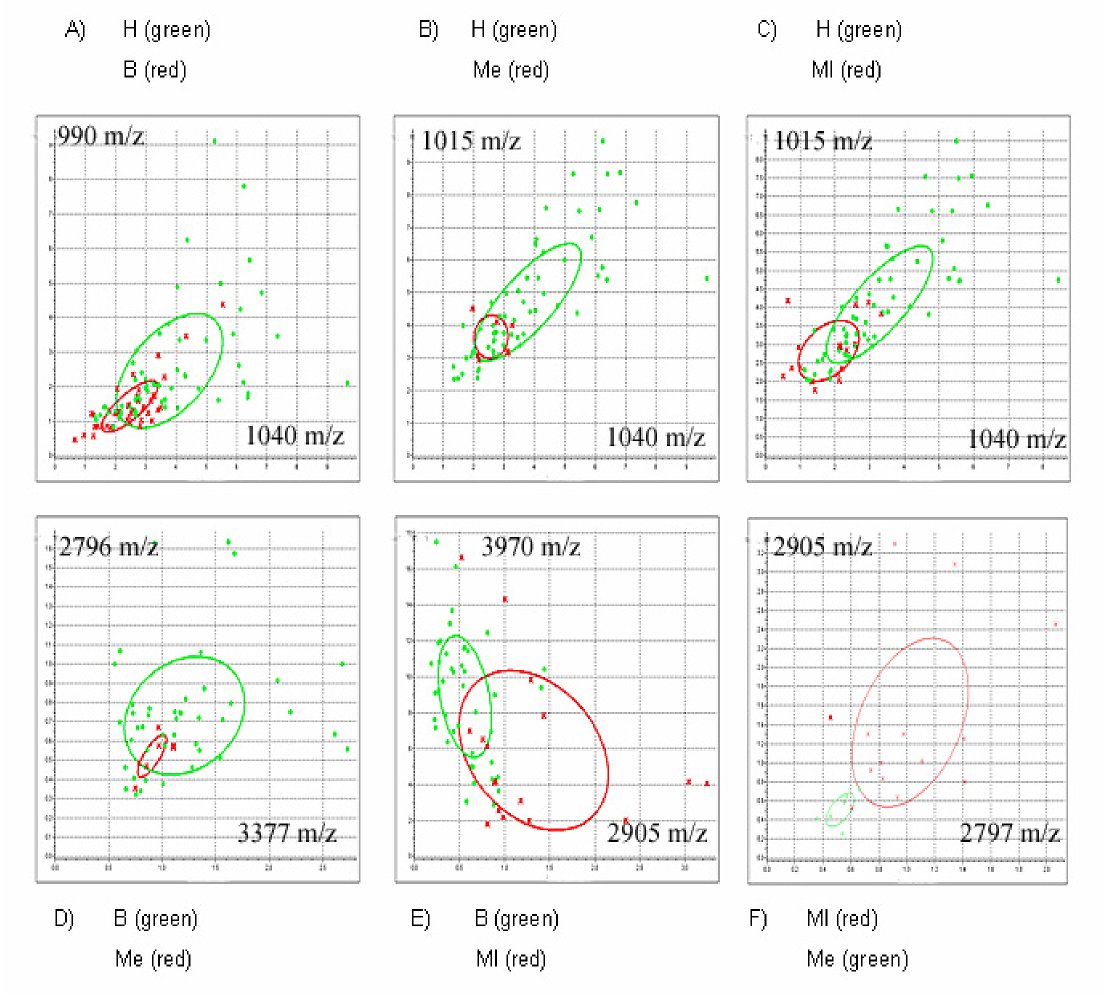


Figure 5.12 2D peak distributions of clinical samples in the LMR. The distributions of the areas of the two most discriminatory peaks in the different clinical conditions are shown. The x-axis shows the peak areas of the most discriminatory peak and the y-axis the peak areas for the second most discriminatory peak. The ellipses represent the standard deviation of the average peak area.

Peak distribution analysis in the HMR showed that the malignant late stage condition separated reasonably well from the healthy, benign and malignant early stage conditions. There was no separation between the healthy, benign and malignant early stage conditions.

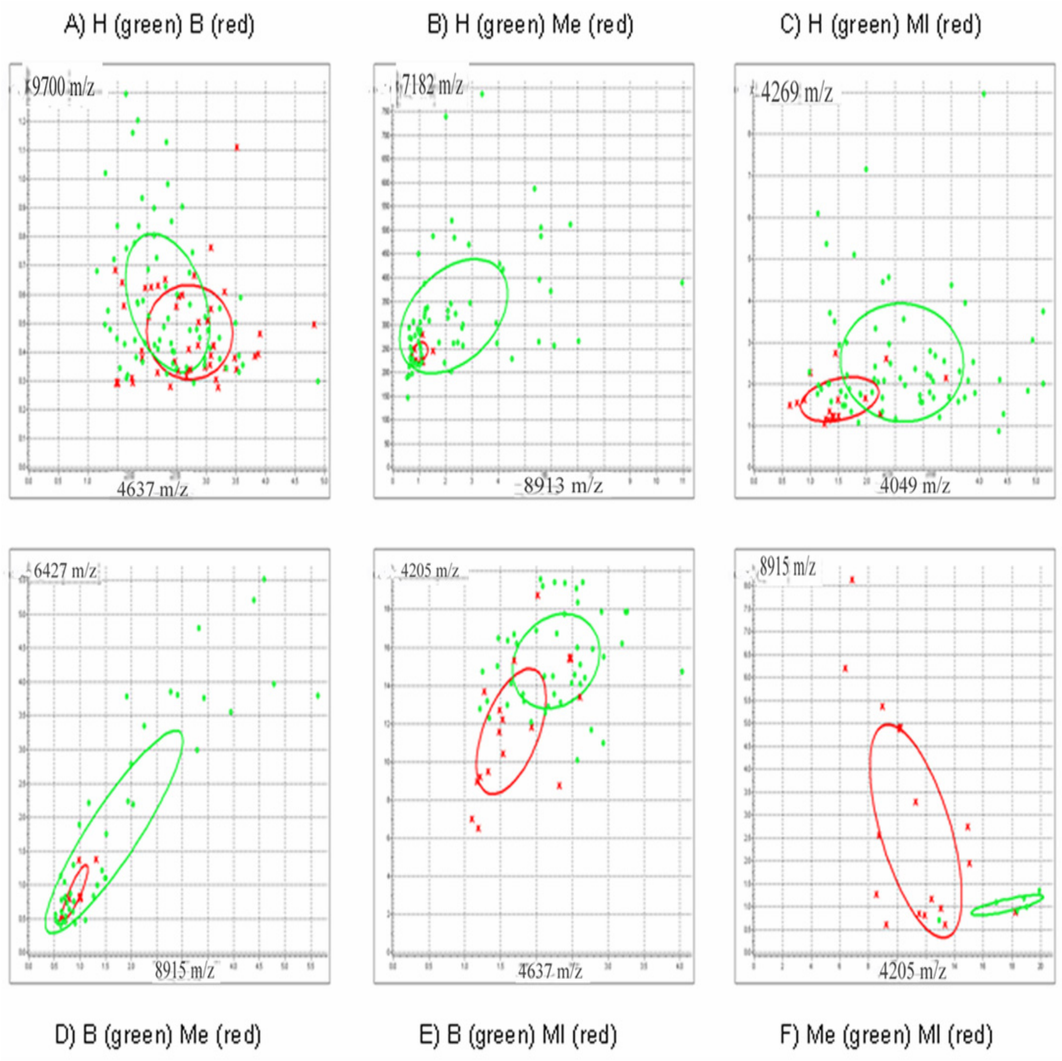
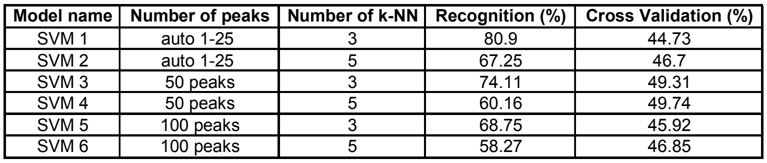


Figure 5.13 2D peak distributions of clinical samples in the HMR. The distributions of the areas of the two most discriminatory peaks in the different clinical conditions are shown. The x-axis shows the peak areas of the most discriminatory peak and the y-axis the peak areas for the second most discriminatory peak. The ellipses represent the standard deviation of the average peak area.

5.4 Classification of clinical samples using SVM

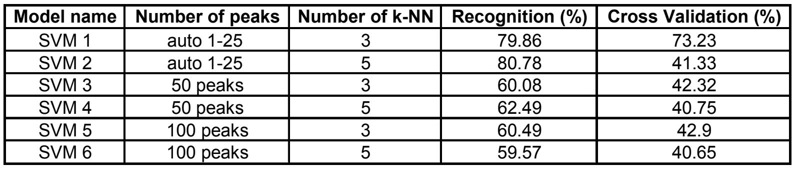
MS spectra consist of high dimensional data which can be used by classification algorithms to classify samples into their respective clinical groups. ClinProTools has a built in function for spectral classification. For the classification of UKOPS samples the Support Vector Machine (SVM) algorithm was chosen. The concept of the SVM was developed by Vladimir Vapnik [Vapnik, 1999]. It has fast become an established pattern recognition tool. In the simplest case, using sophisticated mathematical approaches, the SVM helps to determine an optimal hyperplane separating two clouds of data.

Spectra from the first set of spotting replicates ('set01') were used as a 'training set' to create models. Several tests were done to optimise the recognition capabilities of the model, including the number of peaks used by the model and the number of k-nearest neighbours. The k-nearest neighbours algorithm is amongst the simplest of all machine learning algorithms. Using this algorithm an object is classified by a majority vote of its neighbours, the object is then assigned to the class most common amongst its k nearest neighbors. For spectral recognition, 80% of spectra were used by the model and the remaining 20% were left out for cross-validation as a measure of the reliability of the calculated model (Tables 5.4 & 5.5).



Model name	Number of peaks	Number of k-NN	Recognition (%)	Cross Validation (%)
SVM 1	auto 1-25	3	80.9	44.73
SVM 2	auto 1-25	5	67.25	46.7
SVM 3	50 peaks	3	74.11	49.31
SVM 4	50 peaks	5	60.16	49.74
SVM 5	100 peaks	3	68.75	45.92
SVM 6	100 peaks	5	58.27	46.85

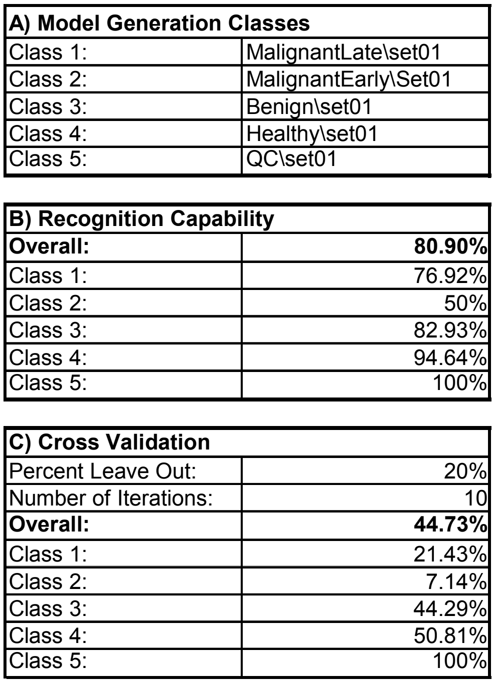
Table 5.4 Optimisation of the SVM's parameters in the LMR. The recognition and cross validation capabilities of the SVM used for classification of clinical samples involved the optimisation of the number of peaks and k-nearest neighbours used as shown.



Model name	Number of peaks	Number of k-NN	Recognition (%)	Cross Validation (%)
SVM 1	auto 1-25	3	79.86	73.23
SVM 2	auto 1-25	5	80.78	41.33
SVM 3	50 peaks	3	60.08	42.32
SVM 4	50 peaks	5	62.49	40.75
SVM 5	100 peaks	3	60.49	42.9
SVM 6	100 peaks	5	59.57	40.65

Table 5.5 Optimisation of the SVM's parameters in the HMR.

The SVM1 model gave the highest recognition and cross validation scores. This model was initially tested on spectral data which included the QC samples. In the LMR, the SVM algorithm automatically selected 24 discriminatory peaks and the results demonstrate that the overall recognition capability of the algorithm was 80.9% and 100% of QC spectra were recognised and cross-validated correctly (Table 5.6).



A) Model Generation Classes	
Class 1:	MalignantLate\set01
Class 2:	MalignantEarly\Set01
Class 3:	Benign\set01
Class 4:	Healthy\set01
Class 5:	QC\set01

B) Recognition Capability	
Overall:	**80.90%**
Class 1:	76.92%
Class 2:	50%
Class 3:	82.93%
Class 4:	94.64%
Class 5:	100%

C) Cross Validation	
Percent Leave Out:	20%
Number of Iterations:	10
Overall:	**44.73%**
Class 1:	21.43%
Class 2:	7.14%
Class 3:	44.29%
Class 4:	50.81%
Class 5:	100%

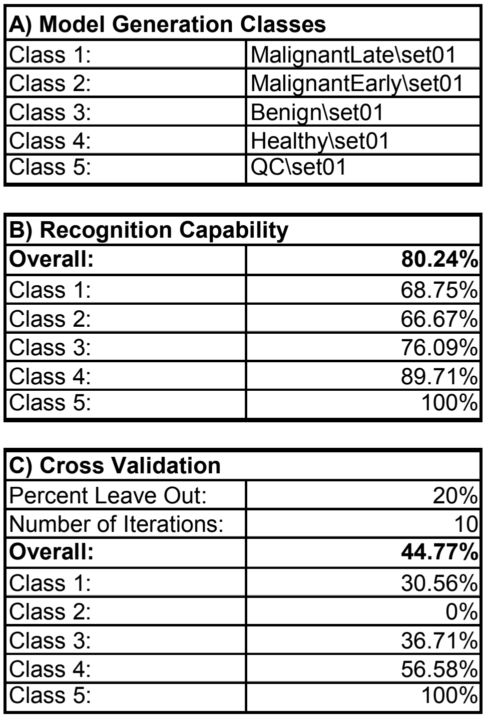
Table 5.6 ClinProTools SVM model generation for the LMR. A) Classes of the clinical conditions, B) the SVM's recognition capability for 80% of spectra from the training set, C) the cross validation results for the remaining 20% of spectra.

D) Integration Regions used for Classification

Index	Mass	Start Mass	End Mass	Weight
1	708.48	704.89	711.06	0.97
5	742.06	737.45	745.85	1.02
7	771.72	766.89	774.05	0.87
22	921.53	915.64	931.04	0.94
32	1064.03	1050.6	1071.58	0.83
54	1391.49	1384.98	1396.63	1.22
71	1643.45	1637.1	1649.41	1.62
77	1724.82	1720.17	1730.94	1.2
87	1912.03	1906.86	1925	0.75
91	1985.93	1978.75	1993.59	1.29
95	2103.39	2094.57	2113.35	1.1
100	2194.48	2190.47	2204.74	0.85
109	2347.43	2339.11	2356.27	0.75
111	2381.07	2369.61	2393.16	0.74
112	2427.27	2412.5	2438.82	0.74
114	2465.08	2454.65	2469.76	1.03
118	2541.01	2531.82	2545.92	0.8
124	2666.82	2653.5	2677.85	0.83
126	2703.61	2697.63	2710.09	0.76
129	2774.16	2752.19	2789.64	0.86
130	2796.24	2789.64	2806.65	0.95
134	2869	2859.76	2877.93	0.9
135	2888.39	2877.93	2898.73	1.77
136	2905.47	2898.73	2915.9	2.77

Table 5.6 continued. ClinProTools SVM model generation for the LMR. D) The integration m/z regions of the 24 peaks automatically selected for classification, here the peak weight refers to the ranking of the peak which is dependent on the separation properties of the peak. Peaks with good separation properties are ranked highly and therefore had a greater weight.

In the HMR, the SVM algorithm automatically selected 23 discriminatory peaks and the results demonstrate that the overall recognition capability of the algorithm was 80.24% all QC spectra were recognised and cross-validated correctly (Table 5.7).



A) Model Generation Classes	
Class 1:	MalignantLate\set01
Class 2:	MalignantEarly\set01
Class 3:	Benign\set01
Class 4:	Healthy\set01
Class 5:	QC\set01

B) Recognition Capability	
Overall:	**80.24%**
Class 1:	68.75%
Class 2:	66.67%
Class 3:	76.09%
Class 4:	89.71%
Class 5:	100%

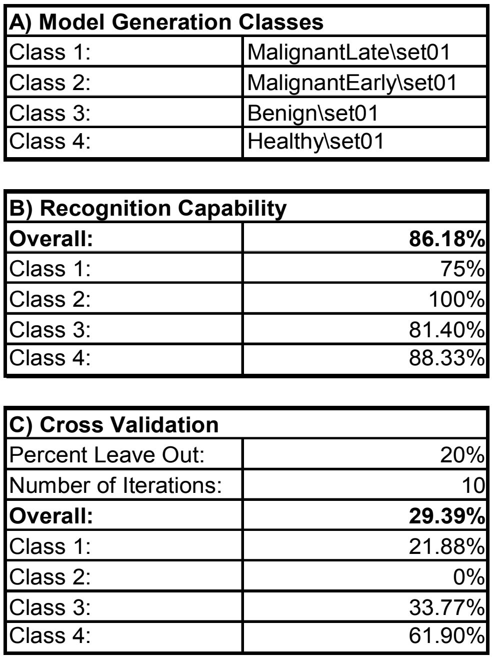
C) Cross Validation	
Percent Leave Out:	20%
Number of Iterations:	10
Overall:	**44.77%**
Class 1:	30.56%
Class 2:	0%
Class 3:	36.71%
Class 4:	56.58%
Class 5:	100%

Table 5.7 ClinProTools SVM model generation for the HMR. A) Classes of the clinical conditions, B) the SVM's recognition capability for 80% of spectra from the training set, C) the cross validation results of the remaining 20% of spectra.

D) Integration Regions used for Classification				
Index	Mass	Start Mass	End Mass	Weight
1	4052.64	4023.84	4071.76	1.53
2	4085.21	4071.76	4108.98	0.77
3	4204.57	4146.51	4243.28	0.79
6	4460.16	4431.15	4498.18	0.74
7	4517.76	4498.18	4543.45	1.24
9	4636.69	4591.79	4667.16	1.82
13	4913.91	4905.28	4931.07	1.29
15	5061.92	5034.37	5113.05	2.13
20	5433.5	5396.42	5454.76	0.85
22	5586.23	5552.22	5607.79	1
24	5745.73	5698.63	5778.98	1
27	6080.9	6054.53	6120.97	0.83
29	6240.24	6213.51	6257.29	2.74
34	6626.33	6551.84	6680.79	1.11
36	6793.12	6756.81	6844.78	1.12
42	7217.61	7198.09	7256.2	0.84
47	7629.24	7617.61	7686.81	1.05
52	8227.55	8176.56	8239.04	1.06
53	8340.4	8306.8	8366.41	0.86
55	8585.3	8498.62	8647.2	0.98
57	8757.14	8718.24	8821.78	1.46
58	8916.4	8821.78	8986.36	0.78
66	9789	9752.14	9804.02	0.95

Table 5.7 continued. ClinProTools SVM model generation for the HMR. D) The integration m/z regions of the 24 peaks automatically selected for classification, here the peak weight refers to the ranking of the peak which is dependent on the separation properties of the peak. Peaks with good separation properties are ranked highly and therefore had a greater weight.

For the classification of clinical spectra (without the QC samples), the SVM algorithm automatically selected 24 peaks in the LMR, based on the T-test p-values to give 86% overall recognition capability for 80% of spectra from the training set and was able to correctly recognise 75%, 100%, 81.4% and 88.3% of the malignant late stage, malignant early stage, benign and healthy samples (Table 5.8). For cross-validation the remaining 20% of the spectra were classified by the model giving an overall score of 29.39% which is considered poor. The algorithm was able to recognise more of the healthy spectra (61.9%) than either of the benign and malignant late stage spectra (33.8% and 21.9%; Table 5.8). Of the 24 peaks selected by the SVM model peaks at 1391.5 m/z, 2796 m/z and 2905 m/z were also reported in Table 5.2 which showed peaks ranked according to Wilcoxon p-values.



A) Model Generation Classes	
Class 1:	MalignantLate\set01
Class 2:	MalignantEarly\set01
Class 3:	Benign\set01
Class 4:	Healthy\set01

B) Recognition Capability	
Overall:	**86.18%**
Class 1:	75%
Class 2:	100%
Class 3:	81.40%
Class 4:	88.33%

C) Cross Validation	
Percent Leave Out:	20%
Number of Iterations:	10
Overall:	**29.39%**
Class 1:	21.88%
Class 2:	0%
Class 3:	33.77%
Class 4:	61.90%

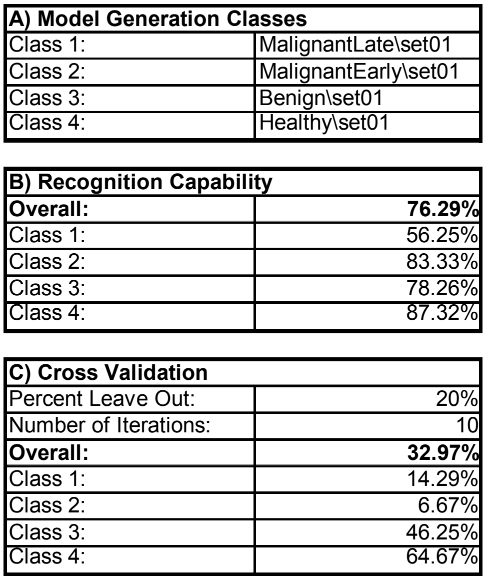
Table 5.8 ClinProTools SVM model generation for the LMR. A) Classes of the clinical conditions, B) the SVM's recognition capability for 80% of spectra from the training set, C) the cross validation results of the remaining 20% of spectra.

D) Integration Regions used for Classification

Index	Mass	Start Mass	End Mass	Weight
1	708.48	704.89	711.06	0.97
5	742.06	737.45	745.85	1.02
7	771.72	766.89	774.05	0.87
22	921.53	915.64	931.04	0.94
32	1064.03	1050.6	1071.58	0.83
54	1391.49	1384.98	1396.63	1.22
71	1643.45	1637.1	1649.41	1.62
77	1724.82	1720.17	1730.94	1.2
87	1912.03	1906.86	1925	0.75
91	1985.93	1978.75	1993.59	1.29
95	2103.39	2094.57	2113.35	1.1
100	2194.48	2190.47	2204.74	0.85
109	2347.43	2339.11	2356.27	0.75
111	2381.07	2369.61	2393.16	0.74
112	2427.27	2412.5	2438.82	0.74
114	2465.08	2454.65	2469.76	1.03
118	2541.01	2531.82	2545.92	0.8
124	2666.82	2653.5	2677.85	0.83
126	2703.61	2697.63	2710.09	0.76
129	2774.16	2752.19	2789.64	0.86
130	2796.24	2789.64	2806.65	0.95
134	2869	2859.76	2877.93	0.9
135	2888.39	2877.93	2898.73	1.77
136	2905.47	2898.73	2915.9	2.77

Table 5.8 continued. ClinProTools SVM model generation for the LMR. D) The integration m/z regions of the 24 peaks automatically selected for classification, here the peak weight refers to the ranking of the peak which is dependent on the separation properties of the peak.

In the HMR (without the QC samples), the SVM automatically selected 24 peaks based on the T-test p-values to give 76.29% overall recognition capability of 80% of the spectra from the training set. For cross-validation the remaining 20% of spectra were classified with an overall score of 32.97% (Table 5.9). Of the 24 peaks selected by the SVM model, peaks at 4049.5 m/z, 4204.6 m/z, 4636.7 m/z and 6241.4 m/z were also reported in Table 5.3 where peaks were ranked according to Wilcoxon p-values.



A) Model Generation Classes	
Class 1:	MalignantLate\set01
Class 2:	MalignantEarly\set01
Class 3:	Benign\set01
Class 4:	Healthy\set01

B) Recognition Capability	
Overall:	**76.29%**
Class 1:	56.25%
Class 2:	83.33%
Class 3:	78.26%
Class 4:	87.32%

C) Cross Validation	
Percent Leave Out:	20%
Number of Iterations:	10
Overall:	**32.97%**
Class 1:	14.29%
Class 2:	6.67%
Class 3:	46.25%
Class 4:	64.67%

Table 5.9 ClinProTools SVM model generation for the HMR. A) Classes of the clinical conditions, B) the SVM's recognition capability for 80% of spectra from the training set, C) the cross validation results of the remaining 20% of spectra.

D) Integration Regions used for Classification				
Index	Mass	Start Mass	End Mass	Weight
1	4049.54	4024.58	4068.91	1.69
2	4085.83	4068.91	4108.55	0.96
3	4204.56	4148.45	4243.37	0.97
4	4269.99	4244.24	4325.33	0.9
8	4520.07	4495.67	4544.24	1.27
10	4636.66	4590.98	4667.54	1.8
13	4884.69	4859.62	4904.08	0.93
14	4915.69	4904.08	4934.56	1.19
17	5061.53	5024.29	5107.07	2.27
18	5129.26	5109.12	5141.4	0.91
24	5481.79	5453.73	5503.71	0.75
25	5584.98	5555.3	5607.82	1.03
27	5745.26	5677.33	5779.39	0.96
32	6241.38	6203.12	6280.6	2.13
34	6426.38	6391.45	6472.98	0.73
36	6623.96	6552.53	6679.81	1.16
38	6792.77	6756.07	6843.53	0.94
50	7647.82	7620.49	7648.96	1.19
56	8231.32	8175.47	8238.88	1.04
57	8343.05	8307.01	8371.67	1
60	8586.01	8498.54	8644.07	1.19
62	8753.87	8717.49	8819.1	1.64
63	8913.51	8819.1	8976.27	0.77
70	9791.25	9746.89	9803.73	0.79

Table 5.9 continued. ClinProTools SVM model generation for the HMR. D) The integration regions of the 24 peaks automatically selected for classification.

The SVM algorithm was then used to classify a second set of data from the other duplicate spots. Results showed that for the LMR data 77% of the healthy, 51% of the benign, 16% of the malignant early stage and 31% of the malignant late stage spectra were classified correctly, this represents sensitivity of the algorithm. In the HMR 63% of the healthy, 67% of the benign, 0% of the malignant early stage and 63% of the malignant late stage spectra were classified correctly. In addition, 92% of QC spectra from both mass ranges were correctly classified (Table 5.10).

Clinical condition	LMR % of spectra classified correctly	HMR % of spectra classified correctly
Mal Late	31	63
Mal Early	16	0
Benign	51	67
Healthy	77	63
QC	92	92

Table 5.10 SVM Classification results. The SVM algorithm was used to classify a ‘test-set’ of spectral data from all conditions. Results for the LMR and HMR are shown.

During the initial model generation stage the SVM results showed that 100% of the QC data were recognised and cross-validated correctly. The SVM was then tested on a second data set from the duplicate spot and results showed that 92% of QC spectra in both the LMR and HMR were classified correctly. This suggests that the sensitivity of the algorithm is greater than 90%. Results from the clinical conditions suggest that the classification algorithm may have been limited by the number of spectra available and the relatively large variation across samples within the conditions. The overall scores for the healthy and benign conditions were better than the scores for the malignant stage conditions where there were fewer samples. In summary, this data shows that the spectra do not contain information for reliably predicting the clinical conditions under investigation.

5.5 Discussion

The human serum proteome is an extremely complex biological sample containing information on numerous biological processes that takes place in the body. Cancer cells can release proteins into the extracellular fluid through secretion of intact or cleaved proteins in response to changes taking place in the cancer tissue microenvironment and due to cancer cell-host interactions. Many of these products will end up in the bloodstream and hence serve as potential serum biomarkers. Therefore, studying the serum proteomes of healthy, benign and malignant donors is the logical starting point for identifying diagnostic biomarkers and therapeutic targets for cancer [Grizzi and Chiriva-Internati, 2006; aoui-Jamali and Xu, 2006]. In recent years, many advances have been made in the field of proteomics. In particular, the use of high-throughput mass spectrometry methods for analysing complex proteomes (e.g. human serum) has become widespread.

The aim of the work presented in this chapter was to analyse clinical samples from the UKOPS collection to determine statistically significant discriminatory peptide/protein peaks and to determine if these peaks could be used for the classification of ovarian cancer. This aim was tackled by processing samples collected in accordance with a standardised protocol on a semi-automated and optimised bead-based serum polypeptide extraction platform followed by MALDI-TOF MS profiling.

Firstly, the intra-assay reproducibility was calculated using QC samples which had been processed alongside the clinical samples. Results demonstrated that the intra-assay reproducibility did not exceed 20% variance. In fact 85% of LMR peaks and 75% of HMR peaks were found to be below a 15% threshold. Despite this, spot-to-spot variations in peak profiles of the same sample were evident. Indeed, one of the major drawbacks in using MALDI-TOF MS for serum profiling is the inherent variability of the sample preparation process including automated sample spotting and inconsistencies in crystallisation [Cohen and Chait, 1996]. The co-crystallisation of the analyte with the matrix is a prerequisite for uniform ionisation of biomolecules.

A major issue is the non-homogeneous distribution of the analyte in the co-crystallite. To minimise the effects of non-homogeneous distribution of the analyte, spectra were generated by accumulating 400 laser shots over 8 different locations on the spots. The QC sample was included in the assay to across for the technical variability of the semi-automated platform at the time of clinical sample analysis.

Herein the comparison of MALDI-TOF MS spectral peak profiles has shown that the intra-condition variation was relatively high (30-60%) which reflects the combination of technical error and the biological heterogeneity between samples. In the LMR, there were significant differences in the abundance of certain peaks between the healthy and malignant late stage, healthy and benign, and benign and malignant late stage conditions. Of these only one statistically significant peak was found to discriminate between the malignant early and late stage. However, no statistically significant peaks were found to discriminate the healthy and benign conditions from malignant early stage condition. In the HMR, no statistically significant peaks were found to discriminate the healthy condition from the benign or malignant early stage conditions, though 5 peaks were found to change significantly between the benign condition and the malignant late stage conditions.

Interestingly, several peaks were found to be common in discriminating clinical conditions. In the LMR, peaks at 989 m/z and 2905 m/z were common discriminatory peaks between the clinical conditions. The average peak area of peak 989 m/z was higher in the healthy condition compared with benign and malignant conditions and the average peak area of 2905 m/z was higher in malignant late stage compared with the other conditions. However, the CVs of both these peaks were high. For the peak at 989 m/z the CV of the average peak area ranged from 37% to 72% and for peak at 2905 m/z from 33% to 64% between the clinical conditions. In the HMR, peaks at m/z 4050 and 4205 were found to be common discriminatory peaks and on average the peak areas of both of these peaks were lower in the malignant late stage condition. Again the CVs of the average peak area were high. For the peak at 4050 m/z the CV ranged from 28% to 46% and for peak at 4205 m/z from 14% to 33% between clinical conditions. The limited number of samples for the malignant conditions and the intra-condition biological variation which was found to be highest in the healthy and benign conditions in both mass ranges could explain the lack of discrimination between the conditions.

Using a combination of peaks it was possible to classify spectra into their respective clinical groups. The results from the SVM analysis demonstrated that in the LMR 77% of healthy, 51% benign, 17% malignant early stage and 31% malignant late stage spectra were classified correctly. In the HMR, 63% healthy, 67% benign, 0% malignant early stage and 63% malignant late stage spectra were classified correctly. These poor classification results could be attributed to the limited number of samples for the malignant conditions as well as a lack of robust discriminatory peaks. Interestingly, none of the benign and malignant early spectra were classified as malignant late stage and none of the malignant late stage as malignant early stage spectra. This was found to be true in both mass ranges.

Results from the discriminatory peak and SVM analysis suggest that a panel of peptide masses at 989 m/z, 1064 m/z, 1392 m/z, 2796 m/z, 2905 m/z, 4049 m/z, 4205 m/z, 4637 m/z, 5062 m/z and 6241 m/z could be potentially useful markers of disease, albeit with poorer performance relative to the existing marker CA125. For CA-125 sensitivities of 85% and specificities of 65% are reported. Combining the discriminatory peaks with CA-125 values could provide better sensitivities and specificities. The average peak area of these masses changed depending on clinical condition. Although not identified, these peaks are likely to be fragments of acute-phase, complement and clotting proteins that are commonly found in serum.

In conclusion, the results demonstrate the feasibility of the technology platform for discriminating clinical samples. However, the intra-condition variation was a major limitation. A larger sample set with more malignant samples would be needed to better validate test results. Standardised operating procedures for donors would also be required to minimise biological variation. Furthermore, direct mass spectrometric serum profiling has a limited dynamic range and difficulties in providing the identification of the distinctive peptides and proteins. It is most likely that the distinctive profiles may result from the differential expression of relatively abundant serum proteins and their fragments associated with the host response to tumours and generated by exoproteases as previously reported by Villanueva et al. However, in the case of ovarian cancer, it appears that these surrogate markers are less informative than for other cancer types. This may be due to different or lower exopeptidase activities in ovarian cancer samples compared with breast, prostate and bladder cancer.

Attempts to isolate masses of interest using 10kDa ultrafiltration devices followed by 1D-SDS PAGE separation for identification of masses of interest were unsuccessful and only very low levels of peptides could be recovered. It is speculated that many low molecular weight peptides and proteins are bound to the high molecular weight and high abundance carrier proteins such as albumin [Lowenthal et al., 2005]. MALDI-QTOF MS analysis is a good alternative method for peptide identification as this allows direct MS/MS sequencing of discriminatory masses. Work is underway to identify these peaks using MALDI-QTOF, although it should be noted that the sensitivity of the technique is lower than that of MALDI-TOF MS with fewer peptides detected, and the observed mass range is reduced.

Finally, MS-based peak pattern recognition is a useful tool for discovery phase research with target masses being identified and characterised prior to the possible translation to the clinic. MALDI-TOF protein profiling however, provides only a limited mass window for putative biomarker analysis. Sensitivity of detection and coverage in clinical proteomics can be effectively improved with extensive pre-fractionation strategies to remove high abundant proteins which can mask the detection of lower abundance protein species. This is the main focus of the next chapter.

Chapter 6: Fluorescence Two-Dimensional Difference Gel Electrophoresis-based profiling of case-control sera for the identification of putative ovarian cancer biomarkers.

6.1 Introduction

Work from the previous chapter demonstrated that the direct analysis of human serum using MALDI-TOF MS profiling has several limitations. Firstly, for serum the method has a limited dynamic range of detection (10^2-10^6). Secondly, only low mass polypeptides are sampled in the MS. Thirdly accuracy of quantification may be compromised by the inherent variability of the crystallisation and ionisation processes. Fourthly, it is difficult to directly identify discriminatory masses of interest without using online tandem MS which is inherently less sensitive. It is well documented that the serum proteome is an extremely complex protein mixture and conceivably contains all proteins (and fragments thereof) that are expressed in the cells and tissues of an organism. The serum proteome also has an exceptionally large dynamic range of protein expression with protein concentrations spanning 10 orders of magnitude [Liang and Chan, 2007]. For putative biomarker discovery it is imperative to employ separation techniques which will facilitate the analysis of differentially expressed proteins across a broad dynamic range.

While two-dimensional electrophoresis (2-DE) is limited in the ability to detect low abundance and hydrophobic proteins, it still remains a valuable method for the separation and profiling of complex mixtures of proteins. Fluorescence two-dimensional Difference Gel Electrophoresis (2D-DIGE) was developed for improved multiplex proteomic profiling based on the spectrally resolvable fluorescent dyes Cy2, Cy3 and Cy5 [Tonge et al., 2001; Gharbi et al., 2002]. Fluorescence 2D-DIGE was employed to complement the MALDI-TOF MS serum profiling work from the previous chapters by extending the mass range of proteins detected and hopefully to increase the dynamic range of detection of serum proteins. However, the complex nature of serum and the presence of a few proteins at very high concentration levels (mg/mL) makes detection of low-abundance proteins by 2-DE challenging, since the sample loading capacity is limited and the presence of high-abundance proteins can mask the detection of low-abundance proteins [Song and Hanash, 2006]. Consequently, depletion and fractionation strategies have become popular for dividing the serum proteome into smaller and simpler subsets for the detection of as many proteins as possible, including those at lower abundance [Echan et al., 2005]. A highly promising first step for most analysis strategies of serum or plasma is to deplete the major proteins. Classically, Cibacron Blue and protein A/G chromatography methods have been used to deplete serum of albumin and the immunoglobulins respectively [Kim and Kim, 2007]. Ideally, for biomarker discovery it is desirable to deplete as many high-abundance proteins as possible while minimising incidental losses of non-targeted proteins. In recent years, a range of methods to deplete high-abundance proteins have been evaluated. For example, a recently commercialised HPLC polyclonal antibody column and its spin column version (MARS, Agilent Technologies) are very promising methods for depleting human serum or plasma samples. Polyclonal antibodies are more likely to deplete multiple structural forms of a protein and thus, these columns enabled 10- to 20-fold higher amounts of depleted serum samples to be applied to 2-D gels [Bjorhall et al., 2004; Echan et al., 2005; Sriyam et al., 2007; Liu et al., 2006]. Alternatively, strategies for compressing the protein dynamic range have also been commercialised. An example of this is the ProteoMiner protein enrichment technology (BioRad), which is based on treatment of complex protein samples with a large, highly diverse library of hexa-peptides bound to chromatographic supports [Guerrier et al. 2006; Guerrier et al. 2008; Boschetti et al. 2008]. In theory, the hexa-peptides can bind to all unique protein sequences in the mixture. Because the bead volume limits binding capacity, high-abundance proteins quickly saturate their ligands and excess protein can be washed out. In contrast, low-abundance proteins are concentrated on their specific ligands, thereby decreasing the dynamic range of proteins in the sample. When analysed using downstream proteomic applications, the number of protein species detected is dramatically increased.

The main focus of this chapter is the use of a 2D-DIGE based protein profiling strategy for the differential analysis of pooled clinical serum samples. Two different fractionation strategies were used to enrich the lower abundance proteins prior to 2D-

DIGE including a commercialised HPLC column containing antibodies to seven of the most abundant serum proteins, and a protein enrichment kit for "dynamic range compression". Each of these fractionation methods were compared with the analysis of unfractionated sera and case control samples (UKOPS) were compared for the discovery of putative biomarkers of ovarian cancer. Differentially expressed protein features were identified using MALDI-TOF peptide mass fingerprinting (PMF) and LC-MS/MS analysis.

6.2 Pooled UKOPS clinical sera

Due to the limited availability of clinical samples equal volumes of each serum sample from the UKOPS collection were pooled into their respective clinical groups. Thus, serum pools were created for the healthy, benign and malignant late stage conditions. In addition, since only a very limited number of malignant early stage samples (n=6) were available, this condition was excluded from the 2D-DIGE experiments. Pooling of biological samples is one method that can allow many samples to be studied simultaneously, while preventing false conclusions based on a limited number of individual samples. However, pooling strategies hide the underlying variation across sample sets and may reveal average data that is possibly skewed by outliers. The obvious advantage of pooling is to save time and money. Analysis of all samples individually could not be accomplished using 2D-DIGE.

The protein concentration of each pool was determined using the Pierce BCA protein assay using BSA to generate a standard curve. For accurate measurement within the linear range of the assay, pooled sera from each clinical condition were diluted 1:100 with HPLC grade deionised water. Results showed that the protein concentrations of the healthy, benign and malignant late stage pools were 93.9 mg/mL, 91.9 mg/mL and 85.3 mg/mL respectively. All pools were then equalised to 85.3 mg/mL with HPLC grade deionised water.

6.2.1 2D-DIGE analysis of pooled serum samples.

The depletion of major proteins is typically accompanied with a significant loss of components of potential interest. For instance, the removal of serum albumin can result in the removal of a multitude of proteins and peptides bound to it, which might be valuable for diagnostic purposes [Granger et al., 2005]. With this in mind, a 6 gel experiment was initially performed to compare the number of differentially expressed protein features between the unfractionated clinical pools. 50µg of protein was labelled with Cy3 or Cy5. In addition, to avoid any protein-specific dye effects, dye swaps were performed and samples were run in quadruplicates as shown in Figure 6.1.

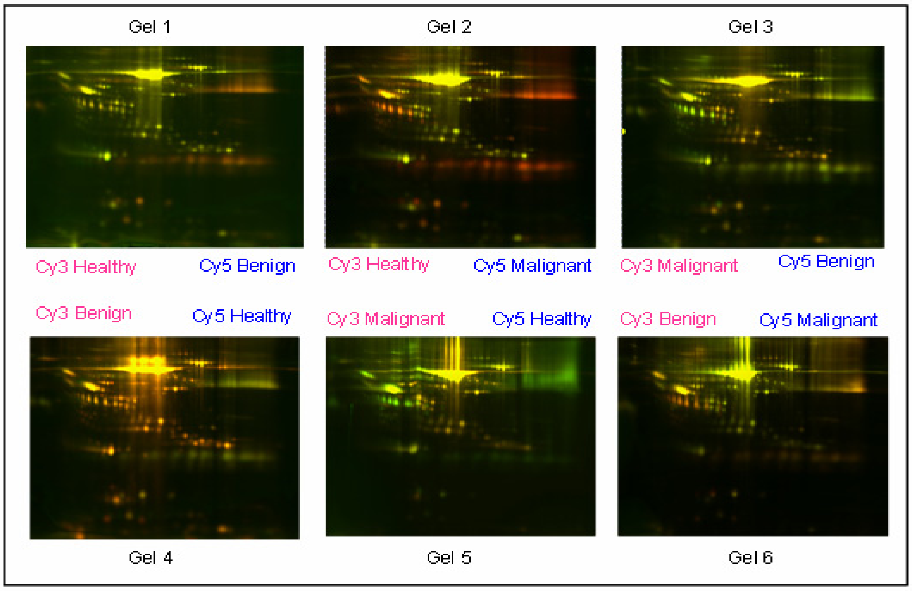


Figure 6.1 Experimental design of DIGE labelling to compare protein profiles across pools of unfractionated UKOPS serum samples. Clinical samples were pooled into healthy, benign and malignant groups and then prepared in NHS-lysis buffer. Equal amounts of protein (50 µg) were labelled with 200 pmol of Cy3 or Cy5 as shown. Cy3 and Cy5 labelled samples were then mixed appropriately and run on six individual 2-D gels and fluorescence images captured. The figure shows the labelling strategy used for the 6 gels analysing unfractionated sera (50µg protein per dye) and superimposed images generated of each of the six gels (using Image Quant software). The same labelling strategy was later used for the 6 gel analyses of MARS fractionated sera.

Samples were diluted in a urea and CHAPS-based 2D-lysis buffer and subjected to 2D-DIGE analysis, as described in Chapter 2. Briefly, equal amounts of protein (50 µg) from each clinical condition were labelled with Cy3 and Cy5 in quadruplicates. The samples were mixed and run on 24 cm pH 3-10 non-linear IPG strips in the first dimension, followed by 12% SDS-PAGE in the second dimension. The gels were scanned at two different excitation/emission wavelength combinations generating two fluorescent images of proteins labelled with Cy3 and Cy5. Images were imported into DeCyder software for image analysis and quantitative comparison of differentially expressed proteins. Paired fluorescent images were processed using the differential in-gel analysis (DIA) module of the software; gel images were normalised and spot boundaries defined to calculate spot abundances and fold-ratios of abundance calculated between the superimposable paired images. The matched images were then imported into the biological variance analysis module (BVA), where protein features from a selected master gel were matched with the corresponding features across the other gel images. A representative superimposed image of two Cy dye images from the unfractionated pooled serum samples is shown in Figure 6.2.

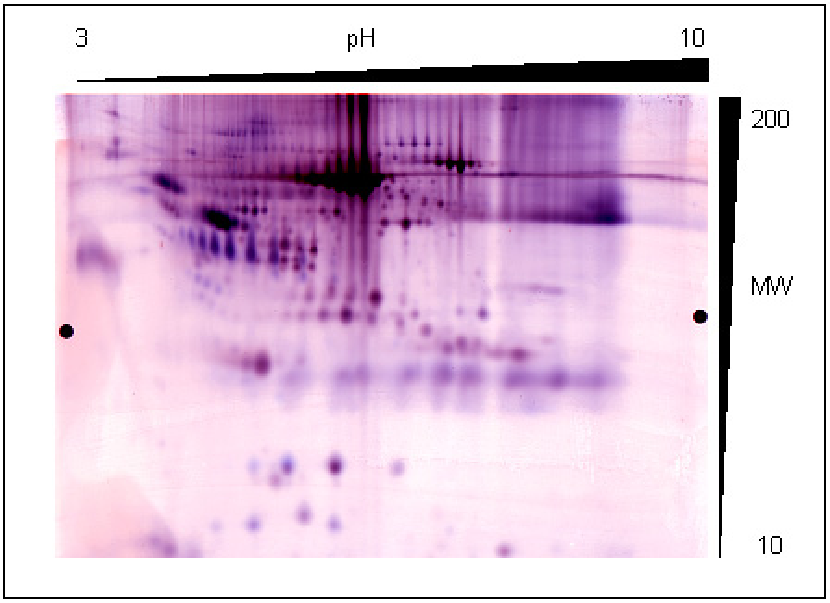


Figure 6.2 Gel image of unfractioned sera used as the ‘master’ gel in DeCyder image analysis. This figure represents the overlaid fluorescent images derived from the unfractionated pooled sera labelled with Cy3 (red-healthy) and Cy5 (blue-malignant). This figure was prepared using Adobe Photoshop.

A total of 934 protein features were detected in the master gel of the unfractionated sera. To keep the cost of the experiment to a minimum, no internal Cy2-labelled control was used. However, this made matching of protein features across the gels more difficult. Despite this, quantitative analysis showed that 39 protein features were up-regulated and 2 were down-regulated in the malignant condition versus the healthy condition (≥ 1.5 average fold-change in abundance, p <0.05, n = 4). Five protein features were found to be differentially expressed between the healthy and benign conditions with all 5 down-regulated in the benign condition. Eight protein features were differentially expressed between the benign and malignant conditions, all of which were up-regulated in the malignant condition. The number of differentially expressed protein features and overlap is represented in Figure 6.3. A total of 48 differentially expressed protein features were found between the 3 conditions out of the 934 detected and matched features. Examples are shown in Figure 6.4.

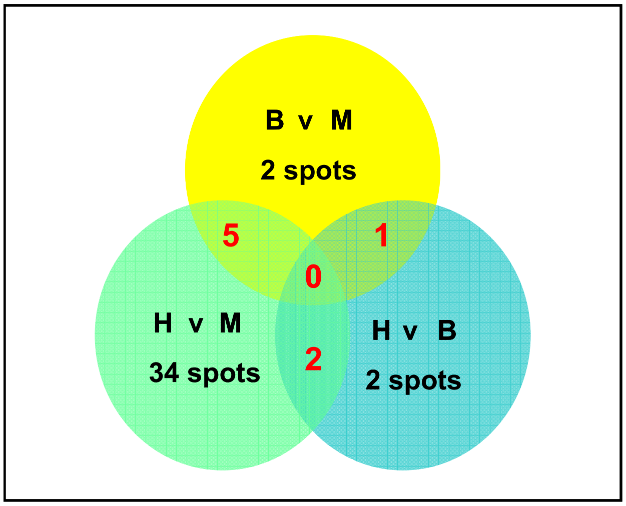


Figure 6.3 Venn diagram showing the number of differentially expressed protein features and their overlap between the three clinical conditions. The ratio in abundance of each protein feature was calculated between each clinical condition). Pairwise comparisons of the clinical conditions were preformed i.e. healthy (H) / malignant (M), benign (B) / malignant (M) and healthy (H) / benign (B) and the numbers of differentially expressed spots that displayed a ≥ 1.5 average fold-change in abundance (p < 0.05) are shown. The numbers shown in red represent protein spots which overlap between the three comparisons.

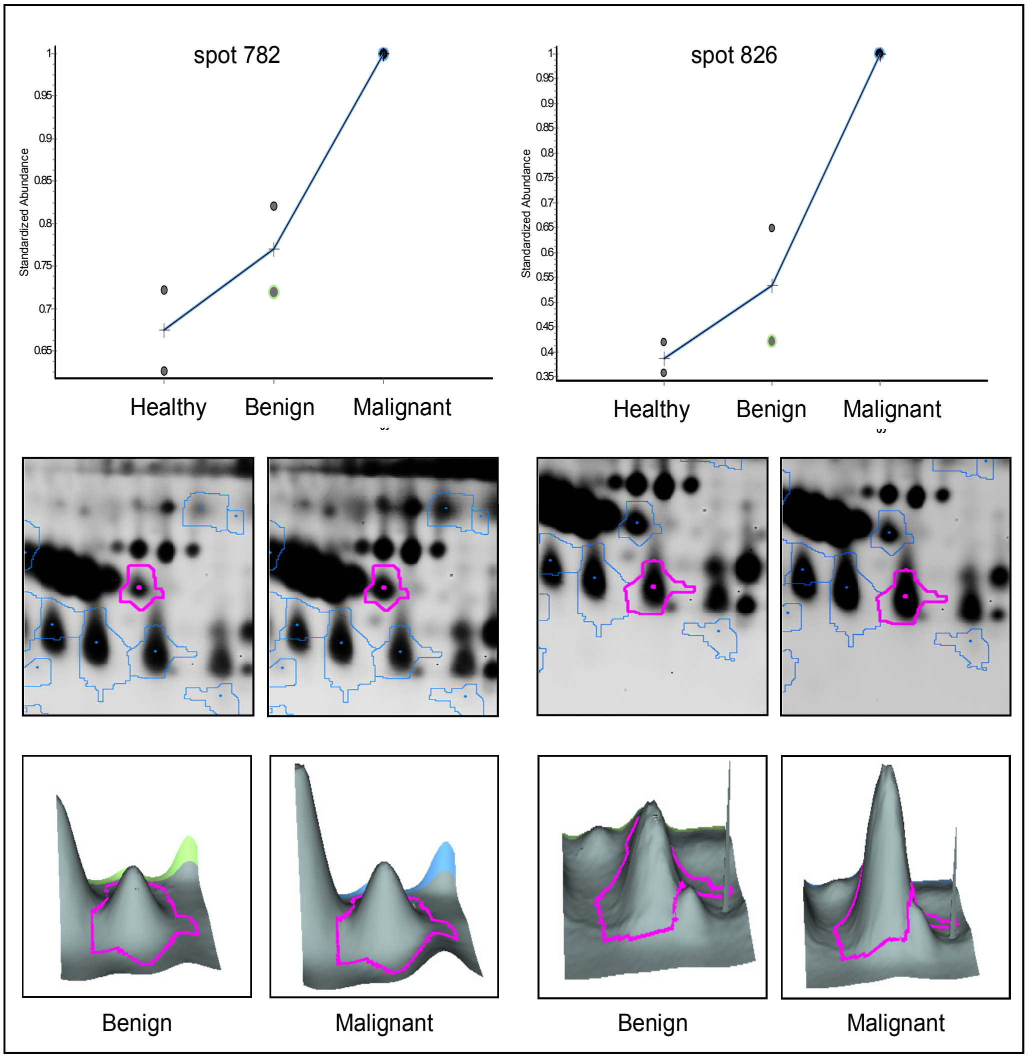
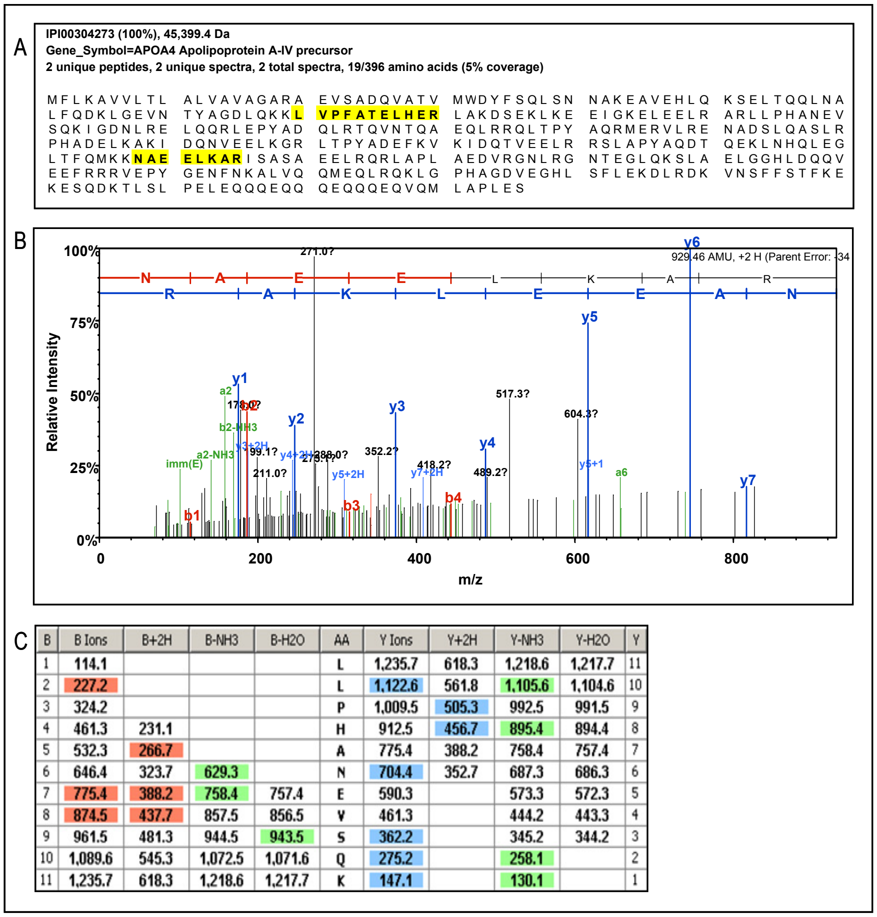


Figure 6.4 Examples of protein features displaying clinical condition-dependent changes in expression from unfractionated samples. Protein spots 782 and 826 displayed clinical stage-dependant differential expression. The peptide mixtures of the trypsin digestion from each of these protein spots were analysed by LC-MS/MS and the experimental peptide fragment masses were searched against theoretical masses using Mascot. Both spots yielded multiple protein identities including haptoglobin (HP), albumin (ALB) and apolipoprotein AIV (APOA4). Graphs were derived from DeCyder image analysis and data points are shown for replicate measurements with lines joining the average values. 3D images of spots are shown for the benign and malignant conditions.

Gels were stained with CCB and protein spots were excised from gels and subjected to trypsin digestion. The peptide mixtures obtained from each trypsin digest were then subjected to PMF by MALDI-TOF MS. When MALDI-TOF PMF analysis and database searching did not return any significant "hits" for a sample, the peptide mixtures were additionally analysed by LC-MS/MS. Of the 48 differentially expressed protein features, 35 were identified with high confidence (Figure 6.5 & Table 6.1). Of the 13 unidentified protein features, most were of low abundance and gave poor spectra, and thus could not be identified from database searches.



IPI00304273 (100%), 45,399.4 Da
Gene_Symbol=APOA4 Apolipoprotein A-IV precursor
2 unique peptides, 2 unique spectra, 2 total spectra, 19/396 amino acids (5% coverage)

```
MFLKAVVLTL  ALVAVAGARA  EVSADQVATV  MWDYFSQLSN  NAKEAVEHLQ  KSELTQQLNA
LFQDKLGEVN  TYAGDLQKKL  VPFATELHER  LAKDSEKLKE  EIGKELEELR  ARLLPHANEV
SQKIGDNLRE  LQQRLEPYAD  QLRTQVNTQA  EQLRRQLTPY  AQRMERVLRE  NADSLQASLR
PHADELKAKI  DQNVEELKGR  LTPYADEFKV  KIDQTVEELR  RSLAPYAQDT  QEKLNHQLEG
LTFQMKKNAE  ELKARISASA  EELRQRLAPL  AEDVRGNLRG  NTEGLQKSLA  ELGGHLDQQV
EEFRRRVEPY  GENFNKALVQ  QMEQLRQKLG  PHAGDVEGHL  SFLEKDLRDK  VNSFFSTFKE
KESQDKTLSL  PELEQQQEQQ  QEQQQEQVQM  LAPLES
```

B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	114.1				L	1,235.7	618.3	1,218.6	1,217.7	11
2	227.2				L	1,122.6	561.8	1,105.6	1,104.6	10
3	324.2				P	1,009.5	505.3	992.5	991.5	9
4	461.3	231.1			H	912.5	456.7	895.4	894.4	8
5	532.3	266.7			A	775.4	388.2	758.4	757.4	7
6	646.4	323.7	629.3		N	704.4	352.7	687.3	686.3	6
7	775.4	388.2	758.4	757.4	E	590.3		573.3	572.3	5
8	874.5	437.7	857.5	856.5	V	461.3		444.2	443.3	4
9	961.5	481.3	944.5	943.5	S	362.2		345.2	344.2	3
10	1,089.6	545.3	1,072.5	1,071.6	Q	275.2		258.1		2
11	1,235.7	618.3	1,218.6	1,217.7	K	147.1		130.1		1

Figure 6.5 An example of a differentially expressed protein feature identified by LC-MS/MS. The trypsin digested peptide mixture of spot number 826 from the unfractionated pooled sera experiment was analysed by LC-MS/MS. A) Apolipoprotein AIV was identified by two unique peptides, (highlighted in yellow in the protein sequence). B) The MS/MS spectrum of peptide NAEELKAR and C) table indicating the masses of identified fragment ions (data produced using Scaffold software, more examples are shown in Appendix 2).

								Healthy / Malignant		Healthy /Benign		Benign / Malignant	
Spot No.	Protein AC	IPI No.	Score	Seq. Cov (%)	No. peptides	Mw	pI	Av. Ratio	T-test	Av. Ratio	T-test	Av. Ratio	T-test
232	Ceruloplasmin 97 kDa protein (CP)	IPI00794184	40	2	2	97007	5.31	-2.88	0.0095	-1.34	-	-1.51	-
232	Albumin (ALB)	IPI00022434	30	8	5	73881	6.33	-2.88	0.0095	-1.34	-	-1.51	-
233	Ceruloplasmin precursor (CP)	IPI00017601	38	2	2	122983	5.44	-2.59	0.0047	-	-	-	-
282	Not identified							-2.62	0.002	-1.1	0.4	-1.69	0.035
283	Not identified							-2.4	0.0095	-1.09	0.59	-1.58	0.081
284	Albumin (ALB)	IPI00022434	29	3	2	73881	6.33	-2.86	0.007	-	-	-	-
303	Not identified							-2.04	0.044	-1.07	0.72	-1.46	0.32
307	Not identified							-2.08	0.0062	-1.4	-	-1.99	-
361	Not identified							-1.63	0.0054	-1.12	-	-1.79	-
392	Albumin (ALB)	IPI00022434	580	39	23	73881	6.33	1.63	0.015	-	-	-	-
458	Immunoglobulin heavy constant mu protein (IGHM)	IPI00385264	54	8	3	43543	5.13	-2.01	0.043	-1.62	0.26	-1.18	0.77
462	Not identified							-1.92	0.011	-1.76	0.21	1.07	0.76
499	Not identified							-1.39	0.25	-2.65	0.015	-2.72	0.031
611	Alpha-1-antichymotrypsin precursor Isoform 1 (SERPINA3)	IPI00847635	65	12	6	47792	5.33	-1.55	0.0092	-1.17	0.28	-1.93	0.03
611	Kininogen-1 precursor Isoform LMW (KNG1)	IPI00215894	37	4	2	48936	6.29	-1.55	0.0092	-1.17	0.28	-1.93	0.03
614	Albumin (ALB)	IPI00022434	251	21	13	71317	5.92	1.6	0.028	-1.03	-	-1.02	-
618	Alpha-1-antichymotrypsin Isoform 1 (SERPINA3)	IPI00550991	65	4	2	50737	5.42	-2.95	0.0093	-1.5	0.21	-1.31	0.34
636	Albumin (ALB)	IPI00022434	229	15	11	71317	5.92	-2.04	0.026	-2.53	-	-1.31	-
636	Immunoglobulin heavy chain constant region alpha 1 protein (IGHA1)	IPI00719233	106	7	4	54150	6.78	-2.04	0.026	-2.53	-	-1.31	-
646	Not identified							-4.09	0.0085	-2.12	-	-1.9	-
648	Albumin (ALB)	IPI00022434	322	24	18	73881	6.33	-1.8	0.17	-1.89	0.04	-1.13	0.67
648	Keratin, type 1 cytoskeletal 9	IPI00019359	68	5	2	62131		-1.8	0.17	-1.89	0.04	-1.13	0.67
654	Immunoglobulin heavy chain constant region alpha 1 protein (IGHA1)	IPI00719233	65	8	5	54150	6.78	-1.99	0.036	-2.2	-	-1.22	-
654	Albumin (ALB)	IPI00022434	51	5	7	73881	6.33	-1.99	0.036	-2.2	-	-1.22	-
678	IGHM protein	IPI00472610	73	4	2	52665		-2.4	0.0065	-2.12	0.15	1.04	0.82
691	IGHG3	IPI00829940	128	11	4	38769	8.36	-4.45	0.0011	-3.62	-	-2.57	-
691	Immunoglobulin heavy constant gamma 3 (G3m marker IGHG3)	IPI00472610	140	13	4	52665		-4.45	0.0011	-3.62	-	-2.57	-
693	Immunoglobulin heavy constant gamma 1 (IGHG1)	IPI00423464	220	16	8	53011	8.84	-4.69	0.0096	-	-	-	-
693	Immunoglobulin heavy constant mu protein (IGHM)	IPI00472610	105	8	3	52665		-4.69	0.0096	-	-	-	
710	Not identified							-2.05	0.041	-1.41	0.18	-1.29	0.37
734	Alpha-2-HS-glycoprotein precursor (AHSG)	IPI00022431	136	16	6	40098	5.43	-1.65	0.01	-1.36	-	1	-
740	Immunoglobulin heavy constant gamma 2 (IGHG2 Fragment)	IPI00399007	32	5	2	46716	7.63	-3.24	0.012	-3.25	0.051	-1.28	0.5
749	Not identified							-1.58		-1.9	0.029	-1.14	

Table 6.1 2D-DIGE based analysis of the unfractionated pooled clinical sera and LC-MS-base protein identification. Protein features displaying differential expression are shown. Values are the average ratio of abundance between different clinical conditions (healthy/malignant, healthy/benign and benign/malignant). T-test *p* values are given as a measure of confidence for each ratio measured. Protein name, IPI accession number, database search score, sequence coverage (%) and the number of matched peptides for each of the identified proteins are shown.

								Healthy / Malignant		Healthy /Benign		Benign / Malignant	
Spot No.	Protein AC	IPI No.	Score	Seq. Cov (%)	No. peptides	Mw	pI	Av. Ratio	T-test	Av. Ratio	T-test	Av. Ratio	T-test
750	Immunoglobulin heavy constant gamma 2 (IGHG2 Fragment)	IPI00399007	166	12	5	46716	7.63	-3.78	0.0033	-3.13	0.042	-1.42	0.33
750	Factor VII active site mutant immunoconjugate	IPI00382606	120	5	8	77386	6.6	-3.78	0.0033	-3.13	0.042	-1.42	0.33
750	Immunoglobulin heavy constant mu protein (IGHM)	IPI00472610	107	7	3	52665		-3.78	0.0033	-3.13	0.042	-1.42	0.33
753	Immunoglobulin heavy constant mu protein (IGHM)	IPI00472610	80	13	4	52633	7.5	-3.87	0.007	-3.94	-	-1.92	-
753	Factor VII active site mutant immunoconjugate	IPI00382606	67	8	6	77386	6.6	-3.87	0.007	-3.94	-	-1.92	-
753	Immunoglobulin heavy constant mu protein (IGHM)	IPI00472610	101	9	4	52665		-3.87	0.007	-3.94	-	-1.92	-
759	Not identified							-4.34	0.0078	-2.85	0.059	-1.8	0.18
763	Immunoglobulin heavy constant gamma 1 (IGHG1)	IPI00384938	97	7	3	53503	8.74	-4.91	0.014	-2.14	0.17	-2.02	0.24
763	Immunoglobulin heavy constant mu protein (IGHM)	IPI00472610	245	9	5	52665		-4.91	0.014	-2.14	0.17	-2.02	0.24
765	Immunoglobulin heavy constant gamma 1 (IGHG1)	IPI00423464	425	29	13	53011	8.84	-4.32	0.0081	-2.67	0.076	-1.85	0.19
765	Immunoglobulin heavy constant mu protein (IGHM)	IPI00472610	315	22	7	52665		-4.32	0.0081	-2.67	0.076	-1.85	0.19
766	Not identified							-4.69	0.0091	-2.17	0.14	-2.02	0.2
782	Haptoglobin (HP)	IPI00431645	98	21	7	31647	8.84	-1.24	0.095	-1.14	0.3	-1.57	0.027
782	Apolipoprotein A-IV precursor (APOA4)	IPI00304273	54	10	4	45371	5.28	-1.24	0.095	-1.14	0.3	-1.57	0.027
782	Albumin (ALB)	IPI00022434	36	4	3	73881	6.33	-1.24	0.095	-1.14	0.3	-1.57	0.027
800	Haptoglobin (HP)	IPI00431645	76	12	5	31647	8.84	-2.92	0.053	-1.29	0.47	-2.58	0.022
808	Haptoglobin (HP)	IPI00431645	363	41	16	31647	8.84	-3.37	0.012	-1.25	0.35	-2.69	0.021
814	Haptoglobin (HP)	IPI00431645	434	41	16	31647	8.84	-2.94	0.019	-1.27	-	-2.43	-
814	Alpha-1-antitrypsin precursor Isoform 1 (SERPINA1)	IPI00553177	159	21	6	46878	5.37	-2.94	0.019	-1.27	-	-2.43	-
822	Haptoglobin (HP)	IPI00431645	162	25	10	31647	8.48	-2.71	0.032	-1.38	0.24	-2.35	0.0086
826	Haptoglobin (HP)	IPI00431645	453	43	19	31647	8.48	-2.29	0.072	-1.38	0.32	-2.21	0.018
826	Albumin (ALB)	IPI00022434	93	7	5	73881	6.33	-2.29	0.072	-1.38	0.32	-2.21	0.018
826	Apolipoprotein A-IV precursor (APOA4)	IPI00304273	37	10	4	45371	5.28	-2.29	0.072	-1.38	0.32	-2.21	0.018
833	Haptoglobin (HP)	IPI00431645	555	43	21	31647	8.48	-1.55	0.014	1.21	0.57	1.08	0.72
842	Complement C3 precursor	IPI00739237	53	7	3	45642	4.94	-2.03	0.011	-1.81	0.018	-1.34	0.11
848	Haptoglobin (HP)	IPI00431645	54	8	5	31647	8.48	-2.46	0.014	1.02	0.87	-1.93	0.36
850	Haptoglobin (HP)	IPI00431645	115	21	8	31647	8.48	-2.2	0.024	-1.29	0.26	-2.31	0.036
850	Albumin (ALB)	IPI00745872	39	2	1	69366	5.92	-2.2	0.024	-1.29	0.26	-2.31	0.036
852	Not identified							-1.55	0.054	1.1	0.65	-2.82	0.025
855	Not identified							-1.75	0.028	1.12	-	-3.15	-
898	Not identified							-3.5	0.024	-4.85	-	-1.81	-
899	Not identified							-3.43	0.015	-4.62	-	-1.68	-
913	Complement C4-A precursor (C4A;C4B)	IPI00032258	18	0	2	192650	6.65	-1.55	0.015	-2.32	-	-1.61	-
928	Not identified							-3.48	0.022	-4.95	-	-1.97	-

Table 6.1 continued

Two protein features that were found to be differentially expressed between healthy and benign samples and also healthy and malignant samples were identified as immunoglobin heavy constant gamma 2 (IGHG2, spot 750 which also contains IGM and factor VII) and complement C3 precursor (spot 842). Three of the five protein features differentially expressed between healthy from malignant and benign and malignant were identified. Of these haptoglobin was identified in two spots (808 & 822) and alpha-1-antichymotrypsin precursor (SERPINA3) and kininogen-1 precursor (Alpha-2-thiol proteinase inhibitor KNG1) were both found in spot 611. The 2D gel migration of the identified differentially expressed protein features is shown in Figure 6.6. Notably, several spots yielded the same protein identification, revealing multiple isoforms of the same gene product.

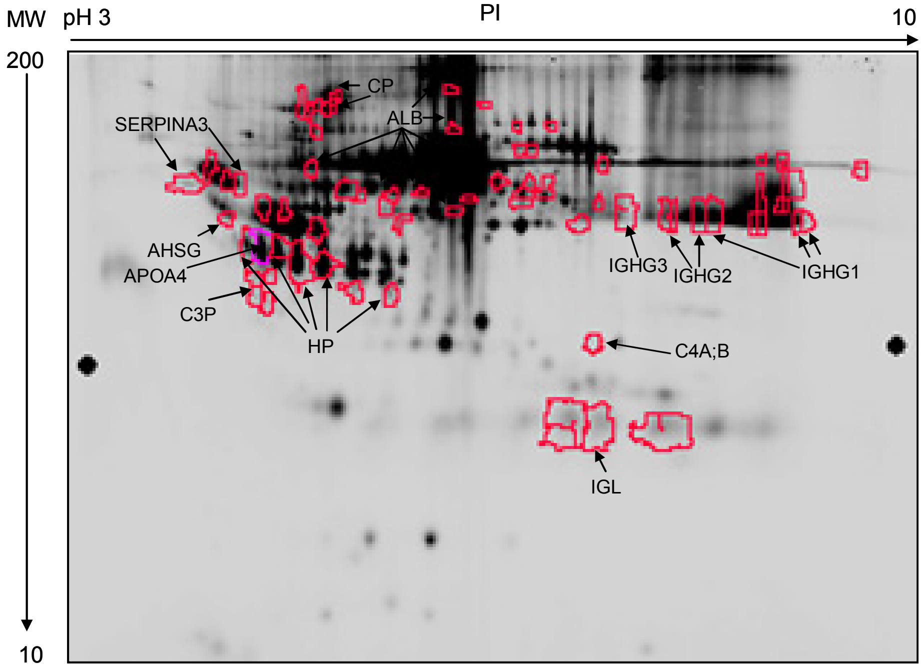


Figure 6.6 Representative gel image displaying positions of differentially expressed proteins identified from the unfractionated samples. Differentially expressed protein features were identified by MS. The locations of several differentially expressed proteins including albumin (ALB), haptoglobin (HP), apolipoprotein AIV (APOA4), isoform 1 of alpha-1-antichymotrypsin (SERPINA3), alpha-2-HS-glycoprotein precursor (AHSG), immunoglobin heavy (IGHG1) and light chain (IGL), complement factor C4 (C4A;B) are shown.

It is apparent from these results that mostly very abundant serum proteins have been identified. Although of interest, these are unlikely to be specific markers of ovarian cancer. The detection of only high abundance proteins in the unfractionated sera is likely to mask lower abundance species and potential disease markers. Indeed, multiple isoforms of the proteins were detected in different spots. Moreover, numerous spots contained more than one protein. The identification of multiple proteins in a single spot makes it very difficult to attribute any changes in spot abundance to a specific protein this is a major drawback of the 2-DE technique. To probe deeper into the serum proteome, the removal of these high-abundance proteins is essential.

6.2.2 Multiple Affinity Removal System (MARS)-based depletion of abundant proteins from pooled sera.

It is known that twenty-two high abundance proteins constitute up to 99% of the total protein content of serum [Fusaro and Stone, 2003]. It is hypothesised that putative biomarkers of interest would be found in the remaining 1%. Thus, to mine deeper into the serum proteome the pooled clinical sera were fractionated with the MARS depletion column (Agilent Technologies), which comprises of polyclonal antibodies designed to effectively remove 85-90% of the top seven most abundant proteins (albumin, IgG, transferrin, haptoglobin, IgA, antitrypsin and fibrinogen) which hinder the detection of lower abundance proteins [Chromy et al., 2004; Echan et al., 2005]. The abundant proteins bind to the column and the unbound fraction of depleted serum was collected as the ‘flow through’. The bound fraction was also recovered by acid elution for analysis.

Briefly, from each pool 30 μL of serum was diluted five times in Buffer A containing protease inhibitors (COMPLETE, Roche) and spun down at 16,000 x g at room temperature for 5 minutes. Automated multiple sample injection (2 per run) on an Agilent 1100 HPLC system was set up for 30 μL of diluted serum sample per injection in Buffer A at a flow rate of 0.25 ml/min for 9 min. Flow-through fractions

(~0.75 ml per injection) containing the unbound protein species were collected manually at 2-4 min into 1 ml Eppendorf tubes and stored at -20°C until further analysis. The bound fractions were eluted with 100% Buffer B at a flow rate of 1 ml/min for 3.5 min. The column was regenerated by equilibrating with Buffer A for 10 min as shown in Figure 6.7.

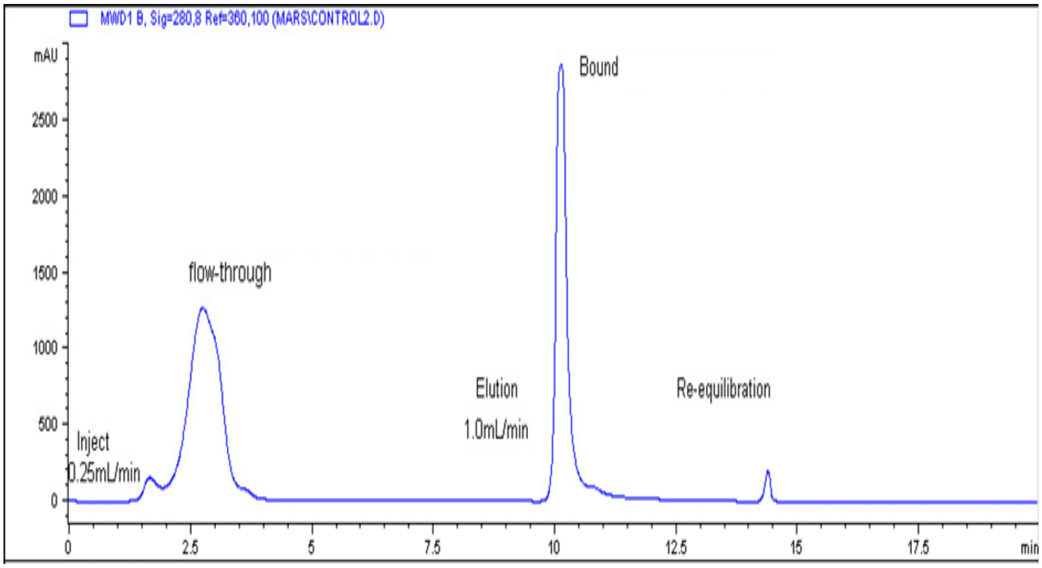


Figure 6.7 HPLC chromatograph of MARS fractionated sera. An aliquot of pooled clinical sera was fractionated on a MARS HPLC column (Agilent Technologies). 30 µL of sample was injected at a flow rate of 0.25 mL/min in Buffer A. The unbound flow through was collected over 3 min (~0.75 ml) and stored at -20°C prior to downstream 2D-DIGE analysis. The bound proteins were eluted at 1ml/min in Buffer B and collected for 1D SDS-PAGE analysis.

Once the abundant proteins had been removed from each pooled serum sample approximately 10-15% of the starting protein concentration remained in the flow-through fraction. In order to obtain sufficient amounts of the depleted protein fraction for 2D-DIGE analysis, 10 aliquots of each clinical condition were fractionated with the MARS column. Overlaid traces of 3 representative runs from each condition analysed are shown in Figure 6.8. The MARS column was found to be highly reproducible. The variability seen in the flow-through peak can be attributed to inconsistencies in the automated injection on the Agilent 1100 HPLC system.

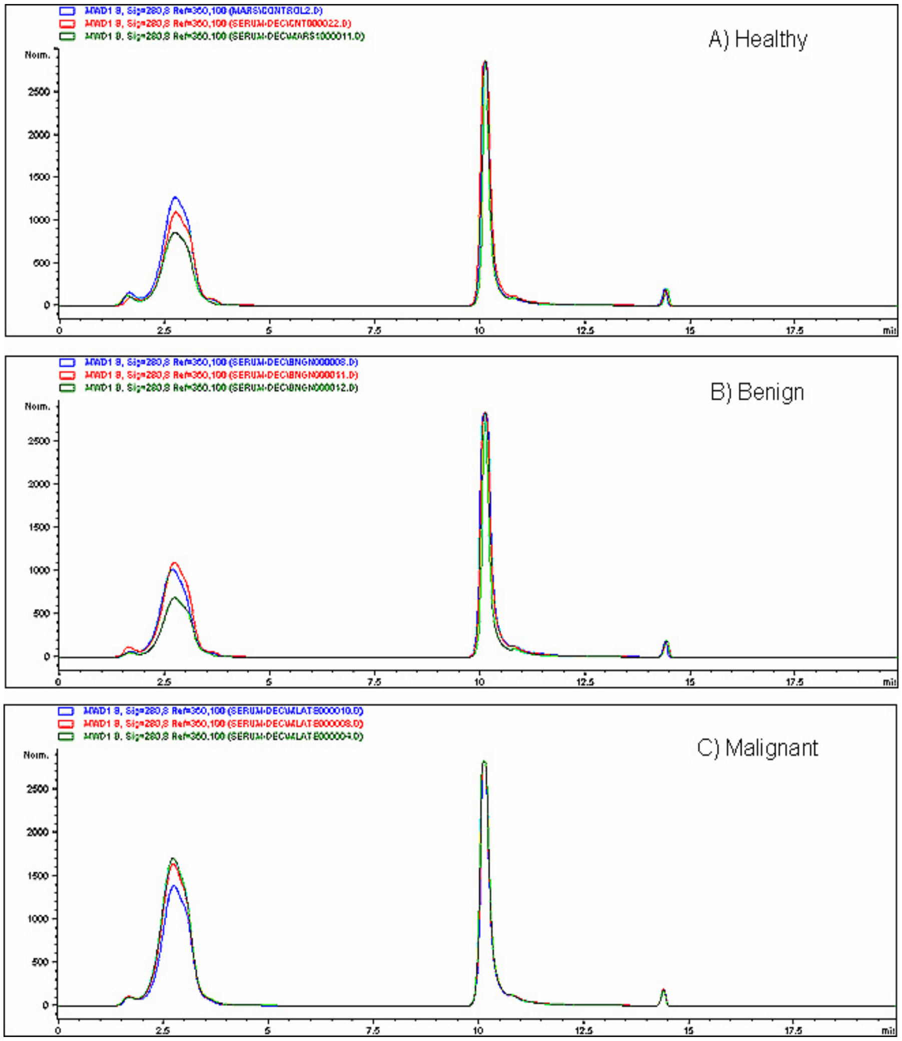


Figure 6.8 HPLC chromatograph overlays. MARS depletion of pooled clinical sera was repeated 10 times. Representative overlays of 3 runs for A) healthy, B) benign and C) malignant late stage conditions are shown.

Prior to 2D-DIGE analysis, the unbound fractions from the repeated MARS runs were pooled into their respective clinical groups. The protein concentration was determined using the Pierce BCA assay. Concentrations of 1.02 mg/mL, 1.20 mg/mL and 1.01 mg/mL were recovered for the healthy, benign and malignant conditions respectively. Fractionated samples were separated on a 1D gel to assess the depletion efficiency of the column. Briefly, equal amounts of protein (25 µg) from the unfractionated sera, unbound and bound fractions from the MARS column were separated on a 12% SDS-PAGE gel. The 1D gel showed enrichment of numerous protein bands particularly above 40 kDa and a reduction in the intensity of the 66 kDa albumin band and 80kDa transferrin band as shown in Figure 6.9.

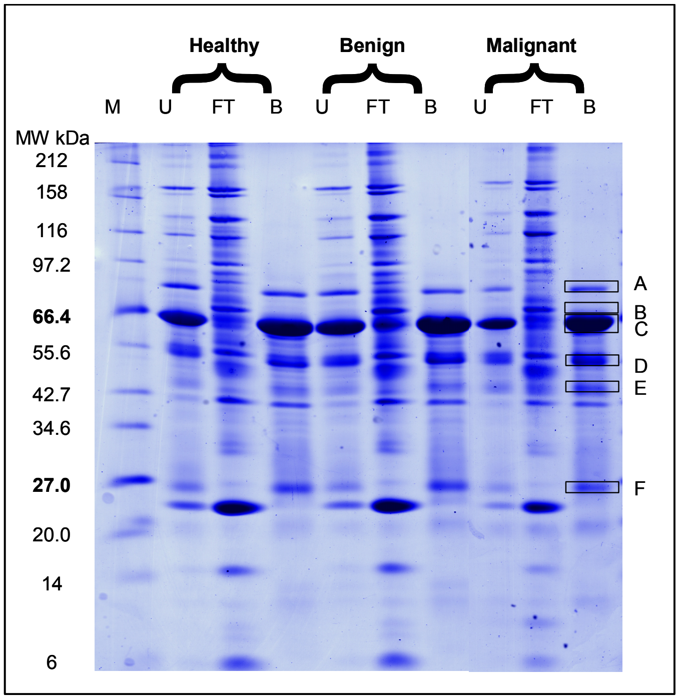


Figure 6.9 1D SDS-PAGE comparison of MARS fractionated samples. 25µg of protein from unfractionated (U), flow-through (FT) and bound (B) fractions for each clinical condition were run on a 12% SDS PAGE gel which was then stained with CCB. Lane 1 contained the MW marker (M). Gel bands A) transferrin, B) alpha-1-antitrypsin, C) albumin, D) IgG, E) and IgA and F) haptoglobin light chain.

Flow through fractions were desalted and concentrated and then subjected to 2D-DIGE analysis in quadruplicate as previous unfractionated samples. A total of 797 features were detected in the master gel (Figure 6.10). Quantitative analysis showed 4 protein spots were up-regulated in the malignant samples versus the healthy and 3 were down-regulated (≥ 1.5 average-fold change in abundance, $p < 0.05$, $n = 4$). Only 1 protein spot was differentially expressed between the healthy and benign conditions and this was up-regulated in the benign condition. Three protein spots were differentially expressed between the benign and malignant conditions. Of these, 1 was up-regulated in the malignant condition and 2 were down-regulated in the malignant condition versus the benign. In total 10 differentially expressed protein features were found between the 3 conditions (Figures 6.11 and 6.12).

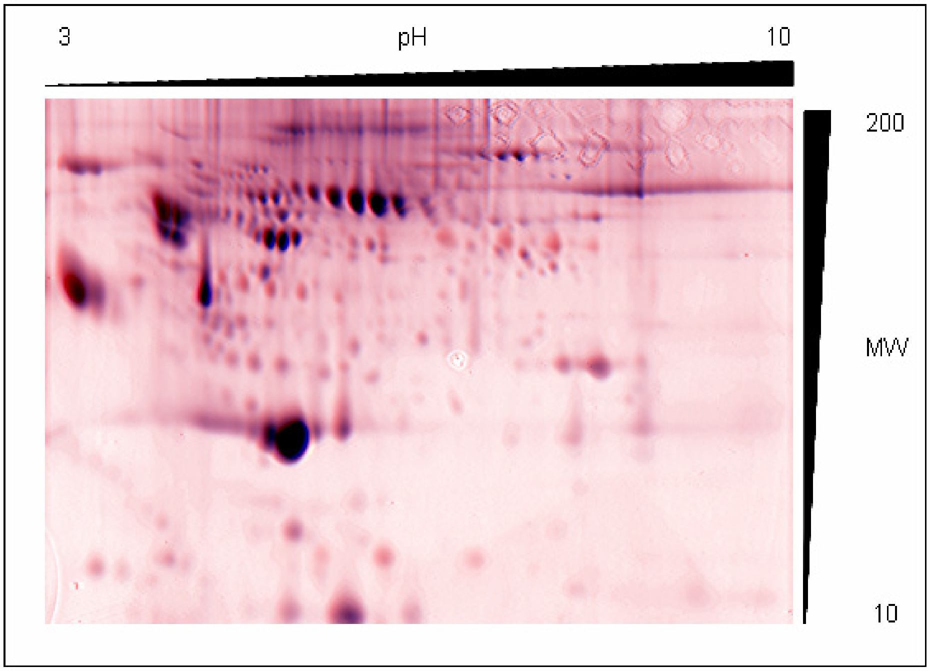


Figure 6.10 Gel image of MARS fractionated sera used as 'master' gel in DeCyder image analysis. This figure represents the overlaid fluorescent images derived from the MARS fractionated sera labelled with Cy3 Malignant (red) and Cy5 Healthy (blue). This image was prepared using Adobe Photoshop.

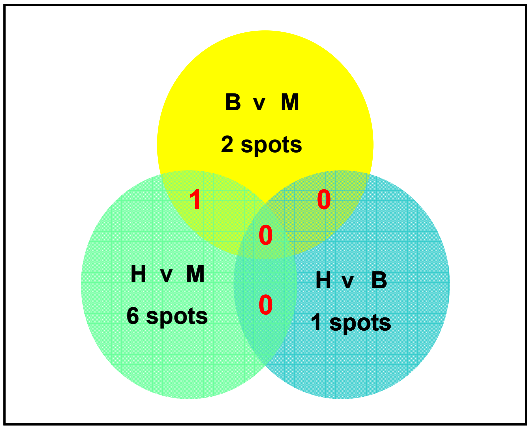


Figure 6.11 Venn diagram showing the number of differentially expressed protein features in MARS depleted sera. The number of protein features that displayed a ≥ 1.5 average fold-change in abundance (p <0.05) and the overlap between the three clinical conditions is shown.

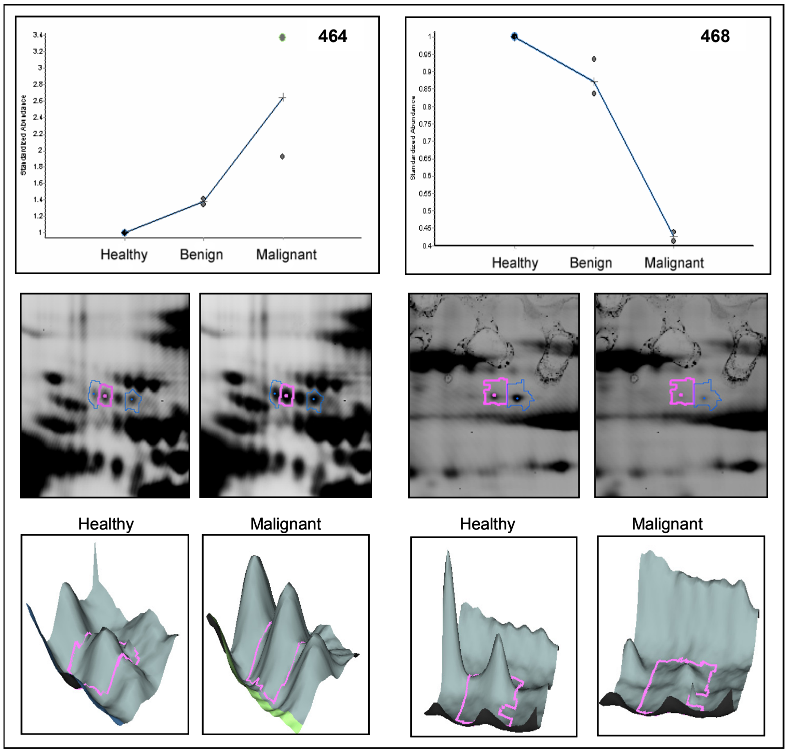


Figure 6.12 Examples of proteins displaying clinical-stage dependent changes in the MARS fractionated samples. Protein spots 464 and 468 displayed differential expression between clinical conditions. Graphs were derived from DeCyder image analysis where the standardised abundance is the ratio of the volume of a gel feature from the healthy condition *versus* the volume of the corresponding gel feature in the malignant condition. Data points are shown for quadruplicate measurements with lines joining the average values. 3D images of spots are shown for the healthy and malignant conditions.

Differentially expressed spots were excised, trypsin digested and the peptide mixtures were analysed by MALDI-TOF PMF or LC-MS/MS. The list of peptide and fragment ions masses generated was then searched against the updated IPI Human database using Mascot. All of the 10 differentially expressed protein features yielded proteins hits of high confidence, although a number yielded more than 1 protein identification (Table 6.3). For example, as shown in Figure 6.12 spot 464 yielded hits for inter-alpha-trypsin inhibitor heavy chain H4 precursor (ITIH4), histidine-rich glycoprotein precursor (HRG), alpha-1B-glycoprotein precursor (A1BG), afamin precursor (AFM), prothrombin precursor (F2 Fragment) and vitamin K-dependent protein S precursor (PROS1). All the differentially expressed protein spots were identified with high confidence (Table 6.2). Examples of MALDI-TOF PMF and LC-MS/MS based protein identifications are shown in Figure 6.13 and 6.14, respectively.

								Healthy / Malignant		Healthy / Benign		Benign / Malignant	
Spot No.	Protein Name	IPI No.	Score	Seq. Cov (%)	No. peptides	Mw	pI	Av. Ratio	T-test	Av. Ratio	T-test	Av. Ratio	T-test
242	Ceruloplasmin precursor (CP)	IPI00017601	70	10	7	122983	5.44	1.05	-	-1.42	0.0086	-1.18	-
412	Inter-alpha-trypsin inhibitor heavy chain H4 precursor (ITIH4) Isoform 1	IPI00294193	62	21	20	103489	6.51	-2.07	0.018	-	-	-	-
459*	Inter-alpha (Globulin) inhibitor H4 (Plasma Kallikrein-sensitive glycoprotein	IPI00556036	401	22	16	76971	5.72	-2.51	0.023	-1.15	0.67	-2.18	0.32
459*	Histidine-rich glycoprotein precursor (HRG)	IPI00022371	137	10	5	60510	7.09	-2.51	0.023	-1.15	0.67	-2.18	0.32
459*	Prothrombin precursor - (F2 Fragment)	IPI00019568	95	7	4	71475	5.64	-2.51	0.023	-1.15	0.67	-2.18	0.32
459*	Alpha-1B-glycoprotein precursor (A1BG)	IPI00022895	52	7	2	54809	5.58	-2.51	0.023	-1.15	0.67	-2.18	0.32
459*	Vitamin K-dependent protein S precursor (PROS1)	IPI00294004	46	4	3	77127	5.48	-2.51	0.023	-1.15	0.67	-2.18	0.32
464*	Inter-alpha-trypsin inhibitor heavy chain H4 precursor (ITIH4) Isoform 1	IPI00294193	289	11	12	103489	6.51	-2.65	0.015	-1.3	0.44	-1.91	0.16
464*	Histidine-rich glycoprotein precursor (HRG)	IPI00022371	242	19	9	60510	7.09	-2.65	0.015	-1.3	0.44	-1.91	0.16
464*	Alpha-1B-glycoprotein precursor (A1BG)	IPI00022895	189	17	6	54809	5.58	-2.65	0.015	-1.3	0.44	-1.91	0.16
464*	Afamin precursor (AFM)	IPI00019943	92	8	3	70963	5.64	-2.65	0.015	-1.3	0.44	-1.91	0.16
464*	Prothrombin precursor (F2 Fragment)	IPI00019568	86	7	4	71475	5.64	-2.65	0.015	-1.3	0.44	-1.91	0.16
464*	Vitamin K-dependent protein S precursor (PROS1)	IPI00294004	61	4	3	77127	5.48	-2.65	0.015	-1.3	0.44	-1.91	0.16
468*	Complement factor B precursor (CFB Isoform 1 of Fragment)	IPI00019591	123	5	4	86847	6.67	2.43	0.0056	1.09	0.22	2.08	0.0074
468*	Alpha-2-macroglobulin precursor (A2M)	IPI00478003	102	5	6	164600	6	2.43	0.0056	1.09	0.22	2.08	0.0074
471*	Histidine-rich glycoprotein precursor (HRG)	IPI00022371	364	21	11	60510	7.09	-1.89	0.044	1.21	0.33	-1.11	0.64
471*	Inter-alpha (Globulin) inhibitor H4 (Plasma Kallikrein-sensitive glycoprotein)	IPI00556036	261	13	10	76971	5.72	-1.89	0.044	1.21	0.33	-1.11	0.64
471*	Complement component C9 precursor	IPI00022395	196	16	9	64615	5.43	-1.89	0.044	1.21	0.33	-1.11	0.64
471*	Alpha-1B-glycoprotein precursor (A1BG)	IPI00022895	86	10	4	54809	5.58	-1.89	0.044	1.21	0.33	-1.11	0.64
472	Inter-alpha-trypsin inhibitor heavy chain H4 precursor (ITIH4) Isoform 2	IPI00218192	64	22	19	101488	6.21	3.83	0.029	-1.78	0.36	1.09	0.72
566	Alpha-1B-glycoprotein precursor (A1BG)	IPI00022895	72	37	16	54809	5.58	-1.38	0.068	-1.08	0.43	-1.45	0.032
566	Complement component C9 precursor	IPI00022395	63	24	18	64615	5.43	-1.38	0.068	-1.08	0.43	-1.45	0.032
657	Complement factor B precursor (CFB Fragment) Isoform 1	IPI00019591	116	3	7	86847	6.67	1.43	0.087	-1	0.99	-1.9	0.015
766	Alpha-2-macroglobulin precursor (A2M)	IPI00478003	62	13	19	164600	6	1.73	0.037	1	0.99	-1.13	0.48

Table 6.2 2D-DIGE based analysis of the MARS-depleted sera and MS-based protein identifications. Protein features displaying differential expression are shown. Values are average ratio of abundance between different clinical conditions (healthy/malignant, benign/malignant, and healthy/benign). T-test *p* values are given as a measure of confidence for each ratio measured. Protein name, IPI accession number, Mascot score, sequence coverage (%) and number of matched peptides for each of the identified proteins are shown. Those marked with * were identified by LC-MS/MS

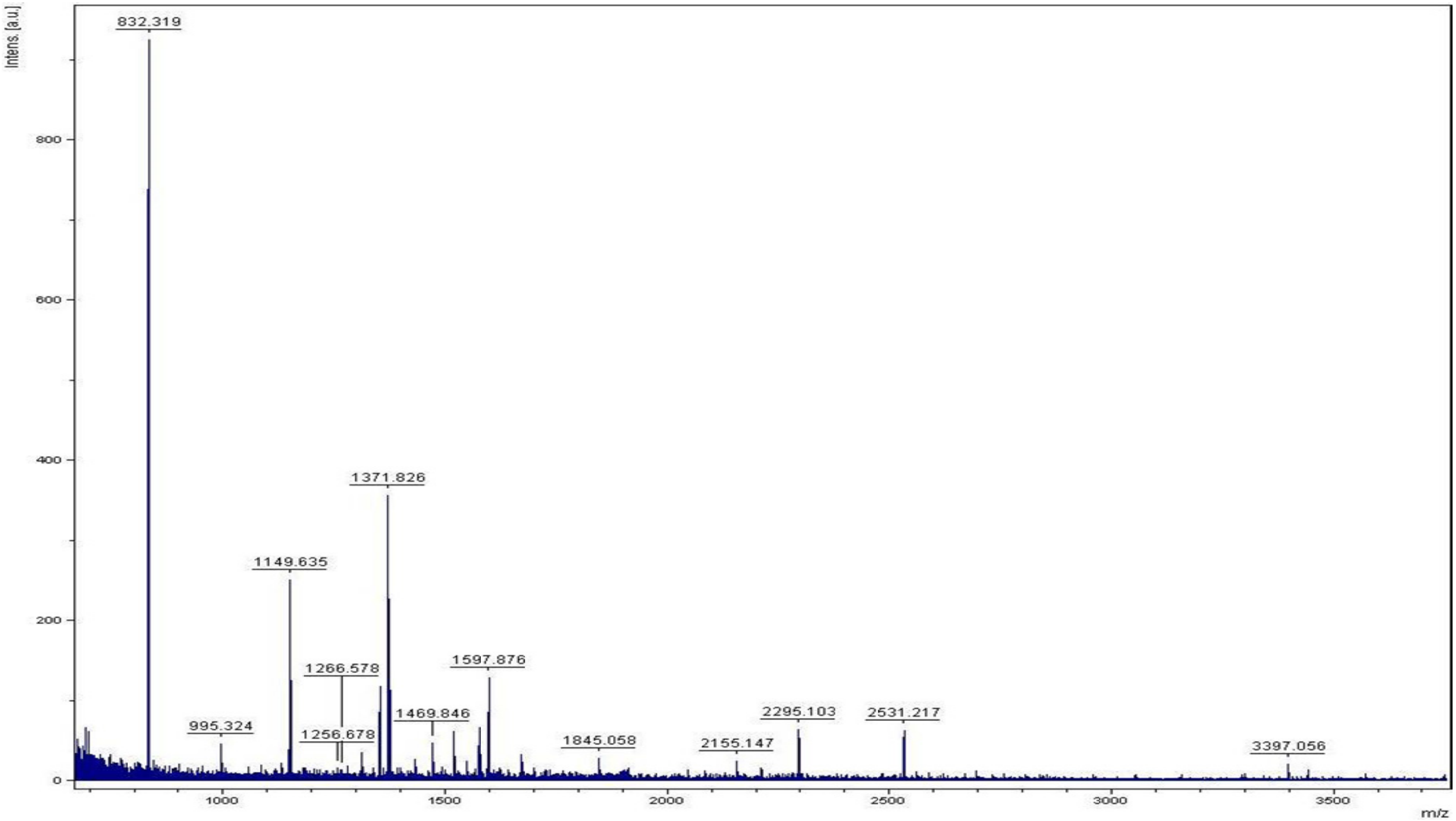
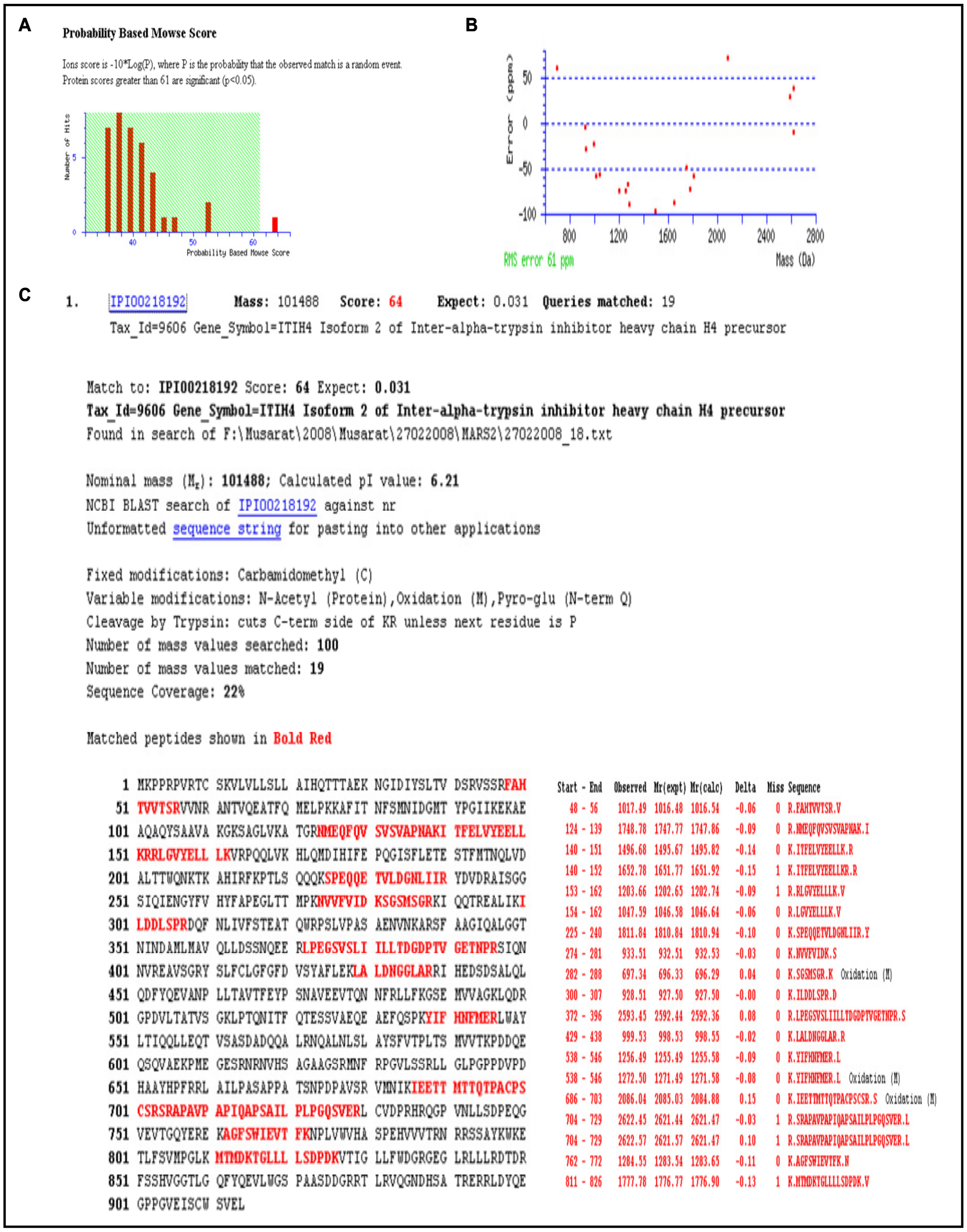


Figure 6.13 A. An example of a differentially expressed protein feature from the MARS fractionated sera identified by MALDI-TOF PMF. The tryspin digested peptides mixture from spot number 472 was analysed by MALDI-TOF PMF. The resulting spectrum was internally calibrated with trypsin autolysis peaks (842.51 & 2211.10). Prominent peaks in the mass range 700-4000 were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine.



Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 61 are significant (p<0.05).

1. IPI00218192 **Mass:** 101488 **Score:** 64 **Expect:** 0.031 **Queries matched:** 19
Tax_Id=9606 Gene_Symbol=ITIH4 Isoform 2 of Inter-alpha-trypsin inhibitor heavy chain H4 precursor

Match to: **IPI00218192** Score: **64** Expect: **0.031**
Tax_Id=9606 Gene_Symbol=ITIH4 Isoform 2 of Inter-alpha-trypsin inhibitor heavy chain H4 precursor
Found in search of F:\Musarat\2008\Musarat\27022008\MARS2\27022008_18.txt

Nominal mass (M_r): **101488**; Calculated pI value: **6.21**
NCBI BLAST search of IPI00218192 against nr
Unformatted sequence string for pasting into other applications

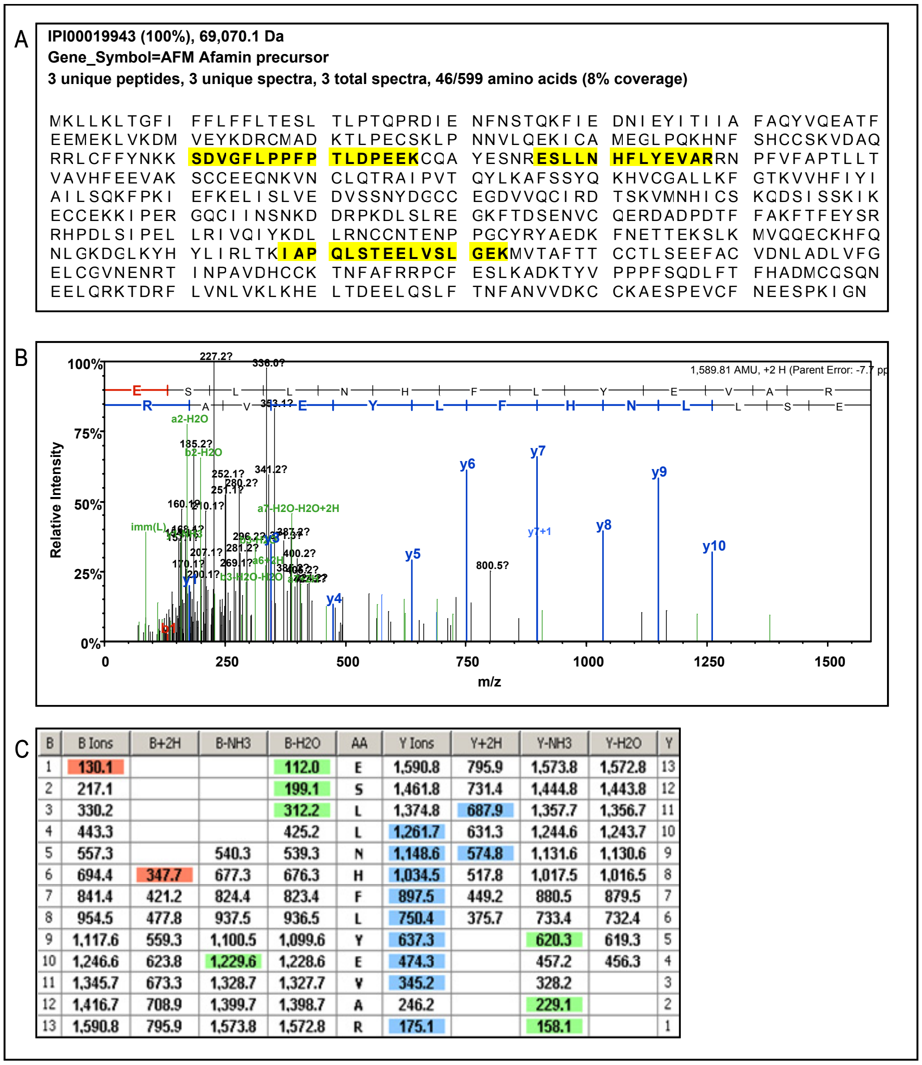
Fixed modifications: Carbamidomethyl (C)
Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: **100**
Number of mass values matched: **19**
Sequence Coverage: **22%**

Matched peptides shown in **Bold Red**

```
  1 MKPPRPVRTC SKVLVLLSLL AIHQTTTAEK NGIDIYSLTV DSRVSSRFAH
 51 TVVTSRVVNR ANTVQEATFQ MELPKKAFIT NFSMNIDGMT YPGIIKEKAE
101 AQAQYSAAVA KGKSAGLVKA TGRNMEQFQV SVSVAPNAKI TFELVYEELL
151 KRRLGVYELL LKVRPQQLVK HLQMDIHIFE PQGISFLETE STFMTNQLVD
201 ALTTWQNKTK AHIRFKPTLS QQQKSPEQQE TVLDGNLIIR YDVDRAISGG
251 SIQIENGYFV HYFAPEGLTT MPKNVVFVID KSGSMSGRKI QQTREALIKI
301 LDDLSPRDQF NLIVFSTEAT QWRPSLVPAS AENVNKARSF AAGIQALGGT
351 NINDAMLMAV QLLDSSNQEE RLPEGSVSLI ILLTDGDPTV GETNPRSIQN
401 NVREAVSGRY SLFCLGFGFD VSYAFLEKLA LDNGGLARRI HEDSDSALQL
451 QDFYQEVANP LLTAVTFEYP SNAVEEVTQN NFRLLFKGSE MVVAGKLQDR
501 GPDVLTATVS GKLPTQNITF QTESSVAEQE AEFQSPKYIF HNFMERLWAY
551 LTIQQLLEQT VSASDADQQA LRNQALNLSL AYSFVTPLTS MVVTKPDDQE
601 QSQVAEKPME GESRNRNVHS AGAAGSRMNF RPGVLSSRLL GLPGPPDVPD
651 HAAYHPFRRL AILPASAPPA TSNPDPAVSR VMNIKIEETT MTTQTPACPS
701 CSRSRAPAVP APIQAPSAIL PLPGQSVERL CVDPRHRQGP VNLLSDPEQG
751 VEVTGQYERE KAGFSWIEVT FKNPLVWVHA SPEHVVVTRN RRSSAYKWKE
801 TLFSVMPGLK MTMDKTGLLL LSDPDKVTIG LLFWDGRGEG LRLLLRDTDR
851 FSSHVGGTLG QFYQEVLWGS PAASDDGRRT LRVQGNDHSA TRERRLDYQE
901 GPPGVEISCW SVEL
```

```
Start - End   Observed  Mr(expt)  Mr(calc)   Delta  Miss Sequence
  48 - 56     1017.49   1016.48   1016.54   -0.06     0  R.FAHTVVTSR.V
 124 - 139    1748.78   1747.77   1747.86   -0.09     0  R.NMEQFQVSVSVAPNAK.I
 140 - 151    1496.68   1495.67   1495.82   -0.14     0  K.ITFELVYEELLK.R
 140 - 152    1652.78   1651.77   1651.92   -0.15     1  K.ITFELVYEELLKR.R
 153 - 162    1203.66   1202.65   1202.74   -0.09     1  R.RLGVYELLLK.V
 154 - 162    1047.59   1046.58   1046.64   -0.06     0  R.LGVYELLLK.V
 225 - 240    1811.84   1810.84   1810.94   -0.10     0  K.SPEQQETVLDGNLIIR.Y
 274 - 281     933.51    932.51    932.53   -0.03     0  K.NVVFVIDK.S
 282 - 288     697.34    696.33    696.29    0.04     0  K.SGSMSGR.K  Oxidation (M)
 300 - 307     928.51    927.50    927.50   -0.00     0  K.ILDDLSPR.D
 372 - 396    2593.45   2592.44   2592.36    0.08     0  R.LPEGSVSLIILLTDGDPTVGETNPR.S
 429 - 438     999.53    998.53    998.55   -0.02     0  K.LALDNGGLAR.R
 538 - 546    1256.49   1255.49   1255.58   -0.09     0  K.YIFHNFMER.L
 538 - 546    1272.50   1271.49   1271.58   -0.08     0  K.YIFHNFMER.L  Oxidation (M)
 686 - 703    2086.04   2085.03   2084.88    0.15     0  K.IEETTMTTQTPACPSCSR.S  Oxidation (M)
 704 - 729    2622.45   2621.44   2621.47   -0.03     1  R.SRAPAVPAPIQAPSAILPLPGQSVER.L
 704 - 729    2622.57   2621.57   2621.47    0.10     1  R.SRAPAVPAPIQAPSAILPLPGQSVER.L
 762 - 772    1284.55   1283.54   1283.65   -0.11     0  K.AGFSWIEVTFK.N
 811 - 826    1777.78   1776.77   1776.90   -0.13     1  K.MTMDKTGLLLLSDPDK.V
```

Figure 6.13B. An example of MALDI-TOF PMF protein identification. The trypsin digested peptides mixture from spot number 472 was analysed by MALDI-TOF PMF. The resulting spectrum was internally calibrated with trypsin autolysis peaks (842.51 & 2211.10). Prominent peaks in the mass range 700-4000 Da were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score, B) RMS based error scores for the peptide mass in the PMF and C) details of the identified protein's score and sequence including matched peptides are shown here (more examples are shown in Appendix 3).



A

IPI00019943 (100%), 69,070.1 Da
Gene_Symbol=AFM Afamin precursor
3 unique peptides, 3 unique spectra, 3 total spectra, 46/599 amino acids (8% coverage)

MKLLKLTGFI FFLFFLTESL TLPTQPRDIE NFNSTQKFIE DNIEYITIIA FAQYVQEATF
EEMEKLVKDM VEYKDRCMAD KTLPECSKLP NNVLQEKICA MEGLPQKHNF SHCCSKVDAQ
RRLCFFYNKK **SDVGFLPPFP** **TLDPEEK**CQA YESNR**ESLLN** **HFLYEVAR**RN PFVFAPTLLT
VAVHFEEVAK SCCEEQNKVN CLQTRAIPVT QYLKAFSSYQ KHVCGALLKF GTKVVHFIYI
AILSQKFPKI EFKELISLVE DVSSNYDGCC EGDVVQCIRD TSKVMNHICS KQDSISSKIK
ECCEKKIPER GQCIINSNKD DRPKDLSLRE GKFTDSENVC QERDADPDTF FAKFTFEYSR
RHPDLSIPEL LRIVQIYKDL LRNCCNTENP PGCYRYAEDK FNETTEKSLK MVQQECKHFQ
NLGKDGLKYH YLIRLTK**IAP** **QLSTEELVSL** **GEK**MVTAFTT CCTLSEEFAC VDNLADLVFG
ELCGVNENRT INPAVDHCCK TNFAFRRPCF ESLKADKTYV PPPFSQDLFT FHADMCQSQN
EELQRKTDRF LVNLVKLKHE LTDEELQSLF TNFANVVDKC CKAESPEVCF NEESPKIGN

C

B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	130.1			112.0	E	1,590.8	795.9	1,573.8	1,572.8	13
2	217.1			199.1	S	1,461.8	731.4	1,444.8	1,443.8	12
3	330.2			312.2	L	1,374.8	687.9	1,357.7	1,356.7	11
4	443.3			425.2	L	1,261.7	631.3	1,244.6	1,243.7	10
5	557.3		540.3	539.3	N	1,148.6	574.8	1,131.6	1,130.6	9
6	694.4	347.7	677.3	676.3	H	1,034.5	517.8	1,017.5	1,016.5	8
7	841.4	421.2	824.4	823.4	F	897.5	449.2	880.5	879.5	7
8	954.5	477.8	937.5	936.5	L	750.4	375.7	733.4	732.4	6
9	1,117.6	559.3	1,100.5	1,099.6	Y	637.3		620.3	619.3	5
10	1,246.6	623.8	1,229.6	1,228.6	E	474.3		457.2	456.3	4
11	1,345.7	673.3	1,328.7	1,327.7	V	345.2		328.2		3
12	1,416.7	708.9	1,399.7	1,398.7	A	246.2		229.1		2
13	1,590.8	795.9	1,573.8	1,572.8	R	175.1		158.1		1

Figure 6.14 An example of a differentially expressed protein feature from the MARS depleted sera identified by LC-MS/MS. The trypsin digested peptide mixture of spot number 464 from the MARS-depleted pooled sera 2D-DIGE experiment was analysed by LC-MS/MS. A) Afamin was identified by three unique peptides, (highlighted in yellow in the protein sequence). B) The MS/MS spectrum of peptide ESLLNHFLYEVAR with the y-ion series is shown in blue, b-ions in red and the immonium ions in green and C) table indicating the masses of identified fragment ions (data produced in Scaffold software).

6.2.3 Comparison of unfractionated sera and MARS-depleted sera

The numbers of expressed features from the unfractionated clinical sera were compared with the MARS-fractionated sera. Results demonstrated that fewer (~100) protein features were detected in the MARS fractionated gels, suggesting that the depleted proteins exist as multiple isoforms or there is a loss of bound protein species. The removal of the high abundant proteins would be expected to result in an enrichment of lower abundance proteins by facilitating a higher load of these less abundant protein species for 2D-DIGE based analysis (Figure 6.15). However, results show that the number of lower abundant proteins identified was modest.

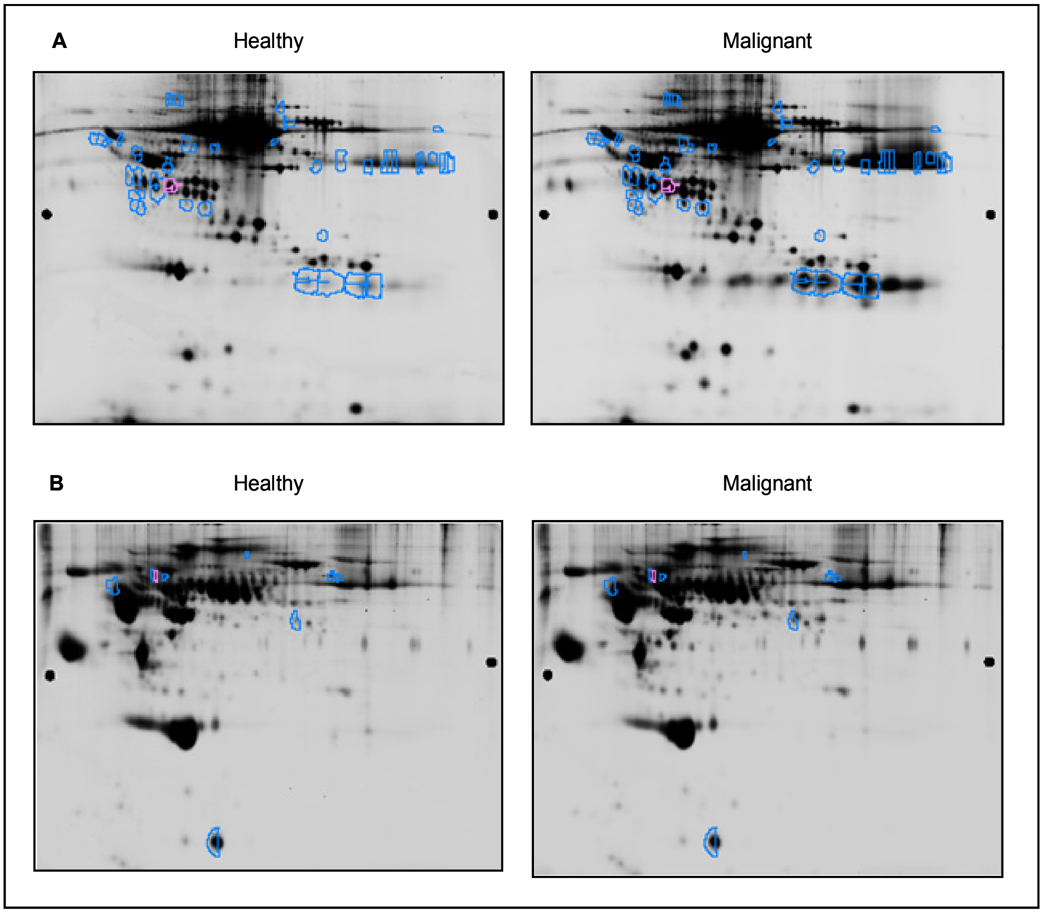


Figure 6.15 Differentially expressed protein features from MARS-depleted and unfractionated pooled clinical sera. Quantitative analysis found 48 differentially expressed protein features in A) the unfractionated samples and 10 in B) the MARS depleted fractions. Cy-dye labelled gel images from the healthy and malignant conditions are shown, created using the DIA module of DeCyder.

6.2.4 ProteoMiner based dynamic range compression and protein enrichment.

The ProteoMiner protein enrichment kit is designed to "compress" the serum protein dynamic range and is comprised of a highly diverse combinatorial peptide library immobilised on beads in a spin column format. Theoretically the library contains binding sites for most, if not all, proteins in a sample. The high abundance proteins saturate their affinity ligands and the excess protein is washed away. The medium and lower abundance proteins are concentrated on their specific affinity ligands. Thus, the dynamic range of protein concentrations is reduced, while representatives of all proteins within the original sample remain intact. Briefly, 1 mL of each pooled serum sample incubated with ProteoMiner beads and unbound material collected. The columns were then and the bound proteins were eluted with 100 µL of 2D lysis buffer (8 M urea, 2 M thiourea, 4% CHAPS, 0.5% NP40 and 10 mM Tris pH 8.3). The elution step was repeated twice to ensure all bound material was collected and the fractions pooled. Protein yields were 632.4 µg, 687.4 µg and 786.75µg for the healthy, benign and malignant pools respectively.

Prior to 2D-DIGE analysis the ProteoMiner fractionated samples were separated on a 1D SDS-PAGE gel to assess the enrichment efficiency. Briefly, equal amounts of protein (25 µg) from the unfractionated pooled sera, bound and unbound fractions from the ProteoMiner column were separated on a 12 % SDS-PAGE gel. The 1D gel showed enrichment of numerous protein bands particularly below 50 kDa and a reduction in the intensity of the 66 kDa albumin band (Figure 6.16).

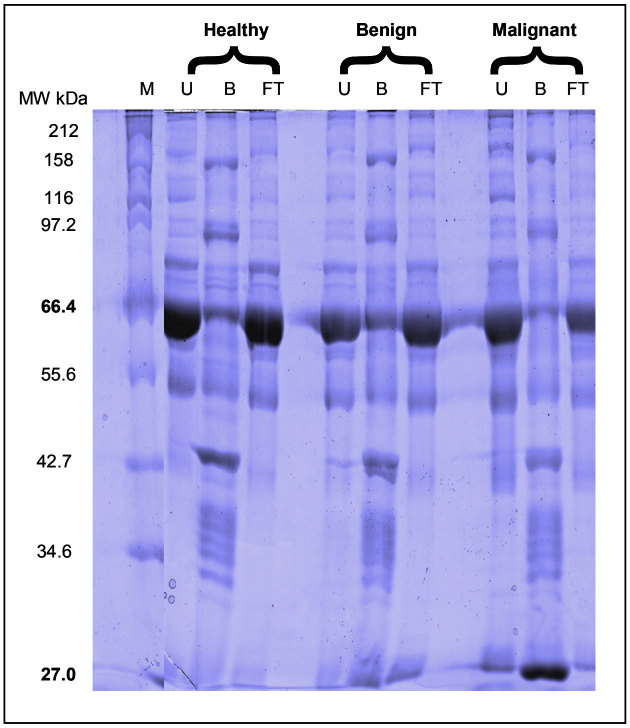


Figure 6.16 1D SDS-PAGE comparison of ProteoMiner fractionated samples. 25μg of protein from unfractionated (U), flowthrough (FT) and bound (B) fractions for each clinical condition were run on a 12% SDS PAGE gel which was then stained with CCB. Lane 1 contained the molecular weight marker (M).

2D-DIGE analysis of the ProteoMiner fractionated sera was performed in parallel with a repeat analysis of freshly prepared MARS-depleted fractions for comparison of the two strategies. To facilitate spot matching pools of fractionated sera were additionally labelled with Cy2 as an internal standard which was run on all gels as shown in Table 6.3.

A) ProteoMiner

Gel No.	Gel 01	Gel 02	Gel 03	Gel 04	Gel 05	Gel 06
Cy dye						
Cy 3	Malignant	Benign	Healthy	Malignant	Healthy	-
Cy 5	Healthy	Malignant	Benign	Benign	Malignant	-
Cy 2	Pool	Pool	Pool	Pool	Pool	-

B) MARS

Gel No.	Gel 01	Gel 02	Gel 03	Gel 04	Gel 05	Gel 06
Cy dye						
Cy 3	Healthy	Healthy	Malignant	Benign	Malignant	Benign
Cy 5	Benign	Malignant	Benign	Healthy	Healthy	Malignant
Cy 2	Pool	Pool	Pool	Pool	Pool	Pool

Table 6.3 Experimental design for 2D-DIGE analysis of fractionated pooled clinical sera. A) 2D-DIGE based comparison of ProteoMiner fractionation of pooled clinical sera. The table shows the sequence of Cy3 and Cy5 labelled sample triplicates for the healthy and benign conditions and quadruplicates for the malignant condition run on each gel (120µg protein per dye), including the Cy2 labelled pool. B) 2D-DIGE based comparison of MARS fractionated pooled sera. The table shows the labelling of quadruplicate samples from each clinical condition with Cy3 or Cy5 (80µg protein per dye), including the Cy2 labelled pool run on each gel.

Image analysis was carried out using DeCyder software. Briefly, fluorescent images from the same 2D gel were automatically curated, normalised, matched and spot abundances calculated in the DIA module. Then, matching and comparison of protein features across different gels was performed in the BVA module using internal landmarks comprising abundant protein features present in all the Cy2 images. Examples of Cy2, Cy3 and Cy5 images from the ProteoMiner set are shown in Figure 6.17.

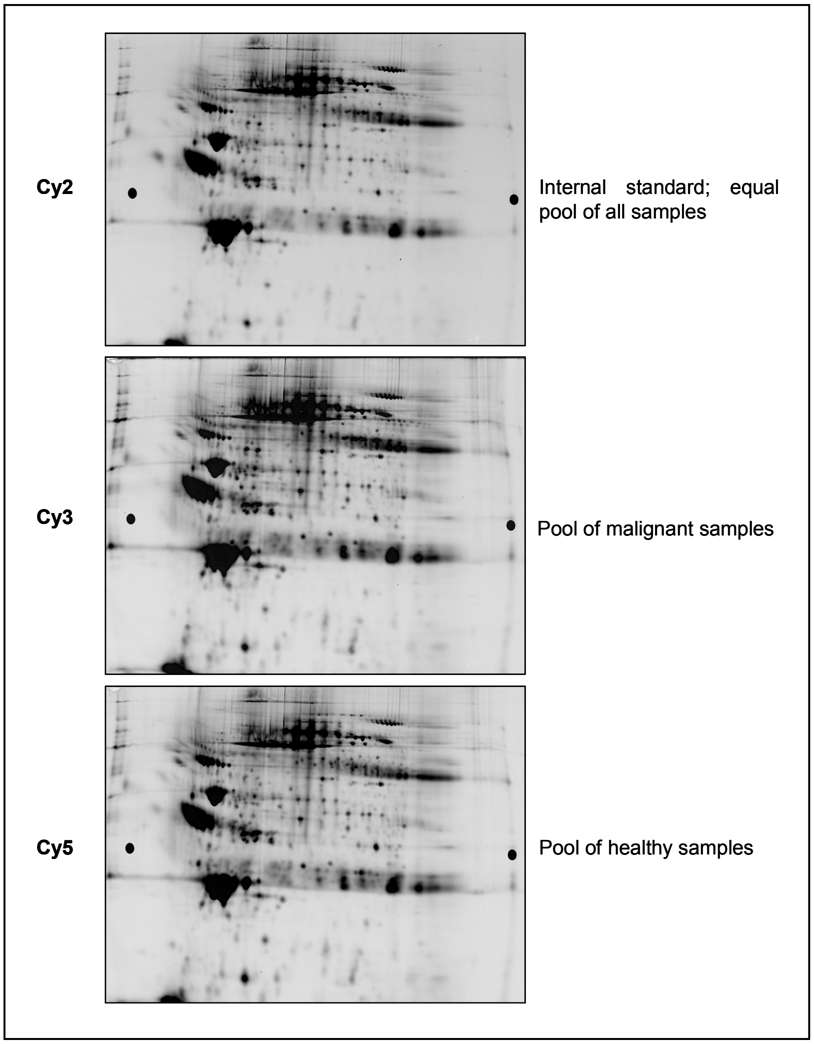

Cy2
Internal standard; equal pool of all samples
Cy3
Pool of malignant samples
Cy5
Pool of healthy samples

Figure 6.17 Representative Cy2, Cy3 and Cy5 fluorescence gel images obtained from a single 2D gel in the ProteoMiner 2D-DIGE experiment. Equal amounts of Cy3 and Cy5 labelled protein samples, derived from pooled clinical sera were mixed with an equal amount of Cy2-internal standard pool. Proteins were separated by 2D gel electrophoresis, and the gel was scanned at the appropriate excitation/emission wavelengths to generate the superimposable set of three images shown.

Statistical analysis was performed where the average abundances of protein features for each condition were compared. In the BVA analysis, 697 and 1468 protein features were found in the master gel of the ProteoMiner and MARS-fractionated samples respectively. However, the high number was due to streaking of gel spots leading to poorer spot definition. Protein features displaying a ≥ 1.5 average fold-change in abundance, displaying reproducible changes ($p < 0.01$) and matching on all images, were selected for MS-based identification; These features are marked on the gel images in Figure 6.18 (the selection criteria were made more stringent in this experiment to reduce the number of false positives).

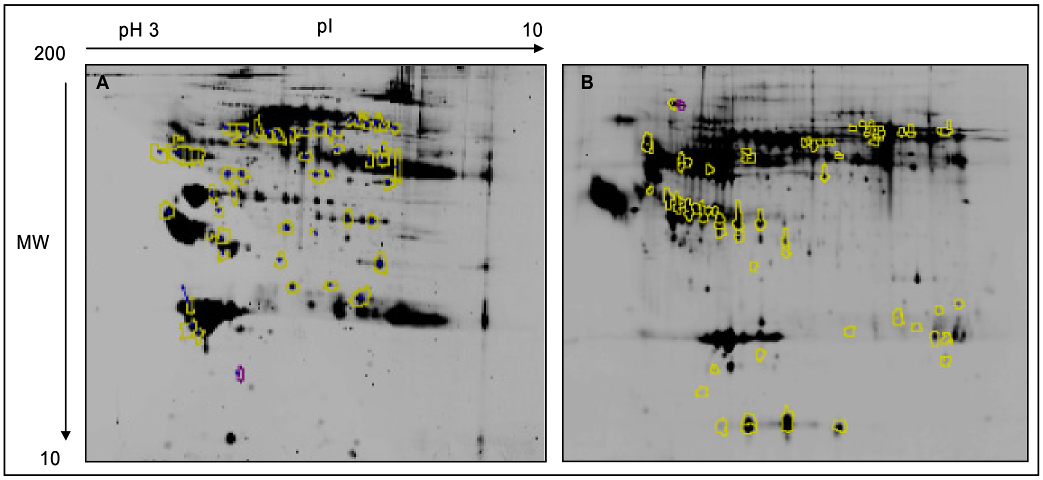


Figure 6.18 Representative gels displaying the position of differentially expressed proteins. Protein features displaying ≥ 1.5 average fold-change in abundance ($p < 0.01$) are shown in yellow in A) ProteoMiner-fractionated and B) MARS-fractionated samples.

Individual analysis and statistical evaluation of differences between the ProteoMiner-fractionated samples and the MARS-depleted samples revealed a total of 65 and 76 differentially expressed protein features, respectively (Figure 6.19). There was considerable overlap in spot numbers in each pairwise comparison of clinical conditions, particularly for the MARS-fractionated samples. The inclusion of the Cy2 labelled internal standard (pool of all samples) made spot matching across gels much easier and enabled the identification of a high number of differentially expressed proteins features which may have otherwise been missed. Gels were post-electrophoretically stained with CCB and gel images matched to the fluorescent images. A pick list of proteins of interest was generated, spots were excised robotically, trypsin digested and subjected to MS-based protein identification (examples are shown in Appendix 4 and 5 for ProteoMiner and MARS 2, respectively).

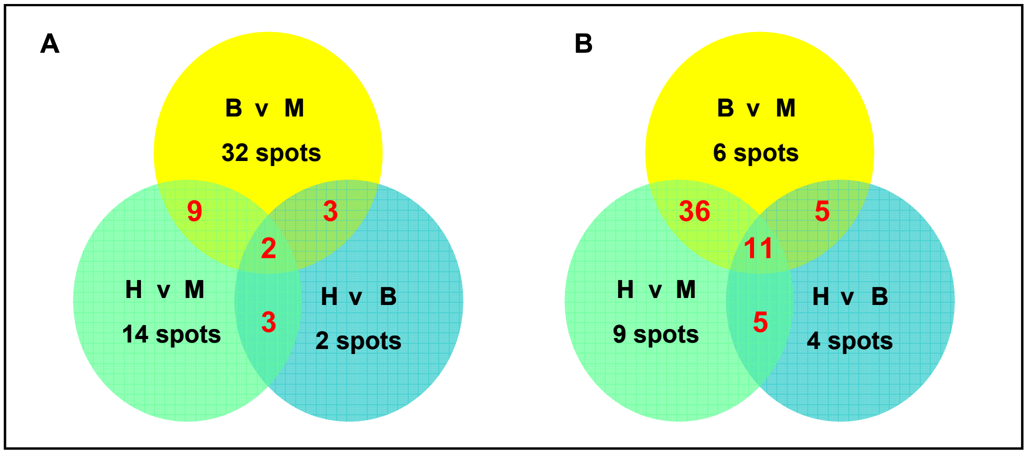


Figure 6.19 Venn diagrams showing the number of differentially expressed protein spots in the fractionated samples and overlapping spots between the three clinical conditions. The average ratio in abundance of each matched protein feature was calculated between clinical conditions. Those displaying a ≥ 1.5 average fold-changes in abundance ($p < 0.01$ n=3) were selected for identification. Numbers of spots are shown for A) ProteoMiner and B) MARS-fractionated samples with the number of overlapping spots are shown in red.

6.2.5 Protein identification by mass spectrometry

Differentially expressed protein spots from the 2D-DIGE experiments were initially subjected to identification by MALDI-TOF MS peptide mass fingerprinting and MASCOT database searching (see Chapter 2). In cases where peptide mass fingerprints could not be confidently matched to available protein sequences, LC-MS/MS was used to obtain peptide sequence information for identification. Of the 65 differentially expressed protein features from the ProteoMiner fractionation, confident protein hits were obtained for 53 of these (Table 6.4). Almost all the protein features yielded single protein identifications, except for 5 spots where 2 protein identifications had significant scores. Of the unidentified protein features, some were of low abundance and gave poor spectra, whilst the others stained well with CCB and gave spectra of good quality, but could not be identified from database searches.

								Healthy/Malignant		Healthy/Benign		Benign/Malignant	
Spot No.	Protein Name	IPI No.	Score	Seq. Cov (%)	No. peptides	Mw	pI	Av. Ratio	T-test	Av. Ratio	T-test	Av. Ratio	T-test
149	Not Identified							1.24	0.051	-1.28	0.018	1.59	0.0019
153	Serotransferrin precursor (TF)	IPI00022463	46	4	2	79280	6.81	1.42	0.0017	-1.16	0.17	1.65	0.0012
157	Not Identified							1.31	0.00016	-1.25	0.014	1.63	0.00018
160*	Serotransferrin (TF)	IPI00022463	59	22	16	79280	6.81	1.53	0.00017	-1.13	0.25	1.73	0.00047
164	Not Identified							1.48	0.00021	-1.28	0.022	1.9	2.50E-05
165	Histidine-rich glycoprotein precursor (HRG)	IPI00022371	59	30	16	60510	7.09	1.31	0.007	-1.22	0.085	1.6	0.0013
174*	Serotransferrin (TF)	IPI00022463	59	20	17	79280	6.81	1.59	6.50E-06	-1.21	0.045	1.92	5.40E-05
192	Not Identified							-1.01	0.92	1.96	0.0014	-1.97	0.0083
208*	Pyruvate kinase L (PKLR)	IPI00743713	63	23	15	64975	7.6	-1.09	0.5	1.72	0.015	-1.88	0.0038
215	Serum albumin precursor Isoform 1 (ALB)	IPI00745872	130	10	5	71317	5.92	1.21	0.02	-1.3	0.013	1.57	0.00023
217	Not Identified							1.26	0.014	-1.24	0.048	1.57	0.0005
221	Serum albumin precursor Isoform 1 (ALB)	IPI00745872	716	31	22	71317	5.92	1.29	0.003	-1.22	0.054	1.56	0.00088
222*	Serum albumin precursor Isoform 1 (ALB)	IPI00745872	94	34	23	71317	5.92	1.33	0.0037	-1.17	0.21	1.56	0.0061
224*	Serum albumin precursor Isoform 1 (ALB)	IPI00745872	32	167	23	71317	5.92	1.31	0.0036	-1.17	0.13	1.54	0.0021
226*	Serum albumin precursor Isoform 1 (ALB)	IPI00745872	152	42	30	71317	5.92	1.25	0.00025	-1.26	0.0054	1.58	9.40E-05
228*	Serum albumin precursor Isoform 1 (ALB)	IPI00745872	145	44	28	71317	5.92	1.25	0.012	-1.41	0.00034	1.75	0.00017
231*	Serum albumin precursor Isoform 1 (ALB)	IPI00745872	145	46	28	71317	5.92	1.32	0.011	-1.22	0.052	1.61	4.60E-05
221	Serum albumin precursor Isoform 1 (ALB)	IPI00745872	813	36	23	71317	5.92	1.3	0.0043	-1.34	0.0072	1.74	0.00017
241	Serum albumin (ALB)	IPI00022434	722	35	20	73881	6.33	1.31	0.0055	-1.28	0.012	1.68	0.0002
243*	Kinesin-like protein (KIFI5)	IPI00024975	68	17	31	161030	5.75	1.29	0.06	-1.37	0.049	1.78	0.0048
243*	Amyotrophic lateral sclerosis 2 chromosomal region Isoform 2 (ALS2CR12)	IPI00044665	63	26	13	29638	11.67	1.29	0.06	-1.37	0.049	1.78	0.0048
246*	Serum albumin precursor Isoform 1 (ALB)	IPI00745872	100	38	23	71317	5.92	1.19	0.12	-1.37	0.0037	1.62	0.0011
247*	Serum albumin precursor Isoform 1 (ALB)	IPI00745872	124	39	26	71317	5.92	1.27	0.066	-1.52	0.01	1.93	0.0007
248	Serum albumin (ALB)	IPI00022434	1051	45	33	73881	6.33	1.25	0.12	-1.54	0.014	1.91	0.0017
249	Serum albumin (ALB)	IPI00022434	958	46	29	73881	6.33	1.25	0.0033	-1.37	0.0015	1.71	1.20E-05
257	Serum albumin (ALB)	IPI00022434	143	33	21	73881	6.33	1.28	0.033	-1.43	0.028	1.84	0.0021
269	Alpha-1-antitrypsin precursor Isoform 1 (SERPINA)	IPI00553177	96	19	6	46878	5.37	-2.82	0.0029	-1.28	0.44	-2.2	0.016
285	Not Identified							-2.19	0.00028	-1.47	0.0072	-1.49	0.0012
289	Immunoglobulin heavy constant gamma 1 (IGHG1)	IPI00384938	124	11	4	53503	8.74	-2.06	2.90E-05	1.28	0.0005	-2.64	9.30E-06
291	Alpha-1-antitrypsin precursor Isoform 1 (SERPINA)	IPI00553177	297	30	11	46878	5.37	-1.97	0.00038	-1.54	0.00016	-1.28	0.026
298*	Antithrombin III variant (SERPINC)	IPI00055812	79	37	19	53114	6.11	-1.67	0.0007	-1.2	0.027	-1.39	0.004
299*	Antithrombin III variant (SERPINC)	IPI00055812	59	33	16	53114	6.11	-1.84	0.00031	-1.55	0.00025	-1.19	0.054
300	Alpha-1-antitrypsin precursor Isoform 1 (SERPINA)	IPI00553177	193	20	8	46878	5.37	-1.99	0.00027	-1.52	0.0062	-1.31	0.0016
303	Vitamin D-binding protein precusor (GC)	IPI00055812	124	9	6	54526	5.4	-1.89	0.00058	-1.47	0.0033	-1.29	0.02

Table 6.4 2D-DIGE based analysis of the ProteoMiner fractionated pooled clinical sera and MS-base protein identification. **P**rotein features displaying differential expression are shown. Values are the average ratio of abundance between different clinical conditions (healthy/malignant, healthy/benign and benign/malignant). T-test *p* values are given as a measure of confidence for each ratio measured. Protein name, IPI accession number, database search score, sequence coverage (%) and the number of matched peptides for each of the identified proteins are shown. Those marked with an * were identified by MALDI-TOF PMF.

								Healthy/Malignant		Healthy/Benign		Benign/Malignant	
Spot No.	Protein Name	IPI No.	Score	Seq. Cov (%)	No. peptides	Mw	pI	Av. Ratio	T-test	Av. Ratio	T-test	Av. Ratio	T-test
306	Antithrombin III variant (SERPINC)	IPI00055812	157	11	7	53114	6.11	-1.62	0.0069	-1.03	0.65	-1.57	0.021
319	Not Identified							2.35	0.0014	1.46	0.0022	1.61	0.012
326	Immunoglobulin heavy constant gamma 1 (IGHG1)	IPI00384938	110	6	3	53503	8.74	1.71	0.00039	1.33	0.017	1.29	0.0034
328	Serum albumin	IPI00022434	845	35	20	73881	6.33	2.65	0.0002	1.34	0.0085	1.98	0.0008
329	Immunoglobulin heavy constant gamma 1 (IGHG1)	IPI00384938	845	35	20	53503	8.74	1.63	0.0011	1.23	0.019	1.32	0.0086
339	Not Identified							-1.35	0.02	1.16	0.13	-1.57	0.0019
360	Not Identified							2.93	9.00E-07	1.59	0.00044	1.84	2.70E-05
371	Fibrinogen Isoform Gamma A/B (FGG)	IPI00021891	206	22	8	52106	5.37	6.69	1.90E-05	1.29	0.0027	5.17	4.40E-05
381	Complement factor H precursor Isoform 2 (CFH)	IPI00218999	430	36	16	52106	6.77	1.32	0.0002	-1.28	0.00092	1.68	1.80E-05
382	Apolipoprotein AIV precursor (APOA4)	IPI00304273	877	50	21	45371	5.28	1.43	0.00048	-1.28	0.0015	1.83	1.30E-05
393*	Apolipoprotein AIV precursor (APOA4)	IPI00304273	73	39	15	45344	5.28	-1.26	0.012	1.32	0.012	-1.66	0.0003
405*	Apolipoprotein AIV precursor (APOA4)	IPI00304273	123	45	21	45344	5.28	1.58	0.0056	1.7	0.00031	-1.08	0.54
405*	Kinectin 1 isoform b (KTN1)	IPI00783726	65	17	32	150545	5.59	1.58	0.0056	1.7	0.00031	-1.08	0.54
406*	Complement factor H-related 1 (CFHR)	IPI00167093	73	33	14	38766	7.38	1.25	0.0051	-1.29	0.0011	1.61	7.40E-05
406*	Periplakin (PPL)	IPI00298057	67	18	33	205096	5.44	1.25	0.0051	-1.29	0.0011	1.61	7.40E-05
431	Complement factor H Isoform 1precursor	IPI00029739	159	5	8	143680	6.21	-1.62	0.0081	-1.23	0.044	-1.32	0.09
431	Complement factor H Isoform 2precursor	IPI00218999	159	14	8	52711	6.77	-1.62	0.0081	-1.23	0.044	-1.32	0.09
435*	Apolipoprotein AIV precursor (APOA4)	IPI00304273	158	57	26	45344	5.28	1.52	0.0017	1.4	0.00037	1.09	0.24
435*	Pericentrin (PCNT)	IPI00479143	68	10	49	380644	5.39	1.52	0.0017	1.4	0.00037	1.09	0.24
444	GTPase-activating Rap/Ran-GAP domain-like 1 Isoform 3	IPI00646904	68	23	15	125169	5.82	-1.24	0.01	1.59	0.00066	-1.97	9.50E-05
445	Not Identified							-1.2	0.019	1.46	0.012	-1.76	0.00042
458	Serum albumin Isoform 2 of precursor	IPI00384697	537	25	14	48641	5.97	1.29	0.00095	-1.35	0.00086	1.75	1.10E-05
461*	Apolipoprotein E precursor (APOE)	IPI00021842	68	41	16	36246	5.65	1.26	0.0015	-1.21	0.0035	1.53	8.90E-05
461*	Adenomatosis polyposis coli 2 (APC2)	IPI00025190	64	11	31	245966	9.08	1.26	0.0015	-1.21	0.0035	1.53	8.90E-05
475	Apolipoprotein E precursor (APOE)	IPI00021842	373	32	12	36246	5.65	-1.53	0.00047	-1.13	0.063	-1.35	0.011
478*	Apolipoprotein E precursor (APOE)	IPI00021842	91	52	22	36841	5.65	-1.77	0.0059	-1.33	0.065	-1.33	0.047
496	Apolipoprotein E precursor (APOE)	IPI00021842	685	48	15	36246	5.65	-1.05	0.37	1.49	7.80E-05	-1.57	0.00027
499	Apolipoprotein E precursor (APOE)	IPI00021842	773	46	14	36246	5.65	-1.2	0.021	1.26	0.017	-1.51	0.0012
500	Not Identified							1.23	0.014	-1.35	0.00065	1.66	0.00029
501	Complement C4-A;B precursor (C4A;B)	IPI00032258	168	5	9	194247	6.65	-2.23	5.50E-05	-1.16	0.048	-1.92	0.00016
506	Not Identified							1.24	0.024	-1.39	0.0089	1.72	0.0012
507	Complement factor H-related protein 2 Isoform long of precursor (CFHR)	IPI00006154	206	15	4	31543	6	-5.38	0.0019	-1.04	0.75	-5.17	0.00053
507	Complement factor H-related protein 2 Isoform short of precursor (CFHR)	IPI00006154	206	17	4	28734	6.52	-5.38	0.0019	-1.04	0.75	-5.17	0.00053
511	Serotransferrin precursor (TF)	IPI00022463	325	26	16	79280	6.81	-1.26	0.002	1.67	0.0006	-2.11	1.90E-05
517	Immunoglobulin Lambda protein (IGL)	IPI00658130	106	13	4	25347	8.14	-1.96	7.40E-05	1.01	0.91	-1.98	2.70E-05
562*	Apolipoprotein A-I precursor (APOA1)	IPI00021841	147	49	15	30759	5.56	1.53	0.0072	1.31	0.094	1.17	0.23
571*	Apolipoprotein A-I precursor (APOA1)	IPI00021841	236	85	32	30759	5.56	1.9	0.00068	1.36	0.05	1.4	0.029
577*	Utrophin fragment (UTRN)	IPI00009329	70	11	49	396472	5.21	-1.39	0.00045	-1.54	8.70E-05	1.11	0.034

Table 6.4 continued

Furthermore, several spots yielded the same protein identification suggesting multiple isoforms of the same gene product. The 2D gel migration of the identified differentially expressed protein features is shown in Figure 6.20. Abundant serum proteins including apolipoprotein AI, apolipoprotein AIV, apolipoprotein E, alpha-1-antitrypsin precursor, antithrombin III variant, serotransferrin, serum albumin precursor, immunoglobin heavy and light chain, fibrinogen and several complement factors were identified as differentially expressed protein features in multiple locations. In addition, cellular proteins such as Pyruvate kinase L (PKLR), Vitamin D-binding protein and a fragment of uthrophin (UTRN) were also identified. Examples of several spots including yielding hits for apolipoprotein A4 (APOA4) are shown in Figure 6.21.

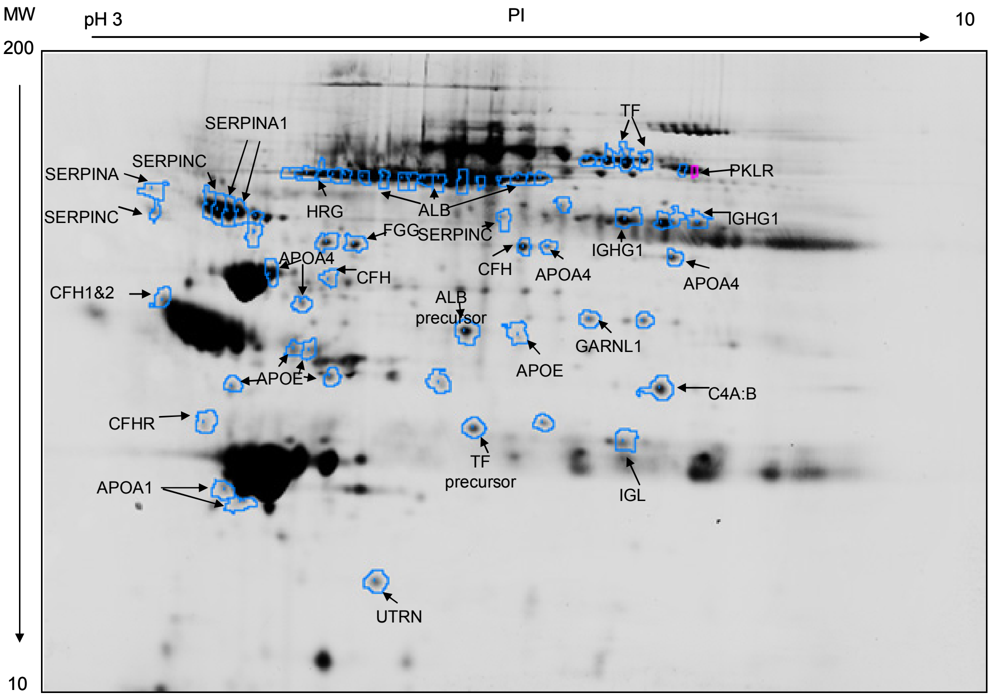


Figure 6.20 Representative gel image displaying positions of differentially expressed proteins from the ProteoMiner-fractionated samples. Differentially expressed protein features were identified by MS. The location of differentially expressed proteins including apolipoprotein AI (APOA1), apolipoprotein AIV (APOA4), apolipoprotein E (APOE), alpha-1-antitrypsin precursor (SERPINA), antithrombin III variant (SERPINC), serotransferrin (TF), serum albumin precursor (ALB), the immunoglobin heavy (IGHG1) and light chain (IGL), fibrinogen (FGG), several complement factors (CFHR, C4A;B), pyruvate kinase L (PKLR), and utrophin (UTRN) are shown.

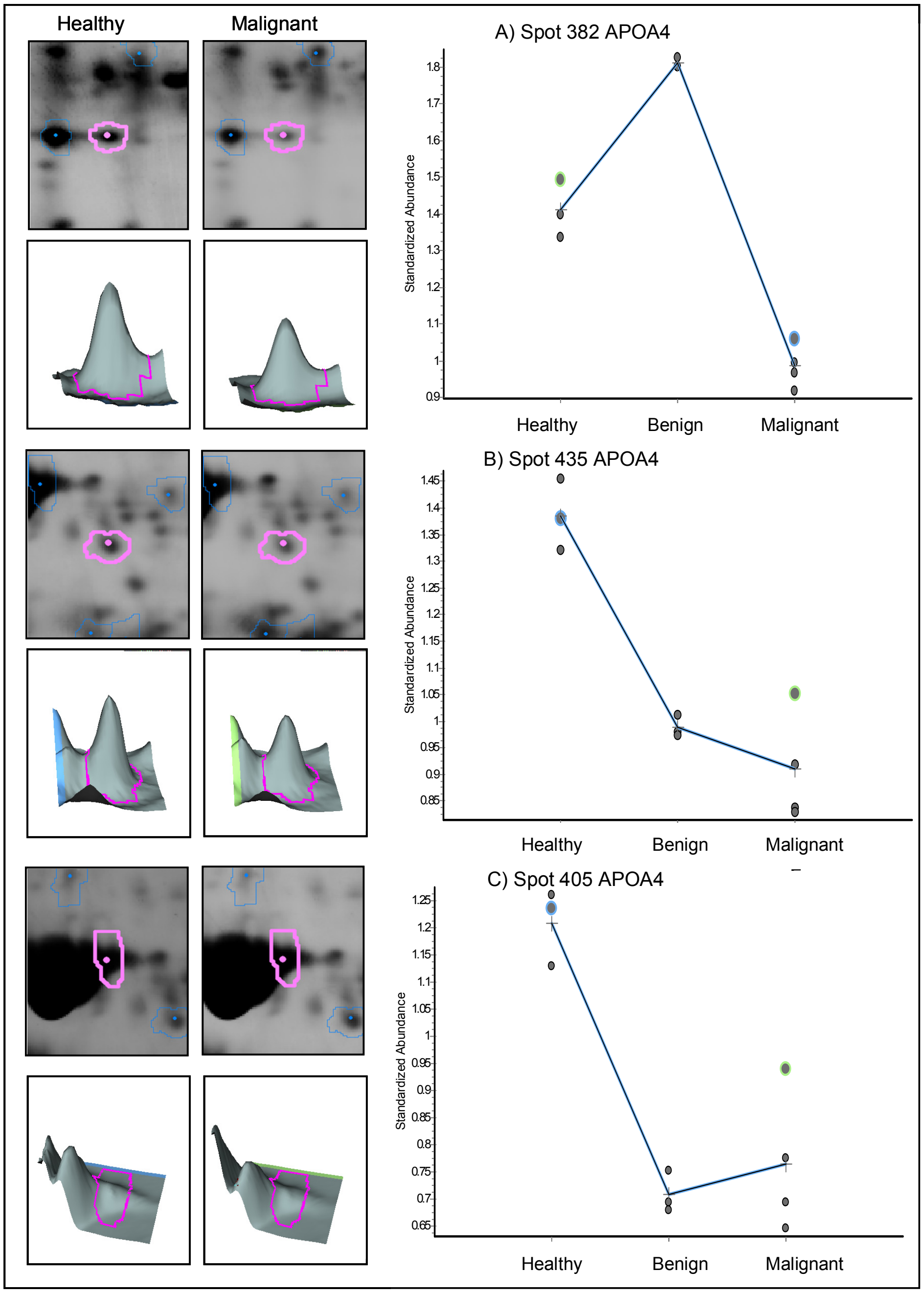


Figure 6.21 Examples of multiple spots displaying clinical-sample dependent changes in the ProteoMiner experiment. Several spots including A) 382, B) 405 and C) 435 yielded hits for Apolipoprotein A4 (APOA4). Differential analysis showed this protein was down-regulated in the malignant condition verses the healthy condition. Graphs were derived from DeCyder image analysis where the standardised abundance is the ratio of the volume of a gel feature in the clinical condition versus the Cy2 standard. Data points are shown for triplicate measurements in the healthy and quadruplicate in the malignant with lines joining the average values. 3D images of spots are shown for the healthy and malignant conditions.

In the second set of MARS-depleted samples, 1468 protein features were detected in the master gel. Quantitative analysis showed 76 differentially expressed protein features (≥ 1.5 average fold-change in abundance, $p < 0.01$). In total, 32 of the 76 differentially expressed protein features were identified with high confidence Table 6.5 examples are shown in Figure 6.22 and MS-based protein identification results are shown in Appendix 5.

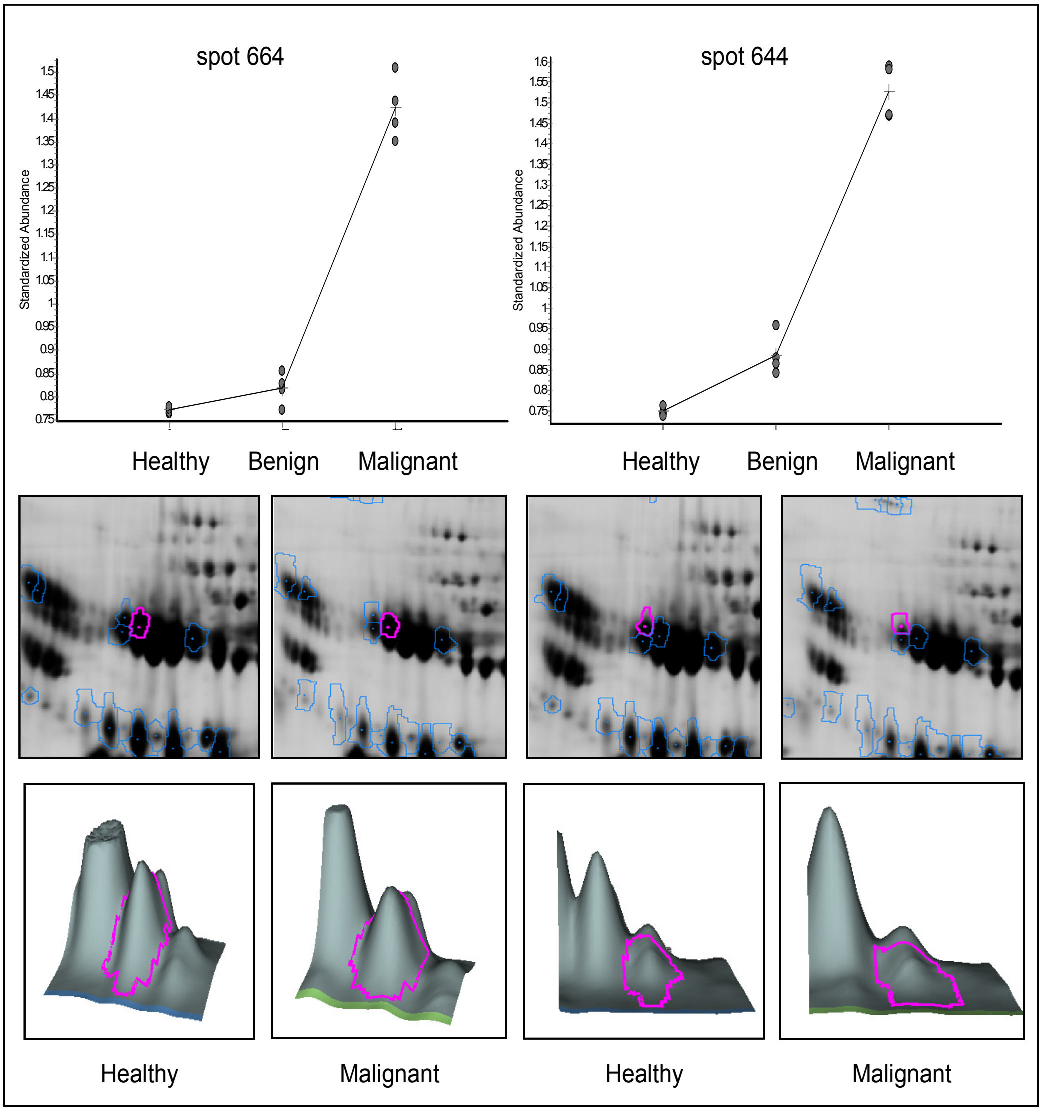


Figure 6.22 Examples of multiple spots displaying clinical-sample dependent changes in the second MARS-fractionated samples. Protein spots 664 and 644 were both identified as alpha-1-antitrypsin (SERPINA) and displayed differential expression. Graphs were derived from DeCyder image and data points are shown for triplicate measurements for the healthy condition and quadruplicate for the malignant condition with lines joining the average values. 3D images of spots are shown for the healthy and malignant conditions.

Spot	Protein name	IPI No.	Score	Seq. Cov.	No.	Mw	pI	Healthy/Malignant		Healthy/Benign		Benign/Malignant	
No.				(%)	Peptides			Av. Ratio	T-test	Av. Ratio	T-test	Av. Ratio	T-test
73*	Alpha-1-antitrypsin precursor Isoform 1 (SERPINA)	IPI00553177	103	40	18	46707	5.37	-1.81	0.0055	-1.09	0.29	-1.66	0.0068
76	Not Identified							-1.65	0.0013	-1.06	0.37	-1.55	0.0031
220	Not Identified							1.44	0.01	1.66	0.00033	-1.15	0.2
250	Not Identified							1.96	7.00E-07	-1.02	0.62	2	2.10E-07
252*	Serotransferrin precursor (TF)	IPI00022463	81	33	24	77000	6.81	2.44	1.70E-08	1.01	0.74	2.41	9.60E-07
279*	Not Identified							2.44	4.20E-07	1.1	0.026	2.21	1.10E-06
280*	Serotransferrin precursor (TF)	IPI00022463	114	38	28	77000	6.81	1.87	0.0027	1.02	0.59	1.82	0.0097
281	Serotransferrin precursor (TF)	IPI00022463	272	20	11	79280	6.81	2.27	5.60E-07	1.08	0.032	2.1	3.50E-07
281	Immunoglobulin heavy constant mu protein (IGHM)	IPI00477090	59	8	4	68052	5.89	2.27	5.60E-07	1.08	0.032	2.1	3.50E-07
289	Immunoglobulin heavy constant gamma 1 (IGHG1)	IPI00384938	124	11	4	53503	6.28	2.23	3.00E-07	1	0.91	2.22	5.60E-07
289	Antithrombin III variant (SERPINC1)	IPI00032179	47	9	4	53114	6.11	2.23	3.00E-07	1	0.91	2.22	5.60E-07
296	Not Identified							2	1.20E-05	1.01	0.79	1.98	6.80E-06
297	Serotransferrin precursor (TF)	IPI00022463	315	16	10	79280	6.81	2.25	4.20E-06	1.01	0.78	2.22	5.90E-07
323*	Serotransferrin precursor (TF)	IPI00022463	146	45	32	77000	6.81	2.42	3.30E-07	1.02	0.66	2.37	8.90E-09
332*	Serotransferrin precursor (TF)	IPI00022463	140	48	32	77000	6.81	2.52	6.60E-07	1.12	0.0072	2.24	2.80E-06
333*	Serotransferrin precursor (TF)	IPI00022463	56	26	20	77000	6.81	2.78	7.50E-08	1.14	0.00069	2.44	6.50E-08
335	Serum albumin (ALB)	IPI00745872	80	32	22	69321	5.92	2.17	0.00011	1.06	0.2	2.05	0.00018
346	Serotransferrin precursor (TF)	IPI00022463	82	31	22	77000	6.81	2.63	3.20E-07	1.05	0.28	2.51	3.30E-07
352	Not Identified							2.14	0.00079	1.27	0.0076	1.68	0.0053
357	Serum albumin (ALB)	IPI00745872	97	7	5	71317	5.92	1.84	2.80E-06	1.36	0.001	1.35	0.00051
357	Serotransferrin precursor (TF)	IPI00022463	92	8	5	79280	6.81	1.84	2.80E-06	1.36	0.001	1.35	0.00051
407	Not Identified							1.77	0.00055	1.07	0.06	1.65	0.00084
452	Not Identified							1.12	0.24	-1.91	0.00046	2.14	2.30E-05
457	Not Identified							-1.7	2.30E-05	-3.34	2.00E-07	1.96	4.70E-07
459	Not Identified							-1.42	0.0003	-2.85	9.40E-07	2.01	3.60E-06
461	Serum albumin precursor Isoform 1 (ALB)	IPI00745872	96	32	22	71317	5.92	-1.77	1.50E-06	-3.94	1.40E-08	2.22	1.30E-06
464	Serum albumin (ALB)	IPI00022434	67	26	19	73881	6.33	-1.04	0.44	-2.09	6.30E-05	2.02	2.10E-05
465	Serum albumin (ALB)	IPI00022434	108	36	32	62166	6.33	-1.4	0.0057	-3.07	4.20E-05	2.2	5.50E-05

Table 6.5 2D-DIGE based analysis of the second MARS-depleted clinical sample set. Protein features displaying differential expression are shown. Values are average ratio of abundance between different clinical conditions (healthy/malignant, benign/malignant, and healthy/benign). T-test *p* values are given as a measure of confidence for each ratio measured. Protein were identified by MS Protein name, IPI accession number Mascot score, sequence coverage (%) and number of matched peptides are given for each of the identified proteins are shown. Those marked with an * were identified by MALDI-TOF PMF.

Spot No.	Protein name	IPI No.	Score	Seq. Cov. (%)	No. Peptides	Mw	pI	Healthy/Malignant		Healthy/Benign		Benign/Malignant	
								Av. Ratio	T-test	Av. Ratio	T-test	Av. Ratio	T-test
477	Serum albumin precursor Isoform 1 (ALB)	IPI00745872	78	32	24	61945	5.92	-1.78	1.90E-06	-4.02	1.10E-07	2.26	5.60E-06
523	Alpha-1-antichymotrypsin precursor Isoform 1 (SERPINA3)	IPI00550991	92	34	14	60850	5.42	-2	2.90E-07	-1.16	0.0081	-1.73	2.20E-06
537	Not Identified							-1.56	7.50E-08	1	0.99	-1.56	2.30E-09
541	Not Identified							1.76	0.0022	-1.37	0.0084	2.41	0.00014
585	Immunoglobulin heavy chain constant region alpha 1 protein (IGHA1)	IPI00386879	97	6	3	54195	6.46	-1.01	0.98	-1.59	0.01	1.57	0.018
609	Not Identified							1.68	0.0065	-1.23	0.00067	2.06	0.0012
644*	Alpha-1-antitrypsin (SERPINA)	IPI00790784	72	29	14	40238	4.98	-2.04	7.50E-08	-1.18	0.0011	-1.72	4.80E-06
644*	Phenylalanyl-tRNA synthetase beta chain (phe T)	IPI00300074	63	22	18	66715	6.4	-2.04	7.50E-08	-1.18	0.0011	-1.72	4.80E-06
664*	Alpha-1-antitrypsin precursor (SERPINA)	IPI00553177	136	43	24	55397	5.37	-1.84	2.60E-07	-1.06	0.041	-1.74	2.60E-06
664*	Protein Daple Isoform 1 (DAPLE)	IPI00740019	82	18	37	229215	5.87	-1.84	2.60E-07	-1.06	0.041	-1.74	2.60E-06
664*	Vinculin Isoform 2 (VCL)	IPI00307162	71	23	26	124292	5.5	-1.84	2.60E-07	-1.06	0.041	-1.74	2.60E-06
664*	Ras-related protein (Rab-2B)	IPI00102896	66	57	11	24427	7.68	-1.84	2.60E-07	-1.06	0.041	-1.74	2.60E-06
680*	Alpha-1-antitrypsin precursor (SERPINA)	IPI00553177	69	32	16	46707	5.37	-1.68	1.60E-07	-1.09	0.04	-1.53	2.40E-05
712*	Alpha-1-antitrypsin precursor (SERPINA)	IPI00553177	145	49	23	46707	5.37	-1.62	9.90E-07	-1.06	0.17	-1.52	6.30E-05
712*	Isoform I of T-box transcription factor TBX3	IPI00298944	76	28	18	77529	8.48	-1.62	9.90E-07	-1.06	0.17	-1.52	6.30E-05
712*	Vinculin Isoform 2 (VCL)	IPI00307162	70	19	25	124292	5.5	-1.62	9.90E-07	-1.06	0.17	-1.52	6.30E-05
769	Not Identified							1.74	6.30E-06	1.39	0.00015	1.25	0.00015
774	Not Identified							-2.05	0.0015	-1.66	0.0065	-1.23	0.012
885	Not Identified							-2.28	2.70E-06	-1.07	0.14	-2.13	1.20E-05
888	Not Identified							1.41	0.048	-1.44	0.0016	2.03	0.009
926	Not Identified							-3.04	1.60E-05	-1.53	0.0001	-1.99	1.40E-07
930	Haptoglobin precursor (HP)	IPI00431645	70	45	16	31362	8.48	-2.09	0.00026	-1.03	0.67	-2.03	0.00024
948	Apolipoprotein A-IV precursor (APOA4)	IPI00304273	83	9	4	45371	5.28	-1.77	1.80E-05	-1.2	0.0031	-1.48	6.40E-05
950	Not Identified					44836	5.14	-2.48	1.60E-07	-1.51	6.10E-05	-1.64	6.80E-06
971	Apolipoprotein A-IV precursor (APOA4)	IPI00304273	190	55	24	44044	5.24	-3.25	1.10E-08	-1.82	6.60E-06	-1.79	4.20E-06
973	Haptoglobin precursor (HP)	IPI00431645	497	39	17	31647	8.48	-1.99	5.20E-06	-1.22	0.0041	-1.63	4.20E-05
974	Alpha-1-acid glycoprotein 2 precursor (AGP 2)							1.44	0.00017	1.59	0.00019	-1.1	0.063
991*	Haptoglobin precursor (HP)	IPI00431645	63	43	14	31647	8.48	-2.81	5.30E-08	-1.52	8.70E-06	-1.85	2.00E-06
1010*	Haptoglobin precursor (HP)	IPI00431645	64	38	14	31362	8.48	-2.58	1.10E-07	-1.51	2.60E-05	-1.71	8.10E-06
1021	Not Identified							-2.25	3.50E-07	-1.47	6.40E-05	-1.53	2.20E-05
1042*	Haptoglobin precursor (HP)	IPI00478493	82	42	17	38941	6.13	-2.69	2.80E-07	-1.51	0.00024	-1.78	5.00E-05
1044*	Haptoglobin precursor (HP)	IPI00431645	61	39	14	31362	8.48	-2.03	1.90E-05	-1.48	0.0005	-1.37	0.0019
1047*	Haptoglobin precursor (HP)	IPI00478493	64	39	14	38427	6.13	-1.74	3.60E-06	-1.43	0.00045	-1.22	0.004
1074	Serum albumin (ALB)	IPI00022434	30	3	2	39529	5.38	-2.54	2.90E-05	-1.44	0.0042	-1.76	4.80E-06
1093	Not Identified							-2.65	3.10E-08	-1.55	5.20E-05	-1.71	9.90E-06

Table 6.5 continued

Spot No.	Protein name	IPI No.	Score	Seq. Cov. (%)	No. Peptides	Mw	pI	Healthy/Malignant		Healthy/Benign		Benign/Malignant	
								Av. Ratio	T-test	Av. Ratio	T-test	Av. Ratio	T-test
1117	Not Identified							-2.08	4.60E-07	-1.5	2.30E-05	-1.39	0.00027
1137	Alpha-1-antitrypsin Isoform 1 of precursor (SERPINA1)	IPI00553177	631	38	21	46878	5.37	-1.61	9.30E-05	-1.68	0.0002	1.04	0.2
1158	Not Identified							1.4	0.0012	-1.31	0.0014	1.83	5.50E-06
1160	Not Identified							-1.4	7.30E-06	-1.55	3.20E-07	1.1	0.011
1186	Not Identified							1.57	0.00029	1.22	0.0023	1.29	0.0057
1193	Alpha-1-antitrypsin Isoform 1 of precursor (SERPINA1)	IPI00553177	40	6	2	46878	5.37	-1.55	4.50E-05	-1.28	0.0072	-1.21	0.014
1249	Not Identified							1.52	0.0033	-1.16	0.11	1.77	9.80E-05
1253	Not Identified							1.6	0.0076	1.32	0.007	1.21	0.15
1265	Not Identified							1.61	0.00037	-1.27	0.0079	2.05	2.00E-06
1280	Immunoglobulin kappa variable 1-5 (IGKV1)	IPI00430820	57	18	3	26034	5.74	1.22	0.0014	-1.41	1.20E-05	1.71	1.40E-06
1313	Not Identified							1.23	0.4	-5.61	2.70E-05	6.89	0.00026
1325	Not Identified							1.27	0.011	-1.39	0.0017	1.77	5.30E-05
1351	Not Identified							1.35	0.00023	-1.19	0.04	1.61	0.00027
1360	Not Identified							1.44	0.00041	-1.17	0.021	1.69	0.00011
1394	Not Identified							2.07	0.0023	1.17	0.18	1.77	0.0055
1402	Not Identified							-1.52	0.0012	-1.43	0.0003	-1.06	0.37
1417	Not Identified							2.59	7.20E-06	-1.04	0.53	2.69	1.90E-05
1425	Haptoglobin related protein Isoform 1 (HRP)	IPI00477597	178	8	6	39496	6.42	1.79	0.003	1.05	0.3	1.69	0.0047
1439	Not Identified							-2.06	1.00E-06	-1.29	0.00071	-1.6	2.30E-06
1441	Not Identified							-2.15	7.50E-06	-1.88	1.20E-05	-1.14	0.099
1442	Not Identified							-2.53	2.10E-08	-1.73	3.00E-06	-1.46	4.40E-06
1445	Not Identified							-2.39	9.50E-09	-1.53	5.10E-06	-1.56	1.10E-06

Table 6.5 continued

The 2D gel migration of the identified differentially expressed proteins in the second MARS experiment is shown in Figure 6.23. Again several abundant serum proteins including alpha-1-antitrypsin precursor (SERPINA), serotransferrin (TF), serum albumin (ALB), the immunoglobins (IGHA) and haptoglobin (HP) were identified in multiple locations.

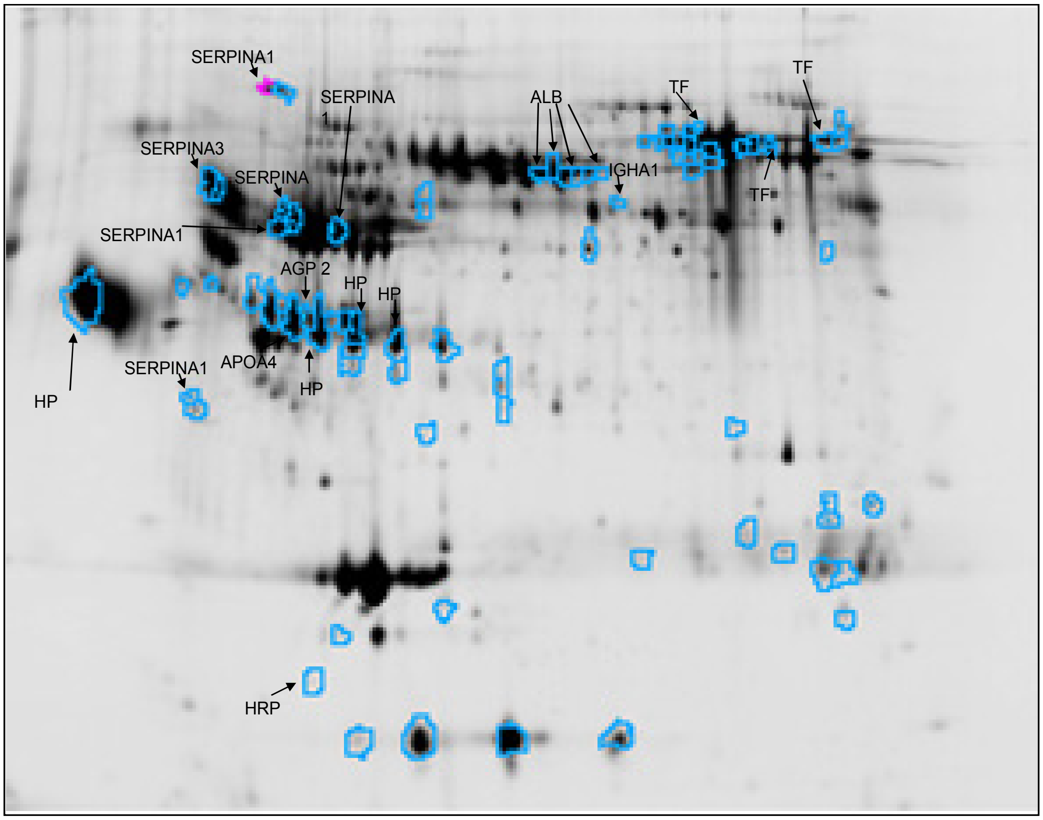


Figure 6.23 Representative gel displaying position of differentially expressed proteins from the MARS-depleted samples. Differentially expressed protein features were identified by MS. The locations of serum albumin precursor (ALB), the immunoglobin heavy (IGHA1), serotransferrin (TF), haptoglobin (HP), apolipoprotein AIV (APOA4), alpha-1-antitrypsin precursor (SERPINA), alpha-1-antichymotrypsin (SERPINA3), alpha-1-acid glycoprotein 2 precursor (AGP2) and Haptoglobin-related protein precursor (HRP) identified as differentially expressed protein features are shown.

6.3 Comparison of 2D-DIGE experiments

Results from the BVA analysis showed that a different number of protein features were detected in each of the 2D-DIGE experiments. The 2D gels showed better protein separation when only 240 µg of protein was loaded (80 µg per Cy dye, second MARS experiment). Indeed, in the second MARS experiment due to streaking 1468 protein features were detected in the master gel used for spot matching. Poor resolution of some peaks e.g. albumin and serotransferrin made it difficult to determine an accurate number of protein features. Furthermore, accurate quantification was often compromised due to the partial co-migration of some protein features. 934 protein features were detected in the unfractionated experiment, while 797 and 697 protein features were detected in the first MARS experiment and ProteoMiner experiments, respectively. Only 32 out of 76 of the differentially expressed features from the second MARS experiment yielded protein identifications compared with 35 out of 48 in the unfractionated, 10 out of 10 in the first MARS and 53 out of 65 in the ProteoMiner experiment. This may be due to the lower protein load used in the second MARS experiment. In the ProteoMiner and second MARS experiments the selection criteria for differentially expressed protein features were made more stringent to reduce the number of false positives [Karp et al., 2007]. The difference in the number of protein features selected for identification using cut-offs of $p < 0.05$ and $p < 0.01$ was modest as shown in Table 6.6.

Differentially expressed protein features		
A) ProteoMiner	p < 0.05	p < 0.01
H v B	12	9
H v M	32	27
B v M	54	46
Total	74	65
B) MARS	p < 0.05	p < 0.01
H v B	61	61
H v M	25	25
B v M	59	58
Total	77	76

Table 6.6 Differentially expressed protein features in the ProteoMiner and second MARS experiments. The number of features displaying ≥ 1.5 average fold-changes in abundance at $p < 0.05$ and $p < 0.01$ are shown.

Results showed a total of 57 gene products were identified with confidence as shown in Table 6.7 which also shows the direction of regulations (up/down) between clinical conditions and the number of features where the protein was identified in each experiment. Ten proteins (AGP2, AHSG, CP, Factor VII, IGHA1, IGHG2, IGHG3, IGHM, IGL, KHG1) were found only in the unfractionated sample. Eight proteins (AFM, A1BG, A2M, CFB, ITIH4, F2 fragment and PROS) were found only in the first MARS-depleted samples. Thirteen proteins (inc; C4A;C4B, APOE, SERPINC1 and Vitamin D binding protein) were only found in the ProteoMiner-fractionated samples and 7 proteins (AGP2, Phe RS, DAPLE, HRP, Rab-2B, TBX3 and VCL) were found only in the second set of MARS-depleted samples. Three proteins including isoform 1 of α-1-antitrypsin precursor (SERPINA1), apolipoprotein A-IV precursor (APOA4) and immunoglobulin heavy constant gamma 1 (IGHG1) were found to be commonly differentially expressed in the unfractionated, MARS2 and ProteoMiner fractionated samples. Of these SERPINA1 was up-regulated in the malignant condition versus healthy and benign conditions in all three experiments (Table 6.7). It is important to note that there are a number of multiple hits per spot which is one of the major drawbacks of the 2-DE technique making attribution of differential expression to a specific protein very difficult.

	Malignant/Healthy				Benign/Healthy				Malignant/Benign			
Protein name	UnF	MARS1	PM	MARS2	UnF	MARS1	PM	MARS2	UnF	MARS1	PM	MARS2
Albumin (ALB)	+(10)			+(3) -(1)	+(10)			+(3) -(1)	+(10)			+(1) -(3)
Adenomatosis polyposis coli 2 (APC2)			+(1)				-(1)				+(1)	
Afamin precursor (AFM)		+(1)				+(1)				+(1)		
Alpha-1-antichymotrypsin precursor (SERPINA3)	+(2)			+(1)	+(2)			+(1)	+(2)			+(1)
Alpha-1-antitrypsin Isoform 1 of precursor (SERPINA1)	+(1)		+(3)	+(7)	+(1)		+(3)	+(7)	+(1)		+(3)	+(6) -(1)
Alpha-1-acid glycoprotein 2 precursor (AGP 2)				-(1)				-(1)				+(1)
Alpha-1B-glycoprotein precursor (A1BG)		+(4)				+(3) -(1)				+(4)		
Alpha-2-HS-glycoprotein precursor (AHSG)	+(1)				+(1)				+(1)			
Alpha-2-macroglobulin precursor (A2M)		-(2)				-(2)				+(1) -(1)		
Antithrombin III variant (SERPINC1)			+(3)	-(1)			+(3)	-(1)			+(3)	-(1)
Amyotrophic lateral sclerosis 2 chromosomal region (ALS2CR12)			-(1)				+(1)				-(1)	
Apolipoprotein A-I precursor (APOA1)			-(2)				-(2)				-(2)	
Apolipoprotein A-IV precursor (APOA4)	+(2)		+(1) -(3)	+(2)	+(2)		+(2) -(2)	+(2)	+(2)		+(1) -(3)	+(2)
Apolipoprotein E precursor (APOE)			+(4) -(1)				+(3) -(2)				+(2) -(2)	
Ceruloplasmin precursor (CP)	+(2)				+(2)				+(2)	+(1)		
Complement C3 precursor (C3)	+(1)				+(1)				+(1)			
Complement C4-A precursor (C4A;C4B)	+(1)		+(1)		+(1)		+(1)		+(1)		+(1)	
Complement component C9 precursor (C9)	+(2)	+(1)			+(2)	+(1)			+(2)	+(1)		
Complement factor B Isoform 1 of precursor (CFB Fragment)		-(2)				+(1) -(1)				-(2)		
Complement factor H Isoform 1 of precursor (CFH)			+(1)				+(1)				+(1)	
Complement factor H Isoform 2 of precursor (CFH)			+(1) -(1)				+(1) -(1)				+(1) -(1)	
Complement factor H-related Isoform short of protein 2 precursor (CFHR2)			+(1)				+(1)				+(1)	
Complement factor H-related Isoform long of protein 2 precursor (CFHR2)			+(1)				+(1)				+(1)	
Factor VII active site mutant immunoconjugate	+(2)				+(2)				+(2)			
Fibrinogen gamma chain Isoform Gamma-A precursor (FGG)			-(1)				-(1)				-(1)	
GTPase-activating Rap/Ran-GAP domain-like 1 isoform 3			+(1)				-(1)				+(1)	
Haptoglobin precursor (HP)	+(9)			+(7)	+(9)			+(7)	+(9)			+(7)
Haptoglobin related protein (HRP)												
Histidine-rich glycoprotein precursor (HRG)		+(3)	+(1)			+(2) -(1)	-(1)			+(3)	-(1)	
Immunoglobulin heavy chain constant region alpha 1 protein (IGHA1)	+(2)				+(2)				+(2)			

Table 6.7 Differentially expressed proteins identified by MS. Proteins from all four experiments displaying differential expression are shown. Values are the number of features up-(+) or down-regulated (-) between different clinical conditions (malignant/healthy, benign/healthy, and malignant/benign).

Protein name	Malignant/Healthy				Benign/Healthy				Malignant/Benign			
	UnF	MARS1	PM	MARS2	UnF	MARS1	PM	MARS2	UnF	MARS1	PM	MARS2
Immunoglobulin heavy constant gamma 1 (IGHG1)	+(3)		+(1) -(2)	-(1)	+(3)		+(1) -(2)	-(1)	+(3)		+(1) -(2)	-(1)
Immunoglobulin heavy constant gamma 2 (IGHG2 Fragment)	+(2)				+(2)				+(2)			
Immunoglobulin heavy constant gamma 3 (G3m marker IGHG3)	+(1)			-(1)	+(1)			-(1)	+(1)			-(1)
Immunoglobulin heavy constant mu protein (IGHM)	+(8)				+(8)				+(8)			
Immunoglobulin kappa variable 1-5 (IGKV1)				-(1)				+(1)				-(1)
Immunoglobulin lambda locus protein (IGL)	+(1)				+(1)				+(1)			
Inter-alpha (Globulin) inhibitor H4 (Plasma Kallikrein sensitive glycoprotein) (ITIH4)		+(1)				-(1)				+(1)		
Inter-alpha-trypsin inhibitor heavy chain H4 Isoform 1 of precursor (ITIH4)		+(2)				+(2)				+(2)		
Inter-alpha-trypsin inhibitor heavy chain H4 Isoform 2 of precursor (ITIH4)		-(1)	-(1)			+(1)	+(1)			-(1)	-(1)	
Kininogen-1 isoform LMW precursor (KHG1)	+(1)				+(1)				+(1)			
Kinesin-like protein (KIFI5)			-(1)				+(1)				-(1)	
Kinectin 1 isoform b (KTN1)			-(1)				-(1)				+(1)	
Pericentrin (PCNT)			+(1)				+(1)				+(1)	
Periplakin (PPL)			-(1)				+(1)				-(1)	
Phenylalanyl-tRNA synthetase beta chain (pheT)				+(1)				+(1)				+(1)
Prothrombin precursor (F2 Fragment)		+(2)				+(2)				+(2)		
Protein Daple (DAPLE) Isoform 2				+(1)				-(1)				+(1)
Pyruvate Kinase L (PKL)			+(1)				+(1)				+(1)	
Ras-related protein (Rab-2B)				+(1)				+(1)				+(1)
Serotransferrin precursor (TF)			+(3) -(1)	-(9)			+(3) -(1)	-(9)			+(1) -(3)	-(9)
Serum albumin precursor Isoform 1 (ALB)			-(14)	+(1)			+(14)	+(1)			-(14)	-(1)
T-box transcription factor isoform 1 (TBX)				-(1)				-(1)				-(1)
Utrophin (UTRN)			+(1)				+(1)				-(1)	
Vinculin (VCL) Isoform 2				+(2)				+(2)				+(2)
Vitamin D-binding protein precursor (GC)			+(1)				+(1)				+(1)	
Vitamin K-dependent protein S precursor (PROS)		+(1)				+(1)				+(1)		

Table 6.7 continued

6.4 Discussion

The human serum proteome is a complex biological mixture potentially containing an archive of patho-physiological information. Serum proteomics is being increasingly used to discover and characterise candidate biomarkers for disease diagnosis, prognosis and treatment response. The most important lesson that has been learned from protein expression profiling is the difficulty of handling the dynamic range of the serum proteome, which is 9-10 orders of magnitude and greater than the quantitative dynamic range of most analytical techniques [Fung et al., 2005].

For the studies presented in this chapter sera from the UKOPS collection was pooled into clinical groups, since analysis of large numbers of samples individually by 2D-DIGE was not possible. Pooling strategies hide the underlying variation within a group and may reveal average data that is skewed by outliers. By pooling larger numbers of samples this could be overcome [Pitteri and Hanash, 2007]. The dynamic range of the proteome was compressed using two complimentary fractionation strategies prior to fluorescence two dimensional difference gel electrophoresis-based profiling. The MARS depletion technique was used to enrich lower abundance protein species by depleting 7 of the most abundant proteins. Additionally, the ProteoMiner protein enrichment kit was used to reduce the dynamic range and capture ‘equalised’ amounts of all constituent proteins. Differentially expressed protein features from whole sera and fractionated sera were excised and trypsin digested. The resultant peptide mixtures were analysed by MALDI-TOF PMF and LC-MS/MS. One of the advantages of LC-MS/MS is that the tryptic peptides are further fragmented in the mass spectrometer (in the collision cell), allowing the determination of amino acid sequence and resulting in less ambiguous protein identifications compared with peptide mass fingerprinting [Koehn and Oehler, 2007].

The work carried out in this section showed that only a few protein species were consistently differentially expressed between the clinical conditions. The results indicate that some of the proteins are related to immuno-host cell defence responses (acute-phase proteins) and may be protein fragments, possibly produced by proteolytic enzymes. In all experiments the majority of the differentially expressed features yielded hits for high abundance proteins.

Proteins identified included SERPINA3 and APOA4, and both of these proteins were up-regulated in the malignant condition. SERPINA3 is an alpha globulin glycoprotein which is a member of the serine proteinase inhibitor (serpin) family. This enzyme is produced in the liver and is known to be an acute-phase protein that is induced during inflammation. SERPINA3 has been shown to be up-regulated in human leukocyte antigen (HLA)-positive tumours in cervical cancer [Kloth et al., 2008]. Gene expression profiling experiments have shown a 2-fold increase in the expression of SERPINA3 in mucinous ovarian cancers compared with normal ovarian surface epithelial cells and to other histotypes [Marquez et al., 2005]. In agreement with this results from both the unfractionation and second MARS experiments showed SERPINA3 was up-regulated in the malignant condition. APOA4 belongs to the apolipoprotein family and is primarily produced in the intestine and secreted into the plasma. APOA4's precise function is unknown but it may serve as a lipid-binding protein and has lecithin:cholesterol acyltransferase (LCAT) activating ability. The APOA4 gene was shown to be up-regulated in familial pancreatic cancer, it has been studied extensively in relation to cardio-vascular disease and is a strong candidate as a breast and ovarian cancer susceptibility gene [Zervos et al., 2006].

To enrich the lower abundant protein species sample pre-fractionation was performed with the multiple affinity removal system. Results showed that 7 of the high abundance proteins were removed with high efficiency. One of the major limitations of 2D-gels is the sample loading capacity. Too much protein sample results in poorly resolved gels and too little protein makes MS based protein identification of low abundant protein features difficult [Sriyam et al., 2007; Huang et al., 2005]. Depleting serum of the abundant proteins facilitated the analysis of lower abundant species since more of these proteins could be loaded on the 2D gels.

Quantitative analysis of the first set of MARS-depleted clinical samples showed only 10 differentially expressed proteins features that were significant ($p < 0.05$). Protein identifications were made for all of these with several spots yielding more than one protein. The identification of more than one protein in a spot is a major limitation of 2D-gel based quantitative analysis, since it is difficult to assign the differential expression to a particular protein found in the spot. It would most likely to have come from the most abundant protein in the spot, particularly if the amount of this protein is far in excess of the others. For example, in the first MARS-depletion experiment spot number 459 yielded multiple identities including ITIH4, HRG, F2, ALGB and PROS1 by LC-MS/MS. Of these ITIH4 was identified by 16 peptides, with a Mascot score of 401. HRG was identified by 5 peptides and a Mascot score of 137, F2 was identified by 4 peptides with a Mascot score of 95, and ALGB was identified by 2 peptides with a Mascot score of 52, and finally, PROS1 was identified by 3 peptides with a Mascot score of 52. This suggests that the likely difference in abundance is attributable to ITIH4. Spot number 459 showed a 2.5-fold increase in the malignant condition versus the healthy conditions, respectively. However, it could be speculated that the co-migration of all these proteins may have generated a false positive.

Interestingly, in the list of proteins identified, ITIH4 and afamin were two proteins which have been previously reported as ‘markers’ for ovarian cancer [Jackson et al., 2007; Zhang et al., 2004]. Both of these were identified in spot number 464 which showed a 2-fold increase in the malignant condition. ITIH4 is an acute-phase protein which is produced by the liver. The levels of ITIH4 have been shown to increase significantly in the sera of patients after different surgical trauma [Pineiro et al., 1999]. Afamin is a vitamin E binding glycoprotein which is part of the albumin super family. Previous studies have reported on the potential of afamin as a putative marker for ovarian cancer recurrence [Jackson et al., 2007]. The study by Jackson et al. showed an inverse relationship between CA-125 and afamin concentrations from the time of treatment to the time of relapse. The potential complementarity of afamin with CA-125 was also shown in three patients in whom CA-125 was relatively uninformative, although the changes in afamin were modest. For diagnostic discrimination, afamin alone was poor, but it was suggested that the potential for the isoforms, in particular isoform 2, for complementing CA-125 or other markers should be explored further in a larger study with an independent test set [Jackson et al., 2007].

The MARS depletion experiment was repeated and several additional proteins which showed differential expression between the clinical pooled sera were identified but there was no overlap between the two MARS experiments. This is likely due to the fact that no Cy2 internal standard was used in the first experiment which made spot matching across gels difficult and only 10 differentially expressed features were identified. The proteins identified in the second MARS experiment included haptoglobin (HP) and isoform 1 of alpha-1-antitrypsin (SERPINA1) both of which were up-regulated in the malignant condition. HP is a blood plasma protein that binds free haemoglobin, preventing the loss of iron through the kidneys and protecting the kidneys from damage by haemoglobin, while making the haemoglobin accessible to degradative enzymes. In the clinic, a haptoglobin assay is used to screen for and monitor haemolytic anemia. HP has been previously reported as an up-regulated marker protein in ovarian cancer patients [Ye et al., 2003]. Previous reports in ovarian cancer have indicated that there is a change of glycosylation on haptoglobin [Turner et al., 1995] and IgG [Gercel-Taylor et al., 2001] in ovarian cancer patients.

SERPINA1 is a plasma protein which acts as an inhibitor of serine proteases. Its primary target is elastase, but it also has a moderate affinity for plasmin, thrombin, trypsin, chymotrypsin and plasminogen activator. The aberrant form increases the rate of coagulation and has proteolytic activity against insulin and plasmin. SERPINA1 has also been previously reported as an acute-phase marker in gyneocological cancers [Kloth et al., 2008; Tatra, 1985]. SERPINA1 was one of three proteins which were differentially expressed in all the unfractionated, MARS2 and ProteoMiner experiments. SERPINA1 was up-regulated in the malignant condition versus the healthy and benign condition in all three experiments. However, HP and SERPINA were in the list of proteins which should have been depleted by the MARS column. In addition, serotransferrin (TF) was identified as a differentially expressed protein feature in a number of locations. This suggested that the depletion efficiency of the column may have declined with repeated use. TF was found to be down-regulated in the malignant condition in 9 different spots. TF is an iron-binding transport protein which is responsible for transporting iron from the sites of absorption and heme degradation to areas of storage and utilisation and has been previously reported to be down-regulated in the serum of ovarian cancer patients [Dumaswala et al., 2000].

Other proteins identified in the second MARS-depletion experiment included alpha-1-acid glycoprotein 2 precursor (AGP2), phenylalanyl-tRNA synthetase beta chain (Phe-RS), protein daple isoform 2 (DAPLE), haptoglobin related protein (HRP), Ras-related protein (Rab-2B), T-box transcription factor isoforms 1 (TBX3) and vinculin isoform 2 (VCL). AGP2 is expressed by the liver, secreted into the plasma and appears to function in modulating the activity of the immune system during the acute-phase reaction. Phenylalanine-tRNA synthetase (Phe-RS) is localised to the mitochondrion and is an essential enzyme which catalyzes the transfer of the amino acid phenylalanine (Phe) to the Phe-specific transfer-RNA. DAPLE is a negative regulator of the canonical Wnt signalling pathway, acting downstream of dishevelled (Dvl) to inhibit βeta-catenin stabilisation. Wnts are a large family of cysteine-rich secreted glycoproteins that control development. The intracellular signalling pathway of Wnt is also conserved evolutionally and regulates cellular proliferation and differentiation [Bienz and Clevers, 2000; Seidensticker and Behrens, 2000; Wodarz and Nusse, 1998] and several components of Wnt signalling are implicated in the genesis of human cancer.

RAB2B is required for protein transport from the endoplasmic reticulum to the Golgi complex. Pyruvate kinase L (PKLR) was also identified as a differentially expressed protein. Pyruvate kinase, is a homotetramer of 50.60 kDa subunit, with two forms, one liver specific (L), the other erythrocyte specific (R), transcribed from a distinct promoter, glycolysis, energy pathway, generating ATP from ADP. T-box transcription factor isoforms 1 (TBX3) is a transcriptional repressor involved in developmental processes which plays a role in limb pattern formation. Tbx3 transcription is activated by ectopic expression of beta-catenin in mouse liver and in human tumor cell lines. Tbx3 has been indicated as a serological marker for ovarian cancer [Souchelnytskyi et al., 2006]. Finally, VCL is involved in cell adhesion and may be involved in the attachment of the actin-based microfilaments to the plasma membrane. It may also play important roles in cell morphology and locomotion. VCL subcellular locations include the cytoplasm and cytoskeleton. These proteins were predominately found in spots which yielded multiple protein identities it is therefore difficult to attribute any specific changes to them. However, several proteins e.g.

TBX3 and DAPLE have previously been reported as markers for ovarian cancer thus these proteins warrant further investigation.

The ProteoMiner-based fractionation showed that compression of the protein dynamic range could be achieved, although mostly high abundance proteins were found to be differently expressed; namely, TF, APOA1, APOA4, APOE and SERPINA1. The results demonstrated the ProteoMiner kit appeared to have a greater affinity for apolipoproteins as previously reported [Pitteri and Hanash, 2007]. Numerous studies have reported on changes in serum lipid and lipoproteins levels in cancer and other diseases. For example, elevated plasma lipoprotein (A) has been associated with an increased risk of cardiovascular disease and significant elevation of total plasma apolipoprotein (A) levels have also been reported in cancer patients compared with hospitalised control patients and normal healthy blood [Wright et al., 1989]. Furthermore, an APOE genetic polymorphism has been shown to modify the risk for a variety of diseases, including breast cancer [Moore et al., 2006; Zhang et al., 2004; Kuesel et al., 1992]. APOA1 is the major protein constituent of the high-density lipoprotein (HDL). Previous reports suggest aberrant serum levels of APOA1 might might be a useful marker of ovarian cancer [Moore et al., 2006; Zhang et al., 2004; Kuesel et al., 1992]. The results demonstrated that APOA1 was down-regulated in the malignant condition in two different spots in the ProteoMiner experiment, APOA4 was up-regulated in the malignant condition in 3 out of 4 spots and down-regulated in the remaining spot in the ProteoMiner experiment and up-regulated in 2 spots in the malignant condition in the unfractionated and MARS 2 experiments. APOE was up-regulated in the malignant condition in 3 out of 4 spots and down-regulated in the remaining spot in the ProteoMiner experiment.

In addition, vitamin D-binding protein (GC) and utrophin (UTRN) were both up-regulated in the malignant condition and have both been previously reported as 'markers' for cancer. GC is a secreted multifunctional protein found in plasma, ascitic fluid, cerebrospinal fluid and urine, and on the surface of many cell types. In plasma, it carries the vitamin D sterols and prevents polymerization of actin by binding its monomers. It associates with membrane-bound immunoglobulin on the surface of B-lymphocytes and with IgG Fc receptor on the membranes of T-lymphocytes. Studies suggest that the exploitation of the unique properties of vitamin D-binding protein could enable the development of important therapeutic agents for the treatment of a variety of diseases [Gomme and Bertolini, 2004]. UTRN may play a role in anchoring the cytoskeleton to the plasma membrane. In normal muscle cells, utrophin is located at the neuromuscular synapse and myotendinous junctions. It is necessary for normal membrane maintenance, and for the clustering of the acetylcholine receptor. Utrophin is known to be intracellular, thus its detection in serum indicates so-called tissue leakage [Souchelnytskyi et al., 2006].

Several cellular proteins were found in spots which yielded multiple hits. These included adenomatosis polyposis coli 2 (APC2), which was identified in a spot with APOE, promotes rapid degradation of cadherin-associated protein and may function as a tumour suppressor and may function in Wnt signalling. Kinectin 1 isoform b (KTN1) which was identified in a spot with APOA4 has been identified as a tumor-associated antigen. The receptor for KTN1 kinesin is involved in kinesin-driven vesicle motility and accumulates in integrin-based adhesion complexes (IAC) upon integrin aggregation by fibronectin. The subcellular locations of KTN1 include the endoplasmic reticulum membrane. High levels of KTN1 are found in peripheral blood lympocytes, testis and ovary, lower levels in spleen, thymus, prostate, small intestine and colon. Defects in KTN1 may be involved in the onset of cancer [Wang et al., 2004b]. Periplakin (PPL) which was identified in a spot with CFHR is a component of the cornified envelope of keratinocytes. PPL may link the cornified envelope to desmosomes, intermediate filaments and may act as a localisation signal in PKB/AKT-mediated cell signalling pathway. Finally, Pericentrin (PCNT) which was also found with APOA4 may be involved in organisation of microtubules during meiosis and mitosis. Again these proteins were predominately found in spots which yielded multiple protein identities and it is therefore difficult to attribute any specific changes to them. It is important to note that only three proteins including SERPINA1, APOA4 and IGHG1 were found to be commonly differentially expressed in the unfractionated, MARS2 and ProteoMiner fractionated samples.

In summary, results have shown that a panel of putative markers consisting of several apolipoproteins (APOA1, A4 & E), serotransferrin (TF), haptoglobin (HP), α-1-antitrypsin precursor isoform 1 (SERPINA1), vitamin D-binding protein (GC), afamin, utrophin and the celluar proteins (APC, Tbx3 and DAPLE) involved in Wnt signalling warrant further investigation. Future experiments may involve a large scale immuno-based study in an independent cohort of UKOPS case control samples which is beyond the time frame of this project. This would be the ideal way to evaluate the diagnostic usefulness of these proteins, in combination with the CA-125 assay, for early detection of ovarian cancer.

Chapter 7: Conclusions and future directions

The primary hypothesis driving the research presented in this thesis is that the human serum proteome contains a source of proteins whose abundances change with disease state. It is proposed that the identified protein abundance changes between the diseased and healthy state could act as early biomarkers of ovarian cancer giving clinicians the opportunity for earlier intervention. The search for putative markers of ovarian cancer was investigated using polypeptide separation methods and mass spectrometry. As previously detailed several collections of clinically relevant serum samples were available for this investigation. However, due to the low incidence of ovarian cancer the number of diseased samples was limited.

The results achieved are presented in four chapters. Chapter three utilised a previously established magnetic bead-based peptide extraction protocol and MALDI-TOF MS profiling which was then used to analyse case-control samples. Chapter four describes the various steps taken to adapt and optimise this protocol in the host laboratory. Chapter five describes the analysis of case-control samples in the host laboratory using the optimised protocols. Finally, chapter six is concerned with the use of additional protein separation techniques designed to increase the dynamic range of detection of the serum proteome for 2D-DIGE-MS profiling.

In Chapter three, one of the interesting findings from the case-control study where samples pre-dating diagnosis of ovarian cancer and control groups were compared using MS analysis was that peaks 4292.5 m/z and 3171.1 m/z could be used in combination with CA-125 levels to detect ovarian cancer up to 12 months prior to diagnosis. Although these peaks were not identified, it could be speculated that they may be fragments of host immune/acute phase proteins (e.g. inter-α-trypsin inhibitor heavy chain (ITIH4)) generated by exo-peptidases as previously reported by Villanueva et al. (2005). Several groups have reported on the use of a combination of host immune/acute phase proteins as potential disease markers [Chen et al., 2005]. It is clear that the presumption of the existence of a single cancer-specific biomarker is outdated and investigations concentrating of finding panels of markers have the potential to produce greater sensitivity and specificity. The fact that cancer cells are themselves ‘deranged’ host cells, tends to suggest that the existence of single and specific cancer biomarkers is improbable. In contrast, the complex proteomic pattern of the tumour–host microenvironment may be unique and may constitute a biomarker amplification cascade. In fact, the most important biomarkers may be normal host response proteins that are aberrantly cleaved through cancer-specific protease activities. Such a proposal lends itself to a pattern analysis approach to investigate the loss or gain of peptide ions within the spectra of disease versus normal samples.

Having identified protocols suited to the high-throughput analysis of serum-based spectral patterns, the overall aim of the experiments presented in chapter 4 was to identify sources of error which may affect the reproducibility of the selected technology platform. Although automation allowed the processing and analysis of a greater number of samples the initial reproducibility of the platform was poor. The reproducibility was improved by reducing the bead loss during the wash steps and diluting the eluate before MS analysis. Interestingly, the beads chosen for general use were found to bind highly abundant proteins including HSA and apolipoproteins which are common serum proteins. It is speculated that the presence of these proteins led to competition and suppression effects during peptide extraction and MALDI ionisation. Diluting the eluate led to improved spectral quality, presumably by improving crystallisation and reducing peptide ion suppression but did not facilitate peptide identification. Subsequent studies showed that despite attempts to deplete the abundant proteins they continued to mask the detection of lower abundant protein species.

In addition, different clinically feasible sample handling and processing protocols were assessed. An interesting finding was that samples kept on ice after collection with a transport time of less than 6 hours were more stable (as defined by serum peptide inter sample profile comparison) than those which had not been kept on ice prior to transport. This provides support for the exopeptidase activity hypothesis [Villanueva et al., 2006]. This states that disease-specific exopeptidases are active on serum proteins after blood samples have been left to clot with the fragmentation patterns generated by these exoprotease activities providing diagnostic information. Fragments of fibrinopeptide A (FPA) and complement 3f (C3f), among others have been identified and reported as surrogate markers of cancer [Villanueva et al., 2006].

More importantly, several groups have also reported on differential expression of abundant serum proteins (or fragments thereof) as markers of disease [Villanueva et al., 2006; Ye et al., 2003; Zhang et al., 2004].

There was also support for the use of older sample collections where samples have been in transit and storage for prolonged periods of time. Case-control studies can be performed as long as all samples have been handled identically. Exopeptidase activity may come to an end point after a certain period depending on the transit and storage conditions. Placing samples on ice would serve to lower the rate of exopeptidase activity and thus, sample integrity remains intact for MS-based analysis. The peptides are themselves the product of specific enzymatic activities, and careful qualitative and quantitative measurements may therefore yield some insights in the protease activities at work. However, steady state measurements can only provide some quantitative estimates of enzyme activity when contributing factors such as specific activity and half-life of the product or metabolite are known. Proteases are well-established components of tumour progression and invasiveness [Matrisian et al., 2003]. As such, enzymes, inactive proteolytic fragments of enzymes or protease inhibitors have become important and promising cancer biomarkers [Landis-Piwowar et al., 2006]. Some efforts are being made towards activity-based proteomic profiling involving the use of chemical probes that selectively label, on a whole-proteome background, certain classes of active enzymes but not their inactive forms [Kato et al., 2005]. Furthermore, a test to compare defined exopeptidase activities by quantifying the peptide products of such enzymes within individual proteomes of two or more groups of biological samples has recently been reported [Villanueva et al., 2008].

The experiments described in chapter 5 utilised the optimised magnetic bead-based serum profiling platform and a panel of peptide masses were found to be useful in (poorly) discriminating between the clinical conditions. However despite defining a cut-off for the technical variance, the biological heterogeneity innate in human-derived clinical samples resulted in large variation within the clinical conditions. The SVM classification algorithm was also limited by the small number of diseased samples. A larger sample set needs to be analysed to validate these results. It would have been ideal to identify some of the discriminatory masses. However, as previously mentioned, direct peptide identification using MALDI-TOF/TOF MS suffers from several limitations. This is especially true for a complex sample such as serum where in the presence of numerous proteins the separation of the parent ion of interest is often difficult. As the data in chapter 4 showed the beads used for peptide extraction also captured larger proteins making the sample eluate complex and generating crystallisation heterogeneity and ion suppression. Although the peaks of interest were not identified it is most likely that the distinctive MS profiles have resulted from the differential expression of relatively abundant serum proteins and their fragments associated with (the response to) tumours, which may have been further cleaved by disease specific ex-vivo exoprotease activity [Villanueva et al., 2008]. Work on isolating and identifying peaks of interest is ongoing. A larger sample set from UKOPS has been analysed using the automated bead-based fractionation protocol in the host laboratory and data is being analysed in collaboration with Professor Gammerman's group at Royal Holloway University. On their own the peptide peaks identified performed poorly when used to classify a test set of samples, but sensitivity is improved when used in combination with CA-125, data analysis is on-going.

The next step would involve the validation of these markers. Validation of biomarkers would require robust diagnostic performance on independent case-control sets. To this end, the platform developed herein could be expanded to incorporate an antibody-based enrichment strategy for capturing peptides on magnetic beads coated with an array of antibodies for several proteins coupled to mass spectrometry-based multiple reaction monitoring (MRM) quantification. Since the majority of markers reported from proteomic studies to date have predominately included host-response (acute-phase) proteins, this would be an ideal way to quantitatively analyse peptides associated with the host immune response [Whiteaker et al., 2007].

The experiments described in chapter six permitted the identification of several proteins that were differentially expressed between pools of clinical sera. In the unfractionated serum experiment, several host response/acute-phase proteins including the immunoglobins and proteins from the SERPIN family were identified as differentially expressed protein features. Of these the SERPIN proteins which are commonly known as positive acute-phase reactants were up-regulated in the malignant condition. SERPIN proteins are involved in down regulation of local inflammation. The inflammatory cascade includes a multitude of constituent proteins with varying functions, including structural proteins, clotting factors, angiogenesis and transport proteins. The host response is generally mediated by the innate immune system and the host response proteins exist at substantially high concentrations within the tumour microenvironment. According to the host-response protein amplification cascade concept, proteins synthesized in the liver enter the circulation and when exposed to a localized disease area, they are processed by the local host response, and modified forms of the proteins re-enter the general circulation. This is the source of amplification of a localised disease signal. Specificity of this process is made possible by the fact that each disease generates a different type of local host response. This may be due to the fact that each disease expresses a different set of antigens (e.g. tumour markers) or that the recruitment of specific inflammatory mediators differs based on the inciting event. However, it is likely that these markers may not be specific for ovarian cancer.

Using an antibody-based enrichment strategy for quantitative measurements of such host response proteins and the peptides associated with them may provide detailed insights on tumour progression and invasiveness. Numerous proteomic studies published to date have identified relatively abundant host response proteins as candidate biomarkers [Moore et al., 2006; Zhang et al., 2004; Zhang et al., 2007]. It would be interesting to measure how their levels fluctuate with tumour progression and to determine if the changes are disease and/or stage specific in samples pre-dating diagnosis.

The systematic detection of low abundance proteins in human blood is complicated by the extremely wide dynamic range of protein abundances. The depletion of major proteins is a popular strategy for enhancing detection sensitivity in serum or plasma. The low abundance proteins enrichment strategies employed herein enabled additional protein species to be detected on 2-DE, although the increase was modest, and most newly visualized spots were minor forms of relatively abundant proteins. The inability to detect low abundance proteins near expected 2-D staining limits was probably due to both the highly heterogeneous nature of most serum proteins and masking of low abundance proteins by the next series of medium abundance proteins, which included the apolipoproteins. Due to the wide dynamic range of the serum proteome, even removing the top 20 most abundant proteins would probably not allow the detection of the lowest abundant proteins. In reality the removal of the top seven proteins enhances mid-range rather than low-range proteins and additional fractionation is required.

Not surprisingly, in both the MARS and ProteoMiner experiments, a number of gel spots yielded identities for the apolipoproteins (APOA1, APOA4 & APOE). Of these APOA4 was also identified in the unfractionated serum experiment. The precise function of APOA4 is unknown. Interestingly, the ProteoMiner protein enrichment kit showed a greater affinity for these ‘medium’ abundant proteins. APOA1 and APOE were only found in the ProteoMiner experiment and both proteins were found to be down-regulated in the malignant condition. APOE is synthesized principally in the liver and is the lipoprotein component of very low-density lipoproteins (VLDLs). APOE combines with lipids and is involved in cholesterol transport, lipid metabolism and protein synthesis. APOE is expressed in significant amounts in the ovaries, testes and the brain, is synthesised by a wide variety of peripheral cells, including macrophages. Its production by extra-hepatic cells has also raised questions regarding its role in peripheral tissues. In addition, APOE is involved in numerous other functions, including tissue repair, the immune response and regulation of cell growth and differentiation. The levels of these proteins in diseased samples may be related to tumour size and future follow up investigations should be able to confirm this as long as the relevant clinical data is available. In summary, the results from chapter 6 showed that a panel of putative markers consisting of several apolipoproteins (APOA1, A4 & E), serotransferrin (TF), haptoglobin (HP), α-1-antitrypsin precursor isoform 1 (SERPINA1), vitamin D-binding protein, afamin and utrophin warrant further investigation. As Table 7.1 shows many of these proteins have previously been reported as markers for ovarian and other cancers.

Biomarker	Cancer type	References
Apolipoprotein A1	Ovarian, pancreatic	Zhang et al., 2004; Kozak et al., 2005
Haptoglobin α-subunit	Ovarian, pancreatic, lung	Ye et al., 2003
Transthyretin fragment	Ovarian	Kozak et al., 2005
Inter-alpha-trypsin inhibitor fragment	Ovarian, pancreatic	Zhang et al., 2004
Vitamin D-binding protein	Prostate, breast	Corder et al.,1993; Pawlik et al., 2006
Serum amyloid A	Nasopharyngeal, pancreatic, ovarian	Orchekowski et al., 2005; Moshkovskii et al.,2005; Helleman et al., 2007
α1-antitrypsin and α1-antichymotrypsin	Pancreatic	Orchekowski et al., 2005: Yu et al., 2005
Haemoglobin-alpha & -beta subunits	Ovarian	Woong-Shick et al., 2005
EPCA-2	Prostate	Leman et al,. 2007
Afamin	Ovarian	Jackson et al., 2007

Table 7.1 Examples of putative serum biomarkers. Adapted from aoui-Jamali and Xu, (2006).

In addition, several cellular proteins involved in Wnt-signalling were found in spots which yielded multiple hits. Wnts are a large family of cysteine-rich secreted glycoproteins that control development and several components of Wnt signalling are implicated in the genesis of human cancer. These included adenomatosis polyposis coli 2 (APC2; identified in a spot with APOE), which promotes rapid degradation of cadherin-associated protein and may function as a tumour suppressor. Protein Daple Isoform 2 (DAPLE) is a negative regulator of the canonical Wnt signalling pathway, acting downstream of Dvl to inhibit β-catenin stabilisation. In addition, Kinectin 1 isoform b (KTN1; identified in a spot with APOA4) has been identified as a tumor-associated antigen. Kinesin, the receptor for KTN1 is involved in kinesin-driven vesicle motility and accumulates in integrin-based adhesion complexes upon integrin aggregation by fibronectin. Previous reports have suggested that defects in KTN1 may be involved in the onset of cancer [Wang et al., 2004].

In conclusion, the work presented in this thesis has revealed several candidate markers that need to be tested on a larger sample set in combination with CA-125 using immuno-based assays. The proteome of an organism, in this case human, is defined as the complete set of proteins that can be expressed by the genetic material. Based on the number of proteins found in the extensive separation and analysis shown by this work, the number of disease specific-biomarkers may well be lower than the postulated number in the literature.

To further improve the chances of finding serum markers for ovarian cancer, a number of parameters could be improved. The proteome could be divided into smaller ‘sub-proteomes’. For example, using fractionation techniques to separate the glycoproteins and phosphoproteins or 2D-LC separation of tryptic peptides would facilitate more detailed analysis of the serum proteome. In addition, quantitative analysis using stable isotope labelling techniques such as isotope-coded affinity tags (ICAT) or isobaric tags for relative and absolute quantitation (iTRAQ) could be performed on these fractions. The use of such sub-proteome fractionation techniques would permit a simplification of the proteome while providing practical steps towards the ultimate dissection of the entire proteome. In essence, these methods are capable of capturing fairly accurately the relative quantitative information from two or more samples in a single analysis, thereby reducing the analysis time and the effect of technical variability. The MARS or ProteoMiner fractionation strategies could also be extended to an LC-MALDI label-free quantitation platform. Additional analysis of the bound fraction from the MARS depletion studies could also yield useful data. As previous studies on the analysis of albumin-associated peptides and proteins from ovarian cancer patients have shown, low-abundance nuclear proteins linked to cancer susceptibility, including BRCA2, were represented in sera as a series of specific fragments bound to albumin [Lowenthal et al., 2005].

8. References

Adam BL, Qu Y, Davis JW, Ward MD, Clements MA, Cazares LH, Semmes OJ, Schellhammer PF, Yasui Y, Feng Z, Wright GL, Jr. (2002) Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res.* **62**: 3609-3614

Anderson NL, Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell Proteomics* **1**: 845-867

Anderson NL, Polanski M, Pieper R, Gatlin T, Tirumalai RS, Conrads TP, Veenstra TD, Adkins JN, Pounds JG, Fagan R, Lobley A (2004) The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol. Cell Proteomics* **3**: 311-326

aoui-Jamali MA, Xu YJ (2006) Proteomic technology for biomarker profiling in cancer: an update. *J. Zhejiang Univ. Sci. B.* **7**: 411-420

Ayhan A, Salman MC, Celik H, Dursun P, Ozyuncu O, Gultekin M (2004) Association between fertility drugs and gynecologic cancers, breast cancer, and childhood cancers. *Acta. Obstet. Gynecol. Scand.* **83**: 1104-1111

Baggerly KA, Morris JS, Coombes KR (2004) Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics* **20**: 777-785

Baggerly KA, Morris JS, Edmonson SR, Coombes KR (2005) Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. *J. Natl. Cancer Inst.* **97**: 307-309

Baker M (2005) In biomarkers we trust? *Nat. Biotechnol.* **23**: 297-304

Bast RC, Jr., Urban N, Shridhar V, Smith D, Zhang Z, Skates S, Lu K, Liu J, Fishman D, Mills G (2002) Early detection of ovarian cancer: promise and reality. *Cancer Treat. Res.* **107**: 61-97

Bast RC, Jr., Xu FJ, Yu YH, Barnhill S, Zhang Z, Mills GB (1998) CA 125: the past and the future. *Int. J. Biol. Markers* **13**: 179-187

Baumann S, Ceglarek U, Fiedler GM, Lembcke J, Leichtle A, Thiery J (2005) Standardized approach to proteome profiling of human serum based on magnetic bead separation and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin. Chem.* **51**: 973-980

Bell DA (2005) Origins and molecular pathology of ovarian cancer. *Mod. Pathol.* **18**: S19-S32

Bienz M, Clevers H (2000) Linking colorectal cancer to Wnt signaling. *Cell* **103**: 311-320

Bjellqvist B, Ek K, Righetti PG, Gianazza E, Gorg A, Westermeier R, Postel W (1982) Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J. Biochem. Biophys. Methods* **6**: 317-339

Bjorhall K, Miliotis T, Davidsson P (2004) Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. *Proteomics* **5**: 307-317

Bosse K, Rhiem K, Wappenschmidt B, Hellmich M, Madeja M, Ortmann M, Mallmann P, Schmutzler R (2006) Screening for ovarian cancer by transvaginal ultrasound and serum CA-125 measurement in women with a familial predisposition: a prospective cohort study. *Gynecol. Oncol.* **103**: 1077-1082

Breedlove G, Busenhart C (2005) Screening and detection of ovarian cancer. *J. Midwifery Womens Health* **50**: 51-54

Cairns J (1975) Mutation selection and the natural history of cancer. *Nature* **255**: 197-200

Chambers AF, Vanderhyden BC (2006) Ovarian cancer biomarkers in urine. *Clin. Cancer Res.* **12**: 323-327

Chan HL, Gharbi S, Gaffney PR, Cramer R, Waterfield MD, Timms JF (2005) Proteomic analysis of redox- and ErbB2-dependent changes in mammary luminal epithelial cells using cysteine- and lysine-labelling two-dimensional difference gel electrophoresis. *Proteomics* **5**: 2908-2926

Chen R, Pan S, Brentnall TA, Aebersold R (2005) Proteomic profiling of pancreatic cancer for biomarker discovery. *Mol. Cell Proteomics* **4**: 523-533

Chromy BA, Gonzales AD, Perkins J, Choi MW, Corzett MH, Chang BC, Corzett CH, Cutchen-Maloney SL (2004) Proteomic analysis of human serum by two-dimensional differential gel electrophoresis after depletion of high-abundant proteins. *J. Proteome Res.* **3**: 1120-1127

Cohen SL, Chait BT (1996) Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins. *Anal. Chem.* **68**: 31-37

Corder EH, Guess HA, Hulka BS, Friedman GD, Sadler M, Vollmer RT, Lobaugh B, Drezner MK, Vogelman JH, Orentreich N (1993) Vitamin D and prostate cancer: a prediagnostic study with stored sera. *Cancer Epidemiol. Biomarkers Prev.* **2**: 467-72

Costanzo ES, Lutgendorf SK, Sood AK, Anderson B, Sorosky J, Lubaroff DM (2005) Psychosocial factors and interleukin-6 among women with advanced ovarian cancer. *Cancer* **104**: 305-313

de Noo ME, Tollenaar RA, Ozalp A, Kuppen PJ, Bladergroen MR, Eilers PH, Deelder AM (2005) Reliability of human serum protein profiles generated with C8 magnetic beads assisted MALDI-TOF mass spectrometry. *Anal. Chem.* **77**: 7232-7241

Drake SK, Bowen RA, Remaley AT, Hortin GL (2004) Potential interferences from blood collection tubes in mass spectrometric analyses of serum polypeptides. *Clin. Chem.* **50**: 2398-2401

Dreisewerd K (2003) The desorption process in MALDI. *Chem. Rev.* **103**: 395-426

Dumaswala UJ, Wilson MJ, Wu YL, Wykle J, Zhuo L, Douglass LM, Daleke DL (2000) Glutathione loading prevents free radical injury in red blood cells after storage. *Free Radic. Res.* **33**: 517-529

Echan LA, Tang HY, li-Khan N, Lee K, Speicher DW (2005) Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. *Proteomics* **5**: 3292-3303

Fallows S, Price J, Atkinson RJ, Johnston PG, Hickey I, Russell SE (2001) P53 mutation does not affect prognosis in ovarian epithelial malignancies. *J. Pathol.* **194**: 68-75

Fung ET, Yip TT, Lomas L, Wang Z, Yip C, Meng XY, Lin S, Zhang F, Zhang Z, Chan DW, Weinberger SR (2005) Classification of cancer types by measuring variants of host response proteins using SELDI serum assays. *Int. J. Cancer* **115**: 783-789

Fusaro VA, Stone JH (2003) Mass spectrometry-based proteomics and analyses of serum: a primer for the clinical investigator. *Clin. Exp. Rheumatol.* **21**: S3-14

Gabay C, Kushner I (1999) Acute-phase proteins and other systemic responses to inflammation. *N. Engl. J. Med.* **340**: 448-454

Gammerman A, Vovk V, Burford B, Nouretdinov I, Luo Z, Chervonenkis A, Waterfield M, Cramer R, Tempst P, Villanueva J, Kabir M, Camuzeaux S, Timms J, Menon U and Jacobs I (2008). Serum proteomic abnormality predating screen detection of ovarian cancer. *The Computer Journal*: **10**: 1093

Gercel-Taylor C, Bazzett LB, Taylor DD (2001) Presence of aberrant tumor-reactive immunoglobulins in the circulation of patients with ovarian cancer. *Gynecol. Oncol.* **81**: 71-76

Gharbi S, Gaffney P, Yang A, Zvelebil MJ, Cramer R, Waterfield MD, Timms JF (2002) Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. *Mol. Cell Proteomics* **1**: 91-98

Gomme PT, Bertolini J (2004) Therapeutic potential of vitamin D-binding protein. *Trends Biotechnol.* **22**: 340-345

Granger J, Siddiqui J, Copeland S, Remick D (2005) Albumin depletion of human plasma also removes low abundance proteins including the cytokines. *Proteomics* **5**: 4713-4718

Grizzi F, Chiriva-Internati M (2006) Cancer: looking for simplicity and finding complexity. *Cancer Cell Int.* **6**: 4

Grizzi F, Di IA, Russo C, Frezza EE, Cobos E, Muzzio PC, Chiriva-Internati M (2006) Cancer initiation and progression: an unsimplifiable complexity. *Theor. Biol. Med. Model.* **3**: 37

Guerrier L, Righetti PG, Boschetti E (2008). Reduction of dynamic protein concentration range of biological extracts for the discovery of low-abundance proteins by means of hexapeptide ligand library. *Nat. Protoc.* **3**: 883-90

Guerrier L, Thulasiraman V, Castagna A, Fortis F, Lin S, Lomas L, Righetti PG, Boschetti E (2006) Reducing protein concentration range of biological samples using solid-phase ligand libraries. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* **20;833**: 33-40

Hanahan D, Weinberg RA (2000) The hallmarks of cancer. *Cell* **100**: 57-70

Hoffmann E de, Stroobant V. Mass Spectrometry: Principles and Applications. 2nd Edn. Chichester: *Wiley* (2002) **1**: 28-32

Holschneider CH, Berek JS (2000) Ovarian cancer: epidemiology, biology, and prognostic factors. *Semin. Surg. Oncol.* **19**: 3-10

Huang HL, Stasyk T, Morandell S, Mogg M, Schreiber M, Feuerstein I, Huck CW, Stecher G, Bonn GK, Huber LA (2005) Enrichment of low-abundant serum proteins by albumin/immunoglobulin G immunoaffinity depletion under partly denaturing conditions. *Electrophoresis* **26**: 2843-2849

Jackson D, Craven RA, Hutson RC, Graze I, Lueth P, Tonge RP, Hartley JL, Nickson JA, Rayner SJ, Johnston C, Dieplinger B, Hubalek M, Wilkinson N, Perren TJ, Kehoe S, Hall GD, Daxenbichler G, Dieplinger H, Selby PJ, Banks RE (2007) Proteomic profiling identifies afamin as a potential biomarker for ovarian cancer. *Clin. Cancer Res.* **13**: 7370-7379

Jacobs IJ, Skates SJ, MacDonald N, Menon U, Rosenthal AN, Davies AP, Woolas R, Jeyarajah AR, Sibley K, Lowe DG, Oram DH (1999) Screening for ovarian cancer: a pilot randomised controlled trial. *Lancet* **353**: 1207-1210

Karas M, Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **60**: 2299-2301

Karp NA, McCormick PS, Russell MR, Lilley KS (2007) Experimental and statistical considerations to avoid false conclusions in proteomics studies using differential in-gel electrophoresis. *Mol. Cell Proteomics* **6**: 1354-1364

Kato D, Boatright KM, Berger AB, Nazif T, Blum G, Ryan C, Chehade KA, Salvesen GS, Bogyo M (2005) Activity-based probes that target diverse cysteine protease families. *Nat. Chem. Biol.* **1**: 33-38

Kim MR, Kim CW (2007) Human blood plasma preparation for two-dimensional gel electrophoresis. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* **849**: 203-210

King MC, Marks JH, Mandell JB (2003) Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science* **302**: 643-646

Kloth J, Gorter A, Fleuren G, Oosting J, Uljee S, Haar NT, Dreef E, Kenter G, Jordanova E (2008) Elevated expression of SerpinA1 and SerpinA3 in HLA-positive cervical carcinoma. *J. Pathol.* **215**: 222-30

Koehn H, Oehler MK (2007) Proteins' promise--progress and challenges in ovarian cancer proteomics. *Menopause Int.* **13**: 148-153

Kozak KR, Amneus MW, Pusey SM, Su F, Luong MN, Luong SA, Reddy ST, Farias-Eisner R (2003). Identification of biomarkers for ovarian cancer using strong anion-exchange proteinChips: Potential use in diagnosis and prognosis. *PNAS*. **100**: 12343–12348

Kuesel AC, Kroft T, Prefontaine M, Smith IC (1992) Lipoprotein(a) and CA-125 levels in the plasma of patients with benign and malignant ovarian disease. *Int. J. Cancer* **52**: 341-346

Landis-Piwowar KR, Milacic V, Chen D, Yang H, Zhao Y, Chan TH, Yan B, Dou QP (2006) The proteasome as a potential target for novel anticancer drugs and chemosensitizers. *Drug Resist. Update* **9**: 263-273

Li J, Zhang Z, Rosenzweig J, Wang YY, Chan DW (2002) Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin. Chem.* **48**: 1296-1304

Liang SL, Chan DW (2007) Enzymes and related proteins as cancer biomarkers: a proteomic approach. *Clin. Chim. Acta.* **381**: 93-97

Lin D, Tabb DL, Yates JR, III (2003) Large-scale protein identification using mass spectrometry 1. *Biochim. Biophys. Acta.* **1646**: 1-10

Liotta LA, Petricoin EF (2006) Serum peptidome for cancer detection: spinning biologic trash into diagnostic gold. *J. Clin. Invest.* **116**: 26-30

Liu T, Qian WJ, Mottaz HM, Gritsenko MA, Norbeck AD, Moore RJ, Purvine SO, Camp DG, Smith RD (2006) Evaluation of multiprotein immunoaffinity subtraction for plasma proteomics and candidate biomarker discovery using mass spectrometry. *Mol. Cell Proteomics* **5**: 2167-2174

Lowenthal MS, Mehta AI, Frogale K, Bandle RW, Araujo RP, Hood BL, Veenstra TD, Conrads TP, Goldsmith P, Fishman D, Petricoin EF, III, Liotta LA (2005) Analysis of albumin-associated peptides and proteins from ovarian cancer patients. *Clin. Chem.* **51**: 1933-1945

Mann M, Hojrup P, Roepstorff P (1993) Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol. Mass Spectrom.* **22**: 338-345

Mann M, Wilm M (1995) Electrospray mass spectrometry for protein characterization. *Trends Biochem. Sci.* **20**: 219-224

Marquez RT, Baggerly KA, Patterson AP, Liu J, Broaddus R, Frumovitz M, Atkinson EN, Smith DI, Hartmann L, Fishman D, Berchuck A, Whitaker R, Gershenson DM, Mills GB, Bast RC, Jr., Lu KH (2005) Patterns of gene expression in different histotypes of epithelial ovarian cancer correlate with those in normal fallopian tube, endometrium, and colon. *Clin. Cancer Res.* **11**: 6116-6126

Martorella A, Robbins R (2007) Serum peptide profiling: identifying novel cancer biomarkers for early disease detection. *Acta. Biomed.* **78 Suppl 1**: 123-128

Matrisian LM, Sledge GW, Jr., Mohla S (2003) Extracellular proteolysis and cancer: meeting summary and future directions. *Cancer Res.* **63**: 6105-6109

Molloy MP, Brzezinski EE, Hang J, McDowell MT, VanBogelen RA (2003) Overcoming technical variation and biological variation in quantitative proteomics. *Proteomics* **3**: 1912-1919

Moore LE, Fung ET, McGuire M, Rabkin CC, Molinaro A, Wang Z, Zhang F, Wang J, Yip C, Meng XY, Pfeiffer RM (2006) Evaluation of apolipoprotein A1 and posttranslationally modified forms of transthyretin as biomarkers for ovarian cancer detection in an independent study population. *Cancer Epidemiol. Biomarkers Prev.* **15**: 1641-1646

Moorman PG, Schildkraut JM, Calingaert B, Halabi S, Berchuck A (2005) Menopausal hormones and risk of ovarian cancer. *Am J Obstet Gynecol.* **193**: 76-82

Mor G, Visintin I, Lai Y, Zhao H, Schwartz P, Rutherford T, Yue L, Bray-Ward P, Ward DC (2005) Serum protein markers for early detection of ovarian cancer. *Proc. Natl. Acad. Sci. U S A.* **102**: 7677-7682

Morita K, Saito T, Ohta M, Ohmori T, Kawai K, Teshima-Kondo S, Rokutan K (2005) Expression analysis of psychological stress-associated genes in peripheral blood leukocytes. *Neurosci. Lett.* **381**: 57-62

Mujumdar RB, Ernst LA, Mujumdar SR, Lewis CJ, Waggoner AS (1993) Cyanine dye labeling reagents: sulfoindocyanine succinimidyl esters. *Bioconjug. Chem.* **4**: 105-111

Mujumdar RB, Ernst LA, Mujumdar SR, Waggoner AS (1989) Cyanine dye labeling reagents containing isothiocyanate groups. *Cytometry.* **10**: 11-19

Ness RB, Grisso JA, Klapper J, Vergona R (2000) Racial differences in ovarian cancer risk. *J. Natl. Med. Assoc.* **92**: 176-182

Neuhoff V, Arold N, Taube D, Ehrhardt W (1988) Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **9**: 255-262

Nossov V, Amneus M, Su F, Lang J, Janco JM, Reddy ST, Farias-Eisner R (2008) The early detection of ovarian cancer: from traditional methods to proteomics. Can we really do better than serum CA-125? *Am. J. Obstet. Gynecol.* **199**: 215-223

O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol. Chem.* **250**: 4007-4021

Orchekowski R, Hamelinck D, Li L, VanBrocklin M, Marrero J, VandeWoude G, Feng Z, Brand RE, Haab BB (2005). Serum-protein alterations in pancreatic cancer patients and their use for disease classification. *Pancreas.* **31**:460.

Pan S, Zhang H, Rush J, Eng J, Zhang N, Patterson D, Comb MJ, Aebersold R (2005) High throughput proteome screening for biomarker detection. *Mol. Cell. Proteomics* **4**: 182-190

Pappin DJ, Hojrup P, Bleasby AJ (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **3**: 327-332

Patton WF (2000) A thousand points of light: the application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics. *Electrophoresis* **21**: 1123-1144

Pawlik TM, Hawke DH, Liu Y, Krishnamurthy S, Fritsche H, Hunt KK, Kuerer HM (2006). Proteomic analysis of nipple aspirate fluid from women with early-stage breast cancer using isotope-coded affinity tags and tandem mass spectrometry reveals differential expression of vitamin D binding protein. *BMC Cancer*. 6: 68.

Pearl DC (2002) Proteomic patterns in serum and identification of ovarian cancer. *Lancet* **360**: 169-170

Petricoin EE, Paweletz CP, Liotta LA (2002) Clinical applications of proteomics: proteomic pattern diagnostics. *J. Mammary Gland. Biol. Neoplasia.* **7**: 433-440

Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC, Liotta LA (2002) Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **359**: 572-577

Petricoin EF, Liotta LA (2002) Proteomic analysis at the bedside: early detection of cancer. *Trends Biotechnol.* **20**: S30-S34

Petricoin EF, Liotta LA (2004) Clinical proteomics: application at the bedside. *Contrib. Nephrol.* **141**: 93-103

Petricoin EF, III, Ornstein DK, Paweletz CP, Ardekani A, Hackett PS, Hitt BA, Velassco A, Trucco C, Wiegand L, Wood K, Simone CB, Levine PJ, Linehan WM, Emmert-Buck MR, Steinberg SM, Kohn EC, Liotta LA (2002) Serum proteomic patterns for detection of prostate cancer. *J. Natl. Cancer Inst.* **94**: 1576-1578

Pineiro M, Alava MA, Gonzalez-Ramon N, Osada J, Lasierra P, Larrad L, Pineiro A, Lampreave F (1999) ITIH4 serum concentration increases during acute-phase processes in human patients and is up-regulated by interleukin-6 in hepatocarcinoma HepG2 cells. *Biochem. Biophys. Res. Commun.* **263**: 224-229

Pitteri SJ, Hanash SM (2007) Proteomic approaches for cancer biomarker discovery in plasma. *Expert Rev. Proteomics* **4**: 589-590

Purdie DM, Bain CJ, Siskind V, Webb PM, Green AC (2003) Ovulation and risk of epithelial ovarian cancer. *Int. J. Cancer.* **104**: 228-232

Qu Y, Adam BL, Yasui Y, Ward MD, Cazares LH, Schellhammer PF, Feng Z, Semmes OJ, Wright GL, Jr. (2002) Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from noncancer patients. *Clin. Chem.* **48**: 1835-1843

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

Rai AJ, Gelfand CA, Haywood BC, Warunek DJ, Yi J, Schuchard MD, Mehigh RJ, Cockrill SL, Scott GB, Tammen H, Schulz-Knappe P, Speicher DW, Vitzthum F, Haab BB, Siest G, Chan DW (2005) HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics* **5**: 3262-3277

Ramus SJ, Harrington PA, Pye C, DiCioccio RA, Cox MJ, Garlinghouse-Jones K, Oakley-Girvan I, Jacobs IJ, Hardy RM, Whittemore AS, Ponder BA, Piver MS, Pharoah PD, Gayther SA (2007) Contribution of BRCA1 and BRCA2 mutations to inherited ovarian cancer. *Hum. Mutat.* **28**: 1207-1215

Ransohoff DF (2005) Lessons from controversy: ovarian cancer screening and serum proteomics. *J. Natl. Cancer Inst.* **97**: 315-319

Righetti PG (1989) Isoelectric focusing in immobilized pH gradients in studies of biochemical polymorphism. *Anim. Genet.* **20**: 343-345

Righetti PG, Gianazza E, Gelfi C (1988) Immobilized pH gradients. *Trends Biochem. Sci.* **13**: 335-338

Robbins RJ, Villanueva J, Tempst P (2005) Distilling cancer biomarkers from the serum peptidome: high technology reading of tea leaves or an insight to clinical systems biology? *J. Clin. Oncol.* **23**: 4835-4837

Rodland KD (2004) Mass spectrometry and biomarker development. *Dis. Markers* **20**: 129-130

Roepstorff P, Fohlman J (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **11**: 601

Rosenthal AN, Menon U, Jacobs IJ (2006) Screening for ovarian cancer. *Clin. Obstet. Gynecol.* **49**: 433-447

Roy R, Wewer UM, Zurakowski D, Pories SE, Moses MA (2004) ADAM 12 cleaves extracellular matrix proteins and correlates with cancer status and stage. *J. Biol. Chem.* **279**: 51323-51330

Scheele GA (1975) Two-dimensional gel analysis of soluble proteins. Charaterization of guinea pig exocrine pancreatic proteins. *J. Biol. Chem.* **250**: 5375-5385

Schulenberg B, Arnold B, Patton WF (2003) An improved mechanically durable electrophoresis gel matrix that is fully compatible with fluorescence-based protein detection technologies. *Proteomics* **3**: 1196-1205

Seidensticker MJ, Behrens J (2000) Biochemical interactions in the wnt pathway. *Biochim. Biophys. Acta.* **1495**: 168-182

Semmes OJ, Feng Z, Adam BL, Banez LL, Bigbee WL, Campos D, Cazares LH, Chan DW, Grizzle WE, Izbicka E, Kagan J, Malik G, McLerran D, Moul JW, Partin A, Prasanna P, Rosenzweig J, Sokoll LJ, Srivastava S, Srivastava S, Thompson I, Welsh MJ, White N, Winget M, Yasui Y, Zhang Z, Zhu L (2005) Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. *Clin. Chem.* **51**: 102-112

Song K, Hanash S (2006) Unraveling the complex proteome for biomarker discovery in gastrointestinal and liver diseases. *Gastroenterology* **131**: 1375-1378

Souchelnytskyi S, Lomnytska M, Dubrovska A, Hellman U, Volodko N (2006) Proteomics success story. Towards early detection of breast and ovarian cancer: plasma proteomics as a tool to find novel markers. *Proteomics* **6 Suppl 2**: 65-68

Srinivas PR, Verma M, Zhao Y, Srivastava S (2002) Proteomics for cancer biomarker discovery. *Clin. Chem.* **48**: 1160-1169

Sriyam S, Sinchaikul S, Tantipaiboonwong P, Tzao C, Phutrakul S, Chen ST (2007) Enhanced detectability in proteome studies. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* **849**: 91-104

Steck B, Grether A, Amsler F, Dillier AS, Romer G, Kappos L, Burgin D (2007) Disease variables and depression affecting the process of coping in families with a somatically ill parent. *Psychopathology* **40**: 394-404

Steinberg TH, Agnew BJ, Gee KR, Leung WY, Goodman T, Schulenberg B, Hendrickson J, Beechem JM, Haugland RP, Patton WF (2003) Global quantitative phosphoprotein analysis using Multiplexed Proteomics technology. *Proteomics* **3**: 1128-1144

Stief TW (2008) The fibrinogen functional turbidimetric assay. *Clin. Appl. Thromb. Hemost.* **14**: 84-96

Tantipaiboonwong P, Sinchaikul S, Sriyam S, Phutrakul S, Chen ST (2005) Different techniques for urinary protein analysis of normal and lung cancer patients. *Proteomics* **5**: 1140-1149

Tatra G (1985) [Acute-phase markers in gynecologic tumors]. *Strahlentherapie* **161**: 487-491

Timms JF, Arslan-Low E, Gentry-Maharaj A, Luo Z, T'Jampens D, Podust VN, Ford J, Fung ET, Gammerman A, Jacobs I, Menon U (2007) Preanalytic influence of sample handling on SELDI-TOF serum protein profiles. *Clin. Chem.* **53**: 645-656

Tonge R, Shaw J, Middleton B, Rowlinson R, Rayner S, Young J, Pognan F, Hawkins E, Currie I, Davison M (2001) Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* **1**: 377-396

Tung KH, Wilkens LR, Wu AH, McDuffie K, Nomura AM, Kolonel LN, Terada KY, Goodman MT (2005) Effect of anovulation factors on pre- and postmenopausal ovarian cancer risk: revisiting the incessant ovulation hypothesis. *Am. J. Epidemiol.* **161**: 321-329

Turner GA, Goodarzi MT, Thompson S (1995) Glycosylation of alpha-1-proteinase inhibitor and haptoglobin in ovarian cancer: evidence for two different mechanisms. *Glycoconj. J.* **12**: 211-218

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

Van HF, Van GA, Neels H, Wauters A, Demedts P, Bruyland K, DeMeester I, Scharpe S, Janca A, Song C, Maes M (1998) The influence of psychological stress on total serum protein and patterns obtained in serum protein electrophoresis. *Psychol. Med.* **28**: 301-309

Vapnik VN (1999) An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **10**: 988-999

Villanueva J (2006) Automated serum peptide profiling. *Nature Protocols.* **1**: 808-91

Villanueva J, Nazarian A, Lawlor K, Yi SS, Robbins RJ, Tempst P (2008) A sequence-specific exopeptidase activity test (SSEAT) for "functional" biomarker discovery. *Mol. Cell Proteomics* **7**: 509-518

Villanueva J, Philip J, Chaparro CA, Li Y, Toledo-Crow R, Denoyer L, Fleisher M, Robbins RJ, Tempst P (2005) Correcting common errors in identifying cancer-specific serum Peptide signatures. *J. Proteome Res.* **4**: 1060-1072

Villanueva J, Philip J, Entenberg D, Chaparro CA, Tanwar MK, Holland EC, Tempst P (2004) Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry. *Anal. Chem.* **76**: 1560-1570

Villanueva J, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, Olshen AB, Fleisher M, Lilja H, Brogi E, Boyd J, Sanchez-Carbayo M, Holland EC, Cordon-Cardo C, Scher HI, Tempst P (2006) Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J. Clin. Invest.* **116**: 271-284

Wang HC, Su YR, Han KJ, Pang XW, Peng JR, Liang B, Wang S, Chen WF (2004) Multiple variants and a differential splicing pattern of kinectin in human hepatocellular carcinoma. *Biochem. Cell Biol.* **82**: 321-327

West-Nielsen M, Hogdall EV, Marchiori E, Hogdall CK, Schou C, Heegaard NH (2005) Sample handling for mass spectrometric proteomic investigations of human sera. *Anal. Chem.* **77**: 5114-5123

West-Norager M, Kelstrup CD, Schou C, Hogdall EV, Hogdall CK, Heegaard NH (2007) Unravelling in vitro variables of major importance for the outcome of mass spectrometry-based serum proteomics. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* **847**: 30-37

Whiteaker JR, Zhao L, Zhang HY, Feng LC, Piening BD, Anderson L, Paulovich AG (2007) Antibody-based enrichment of peptides on magnetic beads for mass-spectrometry-based quantification of serum biomarkers. *Anal. Biochem.* **362**: 44-54

Wodarz A, Nusse R (1998) Mechanisms of Wnt signaling in development. *Annu. Rev. Cell Dev. Biol.* **14**: 59-88

Woong-Shick A, Sung-Pil P, Su-Mi B, Joon-Mo L, Sung-Eun N, Gye-Hyun N, Young-Lae C, Ho-Sun C, Heung-Jae J, Chong-Kook K, Young-Wan K, Byoung-Don H, Hyun-Sun J (2005) Identification of hemoglobin-alpha and -beta subunits as potential serum biomarkers for the diagnosis and prognosis of ovarian cancer. *Cancer Sci.* **96**: 197-201

Wright DD, Whitney J (2006) Multiple hamartoma syndrome (Cowden's syndrome): case report and literature review. *Gen. Dent.* **54**: 417-419

Wright LC, Sullivan DR, Muller M, Dyne M, Tattersall MH, Mountford CE (1989) Elevated apolipoprotein(a) levels in cancer patients. *Int. J. Cancer.* **43**: 241-244

Ye B, Cramer DW, Skates SJ, Gygi SP, Pratomo V, Fu L, Horick NK, Licklider LJ, Schorge JO, Berkowitz RS, Mok SC (2003) Haptoglobin-alpha subunit as potential serum biomarker in ovarian cancer: identification and characterization using proteomic profiling and mass spectrometry. *Clin. Cancer Res.* **9**: 2904-2911

Yu KH, Rustgi AK, Blair IA (2005). Characterization of proteins in human pancreatic cancer serum using differential gel electrophoresis and tandem mass spectrometry. *J. Proteome Res.* **4**: 1742-51

Yu LR, Zhou M, Conrads TP, Veenstra TD (2003) Diagnostic proteomics: serum proteomic patterns for the detection of early stage cancers. *Dis. Markers.* **19**: 209-218

Zervos EE, Tanner SM, Osborne DA, Bloomston M, Rosemurgy AS, Ellison EC, Melvin WS, de la CA (2006) Differential gene expression in patients genetically predisposed to pancreatic cancer. *J. Surg. Res.* **135**: 317-322

Zhang H, Yan W, Aebersold R (2004) Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes. *Curr. Opin. Chem. Biol.* **8**: 66-75

Zhang Z, Bast RC, Jr., Yu Y, Li J, Sokoll LJ, Rai AJ, Rosenzweig JM, Cameron B, Wang YY, Meng XY, Berchuck A, Van Haaften-Day C, Hacker NF, de Bruijn HW, van der Zee AG, Jacobs IJ, Fung ET, Chan DW (2004) Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res.* **64**: 5882-5890

Zhang Z, Yu Y, Xu F, Berchuck A, Van Haaften-Day C, Havrilesky LJ, de Bruijn HW, Van Der Zee AG, Woolas RP, Jacobs IJ, Skates S, Chan DW, Bast RC, Jr. (2007) Combining multiple serum tumor markers improves detection of stage I epithelial ovarian cancer. *Gynecol. Oncol.* **107**: 526-531

Zimmerman LJ, Wernke GR, Caprioli RM, Liebler DC (2005) Identification of protein fragments as pattern features in MALDI-MS analyses of serum. *J. Proteome Res.* **4**: 1672-1680

Zweemer RP. Overview of the aetiology of familiar ovarian cancer. In Jacobs IJ, Shepherd JH, Oram DH, Blackett AD, Luesley DM, Berchuck A, Hudson CN, editors. Ovarian Cancer. 2nd Edn. Great Britian; *Oxford University Press* (2002) Pt (1) 4: 29.

Smooth

```
function [X_new] = smooth(X,windowSize)
start = windowSize+1;
for i=start:size(X,1)-start+1 % changed from size(X,1) - start
X_new(i-start+1,:) = [X(i,1) mean(X(i-
windowSize:i+windowSize,2))];
end
```

Baseline correction

```
function [point_present] =
baseline_correction(base,point_present,X,pd)
% base - baseline function
% pd - percentage difference
counting = 0;
min_X = inf;
for i = 1:size(X,1)
if base(i,2) > X(i,2)
if X(i,2) < min_X
min_X = X(i,2);
min_X_index = i;
end
counting = 1;
else
if counting
counting = 0;
min_X = inf;
for j = 1:size(X)
if point_present(j)
if abs(X(j,1) - X(min_X_index))/X(min_X_index) < pd
point_present(j) = 0;
end
end
end
point_present(min_X_index) = 1;
end
end
end
if counting
counting = 0;
min_X = inf;
for j = 1:size(X)
if point_present(j)
if abs(X(j,1) - X(min_X_index))/X(min_X_index) < pd
point_present(j) = 0;
end
end
end
point_present(min_X_index) = 1;
end
```

```
function [B,point_present] = splinefun(B,point_present)
index = 1;
B1 = B(:,1);
B2 = B(:,2);
xx = B1;
X = B1(point_present==1);
Y = B2(point_present==1);
for i=1:size(B,1)
xx(i) = B(i,1);
if point_present(i)
X(index) = B(i,1);
Y(index) = B(i,2);
index = index + 1;
end
end
yy = pchip(X,Y,xx);
s = size(B,1);
clear B;
B(:,1) = xx(:);
B(:,2) = yy(:);
for i=1:s
B(i,1) = xx(i);
B(i,2) = yy(i);
end
```

Normalise

```
function [X] = normalise(X,C)
point_sum = sum(X(:,2));
%C = 5*10^8;
X(:,2) = (X(:,2)/point_sum)*C;
for i=1:size(X,1)

t = (X(i,2)/point_sum)*C;
  X(i,2) = t;
end
%plot(X(:,1),(point_sum/size(X,1))*ones(size(X,1),1),'-')
```

Peak identification

```
function p=peak_finder(X,ws,thold,sample_index)
% sw - singal-to-noise ratio window size
thold - minimum intensity threshold
% The output 'p':
p(:, 1)    sample index;
p(:, 2)    number of the peak in the initial array;
p(:, 3)    m/z-ratio;
% p(:, 4)  signal-to-noise ratio (intensity divided by the
average intensity in the window);
p(:, 5)    intensity.
if nargin == 3
    sample_index = 0;
end
```

```
slopeSign = diff(X(:,2))>0;
slopeSignChange = diff(slopeSign)<0;
h = find(slopeSignChange) + 1;
h(X(h,2) < thold) = [];
p(:,1) = h;
p(:,2) = X(h,1);
XX = X(:,1);
X2 = X(:,2);
index = 1;
for i=h'
wst = min(ws,length(XX) - i);
wst = min(wst,i-1);
    q = find(XX<=(XX(i+wst)));
q1 = find(XX>=(XX(i-wst)));
    q2 = intersect(q,q1);
p(index,3) = X(i,2)/mean(X2(q2));
index = index + 1;
end
p(:,4) = X(h,2);
p = [sample_index*ones(size(p(:,1))) p];
```

Peak list

```
function peak_list =
genPeakList_001(proc_spec_LMR,proc_spec_HMR,peak_set,lab,peakG
roups,pd)
% peakGroups - max or mean values of the peakalign2-function
output
% pd - mass separation parameter, must be the same as in
peakalign2.m function
% peak_set, lab - not used parameters
peak_list = zeros(1,size(peakGroups,1));
for j = 1:size(peakGroups,1)
    if size(peak_set,1) == 0
        h=[];
    else

        h1 = find(peak_set(:,2)<peakGroups(j,1));
        h2 = find(peak_set(:,2)>peakGroups(j,1));
        h = intersect(h1,h2);
    end
    if size(h,1) == 0
        if 1%peakGroups(j) < max(proc_spec_LMR(:,1))
h1 = find(proc_spec_LMR(:,1)<peakGroups(j,1) +
peakGroups(j,1)*pd);
h2 = find(proc_spec_LMR(:,1)>peakGroups(j,1) -
peakGroups(j,1)*pd);
h = intersect(h1,h2);
ep = max(proc_spec_LMR(h,2));
if length(ep) == 0
 ['no signal, setting to -1, m/z=' num2str(peakGroups(j))]
peak_list(1,j) = -1;
else
peak_list(1,j) = ep;
```

```
end
else
h1 = find(proc_spec_HMR(:,1)<peakGroups(j,1) + 
peakGroups(j,1)*pd);
            h2 = find(proc_spec_HMR(:,1)>peakGroups(j,1) - 
peakGroups(j,1)*pd);
            h = intersect(h1,h2);
            ep = max(proc_spec_HMR(h,2));
if length(ep) == 0
['no signal, setting to -1 m/z = ' num2str(peakGroups(j))]
peak_list(1,j) = -1;
else
peak_list(1,j) = ep;
end
end
else
if size(h,1) == -1
peak_list(1,j) = peak_set(h,5);
else
if size(h,1) > 1
peak_list(1,j) = max(peak_set(h,5));

end
end
peak_list(1,j+1) = lab;
```

Peak alignment

```
function peak_groups = peakalign2(pks, first_SNR, second_SNR, 
massSep)

Input: peaks    -    spectrum ID
peak location in clock ticks
peak location in mass units
signal-to-noise ratio of the peak
normalized baseline-corrected intensity of the peak
first_SNR  -  Minimum signal-to-noise ratio for new cluster
second_SNR -  Minimum signal-to-noise to be included in 
existing cluster
this is not used in this version as it adds no extra peaks 
to a cluster just incriments the count
massSep    -  Mass separation param
Output:
peak_groups - PeakID
Highest peak mass location
Mean Mass location
Min mass
Max mass
Number of pks
Max Intensity
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
pks = sortrows(pks,-5); % descending order of the 5th column
group_index = 1;
peak_groups = [];
```

```
spec_at_peak = zeros(size(pks,1),1);
pks_second_SNR = pks(pks(:,4)<first_SNR,:);
pks_second_SNR(pks_second_SNR(:,4) < second_SNR) = [];
pks(pks(:,4)<first_SNR,:) = [];
size(pks)
for i=1:size(pks,1)
if mod(i,100)==0
i /size(pks,1)
end
if pks(i,4) >= first_SNR
if size(peak_groups,1) == 0
peak_groups(group_index,:) = [ group_index pks(i,3) pks(i,3)
pks(i,3) pks(i,3) 1 pks(i,5)];
spec_at_peak(i) = group_index;
group_index = group_index + 1;
spec_list(group_index).list = pks(i,1);
else
h = [];
for j=1:size(peak_groups,1)
if abs((peak_groups(j,2) - pks(i,3))/peak_groups(j,2)) <=
massSep
h = [h j];
end
end


if size(h,1) == 0 %create new group
peak_groups(group_index,:) = [group_index pks(i,3) pks(i,3)
pks(i,3) pks(i,3) 1 pks(i,5)];
spec_at_peak(i) = group_index;
group_index = group_index + 1;
spec_list(group_index).list = pks(i,1);
else
if size(h,1) >= 1 %add to existing group;
if size(h,1) > 1 then choose closest group
if size(h,1) > 1
ind = h(1);
for j=h
if abs((peak_groups(j,2) - pks(i,3))/peak_groups(j,2)) <=
abs((peak_groups(ind,2) - pks(i,3))/peak_groups(ind,2))
ind = j;
end
end
else
ind = h(1);
end %
end
if size(h,1) > 1
hsp1 = find(spec_at_peak == ind);
hps2 = find(pks(hsp1,1) == pks(i,1));
if size(hps2,1) == 0
%add peak to peak_groups(ind,:)
peak_groups(ind,3) = (peak_groups(ind,6)*peak_groups(ind,3) +
pks(i,3))/(peak_groups(ind,6)+1);
peak_groups(ind,6) = peak_groups(ind,6) + 1;
if pks(i,3) < peak_groups(ind,4)
```

```
peak_groups(ind,4) = pks(i,3);
end
if pks(i,3) > peak_groups(ind,5)
peak_groups(ind,5) = pks(i,3);
end
spec_at_peak(i) = ind;
spec_list(ind).list = [spec_list(ind).list; pks(i,1)];
end %
end
if size(hps2,1) == 0
end %
end
size(h,1) == 0
end
end
size(peak_groups,1) == 0
end %
end
pks(i,4) >= first_SNR
end %
end
for i=1:size(pks,1)
for i=1:size(pks_second_SNR,1)
h= [];
for j=1:size(peak_groups,1)
if abs((peak_groups(j,2) -
pks_second_SNR(i,3))/peak_groups(j,2)) <= massSep
h = [h j];
end
end
if size(h,1) == 0 %do nothing
elseif size(h,1) >= 1
ind = h(1);
if size(h,1) > 1
for j=h
if abs(peak_groups(j,2) - pks_second_SNR(i,3)) <=
abs(peak_groups(ind,2) - pks_second_SNR(i,3))
ind = j;
end
end
end
end
hsp1 = find(spec_list(ind).list == pks_second_SNR(i,1));
if size(hsp1,1) == 0
peak_groups(ind,6) = peak_groups(ind,6) + 1;
spec_list(ind).list = [spec_list(ind).list;
pks_second_SNR(i,1)];
end
end
```

Peak intensity list

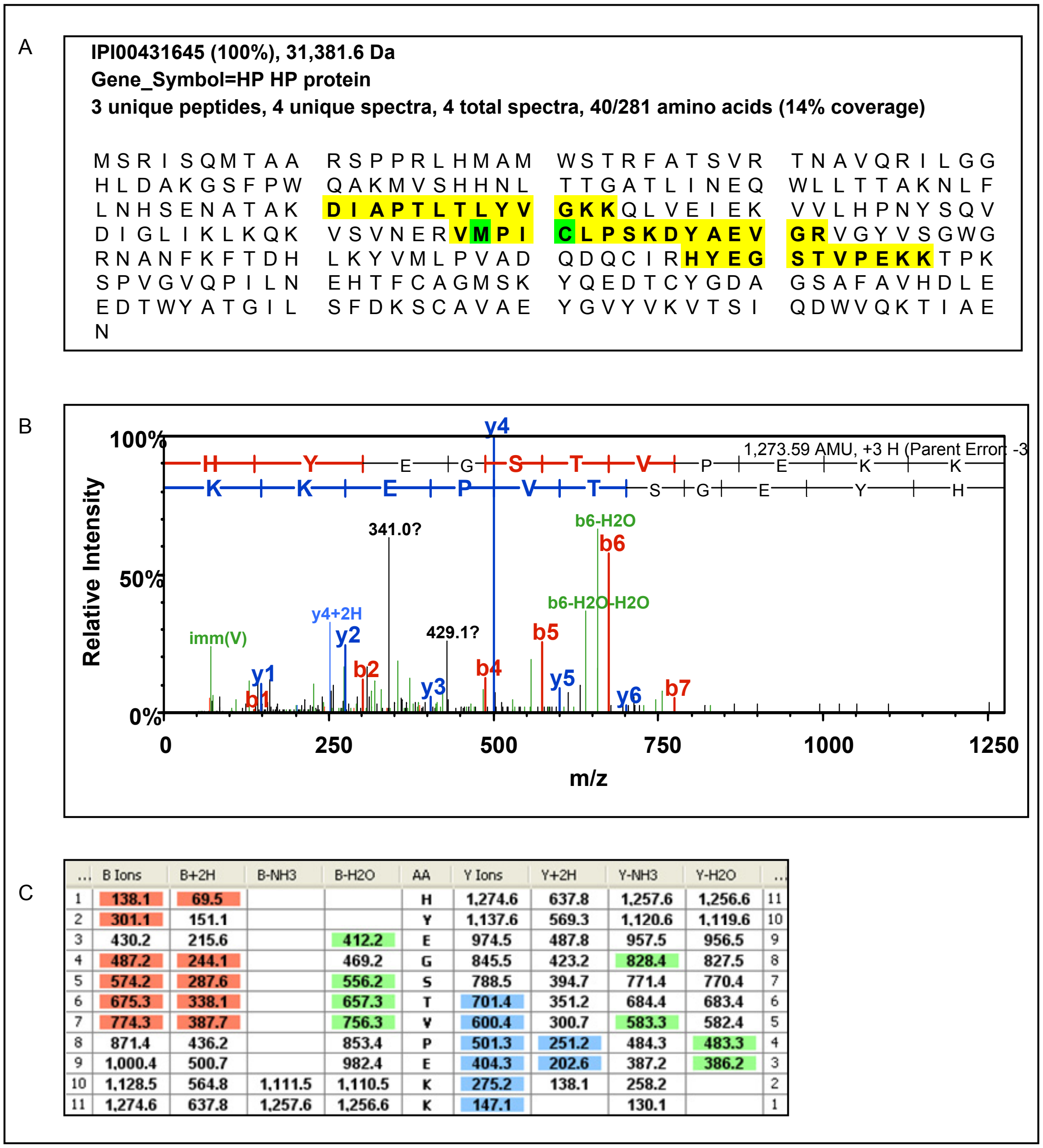
```
function [peak_list f_list] = make_peak_int_list(pg,massSep,
proc_path)

pg - peak_groups, output of peakalign2.m
pg = pg(:, 2); % max value
%pg = pg(:,3); % mean value
%proc_path = '/rmt/csnewton/pgrads/brian/UKCTOCS_GW/proc/';
%proc_path = '/rmt/csnewton/pgrads/brian/Current/D-
drive/Tempst/proc_cal/';
if nargin < 3
    proc_path = [pwd '\res\processed\']
end
f_list = file_finder(proc_path);
%f_list_ctr = file_finder([proc_path 'Ctr\'],'*_1*');%'*_2'
%f_list_bre = file_finder([proc_path 'Bre\'],'*_1*');%'*_2'
%f_list_pro = file_finder([proc_path 'Pro\'],'*_1*');%'*_2'
%f_list_bla = file_finder([proc_path 'Bla\'],'*_1*');%'*_2'
%l = length(f_list_ctr) + length(f_list_bre) +
length(f_list_pro) + length(f_list_bla);
l = length(f_list);
massSep = 0.0015;

peak_list = [];
c=0;
for i=1:length(f_list)
c=c+1;
   c/l
    X = dlmread([proc_path f_list(i).fname]);
peak_list = [peak_list;
genPeakList_001(X,[],[],0,pg,massSep)];
end
%{
for i=1:length(f_list_ctr)
c=c+1;
   c/l
    X = dlmread([proc_path 'Ctr\' f_list_ctr(i).fname]);
peak_list = [peak_list;
genPeakList_001(X,[],[],0,pg,massSep)];
end
for i=1:length(f_list_bre)
c=c+1;
    c/l
    X = dlmread([proc_path 'Bre\' f_list_bre(i).fname]);
peak_list = [peak_list;
genPeakList_001(X,[],[],1,pg,massSep)];
end
for i=1:length(f_list_pro)
c=c+1;
    c/l
    X = dlmread([proc_path 'Pro\' f_list_pro(i).fname]);
peak_list = [peak_list;
genPeakList_001(X,[],[],2,pg,massSep)];
end
for i=1:length(f_list_bla)
c=c+1;
    c/l
    X = dlmread([proc_path 'Bla\' f_list_bla(i).fname]);
peak_list = [peak_list;
genPeakList_001(X,[],[],3,pg,massSep)];
end
%}
```

Spectral calibration

```
function Xc = cal_spec(raw, cal, cal_peaks, massSep,ws)
SNR_low = 1.5;
%ws = 50;
%cal_peaks = [782.402 1047.20 1297.51 1348.66 1620.88 2094.46
2466.73 ...
%         3149.61 6181.05 8565.89 12361.09];
cal_peaks = sort(cal_peaks);
index = 1:length(cal_peaks);
%LMR_index = [1 2 3 7];
%
HMR_index = [10 12 13];
TOFs = sqrt(cal_peaks);
use = [];
length(cal_peaks)
for j = 1:length(cal_peaks)
inds = intersect(find(cal(:,1) > cal_peaks(j)*(1 - massSep)),
find(cal(:,1) < cal_peaks(j)*(1 + massSep)));
[m_1 h_1] = max(cal(inds,2));
peak_mz(j) = cal(inds(h_1),1);
max_point_index = inds(h_1);
XX = cal(:,1);
  X2 = cal(:,2);
q = find(XX<=(XX(max_point_index)+ws));
q1 = find(XX>=(XX(max_point_index)-ws));
q2 = intersect(q,q1);
  snr(j) = cal(max_point_index,2)/mean(X2(q2));
if cal(max_point_index,2)/mean(X2(q2)) > SNR_low
use = [use 1];
else
use = [use 0];
end
end
initial = [1; 1];
1
[peak_mz' snr']
use = find(use == 1);
x = lsqnonlin(@(x)fun1(x, peak_mz(index(use)),
TOFs(index(use))), initial);
raw_mz = ((x(1)*ones(size(raw,1),1) +
sqrt(raw(:,1)/x(2)))).^2;
Xc = [raw_mz raw(:,2)];
```

Graphical output

```
function plot_all(pg,plot_path)
pg = pg(:, 2); % max value %
pg = pg(:,3); % mean value
if nargin < 2
plot_path = [pwd '\res\plots\'];
end
if ~isdir(plot_path)
mkdir([pwd '\res\'], 'plots');
end
if ~isdir([plot_path 'fig\'])
mkdir(plot_path, 'fig');
end
if ~isdir([plot_path 'eps\'])
mkdir(plot_path, 'eps');
end
proc_path = [pwd '\res\processed\'];
proc_path_c = [proc_path 'Ctr\'];
proc_path_bla = [proc_path 'Bla\'];
proc_path_pro = [proc_path 'Pro\'];
proc_path_bre = [proc_path 'Bre\'];
for i=1:length(pg)
figure
plot_peak(pg(i),50,proc_path_c,'m')
plot_peak(pg(i),50,proc_path_bla,'g')
h = gcf; % returns the current figure handle
saveas(h,[plot_path 'eps\' num2str(pg(i)) '_bla.eps'])
saveas(h,[plot_path 'fig\' num2str(pg(i)) '_bla.fig'])
close(h); % deletes the figure
figure % created the figure object
plot_peak(pg(i),50,proc_path_c,'m')
plot_peak(pg(i),50,proc_path_pro,'b')
h = gcf;
saveas(h,[plot_path 'eps\' num2str(pg(i)) '_pro.eps'])
saveas(h,[plot_path 'fig\' num2str(pg(i)) '_pro.fig'])
close(h);
figure
plot_peak(pg(i),50,proc_path_c,'m')
plot_peak(pg(i),50,proc_path_bre,'r')

h = gcf;
saveas(h,[plot_path 'eps\' num2str(pg(i)) '_bre.eps'])

saveas(h,[plot_path 'fig\' num2str(pg(i)) '_bre.fig'])
close(h);
end
```



A

IPI00431645 (100%), 31,381.6 Da
Gene_Symbol=HP HP protein
3 unique peptides, 4 unique spectra, 4 total spectra, 40/281 amino acids (14% coverage)

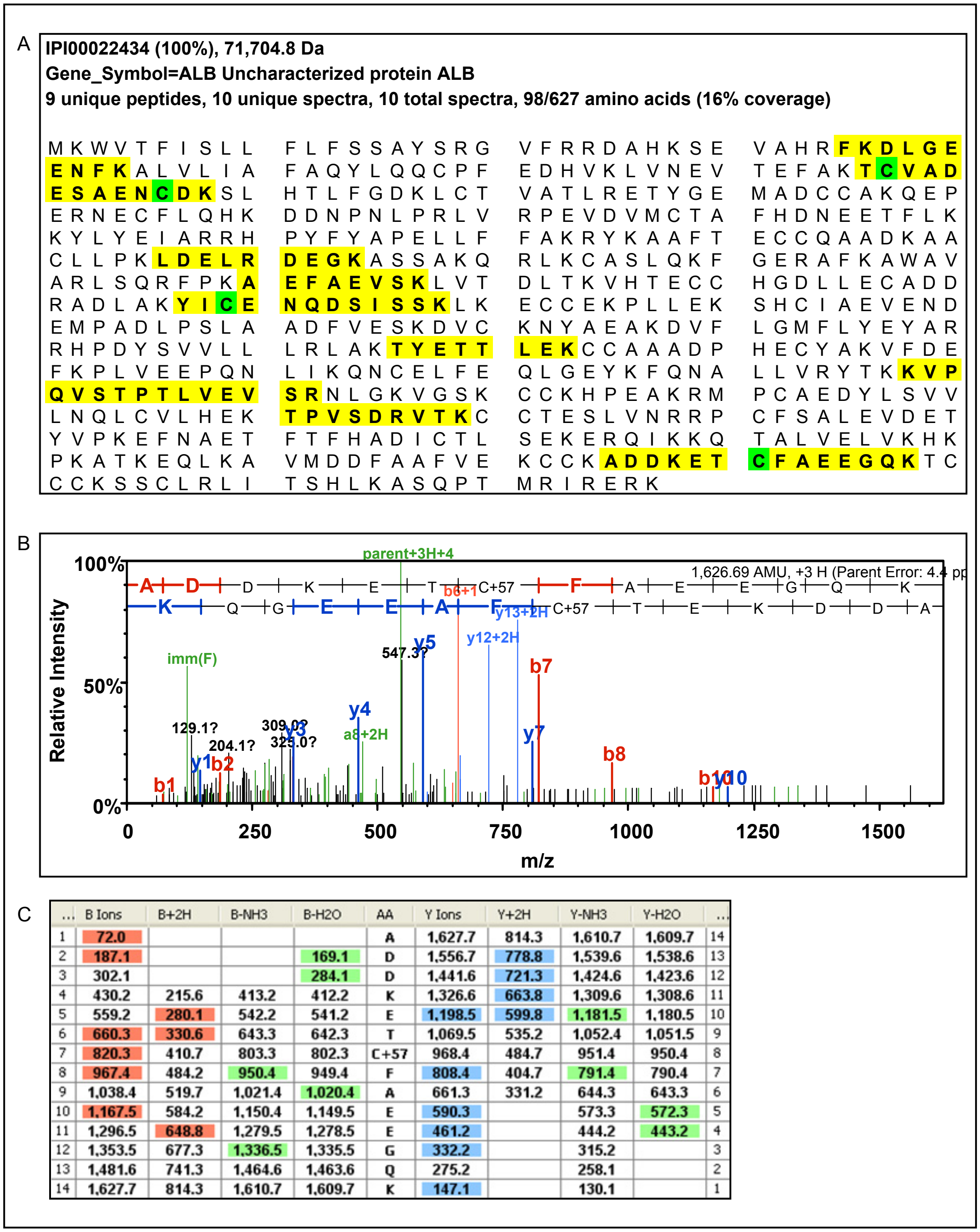
```
MSRISQMTAA  RSPPRLHMAM  WSTRFATSVR  TNAVQRILGG
HLDAKGSFPW  QAKMVSHHNL  TTGATLINEQ  WLLTTAKNLF
LNHSENATAK  DIAPTLTLYV  GKKQLVEIEK  VVLHPNYSQV
DIGLIKLKQK  VSVNERVMPI  CLPSKDYAEV  GRVGYVSGWG
RNANFKFTDH  LKYVMLPVAD  QDQCIRHYEG  STVPEKKTPK
SPVGVQPILN  EHTFCAGMSK  YQEDTCYGDA  GSAFAVHDLE
EDTWYATGIL  SFDKSCAVAE  YGVYVKVTSI  QDWVQKTIAE
N
```

B

C

...	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	...
1	138.1	69.5			H	1,274.6	637.8	1,257.6	1,256.6	11
2	301.1	151.1			Y	1,137.6	569.3	1,120.6	1,119.6	10
3	430.2	215.6		412.2	E	974.5	487.8	957.5	956.5	9
4	487.2	244.1		469.2	G	845.5	423.2	828.4	827.5	8
5	574.2	287.6		556.2	S	788.5	394.7	771.4	770.4	7
6	675.3	338.1		657.3	T	701.4	351.2	684.4	683.4	6
7	774.3	387.7		756.3	V	600.4	300.7	583.3	582.4	5
8	871.4	436.2		853.4	P	501.3	251.2	484.3	483.3	4
9	1,000.4	500.7		982.4	E	404.3	202.6	387.2	386.2	3
10	1,128.5	564.8	1,111.5	1,110.5	K	275.2	138.1	258.2		2
11	1,274.6	637.8	1,257.6	1,256.6	K	147.1		130.1		1

Differently expressed protein feature identified by LC-MS/MS. A) HP protein identified by 3 unique peptides (highlighted in yellow in the protein sequence). B) An MS/MS spectrum and C) table indicating the masses of identified fragment ions.

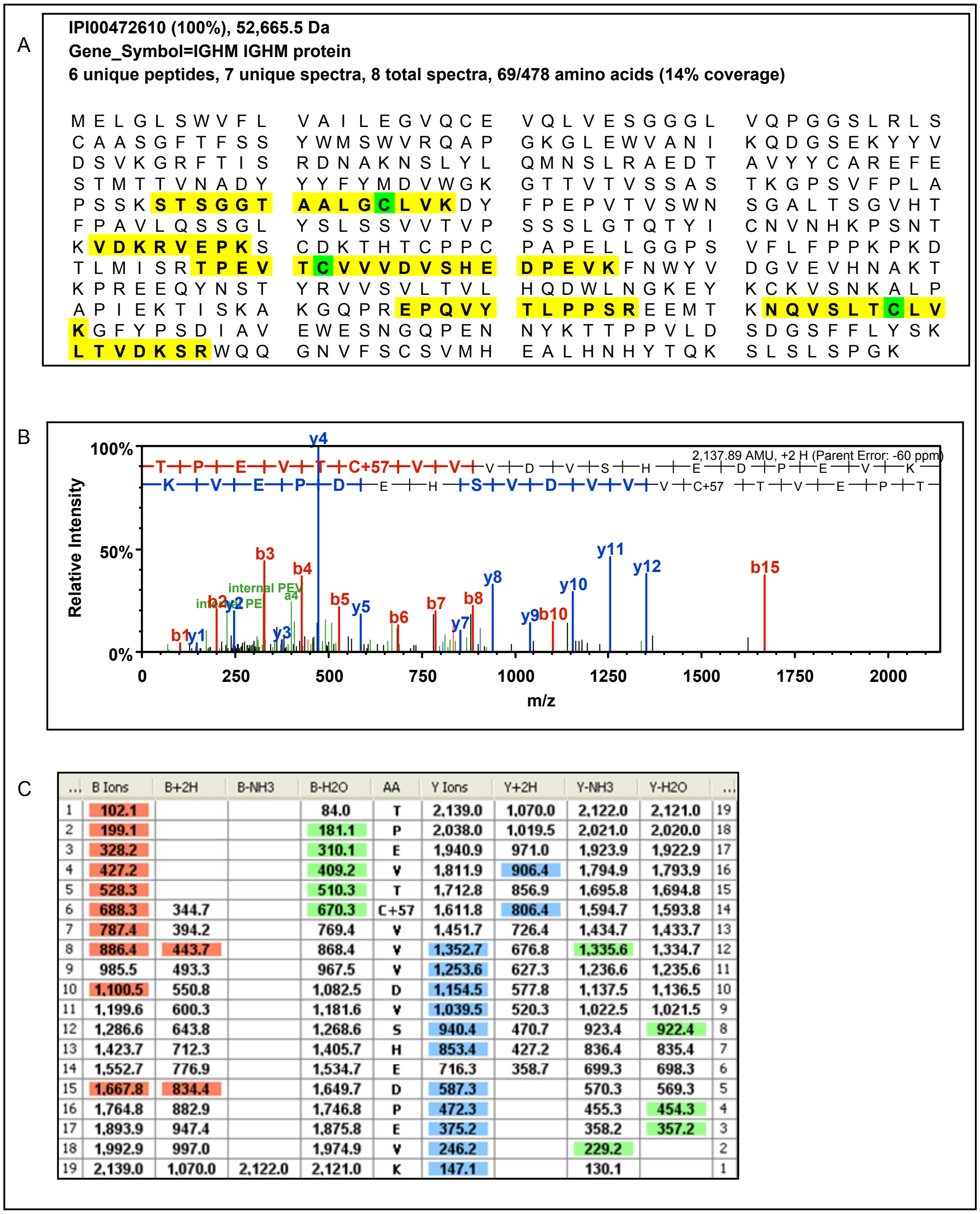


IPI00022434 (100%), 71,704.8 Da
Gene_Symbol=ALB Uncharacterized protein ALB
9 unique peptides, 10 unique spectra, 10 total spectra, 98/627 amino acids (16% coverage)

```
MKWVTFISLL FLFSSAYSRG VFRRDAHKSE VAHRFKDLGE
ENFKALVLIA FAQYLQQCPF EDHVKLVNEV TEFAKTCVAD
ESAENCDKSL HTLFGDKLCT VATLRETYGE MADCCAKQEP
ERNECFLQHK DDNPNLPRLV RPEVDVMCTA FHDNEETFLK
KYLYEIARRH PYFYAPELLF FAKRYKAAFT ECCQAADKAA
CLLPKLDELR DEGKASSAKQ RLKCASLQKF GERAFKAWAV
ARLSQRFPKA EFAEVSKLVT DLTKVHTECC HGDLLECADD
RADLAKYICE NQDSISSKLK ECCEKPLLEK SHCIAEVEND
EMPADLPSLA ADFVESKDVC KNYAEAKDVF LGMFLYEYAR
RHPDYSVVLL LRLAKTYETT LEKCCAAADP HECYAKVFDE
FKPLVEEPQN LIKQNCELFE QLGEYKFQNA LLVRYTKKVP
QVSTPTLVEV SRNLGKVGSK CCKHPEAKRM PCAEDYLSVV
LNQLCVLHEK TPVSDRVTKC CTESLVNRRP CFSALEVDET
YVPKEFNAET FTFHADICTL SEKERQIKKQ TALVELVKHK
PKATKEQLKA VMDDFAAFVE KCCKADDKET CFAEEGQKTC
CCKSSCLRLI TSHLKASQPT MRIRERK
```

...	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	...
1	72.0				A	1,627.7	814.3	1,610.7	1,609.7	14
2	187.1			169.1	D	1,556.7	778.8	1,539.6	1,538.6	13
3	302.1			284.1	D	1,441.6	721.3	1,424.6	1,423.6	12
4	430.2	215.6	413.2	412.2	K	1,326.6	663.8	1,309.6	1,308.6	11
5	559.2	280.1	542.2	541.2	E	1,198.5	599.8	1,181.5	1,180.5	10
6	660.3	330.6	643.3	642.3	T	1,069.5	535.2	1,052.4	1,051.5	9
7	820.3	410.7	803.3	802.3	C+57	968.4	484.7	951.4	950.4	8
8	967.4	484.2	950.4	949.4	F	808.4	404.7	791.4	790.4	7
9	1,038.4	519.7	1,021.4	1,020.4	A	661.3	331.2	644.3	643.3	6
10	1,167.5	584.2	1,150.4	1,149.5	E	590.3		573.3	572.3	5
11	1,296.5	648.8	1,279.5	1,278.5	E	461.2		444.2	443.2	4
12	1,353.5	677.3	1,336.5	1,335.5	G	332.2		315.2		3
13	1,481.6	741.3	1,464.6	1,463.6	Q	275.2		258.1		2
14	1,627.7	814.3	1,610.7	1,609.7	K	147.1		130.1		1

Differently expressed protein feature identified by LC-MS/MS. A) ALB protein identified by 9 unique peptides (highlighted in yellow in the protein sequence). B) An MS/MS spectrum and C) table indicating the masses of identified fragment ions.



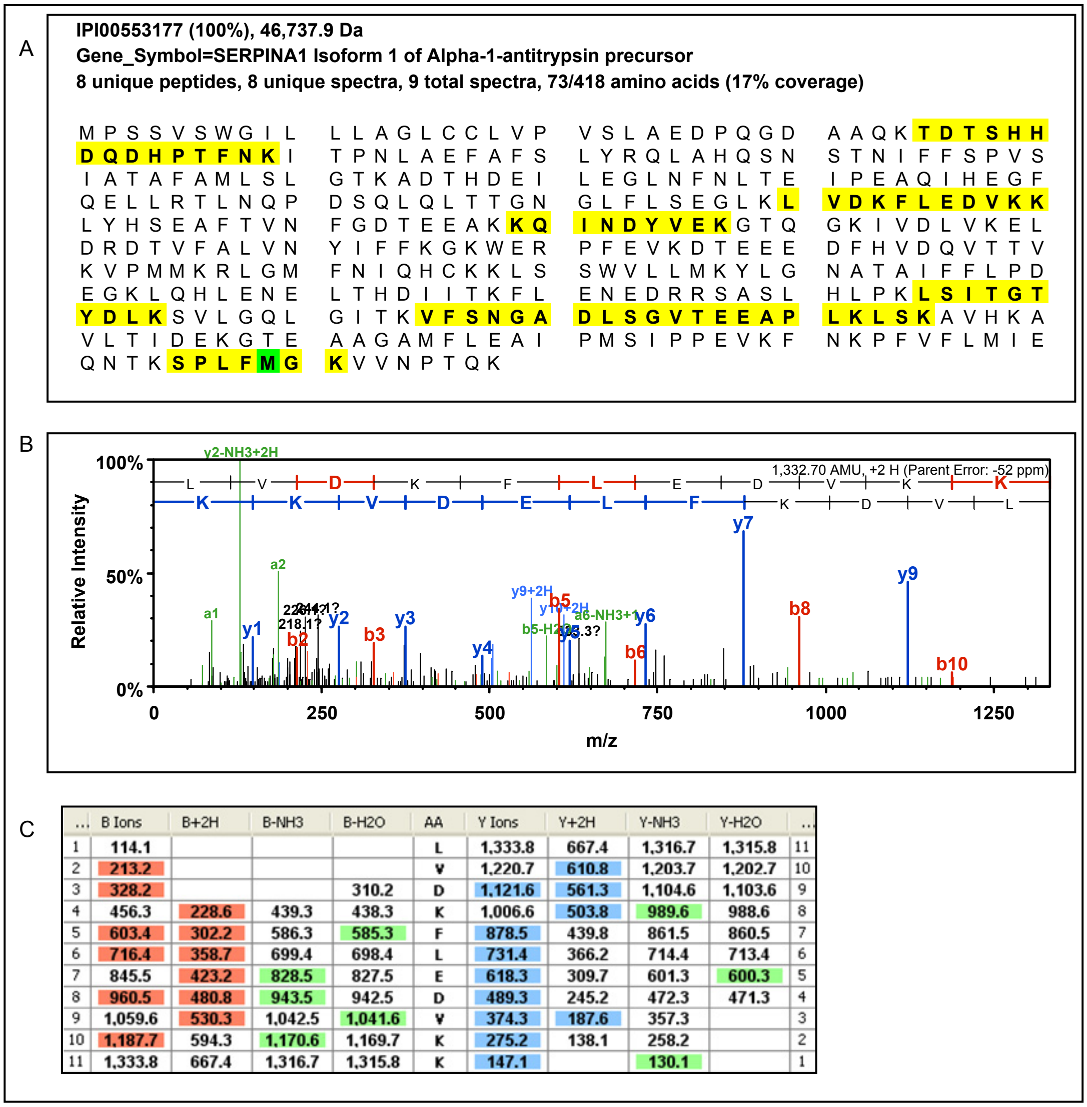
A

IPI00472610 (100%), 52,665.5 Da
Gene_Symbol=IGHM IGHM protein
6 unique peptides, 7 unique spectra, 8 total spectra, 69/478 amino acids (14% coverage)

MELGLSWVFL VAILEGVQCE VQLVESGGGL VQPGGSLRLS
CAASGFTFSS YWMSWVRQAP GKGLEWVANI KQDGSEKYYV
DSVKGRFTIS RDNAKNSLYL QMNSLRAEDT AVYYCAREFE
STMTTVNADY YYFYMDVWGK GTTVTVSSAS TKGPSVFPLA
PSSK**STSGGT** **AALGCLVK**DY FPEPVTVSWN SGALTSGVHT
FPAVLQSSGL YSLSSVVTVP SSSLGTQTYI CNVNHKPSNT
K**VDKRVEPK**S CDKTHTCPPC PAPELLGGPS VFLFPPKPKD
TLMISR**TPEV** **TCVVVDVSHE** **DPEVK**FNWYV DGVEVHNAKT
KPREEQYNST YRVVSVLTVL HQDWLNGKEY KCKVSNKALP
APIEKTISKA KGQPR**EPQVY** **TLPPSR**EEMT K**NQVSLTCLV**
KGFYPSDIAV EWESNGQPEN NYKTTPPVLD SDGSFFLYSK
LTVDKSRWQQ GNVFSCSVMH EALHNHYTQK SLSLSPGK

...	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	...
1	102.1			84.0	T	2,139.0	1,070.0	2,122.0	2,121.0	19
2	199.1			181.1	P	2,038.0	1,019.5	2,021.0	2,020.0	18
3	328.2			310.1	E	1,940.9	971.0	1,923.9	1,922.9	17
4	427.2			409.2	V	1,811.9	906.4	1,794.9	1,793.9	16
5	528.3			510.3	T	1,712.8	856.9	1,695.8	1,694.8	15
6	688.3	344.7		670.3	C+57	1,611.8	806.4	1,594.7	1,593.8	14
7	787.4	394.2		769.4	V	1,451.7	726.4	1,434.7	1,433.7	13
8	886.4	443.7		868.4	V	1,352.7	676.8	1,335.6	1,334.7	12
9	985.5	493.3		967.5	V	1,253.6	627.3	1,236.6	1,235.6	11
10	1,100.5	550.8		1,082.5	D	1,154.5	577.8	1,137.5	1,136.5	10
11	1,199.6	600.3		1,181.6	V	1,039.5	520.3	1,022.5	1,021.5	9
12	1,286.6	643.8		1,268.6	S	940.4	470.7	923.4	922.4	8
13	1,423.7	712.3		1,405.7	H	853.4	427.2	836.4	835.4	7
14	1,552.7	776.9		1,534.7	E	716.3	358.7	699.3	698.3	6
15	1,667.8	834.4		1,649.7	D	587.3		570.3	569.3	5
16	1,764.8	882.9		1,746.8	P	472.3		455.3	454.3	4
17	1,893.9	947.4		1,875.8	E	375.2		358.2	357.2	3
18	1,992.9	997.0		1,974.9	V	246.2		229.2		2
19	2,139.0	1,070.0	2,122.0	2,121.0	K	147.1		130.1		1

Differently expressed protein feature identified by LC-MS/MS. A) IGHM protein identified by 8 unique peptides (highlighted in yellow in the protein sequence). B) An MS/MS spectrum and C) table indicating the masses of identified fragment ions.



A

IPI00553177 (100%), 46,737.9 Da
Gene_Symbol=SERPINA1 Isoform 1 of Alpha-1-antitrypsin precursor
8 unique peptides, 8 unique spectra, 9 total spectra, 73/418 amino acids (17% coverage)

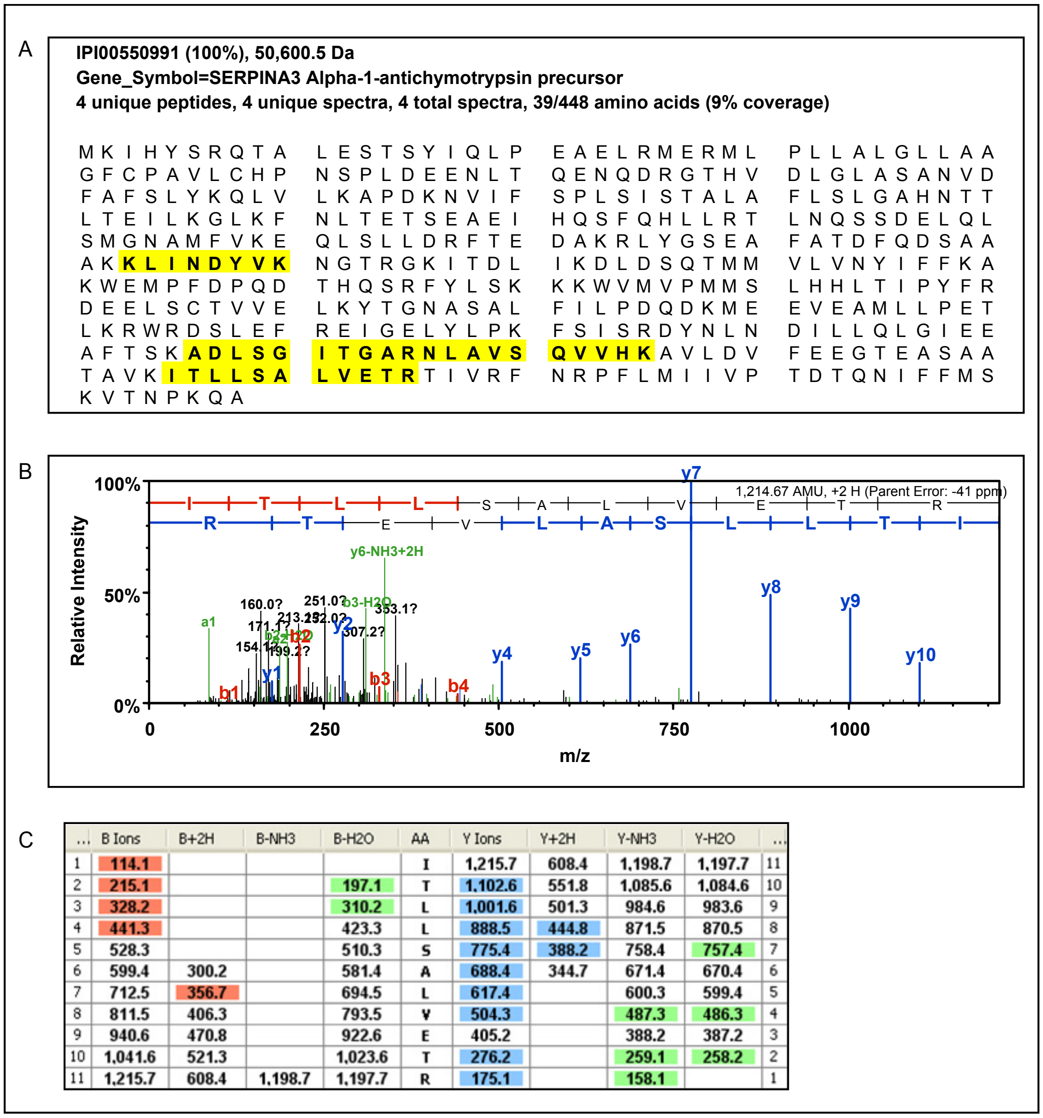
M P S S V S W G I L   L L A G L C C L V P   V S L A E D P Q G D   A A Q K **T D T S H H**
D Q D H P T F N K I   T P N L A E F A F S   L Y R Q L A H Q S N   S T N I F F S P V S
I A T A F A M L S L   G T K A D T H D E I   L E G L N F N L T E   I P E A Q I H E G F
Q E L L R T L N Q P   D S Q L Q L T T G N   G L F L S E G L K **L**   **V D K F L E D V K K**
L Y H S E A F T V N   F G D T E E A K **K Q**   **I N D Y V E K** G T Q   G K I V D L V K E L
D R D T V F A L V N   Y I F F K G K W E R   P F E V K D T E E E   D F H V D Q V T T V
K V P M M K R L G M   F N I Q H C K K L S   S W V L L M K Y L G   N A T A I F F L P D
E G K L Q H L E N E   L T H D I I T K F L   E N E D R R S A S L   H L P K **L S I T G T**
Y D L K S V L G Q L   G I T K **V F S N G A**   **D L S G V T E E A P**   **L K L S K** A V H K A
V L T I D E K G T E   A A G A M F L E A I   P M S I P P E V K F   N K P F V F L M I E
Q N T K **S P L F M G**   **K** V V N P T Q K

B

C

...	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	...
1	114.1				L	1,333.8	667.4	1,316.7	1,315.8	11
2	213.2				V	1,220.7	610.8	1,203.7	1,202.7	10
3	328.2			310.2	D	1,121.6	561.3	1,104.6	1,103.6	9
4	456.3	228.6	439.3	438.3	K	1,006.6	503.8	989.6	988.6	8
5	603.4	302.2	586.3	585.3	F	878.5	439.8	861.5	860.5	7
6	716.4	358.7	699.4	698.4	L	731.4	366.2	714.4	713.4	6
7	845.5	423.2	828.5	827.5	E	618.3	309.7	601.3	600.3	5
8	960.5	480.8	943.5	942.5	D	489.3	245.2	472.3	471.3	4
9	1,059.6	530.3	1,042.5	1,041.6	V	374.3	187.6	357.3		3
10	1,187.7	594.3	1,170.6	1,169.7	K	275.2	138.1	258.2		2
11	1,333.8	667.4	1,316.7	1,315.8	K	147.1		130.1		1

Differently expressed protein feature identified by LC-MS/MS. A) SERPINA1 protein identified by 8 unique peptides (highlighted in yellow in the protein sequence). B) An MS/MS spectrum and C) table indicating the masses of identified fragment ions.



A

IPI00550991 (100%), 50,600.5 Da
Gene_Symbol=SERPINA3 Alpha-1-antichymotrypsin precursor
4 unique peptides, 4 unique spectra, 4 total spectra, 39/448 amino acids (9% coverage)

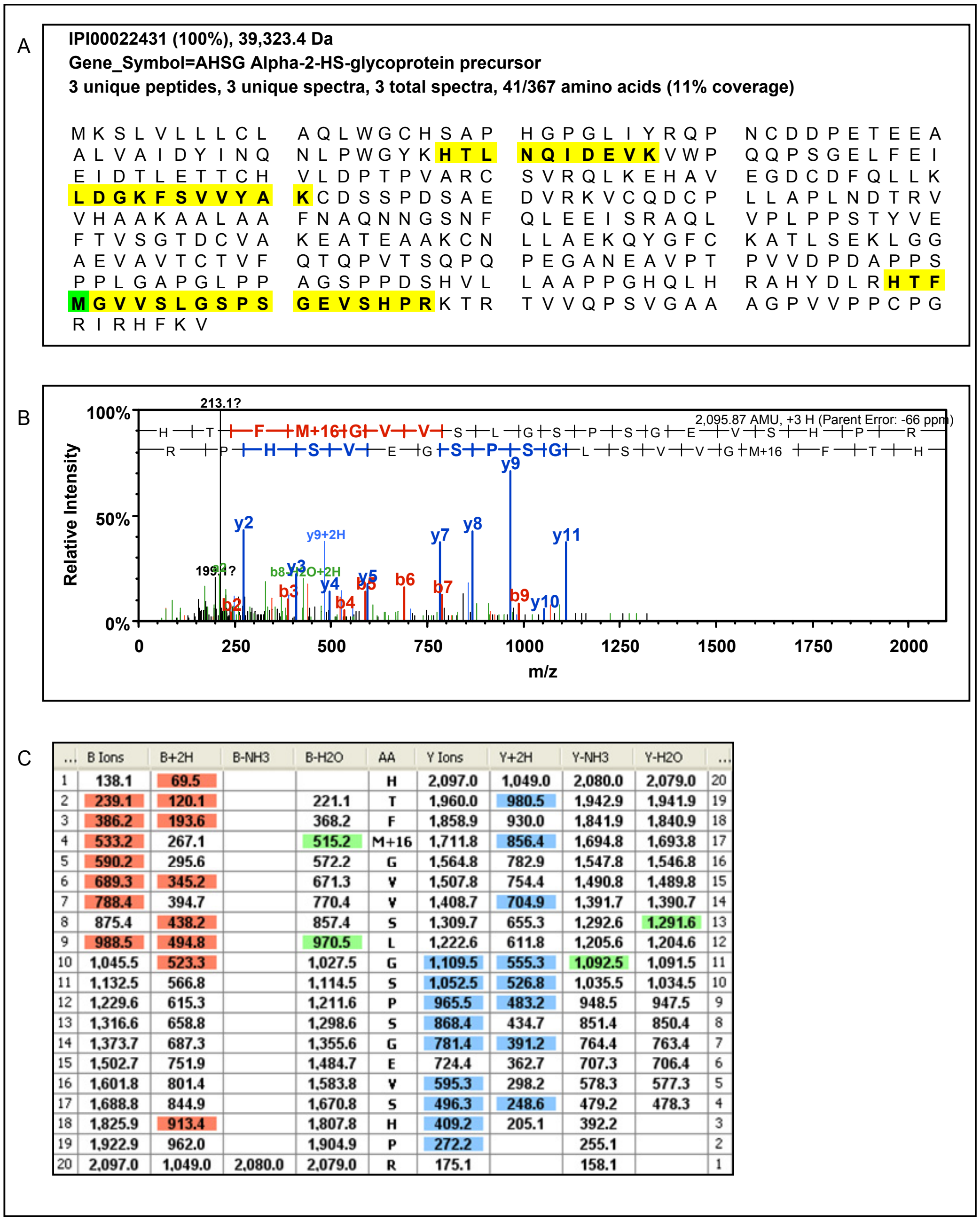
MKIHYSRQTA LESTSYIQLP EAELRMERML PLLALGLLAA
GFCPAVLCHP NSPLDEENLT QENQDRGTHV DLGLASANVD
FAFSLYKQLV LKAPDKNVIF SPLSISTALA FLSLGAHNTT
LTEILKGLKF NLTETSEAEI HQSFQHLLRT LNQSSDELQL
SMGNAMFVKE QLSLLDRFTE DAKRLYGSEA FATDFQDSAA
AK**KLINDYVK** NGTRGKITDL IKDLDSQTMM VLVNYIFFKA
KWEMPFDPQD THQSRFYLSK KKWVMVPMMS LHHLTIPYFR
DEELSCTVVE LKYTGNASAL FILPDQDKME EVEAMLLPET
LKRWRDSLEF REIGELYLPK FSISRDYNLN DILLQLGIEE
AFTSK**ADLSG** **ITGARNLAVS** **QVVHK**AVLDV FEEGTEASAA
TAVK**ITLLSA** **LVETR**TIVRF NRPFLMIIVP TDTQNIFFMS
KVTNPKQA

B

C

...	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	...
1	114.1				I	1,215.7	608.4	1,198.7	1,197.7	11
2	215.1			197.1	T	1,102.6	551.8	1,085.6	1,084.6	10
3	328.2			310.2	L	1,001.6	501.3	984.6	983.6	9
4	441.3			423.3	L	888.5	444.8	871.5	870.5	8
5	528.3			510.3	S	775.4	388.2	758.4	757.4	7
6	599.4	300.2		581.4	A	688.4	344.7	671.4	670.4	6
7	712.5	356.7		694.5	L	617.4		600.3	599.4	5
8	811.5	406.3		793.5	V	504.3		487.3	486.3	4
9	940.6	470.8		922.6	E	405.2		388.2	387.2	3
10	1,041.6	521.3		1,023.6	T	276.2		259.1	258.2	2
11	1,215.7	608.4	1,198.7	1,197.7	R	175.1		158.1		1

Differently expressed protein feature identified by LC-MS/MS. A) SERPINA3 protein identified by 4 unique peptides (highlighted in yellow in the protein sequence). B) An MS/MS spectrum and C) table indicating the masses of identified fragment ions.

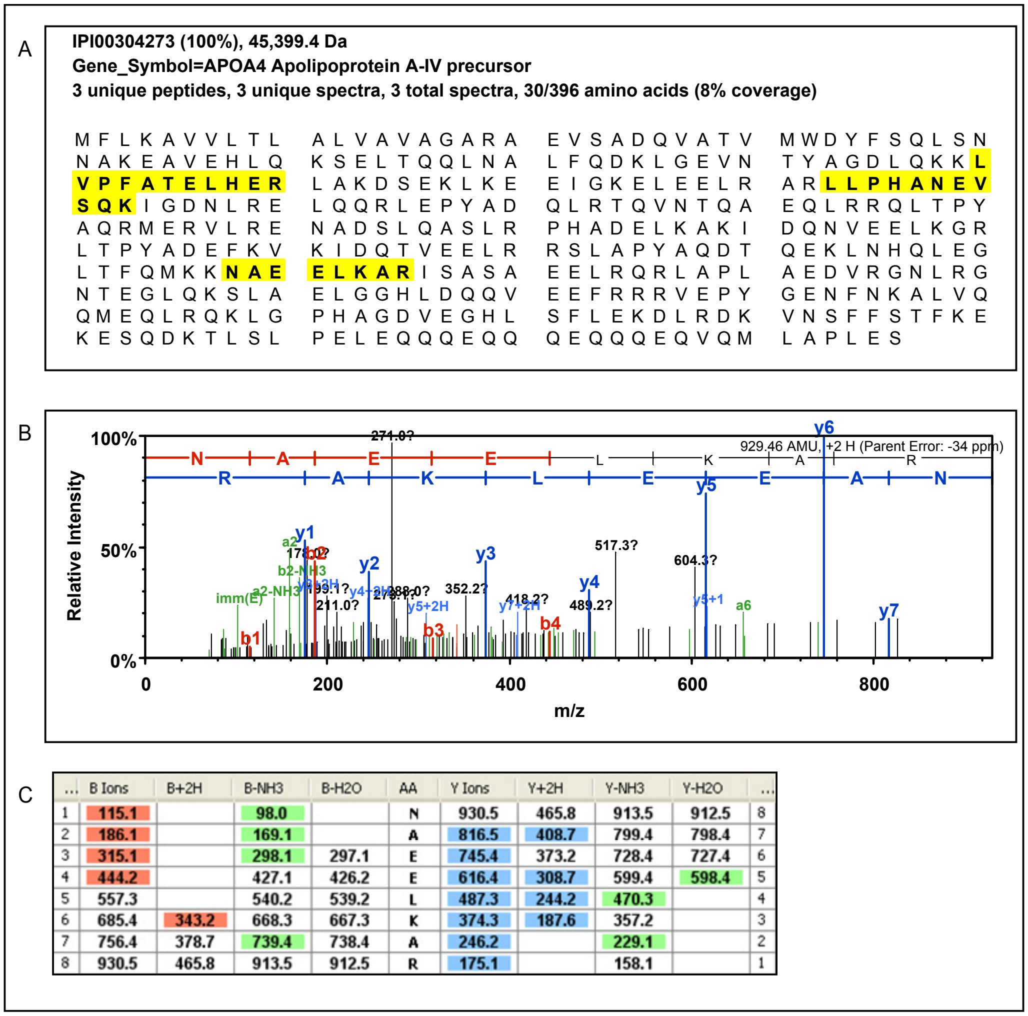


IPI00022431 (100%), 39,323.4 Da
Gene_Symbol=AHSG Alpha-2-HS-glycoprotein precursor
3 unique peptides, 3 unique spectra, 3 total spectra, 41/367 amino acids (11% coverage)

```
MKSLVLLLCL  AQLWGCHSAP  HGPGLIYRQP  NCDDPETEEA
ALVAIDYINQ  NLPWGYKHTL  NQIDEVKVWP  QQPSGELFEI
EIDTLETTCH  VLDPTPVARC  SVRQLKEHAV  EGDCDFQLLK
LDGKFSVVYA  KCDSSPDSAE  DVRKVCQDCP  LLAPLNDTRV
VHAAKAALAA  FNAQNNGSNF  QLEEISRAQL  VPLPPSTYVE
FTVSGTDCVA  KEATEAAKCN  LLAEKQYGFC  KATLSEKLGG
AEVAVTCTVF  QTQPVTSQPQ  PEGANEAVPT  PVVDPDAPPS
PPLGAPGLPP  AGSPPDSHVL  LAAPPGHQLH  RAHYDLRHTF
MGVVSLGSPS  GEVSHPRKTR  TVVQPSVGAA  AGPVVPPCPG
RIRHFKV
```

...	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	...
1	138.1	69.5			H	2,097.0	1,049.0	2,080.0	2,079.0	20
2	239.1	120.1		221.1	T	1,960.0	980.5	1,942.9	1,941.9	19
3	386.2	193.6		368.2	F	1,858.9	930.0	1,841.9	1,840.9	18
4	533.2	267.1		515.2	M+16	1,711.8	856.4	1,694.8	1,693.8	17
5	590.2	295.6		572.2	G	1,564.8	782.9	1,547.8	1,546.8	16
6	689.3	345.2		671.3	V	1,507.8	754.4	1,490.8	1,489.8	15
7	788.4	394.7		770.4	V	1,408.7	704.9	1,391.7	1,390.7	14
8	875.4	438.2		857.4	S	1,309.7	655.3	1,292.6	1,291.6	13
9	988.5	494.8		970.5	L	1,222.6	611.8	1,205.6	1,204.6	12
10	1,045.5	523.3		1,027.5	G	1,109.5	555.3	1,092.5	1,091.5	11
11	1,132.5	566.8		1,114.5	S	1,052.5	526.8	1,035.5	1,034.5	10
12	1,229.6	615.3		1,211.6	P	965.5	483.2	948.5	947.5	9
13	1,316.6	658.8		1,298.6	S	868.4	434.7	851.4	850.4	8
14	1,373.7	687.3		1,355.6	G	781.4	391.2	764.4	763.4	7
15	1,502.7	751.9		1,484.7	E	724.4	362.7	707.3	706.4	6
16	1,601.8	801.4		1,583.8	V	595.3	298.2	578.3	577.3	5
17	1,688.8	844.9		1,670.8	S	496.3	248.6	479.2	478.3	4
18	1,825.9	913.4		1,807.8	H	409.2	205.1	392.2		3
19	1,922.9	962.0		1,904.9	P	272.2		255.1		2
20	2,097.0	1,049.0	2,080.0	2,079.0	R	175.1		158.1		1

Differently expressed protein feature identified by LC-MS/MS. A) AHSG protein identified by 3 unique peptides (highlighted in yellow in the protein sequence). B) An MS/MS spectrum and C) table indicating the masses of identified fragment ions.

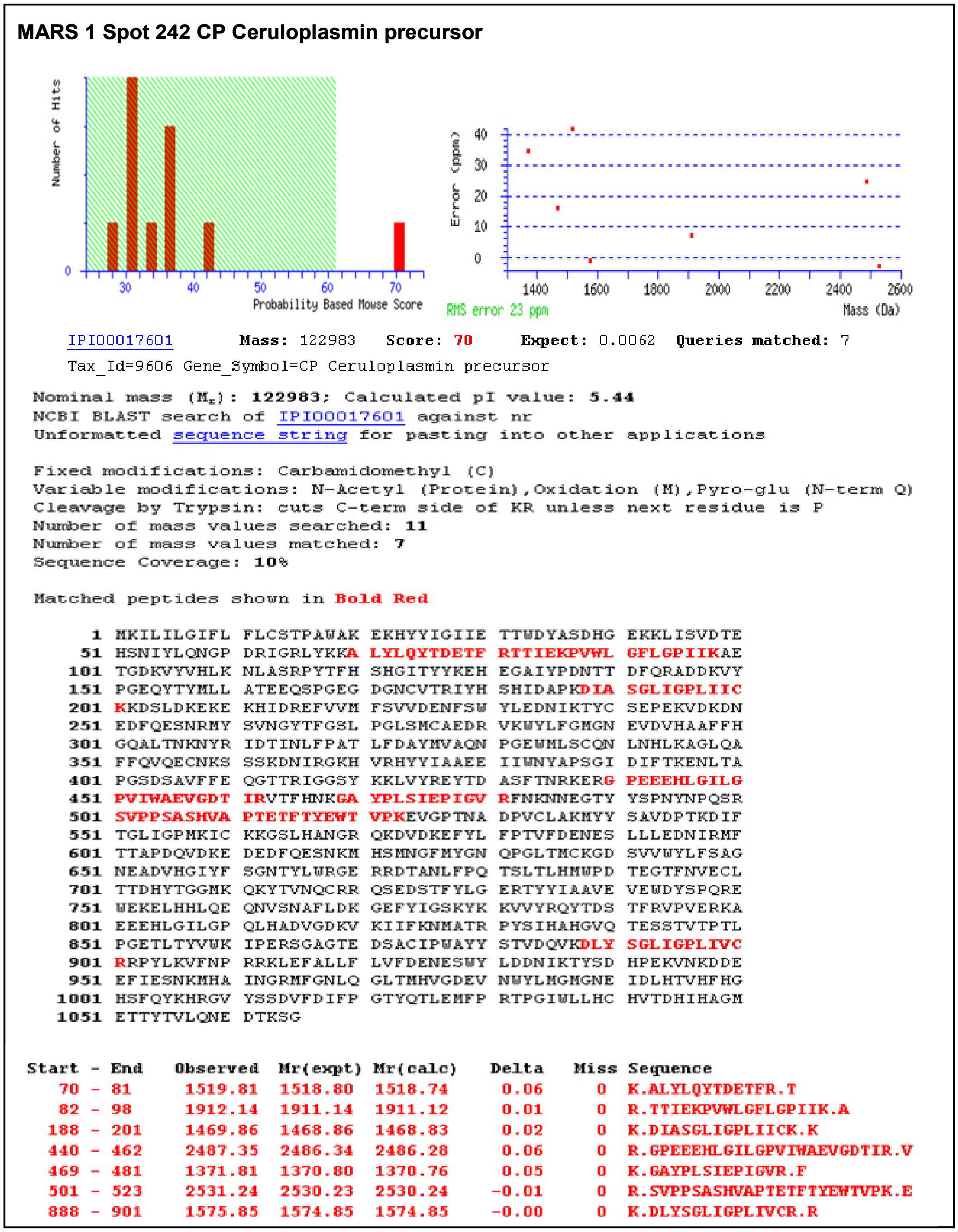


IPI00304273 (100%), 45,399.4 Da
Gene_Symbol=APOA4 Apolipoprotein A-IV precursor
3 unique peptides, 3 unique spectra, 3 total spectra, 30/396 amino acids (8% coverage)

MFLKAVVLTL ALVAVAGARA EVSADQVATV MWDYFSQLSN
NAKEAVEHLQ KSELTQQLNA LFQDKLGEVN TYAGDLQKK**L**
VPFATELHER LAKDSEKLKE EIGKELEELR AR**LLPHANEV**
SQKIGDNLRE LQQRLEPYAD QLRTQVNTQA EQLRRQLTPY
AQRMERVLRE NADSLQASLR PHADELKAKI DQNVEELKGR
LTPYADEFKV KIDQTVEELR RSLAPYAQDT QEKLNHQLEG
LTFQMKK**NAE** **ELKAR**ISASA EELRQRLAPL AEDVRGNLRG
NTEGLQKSLA ELGGHLDQQV EEFRRRVEPY GENFNKALVQ
QMEQLRQKLG PHAGDVEGHL SFLEKDLRDK VNSFFSTFKE
KESQDKTLSL PELEQQQEQQ QEQQQEQVQM LAPLES

...	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	...
1	115.1		98.0		N	930.5	465.8	913.5	912.5	8
2	186.1		169.1		A	816.5	408.7	799.4	798.4	7
3	315.1		298.1	297.1	E	745.4	373.2	728.4	727.4	6
4	444.2		427.1	426.2	E	616.4	308.7	599.4	598.4	5
5	557.3		540.2	539.2	L	487.3	244.2	470.3		4
6	685.4	343.2	668.3	667.3	K	374.3	187.6	357.2		3
7	756.4	378.7	739.4	738.4	A	246.2		229.1		2
8	930.5	465.8	913.5	912.5	R	175.1		158.1		1

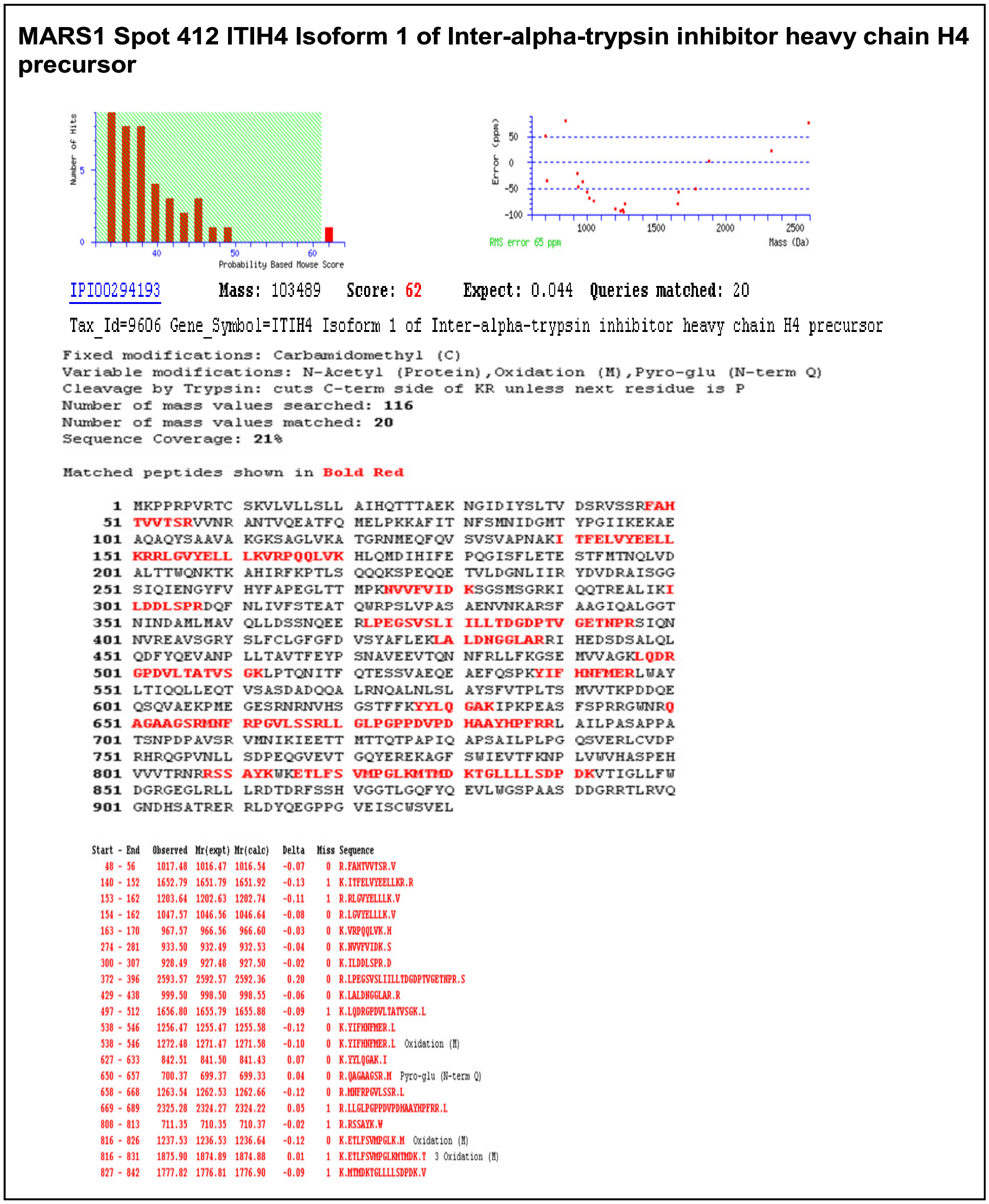
Differently expressed protein feature identified by LC-MS/MS. A) APOA4 protein identified by 3 unique peptides (highlighted in yellow in the protein sequence). B) An MS/MS spectrum and C) table indicating the masses of identified fragment ions.



MARS 1 Spot 242 CP Ceruloplasmin precursor

```
IPI00017601        Mass: 122983    Score: 70     Expect: 0.0062  Queries matched: 7
Tax_Id=9606 Gene_Symbol=CP Ceruloplasmin precursor

Nominal mass (Mr): 122983; Calculated pI value: 5.44
NCBI BLAST search of IPI00017601 against nr
Unformatted sequence string for pasting into other applications

Fixed modifications: Carbamidomethyl (C)
Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 11
Number of mass values matched: 7
Sequence Coverage: 10%

Matched peptides shown in Bold Red

   1 MKILILGIFL FLCSTPAWAK EKHYYIGIIE TTWDYASDHG EKKLISVDTE
  51 HSNIYLQNGP DRIGRLYKKA LYLQYTDETF RTTIEKPVWL GFLGPIIKAE
 101 TGDKVYVHLK NLASRPYTFH SHGITYYKEH EGAIYPDNTT DFQRADDKVY
 151 PGEQYTYMLL ATEEQSPGEG DGNCVTRIYH SHIDAPKDIA SGLIGPLIIC
 201 KKDSLDKEKE KHIDREFVVM FSVVDENFSW YLEDNIKTYC SEPEKVDKDN
 251 EDFQESNRMY SVNGYTFGSL PGLSMCAEDR VKWYLFGMGN EVDVHAAFFH
 301 GQALTNKNYR IDTINLFPAT LFDAYMVAQN PGEWMLSCQN LNHLKAGLQA
 351 FFQVQECNKS SSKDNIRGKH VRHYYIAAEE IIWNYAPSGI DIFTKENLTA
 401 PGSDSAVFFE QGTTRIGGSY KKLVYREYTD ASFTNRKERG PEEEHLGILG
 451 PVIWAEVGDT IRVTFHNKGA YPLSIEPIGV RFNKNNEGTY YSPNYNPQSR
 501 SVPPSASHVA PTETFTYEWT VPKEVGPTNA DPVCLAKMYY SAVDPTKDIF
 551 TGLIGPMKIC KKGSLHANGR QKDVDKEFYL FPTVFDENES LLLEDNIRMF
 601 TTAPDQVDKE DEDFQESNKM HSMNGFMYGN QPGLTMCKGD SVVWYLFSAG
 651 NEADVHGIYF SGNTYLWRGE RRDTANLFPQ TSLTLHMWPD TEGTFNVECL
 701 TTDHYTGGMK QKYTVNQCRR QSEDSTFYLG ERTYYIAAVE VEWDYSPQRE
 751 WEKELHHLQE QNVSNAFLDK GEFYIGSKYK KVVYRQYTDS TFRVPVERKA
 801 EEEHLGILGP QLHADVGDKV KIIFKNMATR PYSIHAHGVQ TESSTVTPTL
 851 PGETLTYVWK IPERSGAGTE DSACIPWAYY STVDQVKDLY SGLIGPLIVC
 901 RRPYLKVFNP RRKLEFALLF LVFDENESWY LDDNIKTYSD HPEKVNKDDE
 951 EFIESNKMHA INGRMFGNLQ GLTMHVGDEV NWYLMGMGNE IDLHTVHFHG
1001 HSFQYKHRGV YSSDVFDIFP GTYQTLEMFP RTPGIWLLHC HVTDHIHAGM
1051 ETTYTVLQNE DTKSG


Start - End     Observed   Mr(expt)  Mr(calc)    Delta   Miss  Sequence
   70 - 81      1519.81   1518.80   1518.74     0.06     0  K.ALYLQYTDETFR.T
   82 - 98      1912.14   1911.14   1911.12     0.01     0  R.TTIEKPVWLGFLGPIIK.A
  188 - 201     1469.86   1468.86   1468.83     0.02     0  K.DIASGLIGPLIICK.K
  440 - 462     2487.35   2486.34   2486.28     0.06     0  R.GPEEEHLGILGPVIWAEVGDTIR.V
  469 - 481     1371.81   1370.80   1370.76     0.05     0  K.GAYPLSIEPIGVR.F
  501 - 523     2531.24   2530.23   2530.24    -0.01     0  R.SVPPSASHVAPTETFTYEWTVPK.E
  888 - 901     1575.85   1574.85   1574.85    -0.00     0  K.DLYSGLIGPLIVCR.R
```

MALDI-TOF PMF protein identification. Prominent peaks in the mass range 700-4000 from the MALDI-TOF spectrum were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score, B) RMS based error scores for the peptide mass in the PMF and C) details of the identified protein's score, sequence including matched peptides are shown.



MARS1 Spot 412 ITIH4 Isoform 1 of Inter-alpha-trypsin inhibitor heavy chain H4 precursor

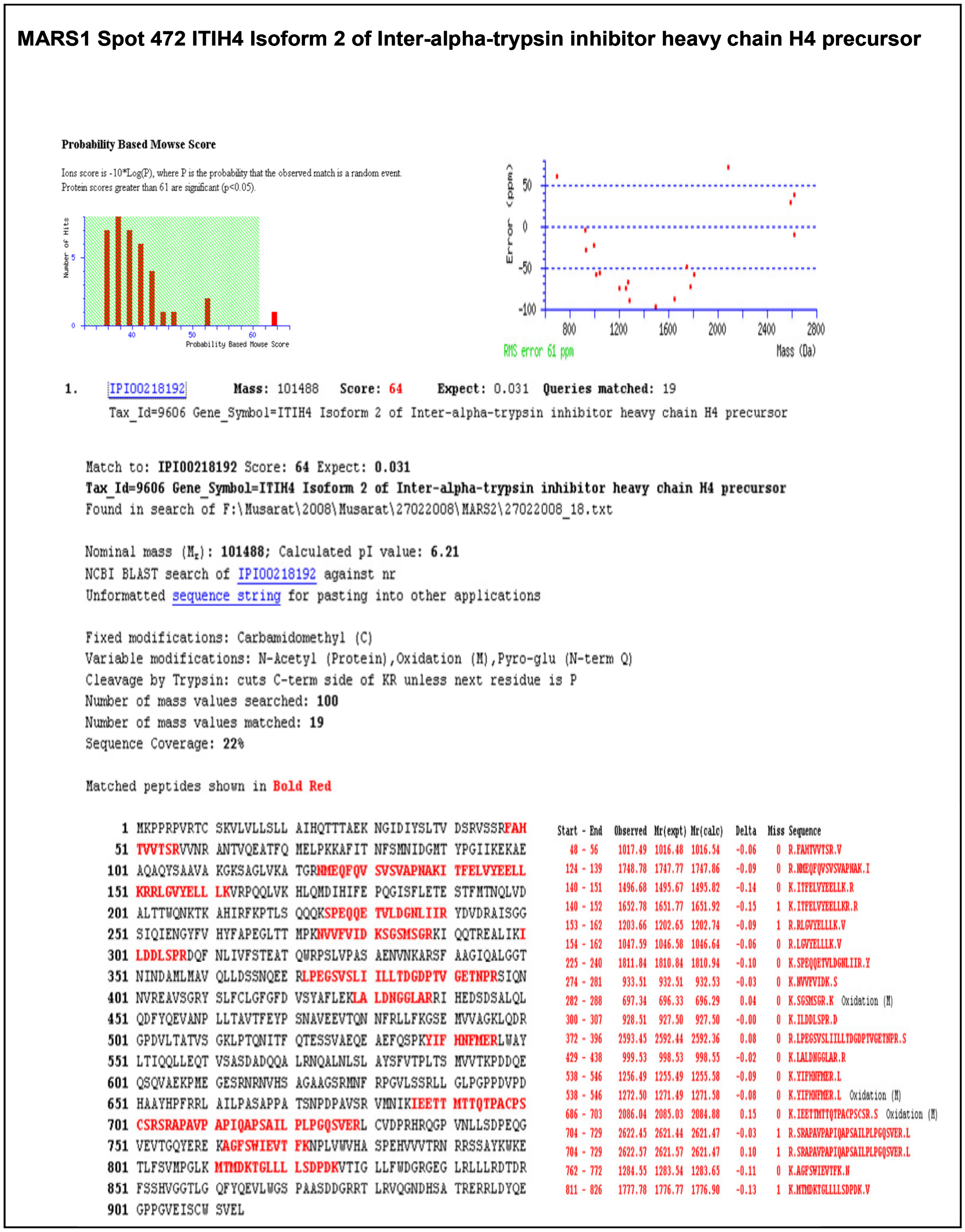
IPI00294193 **Mass:** 103489 **Score:** 62 **Expect:** 0.044 **Queries matched:** 20

Tax_Id=9606 Gene_Symbol=ITIH4 Isoform 1 of Inter-alpha-trypsin inhibitor heavy chain H4 precursor

```
Fixed modifications: Carbamidomethyl (C)
Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 116
Number of mass values matched: 20
Sequence Coverage: 21%

Matched peptides shown in Bold Red

  1 MKPPRPVRTC SKVLVLLSLL AIHQTTTAEK NGIDIYSLTV DSRVSSRFAH
 51 TVVTSRVVNR ANTVQEATFQ MELPKKAFIT NFSMNIDGMT YPGIIKEKAE
101 AQAQYSAAVA KGKSAGLVKA TGRNMEQFQV SVSVAPNAKI TFELVYEELL
151 KRRLGVYELL LKVRPQQLVK HLQMDIHIFE PQGISFLETE STFMTNQLVD
201 ALTTWQNKTK AHIRFKPTLS QQQKSPEQQE TVLDGNLIIR YDVDRAISGG
251 SIQIENGYFV HYFAPEGLTT MPKNVVFVID KSGSMSGRKI QQTREALIKI
301 LDDLSPRDQF NLIVFSTEAT QWRPSLVPAS AENVNKARSF AAGIQALGGT
351 NINDAMLMAV QLLDSSNQEE RLPEGSVSLI ILLTDGDPTV GETNPRSIQN
401 NVREAVSGRY SLFCLGFGFD VSYAFLEKLA LDNGGLARRI HEDSDSALQL
451 QDFYQEVANP LLTAVTFEYP SNAVEEVTQN NFRLLFKGSE MVVAGKLQDR
501 GPDVLTATVS GKLPTQNITF QTESSVAEQE AEFQSPKYIF HNFMERLWAY
551 LTIQQLLEQT VSASDADQQA LRNQALNLSL AYSFVTPLTS MVVTKPDDQE
601 QSQVAEKPME GESRNRNVHS GSTFFKYYLQ GAKIPKPEAS FSPRRGWNRQ
651 AGAAGSRMNF RPGVLSSRLL GLPGPPDVPD HAAYHPFRRL AILPASAPPA
701 TSNPDPAVSR VMNIKIEETT MTTQTPAPIQ APSAILPLPG QSVERLCVDP
751 RHRQGPVNLL SDPEQGVEVT GQYEREKAGF SWIEVTFKNP LVWVHASPEH
801 VVVTRNRRSS AYKWKETLFS VMPGLKMTMD KTGLLLLSDP DKVTIGLLFW
851 DGRGEGLRLL LRDTDRFSSH VGGTLGQFYQ EVLWGSPAAS DDGRRTLRVQ
901 GNDHSATRER RLDYQEGPPG VEISCWSVEL
```

```
Start - End   Observed  Mr(expt)  Mr(calc)   Delta  Miss  Sequence
  48 - 56     1017.48   1016.47   1016.54   -0.07     0  R.FAHTVVTSR.V
 140 - 152    1652.79   1651.79   1651.92   -0.13     1  K.ITFELVYEELLKR.R
 153 - 162    1203.64   1202.63   1202.74   -0.11     1  R.RLGVYELLLK.V
 154 - 162    1047.57   1046.56   1046.64   -0.08     0  R.LGVYELLLK.V
 163 - 170     967.57    966.56    966.60   -0.03     0  K.VRPQQLVK.H
 274 - 281     933.50    932.49    932.53   -0.04     0  K.NVVFVIDK.S
 300 - 307     928.49    927.48    927.50   -0.02     0  K.ILDDLSPR.D
 372 - 396    2593.57   2592.57   2592.36    0.20     0  R.LPEGSVSLIILLTDGDPTVGETNPR.S
 429 - 438     999.50    998.50    998.55   -0.06     0  K.LALDNGGLAR.R
 497 - 512    1656.80   1655.79   1655.88   -0.09     1  K.LQDRGPDVLTATVSGK.L
 538 - 546    1256.47   1255.47   1255.58   -0.12     0  K.YIFHNFMER.L
 538 - 546    1272.48   1271.47   1271.58   -0.10     0  K.YIFHNFMER.L  Oxidation (M)
 627 - 633     842.51    841.50    841.43    0.07     0  K.YYLQGAK.I
 650 - 657     700.37    699.37    699.33    0.04     0  R.QAGAAGSR.M  Pyro-glu (N-term Q)
 658 - 668    1263.54   1262.53   1262.66   -0.12     0  R.MNFRPGVLSSR.L
 669 - 689    2325.28   2324.27   2324.22    0.05     1  R.LLGLPGPPDVPDHAAYHPFRR.L
 808 - 813     711.35    710.35    710.37   -0.02     1  R.RSSAYK.W
 816 - 826    1237.53   1236.53   1236.64   -0.12     0  K.ETLFSVMPGLK.M  Oxidation (M)
 816 - 831    1875.90   1874.89   1874.88    0.01     1  K.ETLFSVMPGLKMTMDK.T  3 Oxidation (M)
 827 - 842    1777.82   1776.81   1776.90   -0.09     1  K.MTMDKTGLLLLSDPDK.V
```

MALDI-TOF PMF protein identification. Prominent peaks in the mass range 700-4000 from the MALDI-TOF spectrum were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score, B) RMS based error scores for the peptide mass in the PMF and C) details of the identified protein's score, sequence including matched peptides are shown.



MARS1 Spot 472 ITIH4 Isoform 2 of Inter-alpha-trypsin inhibitor heavy chain H4 precursor

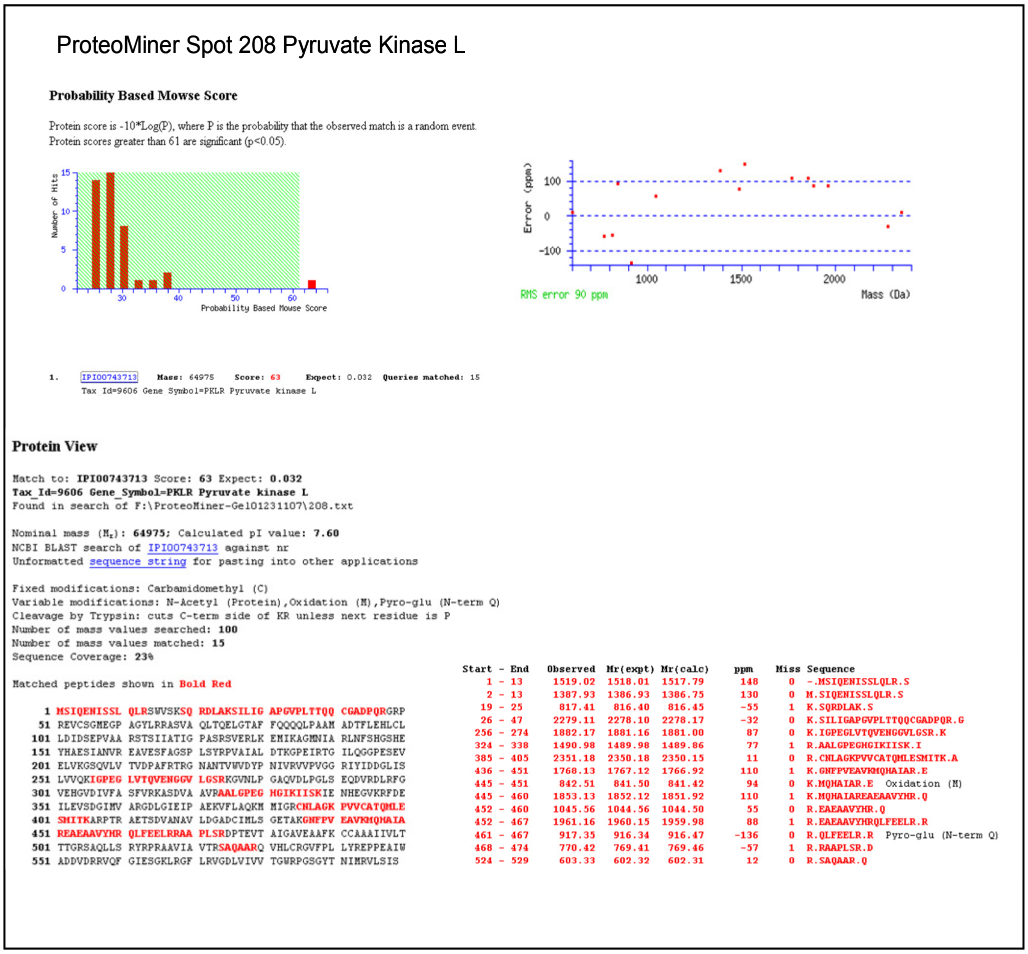
Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 61 are significant (p<0.05).

```
1.   IPI00218192     Mass: 101488    Score: 64     Expect: 0.031   Queries matched: 19
     Tax_Id=9606 Gene_Symbol=ITIH4 Isoform 2 of Inter-alpha-trypsin inhibitor heavy chain H4 precursor

   Match to: IPI00218192 Score: 64 Expect: 0.031
   Tax_Id=9606 Gene_Symbol=ITIH4 Isoform 2 of Inter-alpha-trypsin inhibitor heavy chain H4 precursor
   Found in search of F:\Musarat\2008\Musarat\27022008\MARS2\27022008_18.txt

   Nominal mass (Mr): 101488; Calculated pI value: 6.21
   NCBI BLAST search of IPI00218192 against nr
   Unformatted sequence string for pasting into other applications

   Fixed modifications: Carbamidomethyl (C)
   Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
   Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
   Number of mass values searched: 100
   Number of mass values matched: 19
   Sequence Coverage: 22%

   Matched peptides shown in Bold Red

     1 MKPPRPVRTC SKVLVLLSLL AIHQTTTAEK NGIDIYSLTV DSRVSSRFAH
    51 TVVTSRVVNR ANTVQEATFQ MELPKKAFIT NFSMNIDGMT YPGIIKEKAE
   101 AQAQYSAAVA KGKSAGLVKA TGRNMEQFQV SVSVAPNAKI TFELVYEELL
   151 KRRLGVYELL LKVRPQQLVK HLQMDIHIFE PQGISFLETE STFMTNQLVD
   201 ALTTWQNKTK AHIRFKPTLS QQQKSPEQQE TVLDGNLIIR YDVDRAISGG
   251 SIQIENGYFV HYFAPEGLTT MPKNVVFVID KSGSMSGRKI QQTREALIKI
   301 LDDLSPRDQF NLIVFSTEAT QWRPSLVPAS AENVNKARSF AAGIQALGGT
   351 NINDAMLMAV QLLDSSNQEE RLPEGSVSLI ILLTDGDPTV GETNPRSIQN
   401 NVREAVSGRY SLFCLGFGFD VSYAFLEKLA LDNGGLARRI HEDSDSALQL
   451 QDFYQEVANP LLTAVTFEYP SNAVEEVTQN NFRLLFKGSE MVVAGKLQDR
   501 GPDVLTATVS GKLPTQNITF QTESSVAEQE AEFQSPKYIF HNFMERLWAY
   551 LTIQQLLEQT VSASDADQQA LRNQALNLSL AYSFVTPLTS MVVTKPDDQE
   601 QSQVAEKPME GESRNRNVHS AGAAGSRMNF RPGVLSSRLL GLPGPPDVPD
   651 HAAYHPFRRL AILPASAPPA TSNPDPAVSR VMNIKIEETT MTTQTPACPS
   701 CSRSRAPAVP APIQAPSAIL PLPGQSVERL CVDPRHRQGP VNLLSDPEQG
   751 VEVTGQYERE KAGFSWIEVT FKNPLVWVHA SPEHVVVTRN RRSSAYKWKE
   801 TLFSVMPGLK MTMDKTGLLL LSDPDKVTIG LLFWDGRGEG LRLLLRDTDR
   851 FSSHVGGTLG QFYQEVLWGS PAASDDGRRT LRVQGNDHSA TRERRLDYQE
   901 GPPGVEISCW SVEL
```

```
Start - End   Observed  Mr(expt)  Mr(calc)   Delta   Miss Sequence
  48 - 56     1017.49   1016.48   1016.54    -0.06     0  R.FAHTVVTSR.V
 124 - 139    1748.78   1747.77   1747.86    -0.09     0  R.NMEQFQVSVSVAPNAK.I
 140 - 151    1496.68   1495.67   1495.82    -0.14     0  K.ITFELVYEELLK.R
 140 - 152    1652.78   1651.77   1651.92    -0.15     1  K.ITFELVYEELLKR.R
 153 - 162    1203.66   1202.65   1202.74    -0.09     1  R.RLGVYELLLK.V
 154 - 162    1047.59   1046.58   1046.64    -0.06     0  R.LGVYELLLK.V
 225 - 240    1811.84   1810.84   1810.94    -0.10     0  K.SPEQQETVLDGNLIIR.Y
 274 - 281     933.51    932.51    932.53    -0.03     0  K.NVVFVIDK.S
 282 - 288     697.34    696.33    696.29     0.04     0  K.SGSMSGR.K  Oxidation (M)
 300 - 307     928.51    927.50    927.50    -0.00     0  K.ILDDLSPR.D
 372 - 396    2593.45   2592.44   2592.36     0.08     0  R.LPEGSVSLIILLTDGDPTVGETNPR.S
 429 - 438     999.53    998.53    998.55    -0.02     0  K.LALDNGGLAR.R
 538 - 546    1256.49   1255.49   1255.58    -0.09     0  K.YIFHNFMER.L
 538 - 546    1272.50   1271.49   1271.58    -0.08     0  K.YIFHNFMER.L  Oxidation (M)
 686 - 703    2086.04   2085.03   2084.88     0.15     0  K.IEETTMTTQTPACPSCSR.S  Oxidation (M)
 704 - 729    2622.45   2621.44   2621.47    -0.03     1  R.SRAPAVPAPIQAPSAILPLPGQSVER.L
 704 - 729    2622.57   2621.57   2621.47     0.10     1  R.SRAPAVPAPIQAPSAILPLPGQSVER.L
 762 - 772    1284.55   1283.54   1283.65    -0.11     0  K.AGFSWIEVTFK.N
 811 - 826    1777.78   1776.77   1776.90    -0.13     1  K.MTMDKTGLLLLSDPDK.V
```

MALDI-TOF PMF protein identification. Prominent peaks in the mass range 700-4000 from the MALDI-TOF spectrum were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score. B) RMS based error scores for the peptide mass in the PMF. C) Details of the identified protein's score, sequence including matched peptides are shown.



ProteoMiner Spot 208 Pyruvate Kinase L

Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 61 are significant (p<0.05).

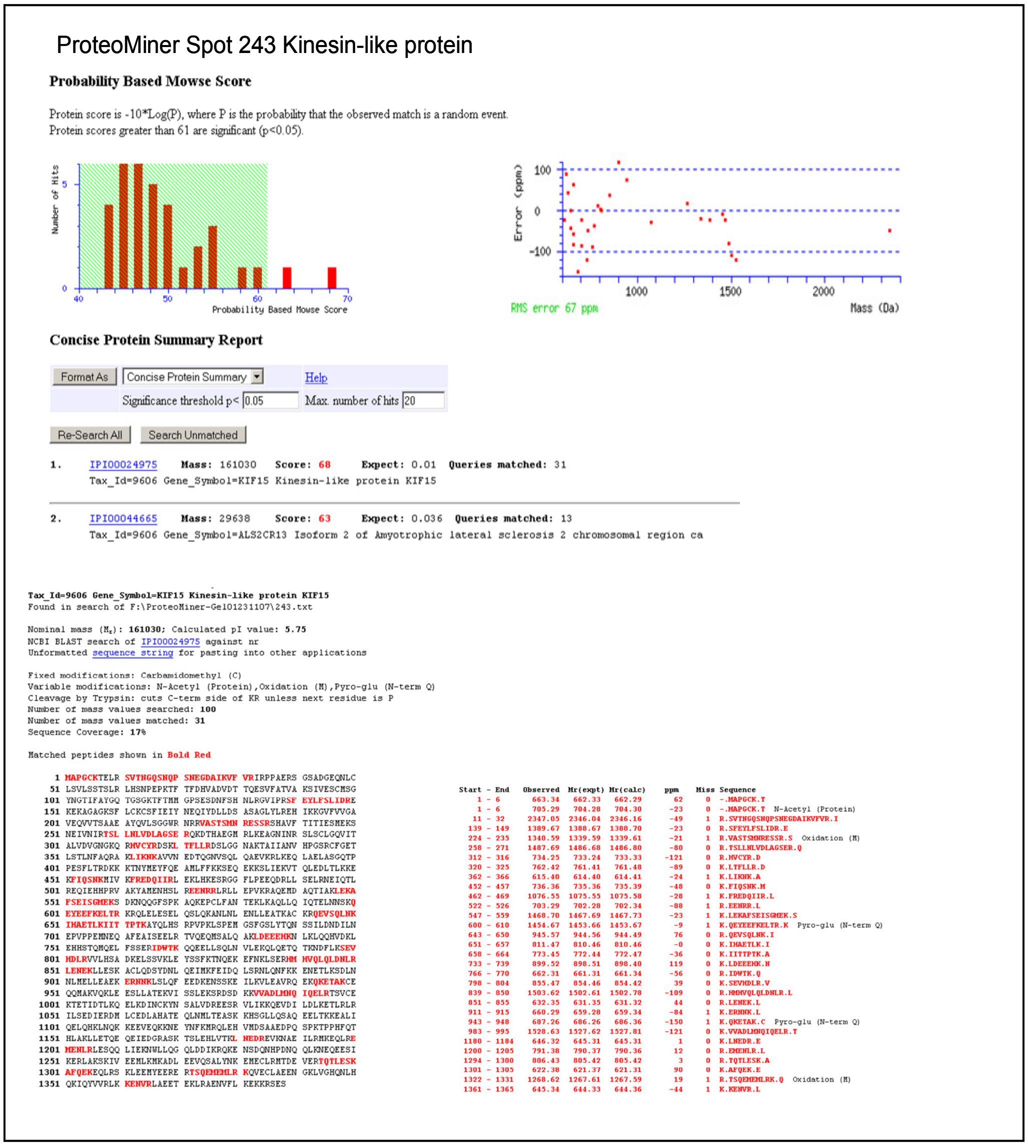
1. IPI00743713 **Mass:** 64975 **Score:** 63 **Expect:** 0.032 **Queries matched:** 15
Tax_Id=9606 Gene_Symbol=PKLR Pyruvate kinase L

Protein View

```
Match to: IPI00743713 Score: 63 Expect: 0.032
Tax_Id=9606 Gene_Symbol=PKLR Pyruvate kinase L
Found in search of F:\ProteoMiner-Gel01231107\208.txt

Nominal mass (Mr): 64975; Calculated pI value: 7.60
NCBI BLAST search of IPI00743713 against nr
Unformatted sequence string for pasting into other applications

Fixed modifications: Carbamidomethyl (C)
Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 100
Number of mass values matched: 15
Sequence Coverage: 23%

Matched peptides shown in Bold Red

  1 MSIQENISSL QLRSWVSKSQ RDLAKSILIG APGVPLTTQQ CGADPQRGRP
 51 REVCSGMEGP AGYLRRASVA QLTQELGTAF FQQQLPAAM ADTFLEHLCL
101 LDIDSEPVAA RSTSIIATIG PASRSVERLK EMIKAGMNIA RLNFSHGSHE
151 YHAESIANVR EAVESFAGSP LSYRPVAIAL DTKGPEIRTG ILQGGPESEV
201 ELVKGSQVLV TVDPAFRTRG NANTVWVDYP NIVRVVPVGG RIYIDDGLIS
251 LVVQKIGPEG LVTQVENGGV LGSRKGVNLP GAQVDLPGLS EQDVRDLRFG
301 VEHGVDIVFA SFVRKASDVA AVRAALGPEG HGIKIISKIE NHEGVKRFDE
351 ILEVSDGIMV ARGDLGIEIP AEKVFLAQKM MIGRCNLAGK PVVCATQMLE
401 SMITKARPTR AETSDVANAV LDGADCIMLS GETAKGNFPV EAVKMQHAIA
451 REAEAAVYHR QLFEELRRAA PLSRDPTEVT AIGAVEAAFK CCAAAIIVLT
501 TTGRSAQLLS RYRPRAAVIA VTRSAQAARQ VHLCRGVFPL LYREPPEAIW
551 ADDVDRRVQF GIESGKLRGF LRVGDLVIVV TGWRPGSGYT NIMRVLSIS
```

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Sequence
1 - 13	1519.02	1518.01	1517.79	148	0	-.MSIQENISSLQLR.S
2 - 13	1387.93	1386.93	1386.75	130	0	M.SIQENISSLQLR.S
19 - 25	817.41	816.40	816.45	-55	1	K.SQRDLAK.S
26 - 47	2279.11	2278.10	2278.17	-32	0	K.SILIGAPGVPLTTQQCGADPQR.G
256 - 274	1882.17	1881.16	1881.00	87	0	K.IGPEGLVTQVENGGVLGSR.K
324 - 338	1490.98	1489.98	1489.86	77	1	R.AALGPEGHGIKIISK.I
385 - 405	2351.18	2350.18	2350.15	11	0	R.CNLAGKPVVCATQMLESMITK.A
436 - 451	1768.13	1767.12	1766.92	110	1	K.GNFPVEAVKMQHAIAR.E
445 - 451	842.51	841.50	841.42	94	0	K.MQHAIAR.E Oxidation (M)
445 - 460	1853.13	1852.12	1851.92	110	1	K.MQHAIAREAEAAVYHR.Q
452 - 460	1045.56	1044.56	1044.50	55	0	R.EAEAAVYHR.Q
452 - 467	1961.16	1960.15	1959.98	88	1	R.EAEAAVYHRQLFEELR.R
461 - 467	917.35	916.34	916.47	-136	0	R.QLFEELR.R Pyro-glu (N-term Q)
468 - 474	770.42	769.41	769.46	-57	1	R.RAAPLSR.D
524 - 529	603.33	602.32	602.31	12	0	R.SAQAAR.Q

MALDI-TOF PMF protein identification. Prominent peaks in the mass range 700-4000 from the MALDI-TOF spectrum were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score, B) RMS based error scores for the peptide mass in the PMF and C) details of the identified protein's score, sequence including matched peptides are shown.



ProteoMiner Spot 243 Kinesin-like protein

Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 61 are significant (p<0.05).

Concise Protein Summary Report

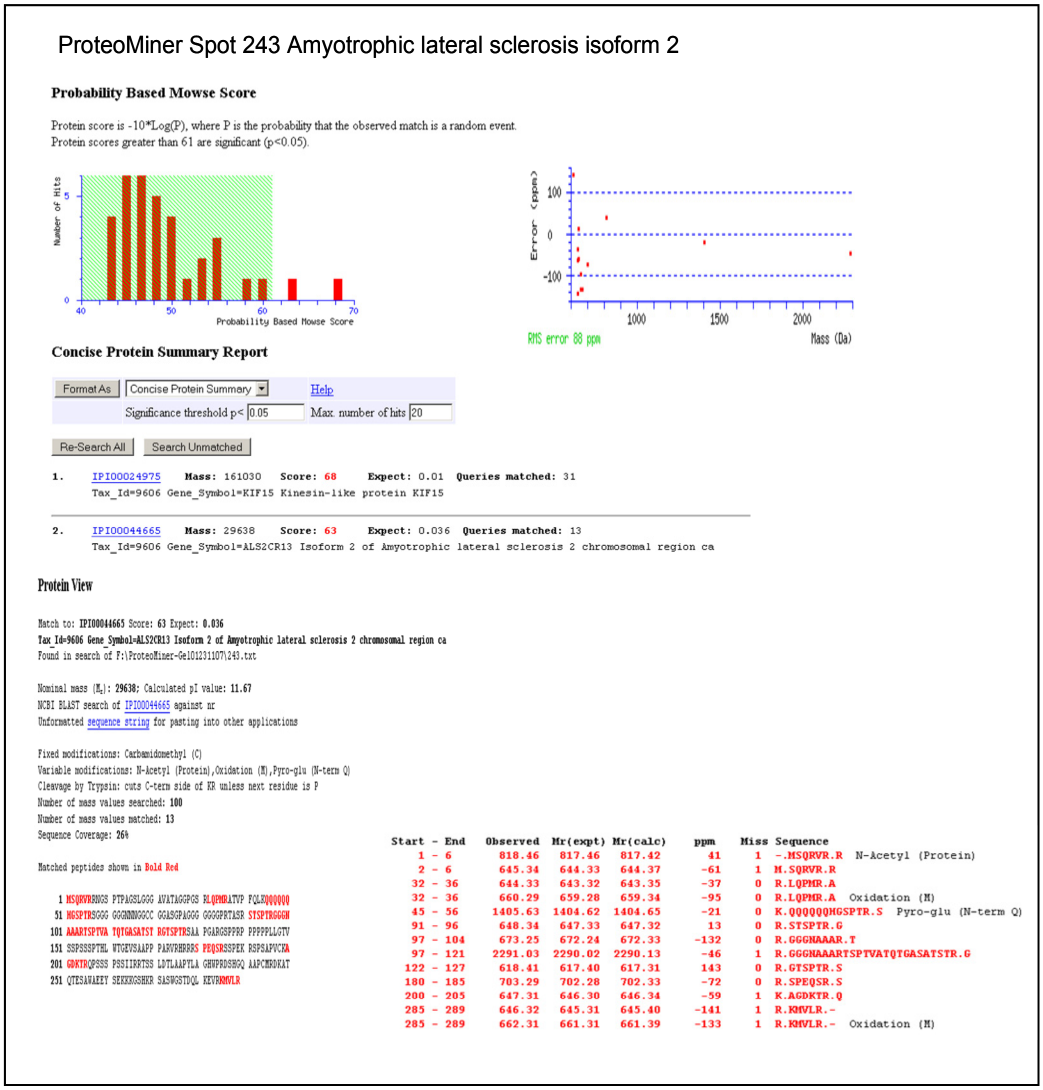
Significance threshold p< 0.05 Max. number of hits 20

```
1.    IPI00024975    Mass: 161030    Score: 68    Expect: 0.01   Queries matched: 31
      Tax_Id=9606 Gene_Symbol=KIF15 Kinesin-like protein KIF15

2.    IPI00044665    Mass: 29638     Score: 63    Expect: 0.036   Queries matched: 13
      Tax_Id=9606 Gene_Symbol=ALS2CR13 Isoform 2 of Amyotrophic lateral sclerosis 2 chromosomal region ca
```

```
Tax_Id=9606 Gene_Symbol=KIF15 Kinesin-like protein KIF15
Found in search of F:\ProteoMiner-Gel01231107\243.txt

Nominal mass (Mr): 161030; Calculated pI value: 5.75
NCBI BLAST search of IPI00024975 against nr
Unformatted sequence string for pasting into other applications

Fixed modifications: Carbamidomethyl (C)
Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 100
Number of mass values matched: 31
Sequence Coverage: 17%

Matched peptides shown in Bold Red

   1 MAPGCKTELR SVTNGQSNQP SNEGDAIKVF VRIRPPAERS GSADGEQNLC
  51 LSVLSSTSLR LHSNPEPKTF TFDHVADVDT TQESVFATVA KSIVESCMSG
 101 YNGTIFAYGQ TGSGKTFTMM GPSESDNFSH NLRGVIPRSF EYLFSLIDRE
 151 KEKAGAGKSF LCKCSFIEIY NEQIYDLLDS ASAGLYLREH IKKGVFVVGA
 201 VEQVVTSAAE AYQVLSGGWR NRRVASTSMN RESSRSHAVF TITIESMEKS
 251 NEIVNIRTSL LNLVDLAGSE RQKDTHAEGM RLKEAGNINR SLSCLGQVIT
 301 ALVDVGNGKQ RHVCYRDSKL TFLLRDSLGG NAKTAIIANV HPGSRCFGET
 351 LSTLNFAQRA KLIKNKAVVN EDTQGNVSQL QAEVKRLKEQ LAELASGQTP
 401 PESFLTRDKK KTNYMEYFQE AMLFFKKSEQ EKKSLIEKVT QLEDLTLKKE
 451 KFIQSNKMIV KFREDQIIRL EKLHKESRGG FLPEEQDRLL SELRNEIQTL
 501 REQIEHHPRV AKYAMENHSL REENRRLRLL EPVKRAQEMD AQTIAKLEKA
 551 FSEISGMEKS DKNQQGFSPK AQKEPCLFAN TEKLKAQLLQ IQTELNNSKQ
 601 EYEEFKELTR KRQLELESEL QSLQKANLNL ENLLEATKAC KRQEVSQLNK
 651 IHAETLKIIT TPTKAYQLHS RPVPKLSPEM GSFGSLYTQN SSILDNDILN
 701 EPVPPEMNEQ AFEAISEELR TVQEQMSALQ AKLDEEEHKN LKLQQHVDKL
 751 EHHSTQMQEL FSSERIDWTK QQEELLSQLN VLEKQLQETQ TKNDFLKSEV
 801 HDLRVVLHSA DKELSSVKLE YSSFKTNQEK EFNKLSERHM HVQLQLDNLR
 851 LENEKLLESK ACLQDSYDNL QEIMKFEIDQ LSRNLQNFKK ENETLKSDLN
 901 NLMELLEAEK ERNNKLSLQF EEDKENSSKE ILKVLEAVRQ EKQKETAKCE
 951 QQMAKVQKLE ESLLATEKVI SSLEKSRDSD KKVVADLMNQ IQELRTSVCE
1001 KTETIDTLKQ ELKDINCKYN SALVDREESR VLIKKQEVDI LDLKETLRLR
1051 ILSEDIERDM LCEDLAHATE QLNMLTEASK KHSGLLQSAQ EELTKKEALI
1101 QELQHKLNQK KEEVEQKKNE YNFKMRQLEH VMDSAAEDPQ SPKTPPHFQT
1151 HLAKLLETQE QEIEDGRASK TSLEHLVTKL NEDREVKNAE ILRMKEQLRE
1201 MENLRLESQQ LIEKNWLLQG QLDDIKRQKE NSDQNHPDNQ QLKNEQEESI
1251 KERLAKSKIV EEMLKMKADL EEVQSALYNK EMECLRMTDE VERTQTLESK
1301 AFQEKEQLRS KLEEMYEERE RTSQEMEMLR KQVECLAEEN GKLVGHQNLH
1351 QKIQYVVRLK KENVRLAEET EKLRAENVFL KEKKRSES
```

```
 Start - End     Observed   Mr(expt) Mr(calc)     ppm    Miss Sequence
    1 - 6        663.34     662.33    662.29      62      0  -.MAPGCK.T
    1 - 6        705.29     704.28    704.30     -23      0  -.MAPGCK.T  N-Acetyl (Protein)
   11 - 32      2347.05    2346.04   2346.16     -49      1  R.SVTNGQSNQPSNEGDAIKVFVR.I
  139 - 149     1389.67    1388.67   1388.70     -23      0  R.SFEYLFSLIDR.E
  224 - 235     1340.59    1339.59   1339.61     -21      1  R.VASTSMNRESSR.S  Oxidation (M)
  258 - 271     1487.69    1486.68   1486.80     -80      0  R.TSLLNLVDLAGSER.Q
  312 - 316      734.25     733.24    733.33    -121      0  R.HVCYR.D
  320 - 325      762.42     761.41    761.48     -89      0  K.LTFLLR.D
  362 - 366      615.40     614.40    614.41     -24      1  K.LIKNK.A
  452 - 457      736.36     735.36    735.39     -48      0  K.FIQSNK.M
  462 - 469     1076.55    1075.55   1075.58     -28      1  K.FREDQIIR.L
  522 - 526      703.29     702.28    702.34     -88      1  R.EENRR.L
  547 - 559     1468.70    1467.69   1467.73     -23      1  K.LEKAFSEISGMEK.S
  600 - 610     1454.67    1453.66   1453.67      -9      1  K.QEYEEFKELTR.K  Pyro-glu (N-term Q)
  643 - 650      945.57     944.56    944.49      76      0  R.QEVSQLNK.I
  651 - 657      811.47     810.46    810.46      -0      0  K.IHAETLK.I
  658 - 664      773.45     772.44    772.47     -36      0  K.IITTPTK.A
  733 - 739      899.52     898.51    898.40     119      0  K.LDEEEHK.N
  766 - 770      662.31     661.31    661.34     -56      0  R.IDWTK.Q
  798 - 804      855.47     854.46    854.42      39      0  K.SEVHDLR.V
  839 - 850     1503.62    1502.61   1502.78    -109      0  R.HMHVQLQLDNLR.L
  851 - 855      632.35     631.35    631.32      44      0  R.LENEK.L
  911 - 915      660.29     659.28    659.34     -84      1  K.ERNNK.L
  943 - 948      687.26     686.26    686.36    -150      1  K.QKETAK.C  Pyro-glu (N-term Q)
  983 - 995     1528.63    1527.62   1527.81    -121      0  K.VVADLMNQIQELR.T
 1180 - 1184     646.32     645.31    645.31       1      0  K.LNEDR.E
 1200 - 1205     791.38     790.37    790.36      12      0  R.EMENLR.L
 1294 - 1300     806.43     805.42    805.42       3      0  R.TQTLESK.A
 1301 - 1305     622.38     621.37    621.31      90      0  K.AFQEK.E
 1322 - 1331    1268.62    1267.61   1267.59      19      1  R.TSQEMEMLRK.Q  Oxidation (M)
 1361 - 1365     645.34     644.33    644.36     -44      1  K.KENVR.L
```

MALDI-TOF PMF protein identification. Prominent peaks in the mass range 700-4000 from the MALDI-TOF spectrum were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score, B) RMS based error scores for the peptide mass in the PMF and C) details of the identified protein's score, sequence including matched peptides are shown.



ProteoMiner Spot 243 Amyotrophic lateral sclerosis isoform 2

Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 61 are significant (p<0.05).

Concise Protein Summary Report

Significance threshold p< 0.05 Max. number of hits 20

```
1.   IPI00024975   Mass: 161030   Score: 68    Expect: 0.01   Queries matched: 31
     Tax_Id=9606 Gene_Symbol=KIF15 Kinesin-like protein KIF15

2.   IPI00044665   Mass: 29638    Score: 63    Expect: 0.036  Queries matched: 13
     Tax_Id=9606 Gene_Symbol=ALS2CR13 Isoform 2 of Amyotrophic lateral sclerosis 2 chromosomal region ca
```

Protein View

Match to: **IPI00044665** Score: **63** Expect: **0.036**
Tax_Id=9606 Gene_Symbol=ALS2CR13 Isoform 2 of Amyotrophic lateral sclerosis 2 chromosomal region ca
Found in search of F:\ProteoMiner-Gel01231107\243.txt

Nominal mass (M_r): **29638**; Calculated pI value: **11.67**
NCBI BLAST search of IPI00044665 against nr
Unformatted sequence string for pasting into other applications

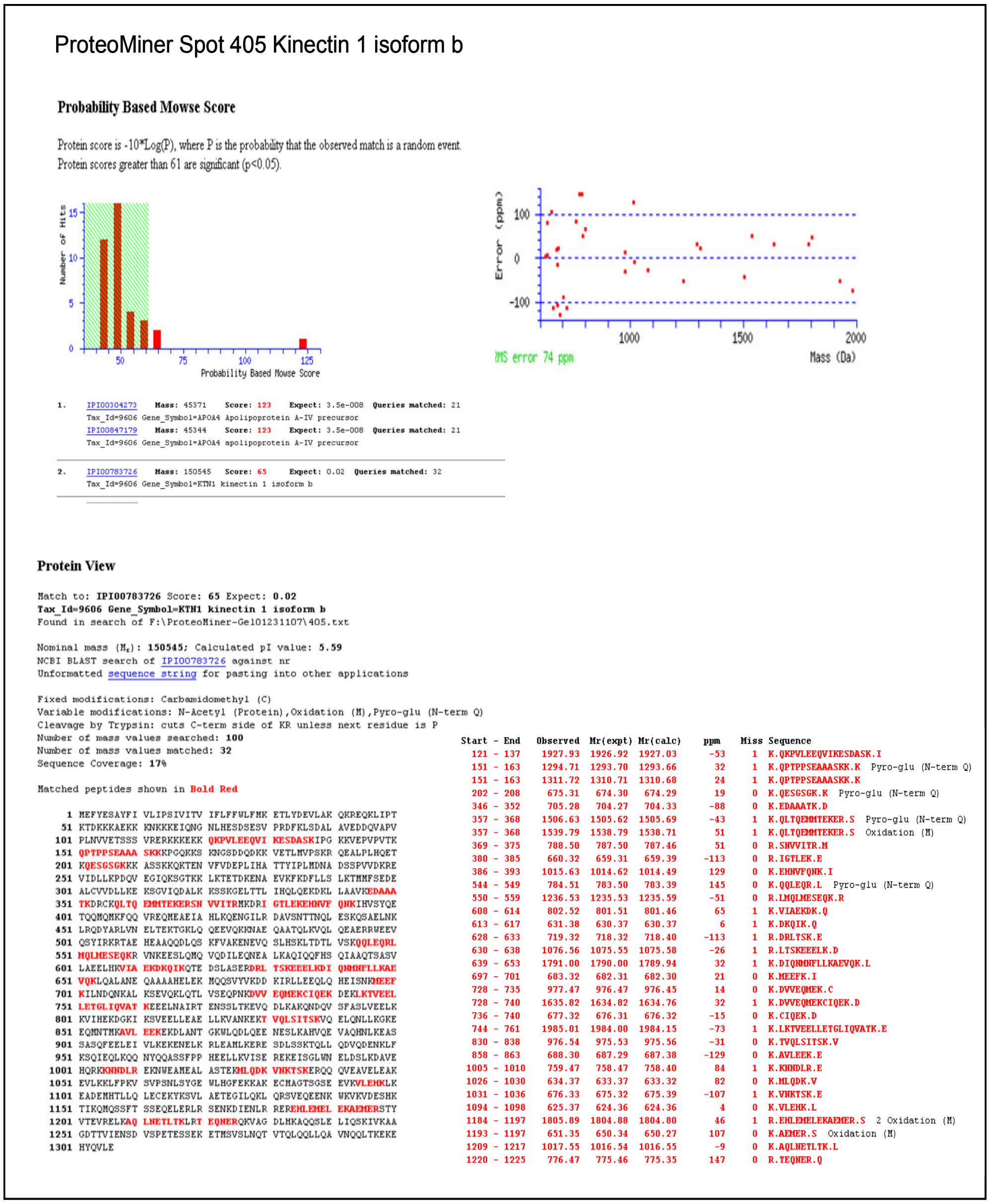
Fixed modifications: Carbamidomethyl (C)
Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: **100**
Number of mass values matched: **13**
Sequence Coverage: **26%**

Matched peptides shown in **Bold Red**

```
  1 MSQRVRRNGS PTPAGSLGGG AVATAGGPGS RLQPMRATVP FQLKQQQQQQ
 51 HGSPTRSGGG GGGNNNGGCC GGASGPAGGG GGGGPRTASR STSPTRGGGN
101 AAARTSPTVA TQTGASATST RGTSPTRSAA PGARGSPPRP PPPPPLLGTV
151 SSPSSSPTHL WTGEVSAAPP PARVRHRRRS PEQSRSSPEK RSPSAPVCKA
201 GDKTRQPSSS PSSIIRRTSS LDTLAAPYLA GHWPRDSHGQ AAPCMRDKAT
251 QTESAWAEEY SEKKKGSHKR SASWGSTDQL KEVRKMVLR
```

```
Start - End   Observed  Mr(expt) Mr(calc)   ppm   Miss Sequence
   1 - 6        818.46    817.46   817.42     41     1  -.MSQRVR.R  N-Acetyl (Protein)
   2 - 6        645.34    644.33   644.37    -61     1  M.SQRVR.R
  32 - 36       644.33    643.32   643.35    -37     0  R.LQPMR.A
  32 - 36       660.29    659.28   659.34    -95     0  R.LQPMR.A  Oxidation (M)
  45 - 56      1405.63   1404.62  1404.65    -21     0  K.QQQQQQHGSPTR.S  Pyro-glu (N-term Q)
  91 - 96       648.34    647.33   647.32     13     0  R.STSPTR.G
  97 - 104      673.25    672.24   672.33   -132     0  R.GGGNAAAR.T
  97 - 121     2291.03   2290.02  2290.13    -46     1  R.GGGNAAARTSPTVATQTGASATSTR.G
 122 - 127      618.41    617.40   617.31    143     0  R.GTSPTR.S
 180 - 185      703.29    702.28   702.33    -72     0  R.SPEQSR.S
 200 - 205      647.31    646.30   646.34    -59     1  K.AGDKTR.Q
 285 - 289      646.32    645.31   645.40   -141     1  R.KMVLR.-
 285 - 289      662.31    661.31   661.39   -133     1  R.KMVLR.-  Oxidation (M)
```

MALDI-TOF PMF protein identification. Prominent peaks in the mass range 700-4000 from the MALDI-TOF spectrum were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score, B) RMS based error scores for the peptide mass in the PMF and C) details of the identified protein's score, sequence including matched peptides are shown.



ProteoMiner Spot 405 Kinectin 1 isoform b

Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 61 are significant (p<0.05).

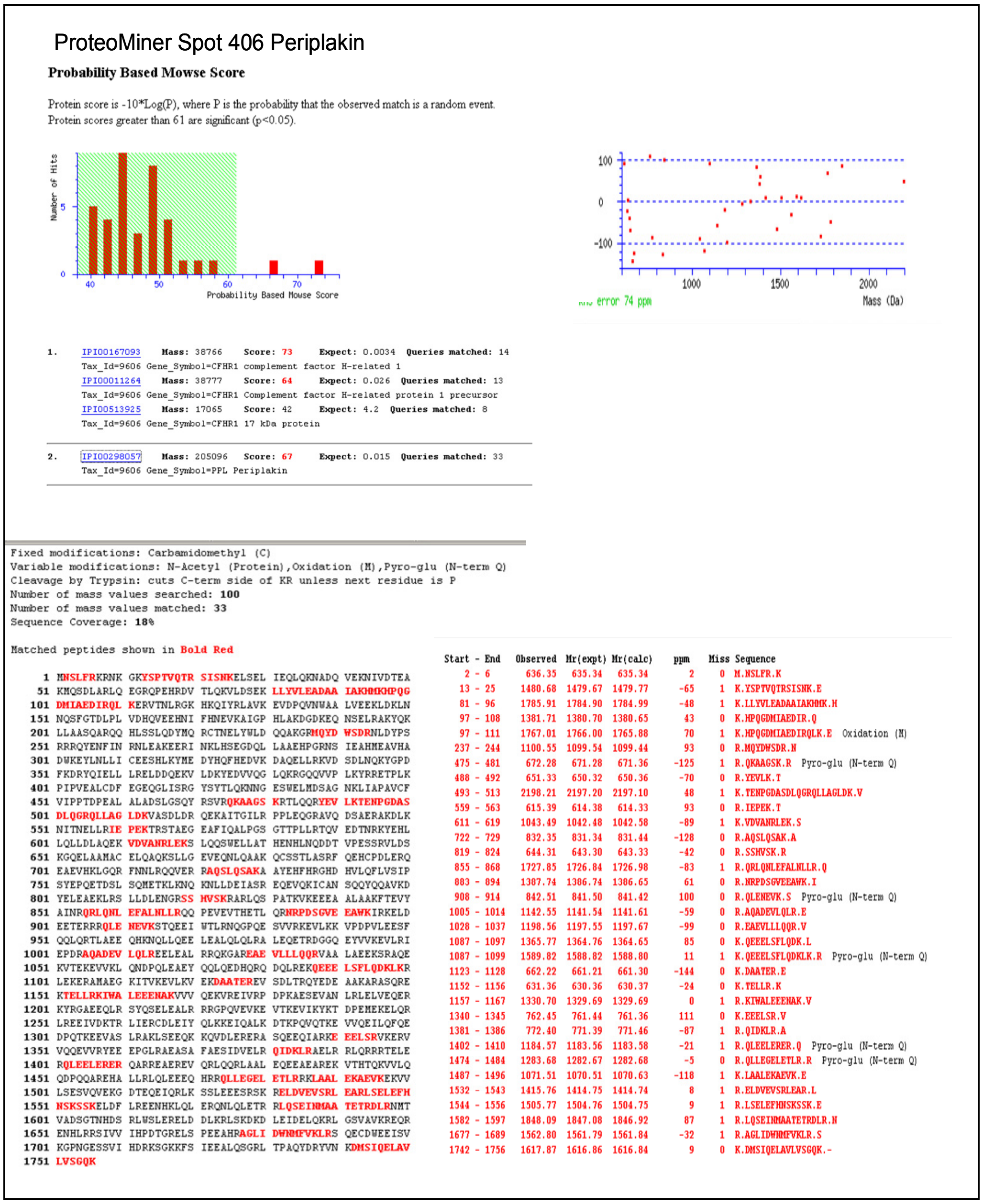
```
1.    IPI00304273    Mass: 45371    Score: 123    Expect: 3.5e-008  Queries matched: 21
      Tax_Id=9606 Gene_Symbol=APOA4 Apolipoprotein A-IV precursor
      IPI00847179    Mass: 45344    Score: 123    Expect: 3.5e-008  Queries matched: 21
      Tax_Id=9606 Gene_Symbol=APOA4 apolipoprotein A-IV precursor

2.    IPI00783726    Mass: 150545   Score: 65     Expect: 0.02  Queries matched: 32
      Tax_Id=9606 Gene_Symbol=KTN1 kinectin 1 isoform b
```

Protein View

```
Match to: IPI00783726 Score: 65 Expect: 0.02
Tax_Id=9606 Gene_Symbol=KTN1 kinectin 1 isoform b
Found in search of F:\ProteoMiner-Gel01231107\405.txt

Nominal mass (Mr): 150545; Calculated pI value: 5.59
NCBI BLAST search of IPI00783726 against nr
Unformatted sequence string for pasting into other applications

Fixed modifications: Carbamidomethyl (C)
Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 100
Number of mass values matched: 32
Sequence Coverage: 17%

Matched peptides shown in Bold Red

   1 MEFYESAYFI VLIPSIVITV IFLFFWLFMK ETLYDEVLAK QKREQKLIPT
  51 KTDKKKAEKK KNKKKEIQNG NLHESDSESV PRDFKLSDAL AVEDDQVAPV
 101 PLNVVETSSS VRERKKKEKK QKPVLEEQVI KESDASKIPG KKVEPVPVTK
 151 QPTPPSEAAA SKKKPGQKKS KNGSDDQDKK VETLMVPSKR QEALPLHQET
 201 KQESGSGKKK ASSKKQKTEN VFVDEPLIHA TTYIPLMDNA DSSPVVDKRE
 251 VIDLLKPDQV EGIQKSGTKK LKTETDKENA EVKFKDFLLS LKTMMFSEDE
 301 ALCVVDLLKE KSGVIQDALK KSSKGELTTL IHQLQEKDKL LAAVKEDAAA
 351 TKDRCKQLTQ EMMTEKERSN VVITRMKDRI GTLEKEHNVF QNKIHVSYQE
 401 TQQMQMKFQQ VREQMEAEIA HLKQENGILR DAVSNTTNQL ESKQSAELNK
 451 LRQDYARLVN ELTEKTGKLQ QEEVQKKNAE QAATQLKVQL QEAERRWEEV
 501 QSYIRKRTAE HEAAQQDLQS KFVAKENEVQ SLHSKLTDTL VSKQQLEQRL
 551 MQLMESEQKR VNKEESLQMQ VQDILEQNEA LKAQIQQFHS QIAAQTSASV
 601 LAEELHKVIA EKDKQIKQTE DSLASERDRL TSKEEELKDI QNMNFLLKAE
 651 VQKLQALANE QAAAAHELEK MQQSVYVKDD KIRLLEEQLQ HEISNKMEEF
 701 KILNDQNKAL KSEVQKLQTL VSEQPNKDVV EQMEKCIQEK DEKLKTVEEL
 751 LETGLIQVAT KEEELNAIRT ENSSLTKEVQ DLKAKQNDQV SFASLVEELK
 801 KVIHEKDGKI KSVEELLEAE LLKVANKEKT VQLSITSKVQ ELQNLLKGKE
 851 EQMNTMKAVL EEKEKDLANT GKWLQDLQEE NESLKAHVQE VAQHNLKEAS
 901 SASQFEELEI VLKEKENELK RLEAMLKERE SDLSSKTQLL QDVQDENKLF
 951 KSQIEQLKQQ NYQQASSFPP HEELLKVISE REKEISGLWN ELDSLKDAVE
1001 HQRKKNNDLR EKNWEAMEAL ASTEKMLQDK VNKTSKERQQ QVEAVELEAK
1051 EVLKKLFPKV SVPSNLSYGE WLHGFEKKAK ECMAGTSGSE EVKVLEHKLK
1101 EADEMHTLLQ LECEKYKSVL AETEGILQKL QRSVEQEENK WKVKVDESHK
1151 TIKQMQSSFT SSEQELERLR SENKDIENLR REREHLEMEL EKAEMERSTY
1201 VTEVRELKAQ LNETLTKLRT EQNERQKVAG DLHKAQQSLE LIQSKIVKAA
1251 GDTTVIENSD VSPETESSEK ETMSVSLNQT VTQLQQLLQA VNQQLTKEKE
1301 HYQVLE
```

```
Start - End   Observed  Mr(expt) Mr(calc)    ppm   Miss Sequence
  121 - 137   1927.93  1926.92  1927.03     -53     1  K.QKPVLEEQVIKESDASK.I
  151 - 163   1294.71  1293.70  1293.66      32     1  K.QPTPPSEAAASKK.K  Pyro-glu (N-term Q)
  151 - 163   1311.72  1310.71  1310.68      24     1  K.QPTPPSEAAASKK.K
  202 - 208    675.31   674.30   674.29      19     0  K.QESGSGK.K  Pyro-glu (N-term Q)
  346 - 352    705.28   704.27   704.33     -88     0  K.EDAAATK.D
  357 - 368   1506.63  1505.62  1505.69     -43     1  K.QLTQEMMTEKER.S  Pyro-glu (N-term Q)
  357 - 368   1539.79  1538.79  1538.71      51     1  K.QLTQEMMTEKER.S  Oxidation (M)
  369 - 375    788.50   787.50   787.46      51     0  R.SNVVITR.M
  380 - 385    660.32   659.31   659.39    -113     0  R.IGTLEK.E
  386 - 393   1015.63  1014.62  1014.49     129     0  K.EHNVFQNK.I
  544 - 549    784.51   783.50   783.39     145     0  K.QQLEQR.L  Pyro-glu (N-term Q)
  550 - 559   1236.53  1235.53  1235.59     -51     0  R.LMQLMESEQK.R
  608 - 614    802.52   801.51   801.46      65     1  K.VIAEKDK.Q
  613 - 617    631.38   630.37   630.37       6     1  K.DKQIK.Q
  628 - 633    719.32   718.32   718.40    -113     1  R.DRLTSK.E
  630 - 638   1076.56  1075.55  1075.58     -26     1  R.LTSKEEELK.D
  639 - 653   1791.00  1790.00  1789.94      32     1  K.DIQNMNFLLKAEVQK.L
  697 - 701    683.32   682.31   682.30      21     0  K.MEEFK.I
  728 - 735    977.47   976.47   976.45      14     0  K.DVVEQMEK.C
  728 - 740   1635.82  1634.82  1634.76      32     1  K.DVVEQMEKCIQEK.D
  736 - 740    677.32   676.31   676.32     -15     0  K.CIQEK.D
  744 - 761   1985.01  1984.00  1984.15     -73     1  K.LKTVEELLETGLIQVATK.E
  830 - 838    976.54   975.53   975.56     -31     0  K.TVQLSITSK.V
  858 - 863    688.30   687.29   687.38    -129     0  K.AVLEEK.E
 1005 - 1010   759.47   758.47   758.40      84     1  K.KNNDLR.E
 1026 - 1030   634.37   633.37   633.32      82     0  K.MLQDK.V
 1031 - 1036   676.33   675.32   675.39    -107     1  K.VNKTSK.E
 1094 - 1098   625.37   624.36   624.36       4     0  K.VLEHK.L
 1184 - 1197  1805.89  1804.88  1804.80      46     1  R.EHLEMELEKAEMER.S  2 Oxidation (M)
 1193 - 1197   651.35   650.34   650.27     107     0  K.AEMER.S  Oxidation (M)
 1209 - 1217  1017.55  1016.54  1016.55      -9     0  K.AQLNETLTK.L
 1220 - 1225   776.47   775.46   775.35     147     0  R.TEQNER.Q
```

MALDI-TOF PMF protein identification. Prominent peaks in the mass range 700-4000 from the MALDI-TOF spectrum were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score, B) RMS based error scores for the peptide mass in the PMF and C) details of the identified protein's score, sequence including matched peptides are shown.



ProteoMiner Spot 406 Periplakin

Probability Based Mowse Score

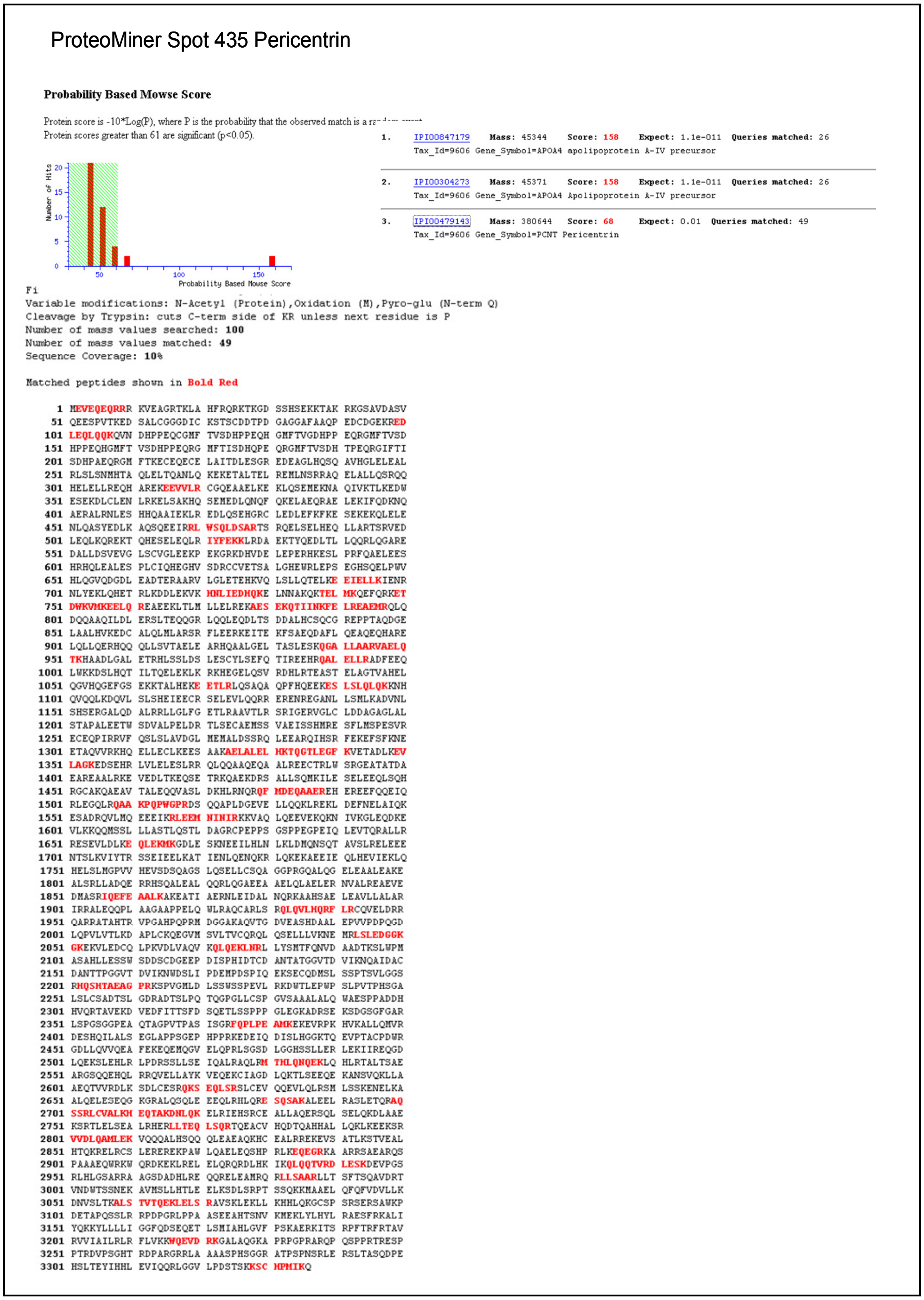
Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 61 are significant (p<0.05).

```
1.    IPI00167093    Mass: 38766    Score: 73    Expect: 0.0034  Queries matched: 14
      Tax_Id=9606 Gene_Symbol=CFHR1 complement factor H-related 1
      IPI00011264    Mass: 38777    Score: 64    Expect: 0.026  Queries matched: 13
      Tax_Id=9606 Gene_Symbol=CFHR1 Complement factor H-related protein 1 precursor
      IPI00513925    Mass: 17065    Score: 42    Expect: 4.2  Queries matched: 8
      Tax_Id=9606 Gene_Symbol=CFHR1 17 kDa protein

2.    IPI00298057    Mass: 205096   Score: 67    Expect: 0.015  Queries matched: 33
      Tax_Id=9606 Gene_Symbol=PPL Periplakin
```

```
Fixed modifications: Carbamidomethyl (C)
Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 100
Number of mass values matched: 33
Sequence Coverage: 18%

Matched peptides shown in Bold Red

   1 MNSLFRKRNK GKYSPTVQTR SISNKELSEL IEQLQKNADQ VEKNIVDTEA
  51 KMQSDLARLQ EGRQPEHRDV TLQKVLDSEK LLYVLEADAA IAKHMKHPQG
 101 DMIAEDIRQL KERVTNLRGK HKQIYRLAVK EVDPQVNWAA LVEEKLDKLN
 151 NQSFGTDLPL VDHQVEEHNI FHNEVKAIGP HLAKDGDKEQ NSELRAKYQK
 201 LLAASQARQQ HLSSLQDYMQ RCTNELYWLD QQAKGRMQYD WSDRNLDYPS
 251 RRRQYENFIN RNLEAKEERI NKLHSEGDQL LAAEHPGRNS IEAHMEAVHA
 301 DWKEYLNLLI CEESHLKYME DYHQFHEDVK DAQELLRKVD SDLNQKYGPD
 351 FKDRYQIELL LRELDDQEKV LDKYEDVVQG LQKRGQQVVP LKYRRETPLK
 401 PIPVEALCDF EGEQGLISRG YSYTLQKNNG ESWELMDSAG NKLIAPAVCF
 451 VIPPTDPEAL ALADSLGSQY RSVRQKAAGS KRTLQQRYEV LKTENPGDAS
 501 DLQGRQLLAG LDKVASDLDR QEKAITGILR PPLEQGRAVQ DSAERAKDLK
 551 NITNELLRIE PEKTRSTAEG EAFIQALPGS GTTPLLRTQV EDTNRKYEHL
 601 LQLLDLAQEK VDVANRLEKS LQQSWELLAT HENHLNQDDT VPESSRVLDS
 651 KGQELAAMAC ELQAQKSLLG EVEQNLQAAK QCSSTLASRF QEHCPDLERQ
 701 EAEVHKLGQR FNNLRQQVER RAQSLQSAKA AYEHFHRGHD HVLQFLVSIP
 751 SYEPQETDSL SQMETKLKNQ KNLLDEIASR EQEVQKICAN SQQYQQAVKD
 801 YELEAEKLRS LLDLENGRSS HVSKRARLQS PATKVKEEEA ALAAKFTEVY
 851 AINRQRLQNL EFALNLLRQQ PEVEVTHETL QRNRPDSGVE EAWKIRKELD
 901 EETERRRQLE NEVKSTQEEI WTLRNQGPQE SVVRKEVLKK VPDPVLEESF
 951 QQLQRTLAEE QHKNQLLQEE LEALQLQLRA LEQETRDGGQ EYVVKEVLRI
1001 EPDRAQADEV LQLREELEAL RRQKGAREAE VLLLQQRVAA LAEEKSRAQE
1051 KVTEKEVVKL QNDPQLEAEY QQLQEDHQRQ DQLREKQEEE LSFLQDKLKR
1101 LEKERAMAEG KITVKEVLKV EKDAATEREV SDLTRQYEDE AAKARASQRE
1151 KTELLRKIWA LEEENAKVVV QEKVREIVRP DPKAESEVAN LRLELVEQER
1201 KYRGAEEQLR SYQSELEALR RRGPQVEVKE VTKEVIKYKT DPEMEKELQR
1251 LREEIVDKTR LIERCDLEIY QLKKEIQALK DTKPQVQTKE VVQEILQFQE
1301 DPQTKEEVAS LRAKLSEEQK KQVDLERERA SQEEQIARKE EELSRVKERV
1351 VQQEVVRYEE EPGLRAEASA FAESIDVELR QIDKLRAELR RLQRRRTELE
1401 RQLEELERER QARREAEREV QRLQQRLAAL EQEEAEAREK VTHTQKVVLQ
1451 QDPQQAREHA LLRLQLEEEQ HRRQLLEGEL ETLRRKLAAL EKAEVKEKVV
1501 LSESVQVEKG DTEQEIQRLK SSLEEESRSK RELDVEVSRL EARLSELEFH
1551 NSKSSKELDF LREENHKLQL ERQNLQLETR RLQSEINMAA TETRDLRNMT
1601 VADSGTNHDS RLWSLERELD DLKRLSKDKD LEIDELQKRL GSVAVKREQR
1651 ENHLRRSIVV IHPDTGRELS PEEAHRAGLI DWNMFVKLRS QECDWEEISV
1701 KGPNGESSVI HDRKSGKKFS IEEALQSGRL TPAQYDRYVN KDMSIQELAV
1751 LVSGQK
```

```
Start - End   Observed  Mr(expt) Mr(calc)   ppm  Miss Sequence
   2 - 6        636.35    635.34   635.34     2    0  M.NSLFR.K
  13 - 25      1480.68   1479.67  1479.77   -65    1  K.YSPTVQTRSISNK.E
  81 - 96      1785.91   1784.90  1784.99   -48    1  K.LLYVLEADAAIAKHMK.H
  97 - 108     1381.71   1380.70  1380.65    43    0  K.HPQGDMIAEDIR.Q
  97 - 111     1767.01   1766.00  1765.88    70    1  K.HPQGDMIAEDIRQLK.E  Oxidation (M)
 237 - 244     1100.55   1099.54  1099.44    93    0  R.MQYDWSDR.N
 475 - 481      672.28    671.28   671.36  -125    1  R.QKAAGSK.R  Pyro-glu (N-term Q)
 488 - 492      651.33    650.32   650.36   -70    0  R.YEVLK.T
 493 - 513     2198.21   2197.20  2197.10    48    1  K.TENPGDASDLQGRQLLAGLDK.V
 559 - 563      615.39    614.38   614.33    93    0  R.IEPEK.T
 611 - 619     1043.49   1042.48  1042.58   -89    1  K.VDVANRLEK.S
 722 - 729      832.35    831.34   831.44  -128    0  R.AQSLQSAK.A
 819 - 824      644.31    643.30   643.33   -42    0  R.SSHVSK.R
 855 - 868     1727.85   1726.84  1726.98   -83    1  R.QRLQNLEFALNLLR.Q
 883 - 894     1387.74   1386.74  1386.65    61    0  R.NRPDSGVEEAWK.I
 908 - 914      842.51    841.50   841.42   100    0  R.QLENEVK.S  Pyro-glu (N-term Q)
1005 - 1014    1142.55   1141.54  1141.61   -59    0  R.AQADEVLQLR.E
1028 - 1037    1198.56   1197.55  1197.67   -99    0  R.EAEVLLLQQR.V
1087 - 1097    1365.77   1364.76  1364.65    85    0  K.QEEELSFLQDK.L
1087 - 1099    1589.82   1588.82  1588.80    11    1  K.QEEELSFLQDKLK.R  Pyro-glu (N-term Q)
1123 - 1128     662.22    661.21   661.30  -144    0  K.DAATER.E
1152 - 1156     631.36    630.36   630.37   -24    0  K.TELLR.K
1157 - 1167    1330.70   1329.69  1329.69     0    1  R.KIWALEEENAK.V
1340 - 1345     762.45    761.44   761.36   111    0  K.EEELSR.V
1381 - 1386     772.40    771.39   771.46   -87    1  R.QIDKLR.A
1402 - 1410    1184.57   1183.56  1183.58   -21    1  R.QLEELERER.Q  Pyro-glu (N-term Q)
1474 - 1484    1283.68   1282.67  1282.68    -5    0  R.QLLEGELETLR.R  Pyro-glu (N-term Q)
1487 - 1496    1071.51   1070.51  1070.63  -118    1  K.LAALEKAEVK.E
1532 - 1543    1415.76   1414.75  1414.74     8    1  R.ELDVEVSRLEAR.L
1544 - 1556    1505.77   1504.76  1504.75     9    1  R.LSELEFHNSKSSK.E
1582 - 1597    1848.09   1847.08  1846.92    87    1  R.LQSEINMAATETRDLR.N
1677 - 1689    1562.80   1561.79  1561.84   -32    1  R.AGLIDWNMFVKLR.S
1742 - 1756    1617.87   1616.86  1616.84     9    0  K.DMSIQELAVLVSGQK.-
```

MALDI-TOF PMF protein identification. Prominent peaks in the mass range 700-4000 from the MALDI-TOF spectrum were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score, B) RMS based error scores for the peptide mass in the PMF and C) details of the identified protein's score, sequence including matched peptides are shown.



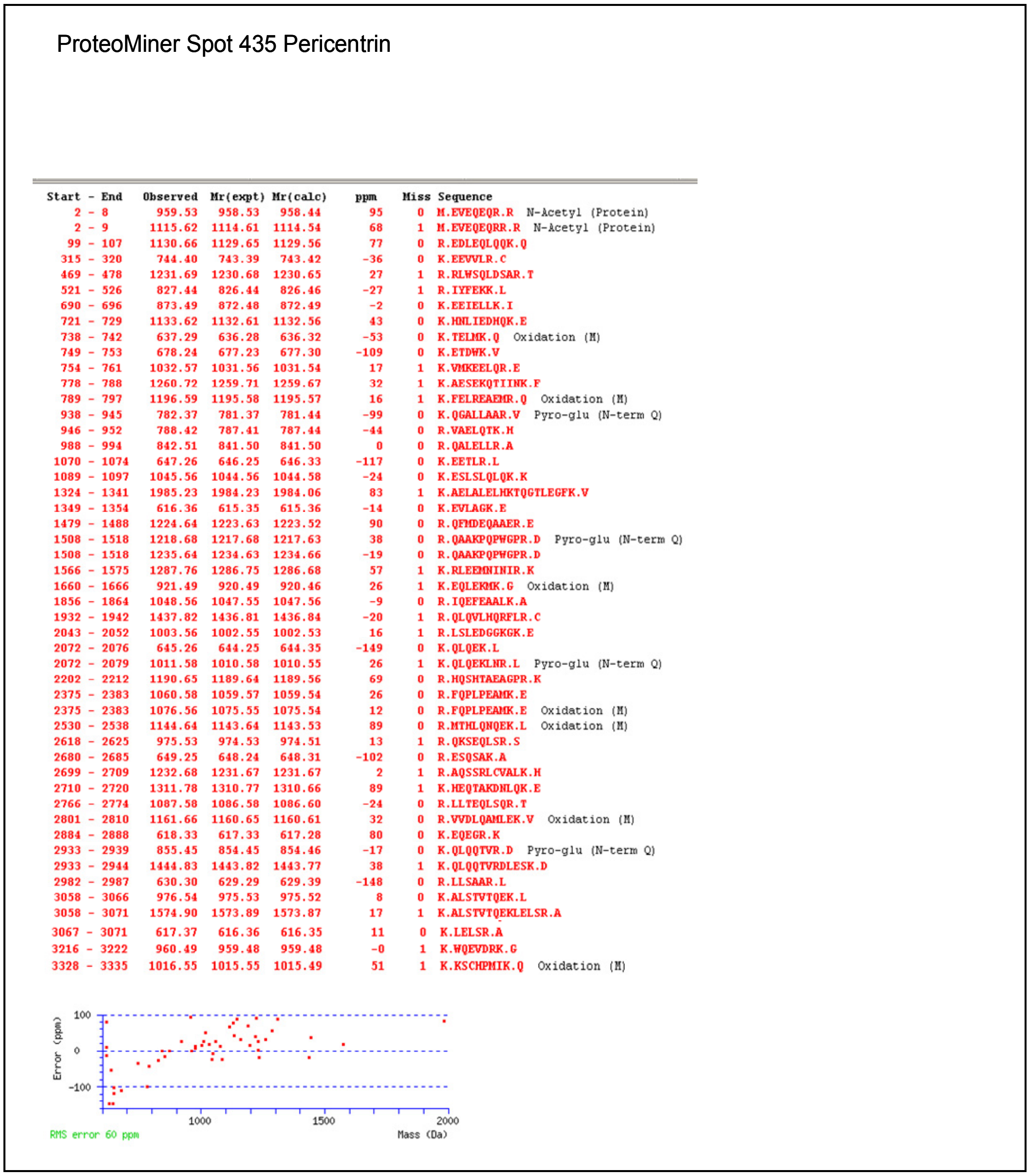
ProteoMiner Spot 435 Pericentrin

Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a ra
Protein scores greater than 61 are significant (p<0.05).

```
1.  IPI00847179   Mass: 45344    Score: 158    Expect: 1.1e-011  Queries matched: 26
    Tax_Id=9606 Gene_Symbol=APOA4 apolipoprotein A-IV precursor

2.  IPI00304273   Mass: 45371    Score: 158    Expect: 1.1e-011  Queries matched: 26
    Tax_Id=9606 Gene_Symbol=APOA4 Apolipoprotein A-IV precursor

3.  IPI00479143   Mass: 380644   Score: 68     Expect: 0.01   Queries matched: 49
    Tax_Id=9606 Gene_Symbol=PCNT Pericentrin
```

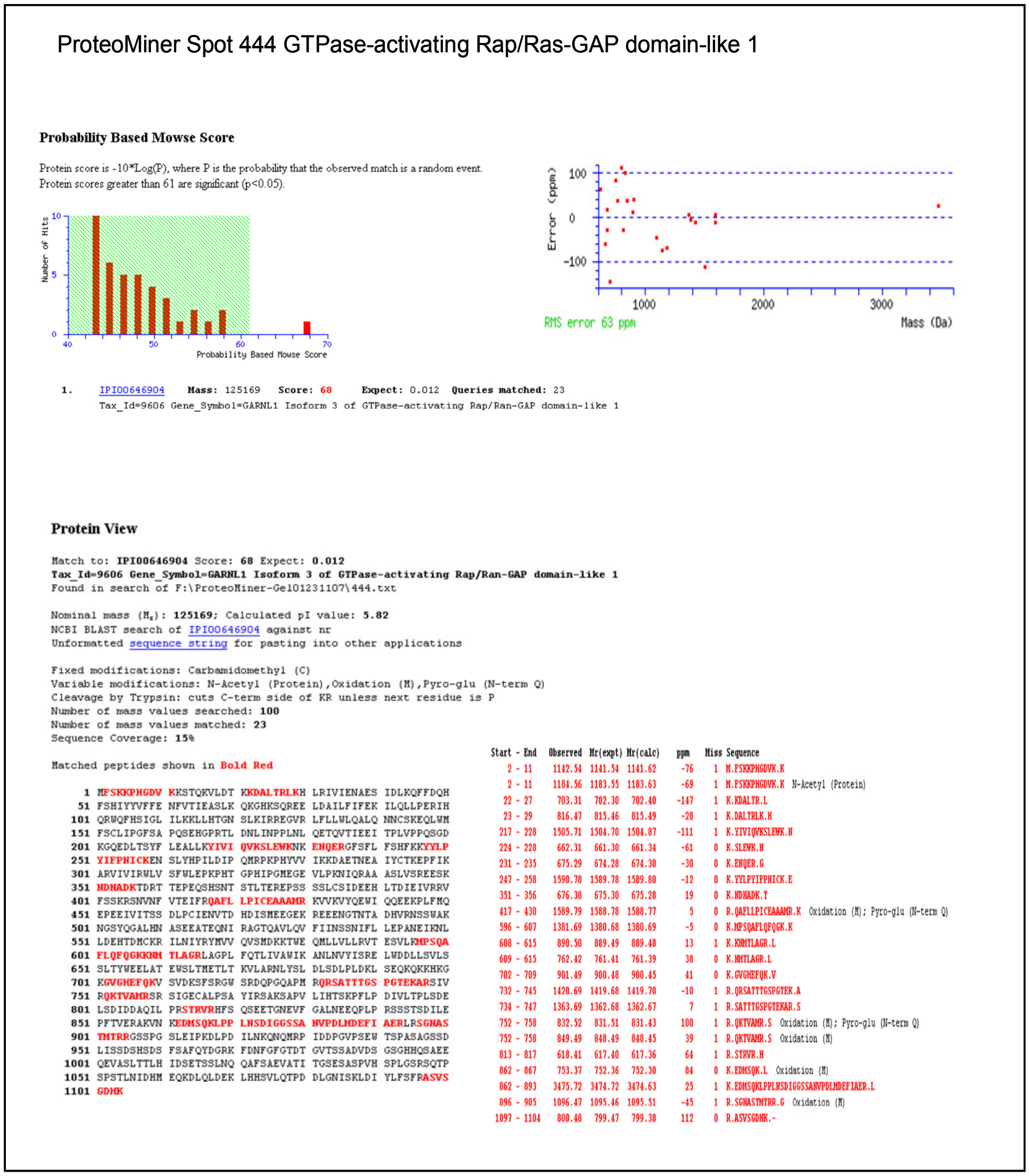
```
Fi
Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 100
Number of mass values matched: 49
Sequence Coverage: 10%

Matched peptides shown in Bold Red

   1 MEVEQEQRRR KVEAGRTKLA HFRQRKTKGD SSHSEKKTAK RKGSAVDASV
  51 QEESPVTKED SALCGGGDIC KSTSCDDTPD GAGGAFAAQP EDCDGEKRED
 101 LEQLQQKQVN DHPPEQCGMF TVSDHPPEQH GMFTVGDHPP EQRGMFTVSD
 151 HPPEQHGMFT VSDHPPEQRG MFTISDHQPE QRGMFTVSDH TPEQRGIFTI
 201 SDHPAEQRGM FTKECEQECE LAITDLESGR EDEAGLHQSQ AVHGLELEAL
 251 RLSLSNMHTA QLELTQANLQ KEKETALTEL REMLNSRRAQ ELALLQSRQQ
 301 HELELLREQH AREKEEVVLR CGQEAAELKE KLQSEMEKNA QIVKTLKEDW
 351 ESEKDLCLEN LRKELSAKHQ SEMEDLQNQF QKELAEQRAE LEKIFQDKNQ
 401 AERALRNLES HHQAAIEKLR EDLQSEHGRC LEDLEFKFKE SEKEKQLELE
 451 NLQASYEDLK AQSQEEIRRL WSQLDSARTS RQELSELHEQ LLARTSRVED
 501 LEQLKQREKT QHESELEQLR IYFEKKLRDA EKTYQEDLTL LQQRLQGARE
 551 DALLDSVEVG LSCVGLEEKP EKGRKDHVDE LEPERHKESL PRFQAELEES
 601 HRHQLEALES PLCIQHEGHV SDRCCVETSA LGHEWRLEPS EGHSQELPWV
 651 HLQGVQDGDL EADTERAARV LGLETEHKVQ LSLLQTELKE EIELLKIENR
 701 NLYEKLQHET RLKDDLEKVK HNLIEDHQKE LNNAKQKTEL MKQEFQRKET
 751 DWKVMKEELQ REAEEKLTLM LLELREKAES EKQTIINKFE LREAEMRQLQ
 801 DQQAAQILDL ERSLTEQQGR LQQLEQDLTS DDALHCSQCG REPPTAQDGE
 851 LAALHVKEDC ALQLMLARSR FLEERKEITE KFSAEQDAFL QEAQEQHARE
 901 LQLLQERHQQ QLLSVTAELE ARHQAALGEL TASLESKQGA LLAARVAELQ
 951 TKHAADLGAL ETRHLSSLDS LESCYLSEFQ TIREEHRQAL ELLRADFEEQ
1001 LWKKDSLHQT ILTQELEKLK RKHEGELQSV RDHLRTEAST ELAGTVAHEL
1051 QGVHQGEFGS EKKTALHEKE ETLRLQSAQA QPFHQEEKES LSLQLQKKNH
1101 QVQQLKDQVL SLSHEIEECR SELEVLQQRR ERENREGANL LSMLKADVNL
1151 SHSERGALQD ALRRLLGLFG ETLRAAVTLR SRIGERVGLC LDDAGAGLAL
1201 STAPALEETW SDVALPELDR TLSECAEMSS VAEISSHMRE SFLMSPESVR
1251 ECEQPIRRVF QSLSLAVDGL MEMALDSSRQ LEEARQIHSR FEKEFSFKNE
1301 ETAQVVRKHQ ELLECLKEES AAKAELALEL HKTQGTLEGF KVETADLKEV
1351 LAGKEDSEHR LVLELESLRR QLQQAAQEQA ALREECTRLW SRGEATATDA
1401 EAREAALRKE VEDLTKEQSE TRKQAEKDRS ALLSQMKILE SELEEQLSQH
1451 RGCAKQAEAV TALEQQVASL DKHLRNQRQF MDEQAAEREH EREEFQQEIQ
1501 RLEGQLRQAA KPQPWGPRDS QQAPLDGEVE LLQQKLREKL DEFNELAIQK
1551 ESADRQVLMQ EEEIKRLEEM NINIRKKVAQ LQEEVEKQKN IVKGLEQDKE
1601 VLKKQQMSSL LLASTLQSTL DAGRCPEPPS GSPPEGPEIQ LEVTQRALLR
1651 RESEVLDLKE QLEKMKGDLE SKNEEILHLN LKLDMQNSQT AVSLRELEEE
1701 NTSLKVIYTR SSEIEELKAT IENLQENQKR LQKEKAEEIE QLHEVIEKLQ
1751 HELSLMGPVV HEVSDSQAGS LQSELLCSQA GGPRGQALQG ELEAALEAKE
1801 ALSRLLADQE RRHSQALEAL QQRLQGAEEA AELQLAELER NVALREAEVE
1851 DMASRIQEFE AALKAKEATI AERNLEIDAL NQRKAAHSAE LEAVLLALAR
1901 IRRALEQQPL AAGAAPPELQ WLRAQCARLS RQLQVLHQRF LRCQVELDRR
1951 QARRATAHTR VPGAHPQPRM DGGAKAQVTG DVEASHDAAL EPVVPDPQGD
2001 LQPVLVTLKD APLCKQEGVM SVLTVCQRQL QSELLLVKNE MRLSLEDGGK
2051 GKEKVLEDCQ LPKVDLVAQV KQLQEKLNRL LYSMTFQNVD AADTKSLWPM
2101 ASAHLLESSW SDDSCDGEEP DISPHIDTCD ANTATGGVTD VIKNQAIDAC
2151 DANTTPGGVT DVIKNWDSLI PDEMPDSPIQ EKSECQDMSL SSPTSVLGGS
2201 RHQSHTAEAG PRKSPVGMLD LSSWSSPEVL RKDWTLEPWP SLPVTPHSGA
2251 LSLCSADTSL GDRADTSLPQ TQGPGLLCSP GVSAAALALQ WAESPPADDH
2301 HVQRTAVEKD VEDFITTSFD SQETLSSPPP GLEGKADRSE KSDGSGFGAR
2351 LSPGSGGPEA QTAGPVTPAS ISGRFQPLPE AMKEKEVRPK HVKALLQMVR
2401 DESHQILALS EGLAPPSGEP HPPRKEDEIQ DISLHGGKTQ EVPTACPDWR
2451 GDLLQVVQEA FEKEQEMQGV ELQPRLSGSD LGGHSSLLER LEKIIREQGD
2501 LQEKSLEHLR LPDRSSLLSE IQALRAQLRM THLQNQEKLQ HLRTALTSAE
2551 ARGSQQEHQL RRQVELLAYK VEQEKCIAGD LQKTLSEEQE KANSVQKLLA
2601 AEQTVVRDLK SDLCESRQKS EQLSRSLCEV QQEVLQLRSM LSSKENELKA
2651 ALQELESEQG KGRALQSQLE EEQLRHLQRE SQSAKALEEL RASLETQRAQ
2701 SSRLCVALKH EQTAKDNLQK ELRIEHSRCE ALLAQERSQL SELQKDLAAE
2751 KSRTLELSEA LRHERLLTEQ LSQRTQEACV HQDTQAHHAL LQKLKEEKSR
2801 VVDLQAMLEK VQQQALHSQQ QLEAEAQKHC EALRREKEVS ATLKSTVEAL
2851 HTQKRELRCS LEREREKPAW LQAELEQSHP RLKEQEGRKA ARRSAEARQS
2901 PAAAEQWRKW QRDKEKLREL ELQRQRDLHK IKQLQQTVRD LESKDEVPGS
2951 RLHLGSARRA AGSDADHLRE QQRELEAMRQ RLLSAARLLT SFTSQAVDRT
3001 VNDWTSSNEK AVMSLLHTLE ELKSDLSRPT SSQKKMAAEL QFQFVDVLLK
3051 DNVSLTKALS TVTQEKLELS RAVSKLEKLL KHHLQKGCSP SRSERSAWKP
3101 DETAPQSSLR RPDPGRLPPA ASEEAHTSNV KMEKLYLHYL RAESFRKALI
3151 YQKKYLLLLI GGFQDSEQET LSMIAHLGVF PSKAERKITS RPFTRFRTAV
3201 RVVIAILRLR FLVKKWQEVD RKGALAQGKA PRPGPRARQP QSPPRTRESP
3251 PTRDVPSGHT RDPARGRRLA AAASPHSGGR ATPSPNSRLE RSLTASQDPE
3301 HSLTEYIHHL EVIQQRLGGV LPDSTSKKSC HPMIKQ
```



ProteoMiner Spot 435 Pericentrin

```
Start - End    Observed  Mr(expt) Mr(calc)    ppm   Miss Sequence
    2 - 8        959.53    958.53   958.44     95     0  M.EVEQEQR.R  N-Acetyl (Protein)
    2 - 9       1115.62   1114.61  1114.54     68     1  M.EVEQEQRR.R  N-Acetyl (Protein)
   99 - 107     1130.66   1129.65  1129.56     77     0  R.EDLEQLQQK.Q
  315 - 320      744.40    743.39   743.42    -36     0  K.EEVVLR.C
  469 - 478     1231.69   1230.68  1230.65     27     1  R.RLWSQLDSAR.T
  521 - 526      827.44    826.44   826.46    -27     1  R.IYFEKK.L
  690 - 696      873.49    872.48   872.49     -2     0  K.EEIELLK.I
  721 - 729     1133.62   1132.61  1132.56     43     0  K.HNLIEDHQK.E
  738 - 742      637.29    636.28   636.32    -53     0  K.TELMK.Q  Oxidation (M)
  749 - 753      678.24    677.23   677.30   -109     0  K.ETDWK.V
  754 - 761     1032.57   1031.56  1031.54     17     1  K.VMKEELQR.E
  778 - 788     1260.72   1259.71  1259.67     32     1  K.AESEKQTIINK.F
  789 - 797     1196.59   1195.58  1195.57     16     1  K.FELREAEMR.Q  Oxidation (M)
  938 - 945      782.37    781.37   781.44    -99     0  K.QGALLAAR.V  Pyro-glu (N-term Q)
  946 - 952      788.42    787.41   787.44    -44     0  R.VAELQTK.H
  988 - 994      842.51    841.50   841.50      0     0  R.QALELLR.A
 1070 - 1074     647.26    646.25   646.33   -117     0  K.EETLR.L
 1089 - 1097    1045.56   1044.56  1044.58    -24     0  K.ESLSLQLQK.K
 1324 - 1341    1985.23   1984.23  1984.06     83     1  K.AELALELHKTQGTLEGFK.V
 1349 - 1354     616.36    615.35   615.36    -14     0  K.EVLAGK.E
 1479 - 1488    1224.64   1223.63  1223.52     90     0  R.QFMDEQAAER.E
 1508 - 1518    1218.68   1217.68  1217.63     38     0  R.QAAKPQPWGPR.D  Pyro-glu (N-term Q)
 1508 - 1518    1235.64   1234.63  1234.66    -19     0  R.QAAKPQPWGPR.D
 1566 - 1575    1287.76   1286.75  1286.68     57     1  K.RLEEMNINIR.K
 1660 - 1666     921.49    920.49   920.46     26     1  K.EQLEKMK.G  Oxidation (M)
 1856 - 1864    1048.56   1047.55  1047.56     -9     0  R.IQEFEAALK.A
 1932 - 1942    1437.82   1436.81  1436.84    -20     1  R.QLQVLHQRFLR.C
 2043 - 2052    1003.56   1002.55  1002.53     16     1  R.LSLEDGGKGK.E
 2072 - 2076     645.26    644.25   644.35   -149     0  K.QLQEK.L
 2072 - 2079    1011.58   1010.58  1010.55     26     1  K.QLQEKLNR.L  Pyro-glu (N-term Q)
 2202 - 2212    1190.65   1189.64  1189.56     69     0  R.HQSHTAEAGPR.K
 2375 - 2383    1060.58   1059.57  1059.54     26     0  R.FQPLPEAMK.E
 2375 - 2383    1076.56   1075.55  1075.54     12     0  R.FQPLPEAMK.E  Oxidation (M)
 2530 - 2538    1144.64   1143.64  1143.53     89     0  R.MTHLQNQEK.L  Oxidation (M)
 2618 - 2625     975.53    974.53   974.51     13     1  R.QKSEQLSR.S
 2680 - 2685     649.25    648.24   648.31   -102     0  R.ESQSAK.A
 2699 - 2709    1232.68   1231.67  1231.67      2     1  R.AQSSRLCVALK.H
 2710 - 2720    1311.78   1310.77  1310.66     89     1  K.HEQTAKDNLQK.E
 2766 - 2774    1087.58   1086.58  1086.60    -24     0  R.LLTEQLSQR.T
 2801 - 2810    1161.66   1160.65  1160.61     32     0  R.VVDLQAMLEK.V  Oxidation (M)
 2884 - 2888     618.33    617.33   617.28     80     0  K.EQEGR.K
 2933 - 2939     855.45    854.45   854.46    -17     0  K.QLQQTVR.D  Pyro-glu (N-term Q)
 2933 - 2944    1444.83   1443.82  1443.77     38     1  K.QLQQTVRDLESK.D
 2982 - 2987     630.30    629.29   629.39   -148     0  R.LLSAAR.L
 3058 - 3066     976.54    975.53   975.52      8     0  K.ALSTVTQEK.L
 3058 - 3071    1574.90   1573.89  1573.87     17     1  K.ALSTVTQEKLELSR.A
 3067 - 3071     617.37    616.36   616.35     11     0  K.LELSR.A
 3216 - 3222     960.49    959.48   959.48     -0     1  K.WQEVDRK.G
 3328 - 3335    1016.55   1015.55  1015.49     51     1  K.KSCHPMIK.Q  Oxidation (M)
```

MALDI-TOF PMF protein identification. Prominent peaks in the mass range 700-4000 from the MALDI-TOF spectrum were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score, B) RMS based error scores for the peptide mass in the PMF and C) details of the identified protein's score, sequence including matched peptides are shown.



ProteoMiner Spot 444 GTPase-activating Rap/Ras-GAP domain-like 1

Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 61 are significant (p<0.05).

1. IPI00646904 **Mass:** 125169 **Score:** 68 **Expect:** 0.012 **Queries matched:** 23
Tax_Id=9606 Gene_Symbol=GARNL1 Isoform 3 of GTPase-activating Rap/Ran-GAP domain-like 1

Protein View

Match to: **IPI00646904** Score: **68** Expect: **0.012**
Tax_Id=9606 Gene_Symbol=GARNL1 Isoform 3 of GTPase-activating Rap/Ran-GAP domain-like 1
Found in search of F:\ProteoMiner-Gel01231107\444.txt

Nominal mass (M_r): **125169;** Calculated pI value: **5.82**
NCBI BLAST search of IPI00646904 against nr
Unformatted sequence string for pasting into other applications

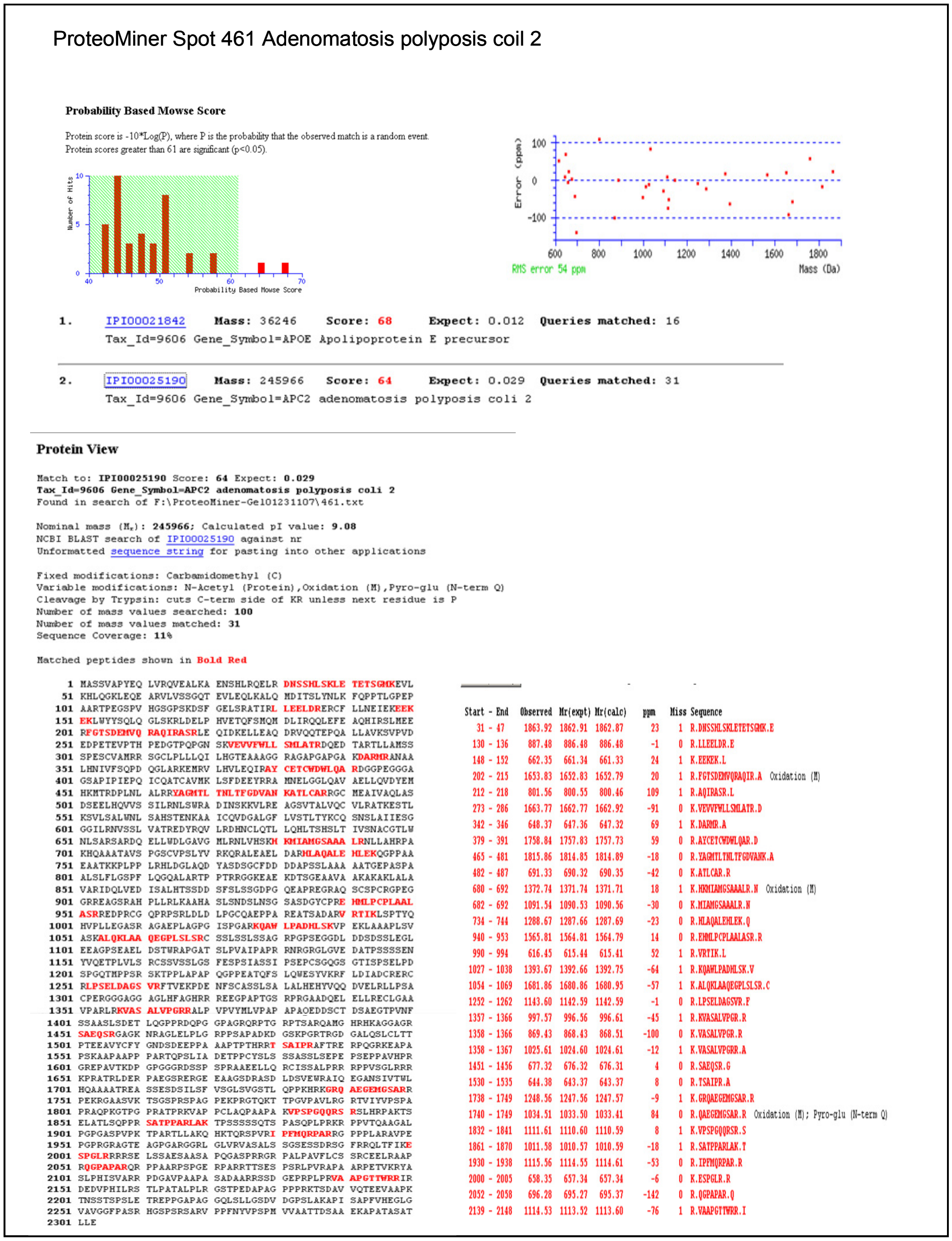
Fixed modifications: Carbamidomethyl (C)
Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: **100**
Number of mass values matched: **23**
Sequence Coverage: **15%**

Matched peptides shown in **Bold Red**

```
   1 MFSKKPHGDV KKSTQKVLDT KKDALTRLKH LRIVIENAES IDLKQFFDQH
  51 FSHIYYVFFE NFVTIEASLK QKGHKSQREE LDAILFIFEK ILQLLPERIH
 101 QRWQFHSIGL ILKKLLHTGN SLKIRREGVR LFLLWLQALQ NNCSKEQLWM
 151 FSCLIPGFSA PQSEHGPRTL DNLINPPLNL QETQVTIEEI TPLVPPQSGD
 201 KGQEDLTSYF LEALLKYIVI QVKSLEWKNK ENQERGFSFL FSHFKKYYLP
 251 YIFPNICKEN SLYHPILDIP QMRPKPHYVV IKKDAETNEA IYCTKEPFIK
 301 ARVIVIRWLV SFWLEPKPHT GPHIPGMEGE VLPKNIQRAA ASLVSREESK
 351 NDNADKTDRT TEPEQSHSNT STLTEREPSS SSLCSIDEEH LTDIEIVRRV
 401 FSSKRSNVNF VTEIFRQAFL LPICEAAAMR KVVKVYQEWI QQEEKPLFMQ
 451 EPEEIVITSS DLPCIENVTD HDISMEEGEK REEENGTNTA DHVRNSSWAK
 501 NGSYQGALHN ASEEATEQNI RAGTQAVLQV FIINSSNIFL LEPANEIKNL
 551 LDEHTDMCKR ILNIYRYMVV QVSMDKKTWE QMLLVLLRVT ESVLKMPSQA
 601 FLQFQGKKNM TLAGRLAGPL FQTLIVAWIK ANLNVYISRE LWDDLLSVLS
 651 SLTYWEELAT EWSLTMETLT KVLARNLYSL DLSDLPLDKL SEQKQKKHKG
 701 KGVGHEFQKV SVDKSFSRGW SRDQPGQAPM RQRSATTTGS PGTEKARSIV
 751 RQKTVAMRSR SIGECALPSA YIRSAKSAPV LIHTSKPFLP DIVLTPLSDE
 801 LSDIDDAQIL PRSTRVRHFS QSEETGNEVF GALNEEQPLP RSSSTSDILE
 851 PFTVERAKVN KEDMSQKLPP LNSDIGGSSA NVPDLMDEFI AERLRSGNAS
 901 TMTRRGSSPG SLEIPKDLPD ILNKQNQMRP IDDPGVPSEW TSPASAGSSD
 951 LISSDSHSDS FSAFQYDGRK FDNFGFGTDT GVTSSADVDS GSGHHQSAEE
1001 QEVASLTTLH IDSETSSLNQ QAFSAEVATI TGSESASPVH SPLGSRSQTP
1051 SPSTLNIDHM EQKDLQLDEK LHHSVLQTPD DLGNISKLDI YLFSFRASVS
1101 GDHK
```

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Sequence
2 - 11	1142.54	1141.54	1141.62	-76	1	M.FSKKPHGDVK.K
2 - 11	1184.56	1183.55	1183.63	-69	1	M.FSKKPHGDVK.K N-Acetyl (Protein)
22 - 27	703.31	702.30	702.40	-147	1	K.KDALTR.L
23 - 29	816.47	815.46	815.49	-28	1	K.DALTRLK.H
217 - 228	1505.71	1504.70	1504.87	-111	1	K.YIVIQVKSLEWK.N
224 - 228	662.31	661.30	661.34	-61	0	K.SLEWK.N
231 - 235	675.29	674.28	674.30	-30	0	K.ENQER.G
247 - 258	1590.78	1589.78	1589.80	-12	0	K.YYLPYIFPNICK.E
351 - 356	676.30	675.30	675.28	19	0	K.NDNADK.T
417 - 430	1589.79	1588.78	1588.77	5	0	R.QAFLLPICEAAAMR.K Oxidation (M); Pyro-glu (N-term Q)
596 - 607	1381.69	1380.68	1380.69	-5	0	K.MPSQAFLQFQGK.K
608 - 615	890.50	889.49	889.48	13	1	K.KNMTLAGR.L
609 - 615	762.42	761.41	761.39	38	0	K.NMTLAGR.L
702 - 709	901.49	900.48	900.45	41	0	K.GVGHEFQK.V
732 - 745	1420.69	1419.68	1419.70	-10	1	R.QRSATTTGSPGTEK.A
734 - 747	1363.69	1362.68	1362.67	7	1	R.SATTTGSPGTEKAR.S
752 - 758	832.52	831.51	831.43	100	1	R.QKTVAMR.S Oxidation (M); Pyro-glu (N-term Q)
752 - 758	849.49	848.49	848.45	39	1	R.QKTVAMR.S Oxidation (M)
813 - 817	618.41	617.40	617.36	64	1	R.STRVR.H
862 - 867	753.37	752.36	752.30	84	0	K.EDMSQK.L Oxidation (M)
862 - 893	3475.72	3474.72	3474.63	25	1	K.EDMSQKLPPLNSDIGGSSANVPDLMDEFIAER.L
896 - 905	1096.47	1095.46	1095.51	-45	1	R.SGNASTMTRR.G Oxidation (M)
1097 - 1104	800.48	799.47	799.38	112	0	R.ASVSGDHK.-

MALDI-TOF PMF protein identification. Prominent peaks in the mass range 700-4000 from the MALDI-TOF spectrum were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score, B) RMS based error scores for the peptide mass in the PMF and C) details of the identified protein's score, sequence including matched peptides are shown.



ProteoMiner Spot 461 Adenomatosis polyposis coil 2

Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 61 are significant (p<0.05).

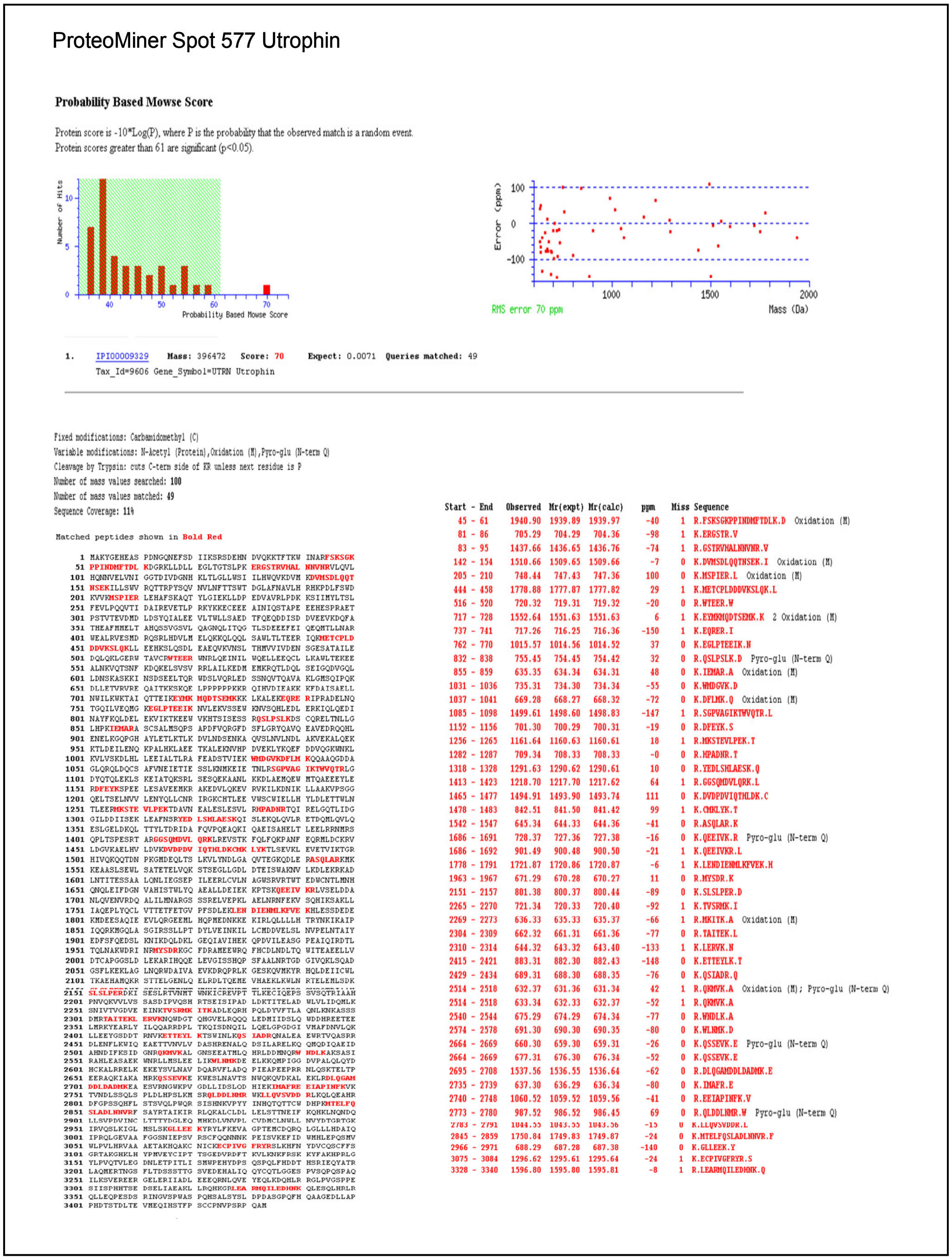
```
1.    IPI00021842    Mass: 36246    Score: 68    Expect: 0.012   Queries matched: 16
      Tax_Id=9606 Gene_Symbol=APOE Apolipoprotein E precursor

2.    IPI00025190    Mass: 245966   Score: 64    Expect: 0.029   Queries matched: 31
      Tax_Id=9606 Gene_Symbol=APC2 adenomatosis polyposis coli 2
```

Protein View

```
Match to: IPI00025190 Score: 64 Expect: 0.029
Tax_Id=9606 Gene_Symbol=APC2 adenomatosis polyposis coli 2
Found in search of F:\ProteoMiner-Gel01231107\461.txt

Nominal mass (Mr): 245966; Calculated pI value: 9.08
NCBI BLAST search of IPI00025190 against nr
Unformatted sequence string for pasting into other applications

Fixed modifications: Carbamidomethyl (C)
Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 100
Number of mass values matched: 31
Sequence Coverage: 11%

Matched peptides shown in Bold Red

   1 MASSVAPYEQ LVRQVEALKA ENSHLRQELR DNSSHLSKLE TETSGMKEVL
  51 KHLQGKLEQE ARVLVSSGQT EVLEQLKALQ MDITSLYNLK FQPPTLGPEP
 101 AARTPEGSPV HGSGPSKDSF GELSRATIRL LEELDRERCF LLNEIEKEEK
 151 EKLWYYSQLQ GLSKRLDELP HVETQFSMQM DLIRQQLEFE AQHIRSLMEE
 201 RFGTSDEMVQ RAQIRASRLE QIDKELLEAQ DRVQQTEPQA LLAVKSVPVD
 251 EDPETEVPTH PEDGTPQPGN SKVEVVFWLL SMLATRDQED TARTLLAMSS
 301 SPESCVAMRR SGCLPLLLQI LHGTEAAAGG RAGAPGAPGA KDARMRANAA
 351 LHNIVFSQPD QGLARKEMRV LHVLEQIRAY CETCWDWLQA RDGGPEGGGA
 401 GSAPIPIEPQ ICQATCAVMK LSFDEEYRRA MNELGGLQAV AELLQVDYEM
 451 HKMTRDPLNL ALRRYAGMTL TNLTFGDVAN KATLCARRGC MEAIVAQLAS
 501 DSEELHQVVS SILRNLSWRA DINSKKVLRE AGSVTALVQC VLRATKESTL
 551 KSVLSALWNL SAHSTENKAA ICQVDGALGF LVSTLTYKCQ SNSLAIIESG
 601 GGILRNVSSL VATREDYRQV LRDHNCLQTL LQHLTSHSLT IVSNACGTLW
 651 NLSARSARDQ ELLWDLGAVG MLRNLVHSKH KMIAMGSAAA LRNLLAHRPA
 701 KHQAAATAVS PGSCVPSLYV RKQRALEAEL DARHLAQALE HLEKQGPPAA
 751 EAATKKPLPP LRHLDGLAQD YASDSGCFDD DDAPSSLAAA AATGEPASPA
 801 ALSLFLGSPF LQGQALARTP PTRRGGKEAE KDTSGEAAVA AKAKAKLALA
 851 VARIDQLVED ISALHTSSDD SFSLSSGDPG QEAPREGRAQ SCSPCRGPEG
 901 GRREAGSRAH PLLRLKAAHA SLSNDSLNSG SASDGYCPRE HMLPCPLAAL
 951 ASRREDPRCG QPRPSRLDLD LPGCQAEPPA REATSADARV RTIKLSPTYQ
1001 HVPLLEGASR AGAEPLAGPG ISPGARKQAW LPADHLSKVP EKLAAAPLSV
1051 ASKALQKLAA QEGPLSLSRC SSLSSLSSAG RPGPSEGGDL DDSDSSLEGL
1101 EEAGPSEAEL DSTWRAPGAT SLPVAIPAPR RNRGRGLGVE DATPSSSSEN
1151 YVQETPLVLS RCSSVSSLGS FESPSIASSI PSEPCSGQGS GTISPSELPD
1201 SPGQTMPPSR SKTPPLAPAP QGPPEATQFS LQWESYVKRF LDIADCRERC
1251 RLPSELDAGS VRFTVEKPDE NFSCASSLSA LALHEHYVQQ DVELRLLPSA
1301 CPERGGGAGG AGLHFAGHRR REEGPAPTGS RPRGAADQEL ELLRECLGAA
1351 VPARLRKVAS ALVPGRRALP VPVYMLVPAP APAQEDDSCT DSAEGTPVNF
1401 SSAASLSDET LQGPPRDQPG GPAGRQRPTG RPTSARQAMG HRHKAGGAGR
1451 SAEQSRGAGK NRAGLELPLG RPPSAPADKD GSKPGRTRGD GALQSLCLTT
1501 PTEEAVYCFY GNDSDEEPPA AAPTPTHRRT SAIPRAFTRE RPQGRKEAPA
1551 PSKAAPAAPP PARTQPSLIA DETPPCYSLS SSASSLSEPE PSEPPAVHPR
1601 GREPAVTKDP GPGGGRDSSP SPRAAEELLQ RCISSALPRR RPPVSGLRRR
1651 KPRATRLDER PAEGSRERGE EAAGSDRASD LDSVEWRAIQ EGANSIVTWL
1701 HQAAAATREA SSESDSILSF VSGLSVGSTL QPPKHRKGRQ AEGEMGSARR
1751 PEKRGAASVK TSGSPRSPAG PEKPRGTQKT TPGVPAVLRG RTVIYVPSPA
1801 PRAQPKGTPG PRATPRKVAP PCLAQPAAPA KVPSPGQQRS RSLHRPAKTS
1851 ELATLSQPPR SATPPARLAK TPSSSSSQTS PASQPLPRKR PPVTQAAGAL
1901 PGPGASPVPK TPARTLLAKQ HKTQRSPVRI PFMQRPARRG PPPLARAVPE
1951 PGPRGRAGTE AGPGARGGRL GLVRVASALS SGSESSDRSG FRRQLTFIKE
2001 SPGLRRRRSE LSSAESAASA PQGASPRRGR PALPAVFLCS SRCEELRAAP
2051 RQGPAPARQR PPAARPSPGE RPARRTTSES PSRLPVRAPA ARPETVKRYA
2101 SLPHISVARR PDGAVPAAPA SADAARRSSD GEPRPLPRVA APGTTWRRIR
2151 DEDVPHILRS TLPATALPLR GSTPEDAPAG PPPRKTSDAV VQTEEVAAPK
2201 TNSSTSPSLE TREPPGAPAG GQLSLLGSDV DGPSLAKAPI SAPFVHEGLG
2251 VAVGGFPASR HGSPSRSARV PPFNYVPSPM VVAATTDSAA EKAPATASAT
2301 LLE
```

```
Start - End   Observed  Mr(expt)  Mr(calc)   ppm  Miss Sequence
   31 - 47     1863.92   1862.91   1862.87    23    1  R.DNSSHLSKLETETSGMK.E
  130 - 136     887.48    886.48    886.48    -1    0  R.LLEELDR.E
  148 - 152     662.35    661.34    661.33    24    1  K.EEKEK.L
  202 - 215    1653.83   1652.83   1652.79    20    1  R.FGTSDEMVQRAQIR.A  Oxidation (M)
  212 - 218     801.56    800.55    800.46   109    1  R.AQIRASR.L
  273 - 286    1663.77   1662.77   1662.92   -91    0  K.VEVVFWLLSMLATR.D
  342 - 346     648.37    647.36    647.32    69    1  K.DARMR.A
  379 - 391    1758.84   1757.83   1757.73    59    0  R.AYCETCWDWLQAR.D
  465 - 481    1815.86   1814.85   1814.89   -18    0  R.YAGMTLTNLTFGDVANK.A
  482 - 487     691.33    690.32    690.35   -42    0  K.ATLCAR.R
  680 - 692    1372.74   1371.74   1371.71    18    1  K.HKMIAMGSAAALR.N  Oxidation (M)
  682 - 692    1091.54   1090.53   1090.56   -30    0  K.MIAMGSAAALR.N
  734 - 744    1288.67   1287.66   1287.69   -23    0  R.HLAQALEHLEK.Q
  940 - 953    1565.81   1564.81   1564.79    14    0  R.EHMLPCPLAALASR.R
  990 - 994     616.45    615.44    615.41    52    1  R.VRTIK.L
 1027 - 1038   1393.67   1392.66   1392.75   -64    1  R.KQAWLPADHLSK.V
 1054 - 1069   1681.86   1680.86   1680.95   -57    1  K.ALQKLAAQEGPLSLSR.C
 1252 - 1262   1143.60   1142.59   1142.59    -1    0  R.LPSELDAGSVR.F
 1357 - 1366    997.57    996.56    996.61   -45    1  R.KVASALVPGR.R
 1358 - 1366    869.43    868.43    868.51  -100    0  K.VASALVPGR.R
 1358 - 1367   1025.61   1024.60   1024.61   -12    1  K.VASALVPGRR.A
 1451 - 1456    677.32    676.32    676.31     4    0  R.SAEQSR.G
 1530 - 1535    644.38    643.37    643.37     8    0  R.TSAIPR.A
 1738 - 1749   1248.56   1247.56   1247.57    -9    1  K.GRQAEGEMGSAR.R
 1740 - 1749   1034.51   1033.50   1033.41    84    0  R.QAEGEMGSAR.R  Oxidation (M); Pyro-glu (N-term Q)
 1832 - 1841   1111.61   1110.60   1110.59     8    1  K.VPSPGQQRSR.S
 1861 - 1870   1011.58   1010.57   1010.59   -18    1  R.SATPPARLAK.T
 1930 - 1938   1115.56   1114.55   1114.61   -53    0  R.IPFMQRPAR.R
 2000 - 2005    658.35    657.34    657.34    -6    0  K.ESPGLR.R
 2052 - 2058    696.28    695.27    695.37  -142    0  R.QGPAPAR.Q
 2139 - 2148   1114.53   1113.52   1113.60   -76    1  R.VAAPGTTWRR.I
```

MALDI-TOF PMF protein identification. Prominent peaks in the mass range 700-4000 from the MALDI-TOF spectrum were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score, B) RMS based error scores for the peptide mass in the PMF and C) details of the identified protein's score, sequence including matched peptides are shown.



ProteoMiner Spot 577 Utrophin

Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 61 are significant (p<0.05).

```
1.   IPI00009329   Mass: 396472   Score: 70    Expect: 0.0071  Queries matched: 49
     Tax_Id=9606 Gene_Symbol=UTRN Utrophin
```

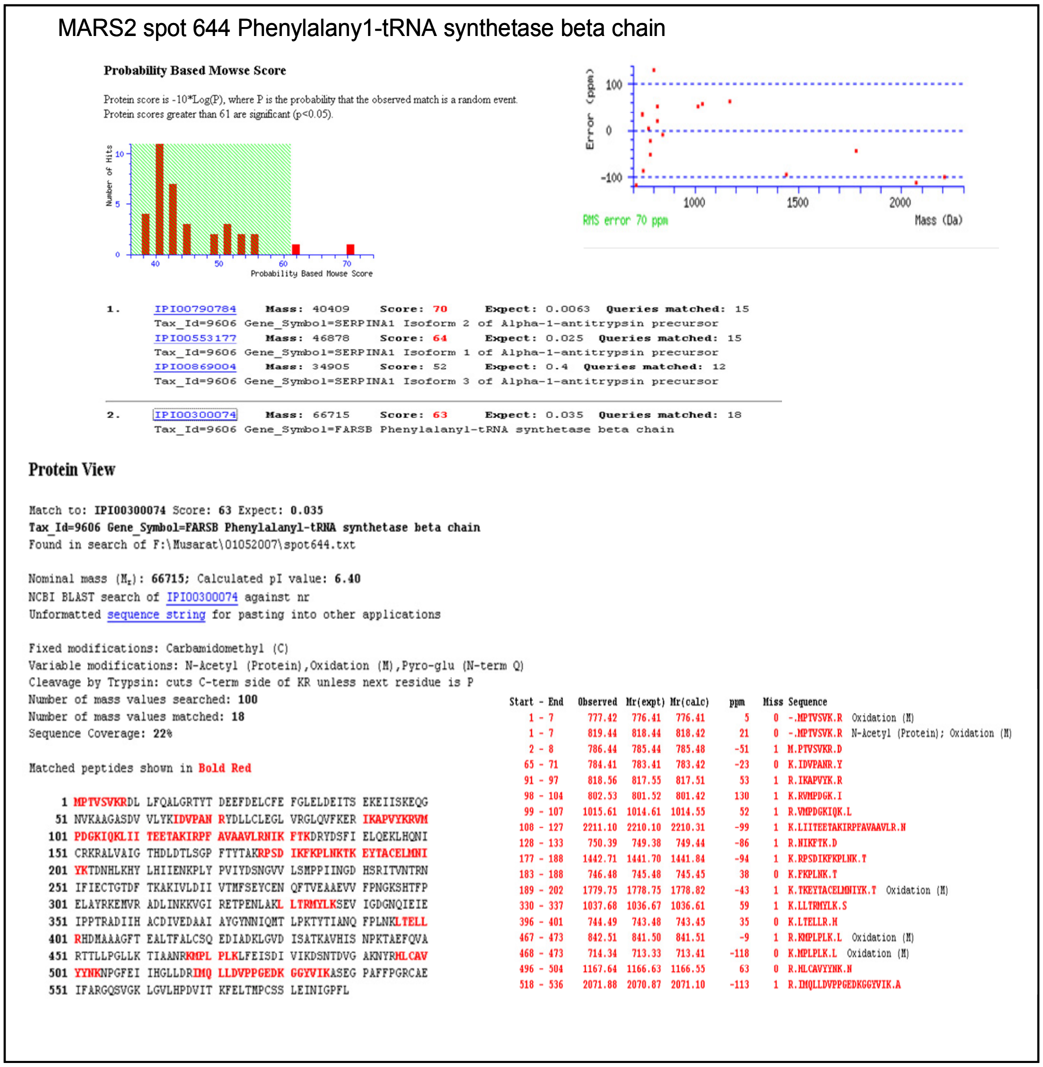
Fixed modifications: Carbamidomethyl (C)
Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: **100**
Number of mass values matched: **49**
Sequence Coverage: **11%**

Matched peptides shown in **Bold Red**

```
   1 MAKYGEHEAS PDNGQNEFSD IIKSRSDEHN DVQKKTFTKW INARFSKSGK
  51 PPINDMFTDL KDGRKLLDLL EGLTGTSLPK ERGSTRVHAL NNVNRVLQVL
 101 HQNNVELVNI GGTDIVDGNH KLTLGLLWSI ILHWQVKDVM KDVMSDLQQT
 151 NSEKILLSWV RQTTRPYSQV NVLNFTTSWT DGLAFNAVLH RHKPDLFSWD
 201 KVVKMSPIER LEHAFSKAQT YLGIEKLLDP EDVAVRLPDK KSIIMYLTSL
 251 FEVLPQQVTI DAIREVETLP RKYKKECEEE AINIQSTAPE EEHESPRAET
 301 PSTVTEVDMD LDSYQIALEE VLTWLLSAED TFQEQDDISD DVEEVKDQFA
 351 THEAFMMELT AHQSSVGSVL QAGNQLITQG TLSDEEEFEI QEQMTLLNAR
 401 WEALRVESMD RQSRLHDVLM ELQKKQLQQL SAWLTLTEER IQKMETCPLD
 451 DDVKSLQKLL EEHKSLQSDL EAEQVKVNSL THMVVIVDEN SGESATAILE
 501 DQLQKLGERW TAVCRWTEER WNRLQEINIL WQELLEEQCL LKAWLTEKEE
 551 ALNKVQTSNF KDQKELSVSV RRLAILKEDM EMKRQTLDQL SEIGQDVGQL
 601 LDNSKASKKI NSDSEELTQR WDSLVQRLED SSNQVTQAVA KLGMSQIPQK
 651 DLLETVRVRE QAITKKSKQE LPPPPPPKKR QIHVDIEAKK KFDAISAELL
 701 NWILKWKTAI QTTEIKEYMK MQDTSEMKKK LKALEKEQRE RIPRADELNQ
 751 TGQILVEQMG KEGLPTEEIK NVLEKVSSEW KNVSQHLEDL ERKIQLQEDI
 801 NAYFKQLDEL EKVIKTKEEW VKHTSISESS RQSLPSLKDS CQRELTNLLG
 851 LHPKIEMARA SCSALMSQPS APDFVQRGFD SFLGRYQAVQ EAVEDRQQHL
 901 ENELKGQPGH AYLETLKTLK DVLNDSENKA QVSLNVLNDL AKVEKALQEK
 951 KTLDEILENQ KPALHKLAEE TKALEKNVHP DVEKLYKQEF DDVQGKWNKL
1001 KVLVSKDLHL LEEIALTLRA FEADSTVIEK WMDGVKDFLM KQQAAQGDDA
1051 GLQRQLDQCS AFVNEIETIE SSLKNMKEIE TNLRSGPVAG IKTWVQTRLG
1101 DYQTQLEKLS KEIATQKSRL SESQEKAANL KKDLAEMQEW MTQAEEEYLE
1151 RDFEYKSPEE LESAVEEMKR AKEDVLQKEV RVKILKDNIK LLAAKVPSGG
1201 QELTSELNVV LENYQLLCNR IRGKCHTLEE VWSCWIELLH YLDLETTWLN
1251 TLEERMKSTE VLPEKTDAVN EALESLESVL RHPADNRTQI RELGQTLIDG
1301 GILDDIISEK LEAFNSRYED LSHLAESKQI SLEKQLQVLR ETDQMLQVLQ
1351 ESLGELDKQL TTYLTDRIDA FQVPQEAQKI QAEISAHELT LEELRRNMRS
1401 QPLTSPESRT ARGGSQMDVL QRKLREVSTK FQLFQKPANF EQRMLDCKRV
1451 LDGVKAELHV LDVKDVDPDV IQTHLDKCMK LYKTLSEVKL EVETVIKTGR
1501 HIVQKQQTDN PKGMDEQLTS LKVLYNDLGA QVTEGKQDLE RASQLARKMK
1551 KEAASLSEWL SATETELVQK STSEGLLGDL DTEISWAKNV LKDLEKRKAD
1601 LNTITESSAA LQNLIEGSEP ILEERLCVLN AGWSRVRTWT EDWCNTLMNH
1651 QNQLEIFDGN VAHISTWLYQ AEALLDEIEK KPTSKQEEIV KRLVSELDDA
1701 NLQVENVRDQ ALILMNARGS SSRELVEPKL AELNRNFEKV SQHIKSAKLL
1751 IAQEPLYQCL VTTETFETGV PFSDLEKLEN DIENMLKFVE KHLESSDEDE
1801 KMDEESAQIE EVLQRGEEML HQPMEDNKKE KIRLQLLLLH TRYNKIKAIP
1851 IQQRKMGQLA SGIRSSLLPT DYLVEINKIL LCMDDVELSL NVPELNTAIY
1901 EDFSFQEDSL KNIKDQLDKL GEQIAVIHEK QPDVILEASG PEAIQIRDTL
1951 TQLNAKWDRI NRMYSDRKGC FDRAMEEWRQ FHCDLNDLTQ WITEAEELLV
2001 DTCAPGGSLD LEKARIHQQE LEVGISSHQP SFAALNRTGD GIVQKLSQAD
2051 GSFLKEKLAG LNQRWDAIVA EVKDRQPRLK GESKQVMKYR HQLDEIICWL
2101 TKAEHAMQKR STTELGENLQ ELRDLTQEME VHAEKLKWLN RTELEMLSDK
2151 SLSLPERDKI SESLRTVNMT WNKICREVPT TLKECIQEPS SVSQTRIAAH
2201 PNVQKVVLVS SASDIPVQSH RTSEISIPAD LDKTITELAD WLVLIDQMLK
2251 SNIVTVGDVE EINKTVSRMK ITKADLEQRH PQLDYVFTLA QNLKNKASSS
2301 DMRTAITEKL ERVKNQWDGT QHGVELRQQQ LEDMIIDSLQ WDDHREETEE
2351 LMRKYEARLY ILQQARRDPL TKQISDNQIL LQELGPGDGI VMAFDNVLQK
2401 LLEEYGSDDT RNVKETTEYL KTSWINLKQS IADRQNALEA EWRTVQASRR
2451 DLENFLKWIQ EAETTVNVLV DASHRENALQ DSILARELKQ QMQDIQAEID
2501 AHNDIFKSID GNRQKMVKAL GNSEEATMLQ HRLDDMNQRW NDLKAKSASI
2551 RAHLEASAEK WNRLLMSLEE LIKWLNMKDE ELKKQMPIGG DVPALQLQYD
2601 HCKALRRELK EKEYSVLNAV DQARVFLADQ PIEAPEEPRR NLQSKTELTP
2651 EERAQKIAKA MRKQSSEVKE KWESLNAVTS NWQKQVDKAL EKLRDLQGAM
2701 DDLDADMKEA ESVRNGWKPV GDLLIDSLQD HIEKIMAFRE EIAPINFKVK
2751 TVNDLSSQLS PLDLHPSLKM SRQLDDLNMR WKLLQVSVDD RLKQLQEAHR
2801 DFGPSSQHFL STSVQLPWQR SISHNKVPYY INHQTQTTCW DHPKMTELFQ
2851 SLADLNNVRF SAYRTAIKIR RLQKALCLDL LELSTTNEIF KQHKLNQNDQ
2901 LLSVPDVINC LTTTYDGLEQ MHKDLVNVPL CVDMCLNWLL NVYDTGRTGK
2951 IRVQSLKIGL MSLSKGLLEE KYRYLFKEVA GPTEMCDQRQ LGLLLHDAIQ
3001 IPRQLGEVAA FGGSNIEPSV RSCFQQNNNK PEISVKEFID WMHLEPQSMV
3051 WLPVLHRVAA AETAKHQAKC NICKECPIVG FRYRSLKHFN YDVCQSCFFS
3101 GRTAKGHKLH YPMVEYCIPT TSGEDVRDFT KVLKNKFRSK KYFAKHPRLG
3151 YLPVQTVLEG DNLETPITLI SMWPEHYDPS QSPQLFHDDT HSRIEQYATR
3201 LAQMERTNGS FLTDSSSTTG SVEDEHALIQ QYCQTLGGES PVSQPQSPAQ
3251 ILKSVEREER GELERIIADL EEEQRNLQVE YEQLKDQHLR RGLPVGSPPE
3301 SIISPHHTSE DSELIAEAKL LRQHKGRLEA RMQILEDHNK QLESQLHRLR
3351 QLLEQPESDS RINGVSPWAS PQHSALSYSL DPDASGPQFH QAAGEDLLAP
3401 PHDTSTDLTE VMEQIHSTFP SCCPNVPSRP QAM
```

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Sequence
45 - 61	1940.90	1939.89	1939.97	-40	1	R.FSKSGKPPINDMFTDLK.D Oxidation (M)
81 - 86	705.29	704.29	704.36	-98	1	K.ERGSTR.V
83 - 95	1437.66	1436.65	1436.76	-74	1	R.GSTRVHALNNVNR.V
142 - 154	1510.66	1509.65	1509.66	-7	0	K.DVMSDLQQTNSEK.I Oxidation (M)
205 - 210	748.44	747.43	747.36	100	0	K.MSPIER.L Oxidation (M)
444 - 458	1778.88	1777.87	1777.82	29	1	K.METCPLDDDVKSLQK.L
516 - 520	720.32	719.31	719.32	-20	0	R.WTEER.W
717 - 728	1552.64	1551.63	1551.63	6	1	K.EYMKMQDTSEMK.K 2 Oxidation (M)
737 - 741	717.26	716.25	716.36	-150	1	K.EQRER.I
762 - 770	1015.57	1014.56	1014.52	37	0	K.EGLPTEEIK.N
832 - 838	755.45	754.45	754.42	32	0	R.QSLPSLK.D Pyro-glu (N-term Q)
855 - 859	635.35	634.34	634.31	48	0	K.IEMAR.A Oxidation (M)
1031 - 1036	735.31	734.30	734.34	-55	0	K.WMDGVK.D
1037 - 1041	669.28	668.27	668.32	-72	0	K.DFLMK.Q Oxidation (M)
1085 - 1098	1499.61	1498.60	1498.83	-147	1	R.SGPVAGIKTWVQTR.L
1152 - 1156	701.30	700.29	700.31	-19	0	R.DFEYK.S
1256 - 1265	1161.64	1160.63	1160.61	18	1	R.MKSTEVLPEK.T
1282 - 1287	709.34	708.33	708.33	-0	0	R.HPADNR.T
1318 - 1328	1291.63	1290.62	1290.61	10	0	R.YEDLSHLAESK.Q
1413 - 1423	1218.70	1217.70	1217.62	64	1	R.GGSQMDVLQRK.L
1465 - 1477	1494.91	1493.90	1493.74	111	0	K.DVDPDVIQTHLDK.C
1478 - 1483	842.51	841.50	841.42	99	1	K.CMKLYK.T
1542 - 1547	645.34	644.33	644.36	-41	0	R.ASQLAR.K
1686 - 1691	728.37	727.36	727.38	-16	0	K.QEEIVK.R Pyro-glu (N-term Q)
1686 - 1692	901.49	900.48	900.50	-21	1	K.QEEIVKR.L
1778 - 1791	1721.87	1720.86	1720.87	-6	1	K.LENDIENMLKFVEK.H
1963 - 1967	671.29	670.28	670.27	11	0	R.MYSDR.K
2151 - 2157	801.38	800.37	800.44	-89	0	K.SLSLPER.D
2265 - 2270	721.34	720.33	720.40	-92	1	K.TVSRMK.I
2269 - 2273	636.33	635.33	635.37	-66	1	R.MKITK.A Oxidation (M)
2304 - 2309	662.32	661.31	661.36	-77	0	R.TAITEK.L
2310 - 2314	644.32	643.32	643.40	-133	1	K.LERVK.N
2415 - 2421	883.31	882.30	882.43	-148	0	K.ETTEYLK.T
2429 - 2434	689.31	688.30	688.35	-76	0	K.QSIADR.Q
2514 - 2518	632.37	631.36	631.34	42	1	R.QKMVK.A Oxidation (M); Pyro-glu (N-term Q)
2514 - 2518	633.34	632.33	632.37	-52	1	R.QKMVK.A
2540 - 2544	675.29	674.29	674.34	-77	0	R.WNDLK.A
2574 - 2578	691.30	690.30	690.35	-80	0	K.WLNMK.D
2664 - 2669	660.30	659.30	659.31	-26	0	K.QSSEVK.E Pyro-glu (N-term Q)
2664 - 2669	677.31	676.30	676.34	-52	0	K.QSSEVK.E
2695 - 2708	1537.56	1536.55	1536.64	-62	0	R.DLQGAMDDLDADMK.E
2735 - 2739	637.30	636.29	636.34	-80	0	K.IMAFR.E
2740 - 2748	1060.52	1059.52	1059.56	-41	0	R.EEIAPINFK.V
2773 - 2780	987.52	986.52	986.45	69	0	R.QLDDLNMR.W Pyro-glu (N-term Q)
2783 - 2791	1044.55	1043.55	1043.56	-15	0	K.LLQVSVDDR.L
2845 - 2859	1750.84	1749.83	1749.87	-24	0	K.MTELFQSLADLNNVR.F
2966 - 2971	688.29	687.28	687.38	-140	0	K.GLLEEK.Y
3075 - 3084	1296.62	1295.61	1295.64	-24	1	K.ECPIVGFRYR.S
3328 - 3340	1596.80	1595.80	1595.81	-8	1	R.LEARMQILEDHNK.Q

MALDI-TOF PMF protein identification. Prominent peaks in the mass range 700-4000 from the MALDI-TOF spectrum were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score, B) RMS based error scores for the peptide mass in the PMF and C) details of the identified protein's score, sequence including matched peptides are shown.



MARS2 spot 644 Phenylalany1-tRNA synthetase beta chain

Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 61 are significant (p<0.05).

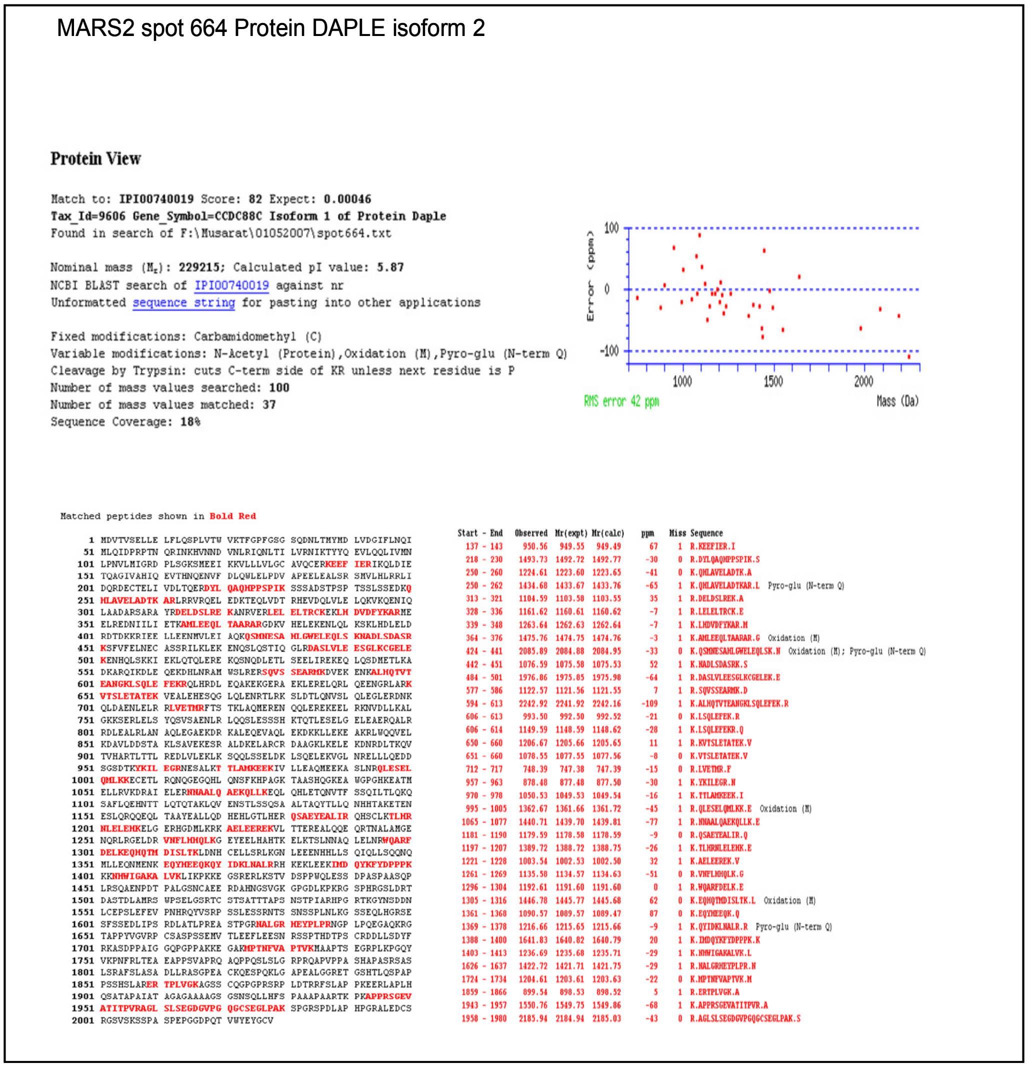
```
1.    IPI00790784    Mass: 40409    Score: 70    Expect: 0.0063  Queries matched: 15
      Tax_Id=9606 Gene_Symbol=SERPINA1 Isoform 2 of Alpha-1-antitrypsin precursor
      IPI00553177    Mass: 46878    Score: 64    Expect: 0.025  Queries matched: 15
      Tax_Id=9606 Gene_Symbol=SERPINA1 Isoform 1 of Alpha-1-antitrypsin precursor
      IPI00869004    Mass: 34905    Score: 52    Expect: 0.4  Queries matched: 12
      Tax_Id=9606 Gene_Symbol=SERPINA1 Isoform 3 of Alpha-1-antitrypsin precursor

2.    IPI00300074    Mass: 66715    Score: 63    Expect: 0.035  Queries matched: 18
      Tax_Id=9606 Gene_Symbol=FARSB Phenylalanyl-tRNA synthetase beta chain
```

Protein View

```
Match to: IPI00300074 Score: 63 Expect: 0.035
Tax_Id=9606 Gene_Symbol=FARSB Phenylalanyl-tRNA synthetase beta chain
Found in search of F:\Musarat\01052007\spot644.txt

Nominal mass (Mr): 66715; Calculated pI value: 6.40
NCBI BLAST search of IPI00300074 against nr
Unformatted sequence string for pasting into other applications

Fixed modifications: Carbamidomethyl (C)
Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 100
Number of mass values matched: 18
Sequence Coverage: 22%

Matched peptides shown in Bold Red

  1 MPTVSVKRDL LFQALGRTYT DEEFDELCFE FGLELDEITS EKEIISKEQG
 51 NVKAAGASDV VLYKIDVPAN RYDLLCLEGL VRGLQVFKER IKAPVYKRVM
101 PDGKIQKLII TEETAKIRPF AVAAVLRNIK FTKDRYDSFI ELQEKLHQNI
151 CRKRALVAIG THDLDTLSGP FTYTAKRPSD IKFKPLNKTK EYTACELMNI
201 YKTDNHLKHY LHIIENKPLY PVIYDSNGVV LSMPPIINGD HSRITVNTRN
251 IFIECTGTDF TKAKIVLDII VTMFSEYCEN QFTVEAAEVV FPNGKSHTFP
301 ELAYRKEMVR ADLINKKVGI RETPENLAKL LTRMYLKSEV IGDGNQIEIE
351 IPPTRADIIH ACDIVEDAAI AYGYNNIQMT LPKTYTIANQ FPLNKLTELL
401 RHDMAAAGFT EALTFALCSQ EDIADKLGVD ISATKAVHIS NPKTAEFQVA
451 RTTLLPGLLK TIAANRKMPL PLKLFEISDI VIKDSNTDVG AKNYRHLCAV
501 YYNKNPGFEI IHGLLDRIMQ LLDVPPGEDK GGYVIKASEG PAFFPGRCAE
551 IFARGQSVGK LGVLHPDVIT KFELTMPCSS LEINIGPFL
```

```
Start - End   Observed  Mr(expt)  Mr(calc)   ppm   Miss Sequence
    1 - 7       777.42    776.41    776.41      5    0  -.MPTVSVK.R  Oxidation (M)
    1 - 7       819.44    818.44    818.42     21    0  -.MPTVSVK.R  N-Acetyl (Protein); Oxidation (M)
    2 - 8       786.44    785.44    785.48    -51    1  M.PTVSVKR.D
   65 - 71      784.41    783.41    783.42    -23    0  K.IDVPANR.Y
   91 - 97      818.56    817.55    817.51     53    1  R.IKAPVYK.R
   98 - 104     802.53    801.52    801.42    130    1  K.RVMPDGK.I
   99 - 107    1015.61   1014.61   1014.55     52    1  R.VMPDGKIQK.L
  108 - 127    2211.10   2210.10   2210.31    -99    1  K.LIITEETAKIRPFAVAAVLR.N
  128 - 133     750.39    749.38    749.44    -86    1  R.NIKFTK.D
  177 - 188    1442.71   1441.70   1441.84    -94    1  K.RPSDIKFKPLNK.T
  183 - 188     746.48    745.48    745.45     38    0  K.FKPLNK.T
  189 - 202    1779.75   1778.75   1778.82    -43    1  K.TKEYTACELMNIYK.T  Oxidation (M)
  330 - 337    1037.68   1036.67   1036.61     59    1  K.LLTRMYLK.S
  396 - 401     744.49    743.48    743.45     35    0  K.LTELLR.H
  467 - 473     842.51    841.50    841.51     -9    1  R.KMPLPLK.L  Oxidation (M)
  468 - 473     714.34    713.33    713.41   -118    0  K.MPLPLK.L  Oxidation (M)
  496 - 504    1167.64   1166.63   1166.55     63    0  R.HLCAVYYNK.N
  518 - 536    2071.88   2070.87   2071.10   -113    1  R.IMQLLDVPPGEDKGGYVIK.A
```

MALDI-TOF PMF protein identification. Prominent peaks in the mass range 700-4000 from the MALDI-TOF spectrum were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score, B) RMS based error scores for the peptide mass in the PMF and C) details of the identified protein's score, sequence including matched peptides are shown.



MARS2 spot 664 Protein DAPLE isoform 2

Protein View

Match to: **IPI00740019** Score: **82** Expect: **0.00046**
Tax_Id=9606 Gene_Symbol=CCDC88C Isoform 1 of Protein Daple
Found in search of F:\Musarat\01052007\spot664.txt

Nominal mass (M_r): **229215**; Calculated pI value: **5.87**
NCBI BLAST search of IPI00740019 against nr
Unformatted sequence string for pasting into other applications

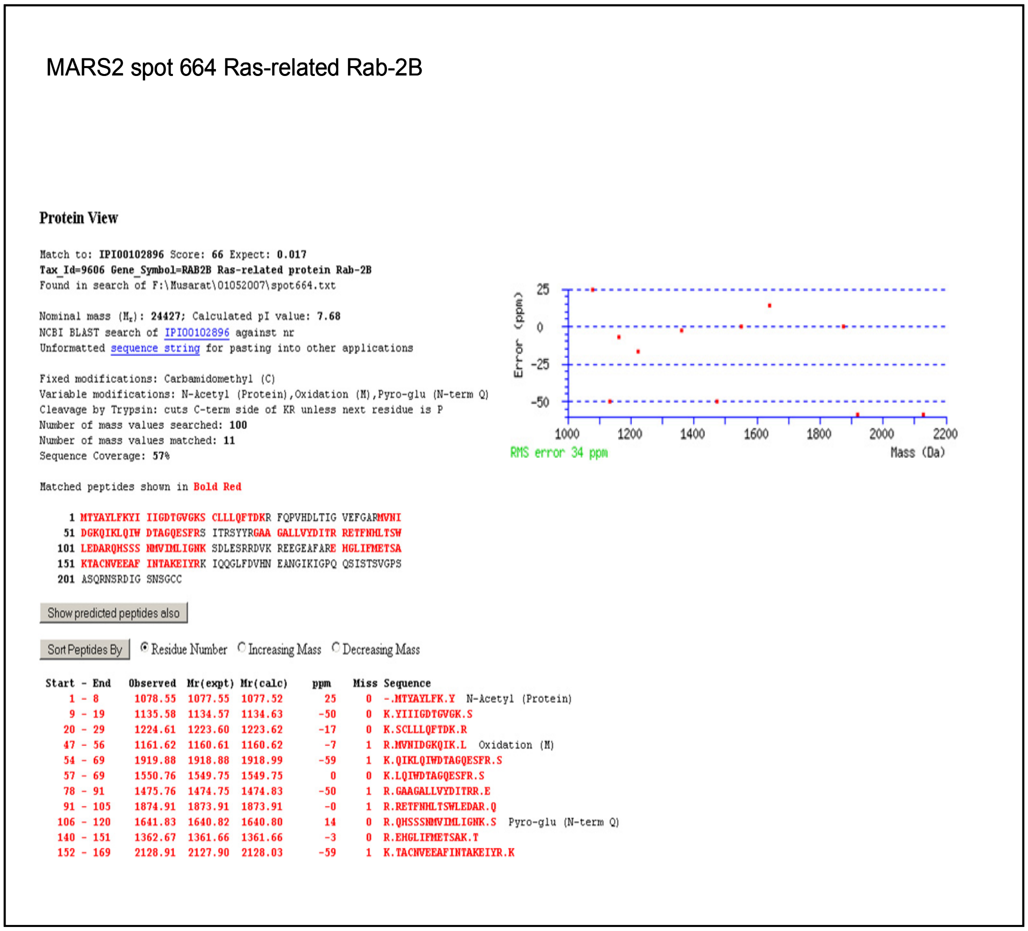
Fixed modifications: Carbamidomethyl (C)
Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: **100**
Number of mass values matched: **37**
Sequence Coverage: **18%**

Matched peptides shown in **Bold Red**

```
   1 MDVTVSELLE LFLQSPLVTW VKTFGPFGSG SQDNLTMYMD LVDGIFLNQI
  51 MLQIDPRPTN QRINKHVNND VNLRIQNLTI LVRNIKTYYQ EVLQQLIVMN
 101 LPNVLMIGRD PLSGKSMEEI KKVLLLVLGC AVQCERKEEF IERIKQLDIE
 151 TQAGIVAHIQ EVTHNQENVF DLQWLELPDV APEELEALSR SMVLHLRRLI
 201 DQRDECTELI VDLTQERDYL QAQHPPSPIK SSSADSTPSP TSSLSSEDKQ
 251 HLAVELADTK ARLRRVRQEL EDKTEQLVDT RHEVDQLVLE LQKVKQENIQ
 301 LAADARSARA YRDELDSLRE KANRVERLEL ELTRCKEKLH DVDFYKARME
 351 ELREDNIILI ETKAMLEEQL TAARARGDKV HELEKENLQL KSKLHDLELD
 401 RDTDKKRIEE LLEENMVLEI AQKQSMNESA HLGWELEQLS KNADLSDASR
 451 KSFVFELNEC ASSRILKLEK ENQSLQSTIQ GLRDASLVLE ESGLKCGELE
 501 KENHQLSKKI EKLQTQLERE KQSNQDLETL SEELIREKEQ LQSDMETLKA
 551 DKARQIKDLE QEKDHLNRAM WSLRERSQVS SEARMKDVEK ENKALHQTVT
 601 EANGKLSQLE FEKRQLHRDL EQAKEKGERA EKLERELQRL QEENGRLARK
 651 VTSLETATEK VEALEHESQG LQLENRTLRK SLDTLQNVSL QLEGLERDNK
 701 QLDAENLELR RLVETMRFTS TKLAQMEREN QQLEREKEEL RKNVDLLKAL
 751 GKKSERLELS YQSVSAENLR LQQSLESSSH KTQTLESELG ELEAERQALR
 801 RDLEALRLAN AQLEGAEKDR KALEQEVAQL EKDKKLLEKE AKRLWQQVEL
 851 KDAVLDDSTA KLSAVEKESR ALDKELARCR DAAGKLKELE KDNRDLTKQV
 901 TVHARTLTTL REDLVLEKLK SQQLSSELDK LSQELEKVGL NRELLLQEDD
 951 SGSDTKYKIL EGRNESALKT TLAMKEEKIV LLEAQMEEKA SLNRQLESEL
1001 QMLKKECETL RQNQGEGQHL QNSFKHPAGK TAASHQGKEA WGPGHKEATM
1051 ELLRVKDRAI ELERNNAALQ AEKQLLKEQL QHLETQNVTF SSQILTLQKQ
1101 SAFLQEHNTT LQTQTAKLQV ENSTLSSQSA ALTAQYTLLQ NHHTAKETEN
1151 ESLQRQQEQL TAAYEALLQD HEHLGTLHER QSAEYEALIR QHSCLKTLHR
1201 NLELEHKELG ERHGDMLKRK AELEEREKVL TTEREALQQE QRTNALAMGE
1251 NQRLRGELDR VNFLHHQLKG EYEELHAHTK ELKTSLNNAQ LELNRWQARF
1301 DELKEQHQTM DISLTKLDNH CELLSRLKGN LEEENHHLLS QIQLLSQQNQ
1351 MLLEQNMENK EQYHEEQKQY IDKLNALRRH KEKLEEKIMD QYKFYDPPPK
1401 KKNHWIGAKA LVKLIKPKKE GSRERLKSTV DSPPWQLESS DPASPAASQP
1451 LRSQAENPDT PALGSNCAEE RDAHNGSVGK GPGDLKPKRG SPHRGSLDRT
1501 DASTDLAMRS WPSELGSRTC STSATTTAPS NSTPIARHPG RTKGYNSDDN
1551 LCEPSLEFEV PNHRQYVSRP SSLESSRNTS SNSSPLNLKG SSEQLHGRSE
1601 SFSSEDLIPS RDLATLPREA STPGRNALGR HEYPLPRNGP LPQEGAQKRG
1651 TAPPYVGVRP CSASPSSEMV TLEEFLEESN RSSPTHDTPS CRDDLLSDYF
1701 RKASDPPAIG GQPGPPAKKE GAKMPTNFVA PTVKMAAPTS EGRPLKPGQY
1751 VKPNFRLTEA EAPPSVAPRQ AQPPQSLSLG RPRQAPVPPA SHAPASRSAS
1801 LSRAFSLASA DLLRASGPEA CKQESPQKLG APEALGGRET GSHTLQSPAP
1851 PSSHSLARER TPLVGKAGSS CQGPGPRSRP LDTRRFSLAP PKEERLAPLH
1901 QSATAPAIAT AGAGAAAAGS GSNSQLLHFS PAAAPAARTK PKAPPRSGEV
1951 ATITPVRAGL SLSEGDGVPG QGCSEGLPAK SPGRSPDLAP HPGRALEDCS
2001 RGSVSKSSPA SPEPGGDPQT VWYEYGCV
```

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Sequence
137 - 143	950.56	949.55	949.49	67	1	R.KEEFIER.I
218 - 230	1493.73	1492.72	1492.77	-30	0	R.DYLQAQHPPSPIK.S
250 - 260	1224.61	1223.60	1223.65	-41	0	K.QHLAVELADTK.A
250 - 262	1434.68	1433.67	1433.76	-65	1	K.QHLAVELADTKAR.L Pyro-glu (N-term Q)
313 - 321	1104.59	1103.58	1103.55	35	1	R.DELDSLREK.A
328 - 336	1161.62	1160.61	1160.62	-7	1	R.LELELTRCK.E
339 - 348	1263.64	1262.63	1262.64	-7	1	K.LHDVDFYKAR.M
364 - 376	1475.76	1474.75	1474.76	-3	1	K.AMLEEQLTAARAR.G Oxidation (M)
424 - 441	2085.89	2084.88	2084.95	-33	0	K.QSMNESAHLGWELEQLSK.N Oxidation (M); Pyro-glu (N-term Q)
442 - 451	1076.59	1075.58	1075.53	52	1	K.NADLSDASRK.S
484 - 501	1976.86	1975.85	1975.98	-64	1	R.DASLVLEESGLKCGELEK.E
577 - 586	1122.57	1121.56	1121.55	7	1	R.SQVSSEARMK.D
594 - 613	2242.92	2241.92	2242.16	-109	1	K.ALHQTVTEANGKLSQLEFEK.R
606 - 613	993.50	992.50	992.52	-21	0	K.LSQLEFEK.R
606 - 614	1149.59	1148.59	1148.62	-28	1	K.LSQLEFEKR.Q
650 - 660	1206.67	1205.66	1205.65	11	1	R.KVTSLETATEK.V
651 - 660	1078.55	1077.55	1077.56	-8	0	K.VTSLETATEK.V
712 - 717	748.39	747.38	747.39	-15	0	R.LVETMR.F
957 - 963	878.48	877.48	877.50	-30	1	K.YKILEGR.N
970 - 978	1050.53	1049.53	1049.54	-16	1	K.TTLAMKEEK.I
995 - 1005	1362.67	1361.66	1361.72	-45	1	R.QLESELQMLKK.E Oxidation (M)
1065 - 1077	1440.71	1439.70	1439.81	-77	1	R.NNAALQAEKQLLK.E
1181 - 1190	1179.59	1178.58	1178.59	-9	0	R.QSAEYEALIR.Q
1197 - 1207	1389.72	1388.72	1388.75	-26	1	K.TLHRNLELEHK.E
1221 - 1228	1003.54	1002.53	1002.50	32	1	K.AELEEREK.V
1261 - 1269	1135.58	1134.57	1134.63	-51	0	R.VNFLHHQLK.G
1296 - 1304	1192.61	1191.60	1191.60	0	1	R.WQARFDELK.E
1305 - 1316	1446.78	1445.77	1445.68	62	0	K.EQHQTMDISLTK.L Oxidation (M)
1361 - 1368	1090.57	1089.57	1089.47	87	0	K.EQYHEEQK.Q
1369 - 1378	1216.66	1215.65	1215.66	-9	1	K.QYIDKLNALR.R Pyro-glu (N-term Q)
1388 - 1400	1641.83	1640.82	1640.79	20	1	K.IMDQYKFYDPPPK.K
1403 - 1413	1236.69	1235.68	1235.71	-29	1	K.NHWIGAKALVK.L
1626 - 1637	1422.72	1421.71	1421.75	-29	1	R.NALGRHEYPLPR.N
1724 - 1734	1204.61	1203.61	1203.63	-22	0	K.MPTNFVAPTVK.M
1859 - 1866	899.54	898.53	898.52	5	1	R.ERTPLVGK.A
1943 - 1957	1550.76	1549.75	1549.86	-68	1	K.APPRSGEVATITPVR.A
1958 - 1980	2185.94	2184.94	2185.03	-43	0	R.AGLSLSEGDGVPGQGCSEGLPAK.S

MALDI-TOF PMF protein identification. Prominent peaks in the mass range 700-4000 from the MALDI-TOF spectrum were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score, B) RMS based error scores for the peptide mass in the PMF and C) details of the identified protein's score, sequence including matched peptides are shown.



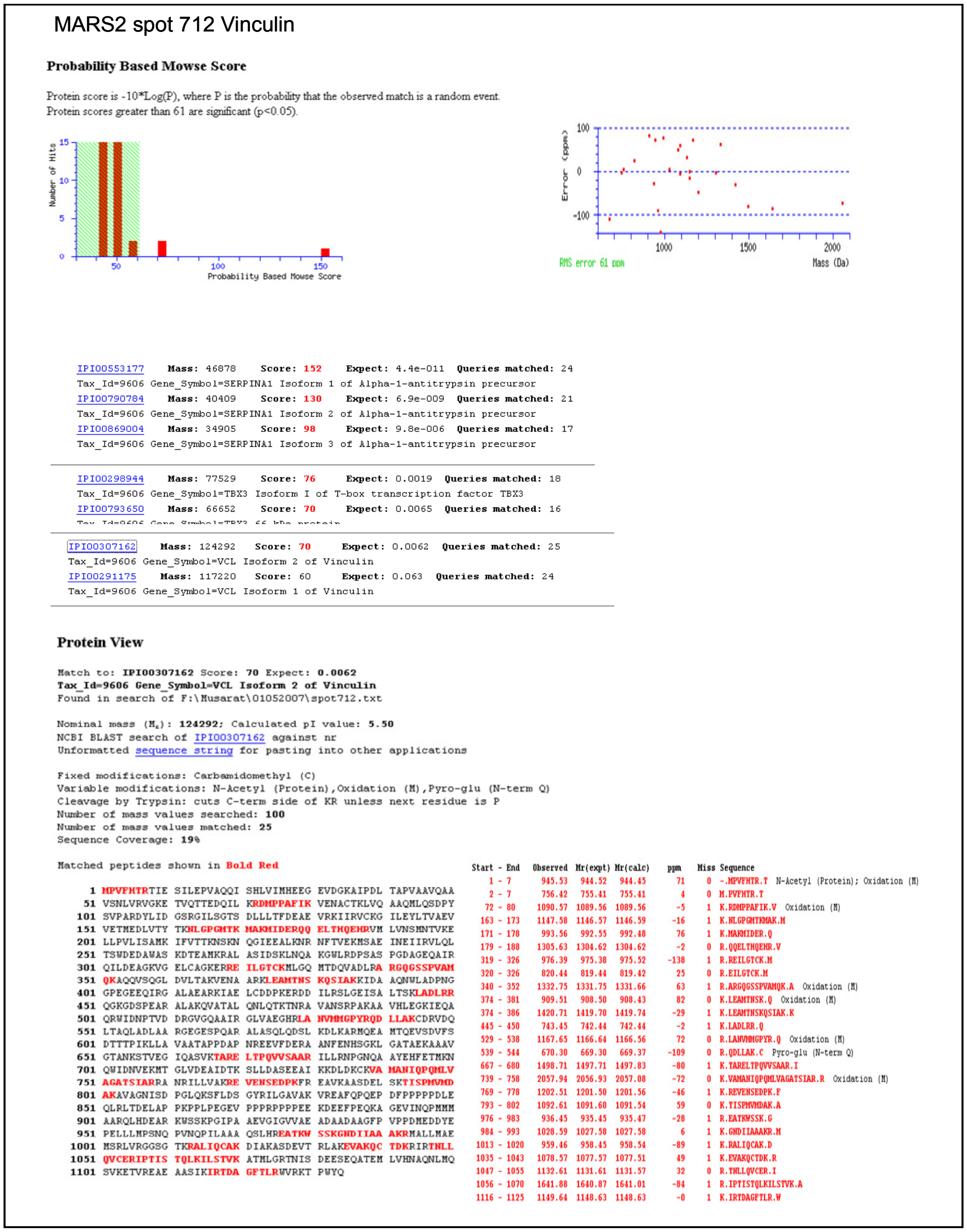
MARS2 spot 664 Ras-related Rab-2B

Protein View

```
Match to: IPI00102896 Score: 66 Expect: 0.017
Tax_Id=9606 Gene_Symbol=RAB2B Ras-related protein Rab-2B
Found in search of F:\Musarat\01052007\spot664.txt

Nominal mass (Mr): 24427; Calculated pI value: 7.68
NCBI BLAST search of IPI00102896 against nr
Unformatted sequence string for pasting into other applications

Fixed modifications: Carbamidomethyl (C)
Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 100
Number of mass values matched: 11
Sequence Coverage: 57%

Matched peptides shown in Bold Red

  1 MTYAYLFKYI IIGDTGVGKS CLLLQFTDKR FQPVHDLTIG VEFGARMVNI
 51 DGKQIKLQIW DTAGQESFRS ITRSYYRGAA GALLVYDITR RETFNHLTSW
101 LEDARQHSSS NMVIMLIGNK SDLESRRDVK REEGEAFARE HGLIFMETSA
151 KTACNVEEAF INTAKEIYRK IQQGLFDVHN EANGIKIGPQ QSISTSVGPS
201 ASQRNSRDIG SNSGCC
```

```
Start - End   Observed  Mr(expt) Mr(calc)    ppm   Miss Sequence
   1 - 8      1078.55   1077.55  1077.52      25     0  -.MTYAYLFK.Y  N-Acetyl (Protein)
   9 - 19     1135.58   1134.57  1134.63     -50     0  K.YIIIGDTGVGK.S
  20 - 29    1224.61   1223.60  1223.62     -17     0  K.SCLLLQFTDK.R
  47 - 56    1161.62   1160.61  1160.62      -7     1  R.MVNIDGKQIK.L  Oxidation (M)
  54 - 69    1919.88   1918.88  1918.99     -59     1  K.QIKLQIWDTAGQESFR.S
  57 - 69    1550.76   1549.75  1549.75       0     0  K.LQIWDTAGQESFR.S
  78 - 91    1475.76   1474.75  1474.83     -50     1  R.GAAGALLVYDITRR.E
  91 - 105   1874.91   1873.91  1873.91      -0     1  R.RETFNHLTSWLEDAR.Q
 106 - 120   1641.83   1640.82  1640.80      14     0  R.QHSSSNMVIMLIGNK.S  Pyro-glu (N-term Q)
 140 - 151   1362.67   1361.66  1361.66      -3     0  R.EHGLIFMETSAK.T
 152 - 169   2128.91   2127.90  2128.03     -59     1  K.TACNVEEAFINTAKEIYR.K
```

MALDI-TOF PMF protein identification. Prominent peaks in the mass range 700-4000 from the MALDI-TOF spectrum were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score, B) RMS based error scores for the peptide mass in the PMF and C) details of the identified protein's score, sequence including matched peptides are shown.



MARS2 spot 712 Vinculin

Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 61 are significant (p<0.05).

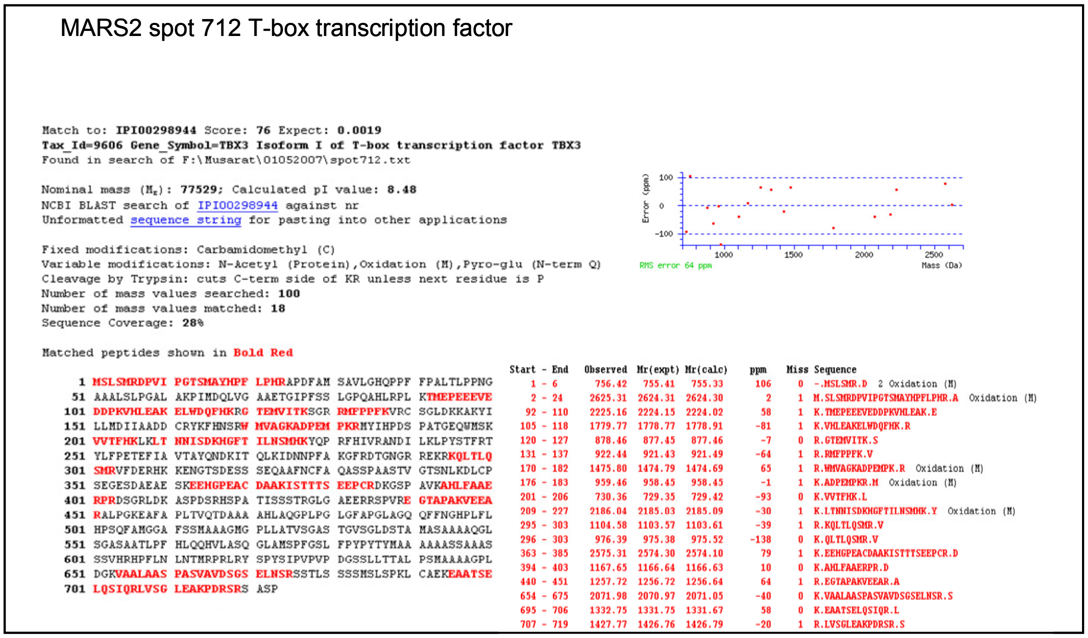
```
IPI00553177    Mass: 46878    Score: 152    Expect: 4.4e-011  Queries matched: 24
Tax_Id=9606 Gene_Symbol=SERPINA1 Isoform 1 of Alpha-1-antitrypsin precursor
IPI00790784    Mass: 40409    Score: 130    Expect: 6.9e-009  Queries matched: 21
Tax_Id=9606 Gene_Symbol=SERPINA1 Isoform 2 of Alpha-1-antitrypsin precursor
IPI00869004    Mass: 34905    Score: 98     Expect: 9.8e-006  Queries matched: 17
Tax_Id=9606 Gene_Symbol=SERPINA1 Isoform 3 of Alpha-1-antitrypsin precursor

IPI00298944    Mass: 77529    Score: 76     Expect: 0.0019  Queries matched: 18
Tax_Id=9606 Gene_Symbol=TBX3 Isoform I of T-box transcription factor TBX3
IPI00793650    Mass: 66652    Score: 70     Expect: 0.0065  Queries matched: 16
Tax_Id=9606 Gene_Symbol=TBX3 66 kDa protein

IPI00307162    Mass: 124292    Score: 70     Expect: 0.0062  Queries matched: 25
Tax_Id=9606 Gene_Symbol=VCL Isoform 2 of Vinculin
IPI00291175    Mass: 117220    Score: 60     Expect: 0.063  Queries matched: 24
Tax_Id=9606 Gene_Symbol=VCL Isoform 1 of Vinculin
```

Protein View

```
Match to: IPI00307162 Score: 70 Expect: 0.0062
Tax_Id=9606 Gene_Symbol=VCL Isoform 2 of Vinculin
Found in search of F:\Musarat\01052007\spot712.txt

Nominal mass (Mr): 124292; Calculated pI value: 5.50
NCBI BLAST search of IPI00307162 against nr
Unformatted sequence string for pasting into other applications

Fixed modifications: Carbamidomethyl (C)
Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 100
Number of mass values matched: 25
Sequence Coverage: 19%

Matched peptides shown in Bold Red

   1 MPVFHTRTIE SILEPVAQQI SHLVIMHEEG EVDGKAIPDL TAPVAAVQAA
  51 VSNLVRVGKE TVQTTEDQIL KRDMPPAFIK VENACTKLVQ AAQMLQSDPY
 101 SVPARDYLID GSRGILSGTS DLLLTFDEAE VRKIIRVCKG ILEYLTVAEV
 151 VETMEDLVTY TKNLGPGMTK MAKMIDERQQ ELTHQEHRVM LVNSMNTVKE
 201 LLPVLISAMK IFVTTKNSKN QGIEEALKNR NFTVEKMSAE INEIIRVLQL
 251 TSWDEDAWAS KDTEAMKRAL ASIDSKLNQA KGWLRDPSAS PGDAGEQAIR
 301 QILDEAGKVG ELCAGKERRE ILGTCKMLGQ MTDQVADLRA RGQGSSPVAM
 351 QKAQQVSQGL DVLTAKVENA ARKLEAMTNS KQSIAKKIDA AQNWLADPNG
 401 GPEGEEQIRG ALAEARKIAE LCDDPKERDD ILRSLGEISA LTSKLADLRR
 451 QGKGDSPEAR ALAKQVATAL QNLQTKTNRA VANSRPAKAA VHLEGKIEQA
 501 QRWIDNPTVD DRGVGQAAIR GLVAEGHRLA NVMMGPYRQD LLAKCDRVDQ
 551 LTAQLADLAA RGEGESPQAR ALASQLQDSL KDLKARMQEA MTQEVSDVFS
 601 DTTTPIKLLA VAATAPPDAP NREEVFDERA ANFENHSGKL GATAEKAAAV
 651 GTANKSTVEG IQASVKTARE LTPQVVSAAR ILLRNPGNQA AYEHFETMKN
 701 QWIDNVEKMT GLVDEAIDTK SLLDASEEAI KKDLDKCKVA MANIQPQMLV
 751 AGATSIARRA NRILLVAKRE VENSEDPKFR EAVKAASDEL SKTISPMVMD
 801 AKAVAGNISD PGLQKSFLDS GYRILGAVAK VREAFQPQEP DFPPPPPDLE
 851 QLRLTDELAP PKPPLPEGEV PPPRPPPPEE KDEEFPEQKA GEVINQPMMM
 901 AARQLHDEAR KWSSKPGIPA AEVGIGVVAE ADAADAAGFP VPPDMEDDYE
 951 PELLLMPSNQ PVNQPILAAA QSLHREATKW SSKGNDIIAA AKRMALLMAE
1001 MSRLVRGGSG TKRALIQCAK DIAKASDEVT RLAKEVAKQC TDKRIRTNLL
1051 QVCERIPTIS TQLKILSTVK ATMLGRTNIS DEESEQATEM LVHNAQNLMQ
1101 SVKETVREAE AASIKIRTDA GFTLRWVRKT PWYQ
```

```
Start - End   Observed  Mr(expt)  Mr(calc)    ppm  Miss Sequence
    1 - 7       945.53    944.52    944.45     71    0  -.MPVFHTR.T  N-Acetyl (Protein); Oxidation (M)
    2 - 7       756.42    755.41    755.41      4    0  M.PVFHTR.T
   72 - 80     1090.57   1089.56   1089.56     -5    1  K.RDMPPAFIK.V  Oxidation (M)
  163 - 173    1147.58   1146.57   1146.59    -16    1  K.NLGPGMTKMAK.M
  171 - 178     993.56    992.55    992.48     76    1  K.MAKMIDER.Q
  179 - 188    1305.63   1304.62   1304.62     -2    0  R.QQELTHQEHR.V
  319 - 326     976.39    975.38    975.52   -138    1  R.REILGTCK.M
  320 - 326     820.44    819.44    819.42     25    0  R.EILGTCK.M
  340 - 352    1332.75   1331.75   1331.66     63    1  R.ARGQGSSPVAMQK.A  Oxidation (M)
  374 - 381     909.51    908.50    908.43     82    0  K.LEAMTNSK.Q  Oxidation (M)
  374 - 386    1420.71   1419.70   1419.74    -29    1  K.LEAMTNSKQSIAK.K
  445 - 450     743.45    742.44    742.44     -2    1  K.LADLRR.Q
  529 - 538    1167.65   1166.64   1166.56     72    0  R.LANVMMGPYR.Q  Oxidation (M)
  539 - 544     670.30    669.30    669.37   -109    0  R.QDLLAK.C  Pyro-glu (N-term Q)
  667 - 680    1498.71   1497.71   1497.83    -80    1  K.TARELTPQVVSAAR.I
  739 - 758    2057.94   2056.93   2057.08    -72    0  K.VAMANIQPQMLVAGATSIAR.R  Oxidation (M)
  769 - 778    1202.51   1201.50   1201.56    -46    1  K.REVENSEDPK.F
  793 - 802    1092.61   1091.60   1091.54     59    0  K.TISPMVMDAK.A
  976 - 983     936.45    935.45    935.47    -28    1  R.EATKWSSK.G
  984 - 993    1028.59   1027.58   1027.58      6    1  K.GNDIIAAAKR.M
 1013 - 1020    959.46    958.45    958.54    -89    1  K.RALIQCAK.D
 1035 - 1043   1078.57   1077.57   1077.51     49    1  K.EVAKQCTDK.R
 1047 - 1055   1132.61   1131.61   1131.57     32    0  R.TNLLQVCER.I
 1056 - 1070   1641.88   1640.87   1641.01    -84    1  R.IPTISTQLKILSTVK.A
 1116 - 1125   1149.64   1148.63   1148.63     -0    1  K.IRTDAGFTLR.W
```

MALDI-TOF PMF protein identification. Prominent peaks in the mass range 700-4000 from the MALDI-TOF spectrum were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score, B) RMS based error scores for the peptide mass in the PMF and C) details of the identified protein's score, sequence including matched peptides are shown.



MARS2 spot 712 T-box transcription factor

```
Match to: IPI00298944 Score: 76 Expect: 0.0019
Tax_Id=9606 Gene_Symbol=TBX3 Isoform I of T-box transcription factor TBX3
Found in search of F:\Musarat\01052007\spot712.txt

Nominal mass (Mr): 77529; Calculated pI value: 8.48
NCBI BLAST search of IPI00298944 against nr
Unformatted sequence string for pasting into other applications

Fixed modifications: Carbamidomethyl (C)
Variable modifications: N-Acetyl (Protein),Oxidation (M),Pyro-glu (N-term Q)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 100
Number of mass values matched: 18
Sequence Coverage: 28%

Matched peptides shown in Bold Red

  1 MSLSMRDPVI PGTSMAYHPF LPHRAPDFAM SAVLGHQPPF FPALTLPPNG
 51 AAALSLPGAL AKPIMDQLVG AAETGIPFSS LGPQAHLRPL KTMEPEEEVE
101 DDPKVHLEAK ELWDQFHKRG TEMVITKSGR RMFPPFKVRC SGLDKKAKYI
151 LLMDIIAADD CRYKFHNSRW MVAGKADPEM PKRMYIHPDS PATGEQWMSK
201 VVTFHKLKLT NNISDKHGFT ILNSMHKYQP RFHIVRANDI LKLPYSTFRT
251 YLFPETEFIA VTAYQNDKIT QLKIDNNPFA KGFRDTGNGR REKRKQLTLQ
301 SMRVFDERHK KENGTSDESS SEQAAFNCFA QASSPAASTV GTSNLKDLCP
351 SEGESDAEAE SKEEHGPEAC DAAKISTTTS EEPCRDKGSP AVKAHLFAAE
401 RPRDSGRLDK ASPDSRHSPA TISSSTRGLG AEERRSPVRE GTAPAKVEEA
451 RALPGKEAFA PLTVQTDAAA AHLAQGPLPG LGFAPGLAGQ QFFNGHPLFL
501 HPSQFAMGGA FSSMAAAGMG PLLATVSGAS TGVSGLDSTA MASAAAAQGL
551 SGASAATLPF HLQQHVLASQ GLAMSPFGSL FPYPYTYMAA AAASSAAAS
601 SSVHRHPFLN LNTMRPRLRY SPYSIPVPVP DGSSLLTTAL PSMAAAAGPL
651 DGKVAALAAS PASVAVDSGS ELNSRSSTLS SSSMSLSPKL CAEKEAATSE
701 LQSIQRLVSG LEAKPDRSRS ASP
```

```
Start - End   Observed  Mr(expt) Mr(calc)    ppm   Miss Sequence
   1 - 6        756.42    755.41   755.33    106     0  -.MSLSMR.D  2 Oxidation (M)
   2 - 24      2625.31   2624.31  2624.30      2     1  M.SLSMRDPVIPGTSMAYHPFLPHR.A  Oxidation (M)
  92 - 110     2225.16   2224.15  2224.02     58     1  K.TMEPEEEVEDDPKVHLEAK.E
 105 - 118     1779.77   1778.77  1778.91    -81     1  K.VHLEAKELWDQFHK.R
 120 - 127      878.46    877.45   877.46     -7     0  R.GTEMVITK.S
 131 - 137      922.44    921.43   921.49    -64     1  R.RMFPPFK.V
 170 - 182     1475.80   1474.79  1474.69     65     1  R.WMVAGKADPEMPK.R  Oxidation (M)
 176 - 183      959.46    958.45   958.45     -1     1  K.ADPEMPKR.M  Oxidation (M)
 201 - 206      730.36    729.35   729.42    -93     0  K.VVTFHK.L
 209 - 227     2186.04   2185.03  2185.09    -30     1  K.LTNNISDKHGFTILNSMHK.Y  Oxidation (M)
 295 - 303     1104.58   1103.57  1103.61    -39     1  R.KQLTLQSMR.V
 296 - 303      976.39    975.38   975.52   -138     0  K.QLTLQSMR.V
 363 - 385     2575.31   2574.30  2574.10     79     1  K.EEHGPEACDAAKISTTTSEEPCR.D
 394 - 403     1167.65   1166.64  1166.63     10     0  K.AHLFAAERPR.D
 440 - 451     1257.72   1256.72  1256.64     64     1  R.EGTAPAKVEEAR.A
 654 - 675     2071.98   2070.97  2071.05    -40     0  K.VAALAASPASVAVDSGSELNSR.S
 695 - 706     1332.75   1331.75  1331.67     58     0  K.EAATSELQSIQR.L
 707 - 719     1427.77   1426.76  1426.79    -20     1  R.LVSGLEAKPDRSR.S
```

MALDI-TOF PMF protein identification. Prominent peaks in the mass range 700-4000 from the MALDI-TOF spectrum were used to generate a peptide mass fingerprint which was searched against the IPI-Human database using the Mascot search engine. A) Probability based Mowse score, B) RMS based error scores for the peptide mass in the PMF and C) details of the identified protein's score, sequence including matched peptides are shown.