FSA Project Code M03064
Project Title: Development of a multi-marker live animal diagnostic specific to TSE disease in blood plasma which is not reliant on PrP^{Sc}

TABLE OF CONTENTS

1. Introduction

Transmissible spongiform encephalopathy describes a group of diseases affecting humans (Creutzfeldt-Jacob disease (vCJD) and animals (e.g. scrapie in sheep, bovine spongiform encephalopathy (BSE) in cattle, chronic wasting disease CWD in deer). TSE disease affects the central nervous system and lymphoreticular system of animals and is characterized by a long asymptomatic phase, followed by a short clinical phase, before certain death of the animal. Brain pathology reveals the hallmarks of these diseases with spongiform changes in the brain parenchyma, gliosis and the deposition of an abnormal protease resistant form (PrP^{d}) of the host encoded prion protein (PrP^{c}).

In 1986 BSE was reported as a newly emerging disease affecting cattle [1]. Surveillance for the presence of TSE diseases became necessary after an outbreak of BSE in cattle reached epidemic levels in the UK in the 1980s and 1990s and was found to have been transmitted to humans in 1996[2]. Active and passive surveillance of carcasses reaching the food chain are now mandatory under UK and European legislation.

Current rapid TSE diagnostic assays[3] used for surveillance all rely on detection of the abnormal isoform of the host prion protein (PrP^{d}) in brain by either immunoassay or ligand affinity. Most assays also utilise proteinase K digestion to enrich for PrP^{d} and increase specificity and sensitivity. These assay systems are only reliable if PrP^{d} always associates with TSE infectivity. Recent evidence from experimental animal models [4-5] and the emergence of atypical forms of the disease (atypical scrapie and BASE) indicate that PrP^{d} may not be a reliable marker of these newly emerging forms of the disease. There is therefore an urgent need for new types of diagnostic tests preferably in the live animal which can be carried out before reaching the abattoir.

For a pre-clinical diagnostic assay to be successfully applied to TSE surveillance, the sample must be easily collected from the live animal, preferably by non-invasive means, and analyzed quickly, with high levels of sensitivity and specificity. Blood samples are the most obvious choice for ease of sampling and could be taken before the animal reaches the abattoir. The complexity of blood and the variability between individuals, limits single marker assays which fail to achieve the stringent sensitivity and specificity levels required for application to large populations. TSE disease manifests in different clinical and pathological profiles within and between species making it particularly difficult to find a universal diagnostic test. For these reasons in this project we have focused on the development of a multi-marker test which by definition is a more robust approach for application to blood samples.

In previous studies we have demonstrated the potential use of differential protein expression profiling in brain tissue samples using SELDI-TOF technology for the detection of TSE disease [6-7].

In this study we aimed to further our previous studies to establish a blood plasma data driven assay based on the temporal protein expression differences found between normal and TSE diseased animals. Plasma samples were obtained from ongoing projects (Professor N. Hunter, Dr F. Houston, The Roslin Institute and R(D)SVS) which had been taken at intervals throughout the course of disease from groups of sheep of the susceptible genotype VRQ/VRQ and ARQ/VRQ infected with scrapie and age matched control groups. A further temporal study examined plasma samples from an ongoing project (Professor J. Manson, Dr S McCutcheon, The Roslin Institute and R(D)SVS) of groups of sheep (ARQ/ARQ genotype) infected with BSE and control animals.

Surface enhanced laser desorption and ionisation time of flight mass spectrometry (SELDI–TOF MS) technology, combines protein chromatography with mass spectroscopy for the analysis of complex biological samples. A solid phase surface (in array format) captures subsets of proteins from crude biological extracts based on their chemical or biological properties. Washes eliminate unbound proteins and the array is then placed directly into a time-of-flight mass spectrometer for proteomic analysis. The resultant data can be displayed as a pattern of peaks representative (m/z, mass to charge ratio) of captured proteins. Specialized software identifies protein peaks common to all spectra in the groups to be compared (such as normal and diseased tissue samples) and clusters peaks which are differentially expressed. These clusters are then used to build a predictive classification model or “profile” of disease specific markers. The protein clusters are divided into a training set and a test set of samples. The training set where the disease status is known is used to build a classification model using one or more statistical learning techniques [8]. The model is then tested for predictive power on the test set where the disease status labels are removed. For large data sets a single division into training and test sets may be used, but for smaller numbers as here, cross validation using the majority of samples to build the model is preferred.

In summary we report in this study a methodology which includes robust statistical learning techniques to identify the potential of panels of protein markers with predictive power to detect TSE disease in ovine blood plasma.

2. Methods

2.1 Instrument

The Bio-Rad ProteinChip Enterprise 4000 SELDI-TOF instrument was purchased for the purposes of this project. This was equipped with ProteinChip Data Manager Database and server software which provides data handling and data mining and analysis capability. J. Barr was trained, in addition to her expertise on the previous system, by a field scientist from Bio-Rad specifically using the new instrument as part of commissioning the new system.

2.2 Samples

Scrapie

Ovine blood samples were kindly donated by Prof. N. Hunter and Dr F. Houston from a maternal transmission study carried out at the Institute for Animal Health (IAH), Compton.

Two groups of five sheep homozygous and heterozygous for a scrapie susceptible genotype (VRQ/VRQ, VRQ/ARQ) were sub-cutaneously challenged with scrapie (SSBP1) and two similar control groups (three sheep per group) sub-cutaneously challenged with normal brain homogenate. Blood was taken before inoculation and then at monthly intervals throughout the course of infection until the animals was deemed to be at the terminal stage of disease. Samples from a further nine scrapie infected sheep with the genotype VRQ/VRQ, were available which were bled at the terminal stage of disease only. A total of 70 scrapie infected samples and 28 control samples at time points between 1 - 8 months were analysed.

BSE

Ovine blood plasma samples were kindly donated from a TSE blood transfusion project conducted by Prof. Jean Manson and Dr Sandra McCutcheon (TSE Transmission group, The Roslin Institute). Blood was taken from sheep (ARQ/ARQ genotype) orally infected with BSE at intervals throughout the time course of disease. A group of 10 sheep were similarly challenged with a normal brain homogenate. A summary of the groups and numbers of samples can be seen in Table 2.1.

Table 2. 1 BSE - numbers of plasma samples

Time (months post inoculation)	0	2	4	6	8	10	12	18	22	25	27	28	Clinical
Group	**1**	**2**	**3**	**4**	**5**	**6**		**7**					**8**
BSE infected (number)	39	40	40	40	40	35	36	33	0	9	8	3	17
non-infected controls(number)	10	10	10	10	10	9	9	0	1	5	2	0	0

Toxiplasma gondi samples

The aim of this project is to build a classification model which is predictive for the presence of TSE disease. As a test of specificity, the classifiers from each time point were applied to groups of ovine plasma infected with a disease unrelated to TSE disease. Plasma samples from groups of sheep (5 x high, 5 x low *Toxiplasma gondi,* 5 sentinels and 5 controls), included in a project investigating *Toxoplasma gondi,* were obtained from Dr Francesca Chianini (Moredun Research Institute, Edinburgh). Blood samples with the addition of EDTA were fractionated at MRI and aliquots of plasma samples were received for storage at -80°C.

Blood sample preparation

Ovine blood samples had been taken from the sheep and collected in tubes containing ~1% EDTA. The samples were spun at 1000g for 30 minutes at RT. The subsequent fractions of blood were separated with the top plasma fraction removed to a clean 50ml falcon tube. Samples in 2ml microcentrifuge tube aliquots were then frozen and stored at -80°C. Each aliquot was divided into two further 150µl aliquots (1 for each type of array) in 0.5ml tubes to avoid further freeze/ thaw degradation of proteins. A pooled sample for each TSE batch of samples (BSE/controls and scrapie/controls) was created by combining 100µl from each sample, mixing and dividing into 50 x 200µl aliquots. One aliquot per bioprocessor was used and applied to a randomised spot on each array included in the bioprocessor. This acted as a quality control for each array and sample processing.

2.3 Optimisation of instrument conditions

Extensive experiments were carried out to achieve the optimal conditions for analysis on the ProteinChip instrument (as seen in section headings 2.3a-f). A balance of the maximum number of peaks with the best resolution is desirable. In order to achieve this, conditions such as protein concentration, pH of buffers, array chemistry and laser power is assessed.

For these initial stages of development, aliquots of blood from sheep housed on the IAH Edinburgh Farm were utilised to avoid over use of the experimental samples. These conditions were then finalised with the use of aliquots of a pooled sample consisting of a mixture of all the scrapie plasma samples from each year group. This achieved a mixture of all possible proteins expressed in all the samples.

a) Protein Determination

A typical protein concentration determination was established using aliquots of two plasma samples which were analysed using a BCA Bradford Protein Determination Kit (Thermo Scientific/ Pierce) as described in the kit instructions. 10ul of sample was diluted to 100ul then 10ul serially diluted to 100ul. A 10ul aliquot of this solution was placed in a microplate and 300ul of BCA solution added. The samples were similar in concentration with an average concentration of 76,222µg/ml.

b) Albumin Depletion

Some studies using SELDI-TOF for the analysis of serum or plasma have recommended the depletion of albumin [9]. There are two schools of thought on the matter; (i), that the albumin is so abundant that it obscures the less concentrated small proteins seen in the spectra and (ii); that in taking the albumin out there is a risk that small proteins bound to albumin will also be removed.

To investigate this in relation to the sheep plasma samples we used an albumin depletion column kit (Calbiocem). 40µl plasma plus 360µl of buffer were added to the column and allowed to flow through. The column was washed twice and the flow-through analysed on CM10 and Q10 arrays by SELDI-TOF. At the sample incubation step of array preparation 100µl plus 100µl of appropriate buffer was added to the bio-proccessor containing the arrays.

c) Denaturation of proteins

Initial studies [10] prepared plasma and serum samples prior to analysis by denaturing with catatrophic buffers. In our study we aimed to minimise the preparation of samples whilst ensuring reproducibility. The number of preparative steps involved would be critical for the success of the methodology in a future population wide screening programme. We tested denatured plasma, neat plasma and diluted plasma samples as follows:-

i) 20µl plasma was added to 30µl U9 buffer (containing 9M Urea, Bio-Rad Laboratories), the sample then incubated for 30 minutes. ii)Two array chemistries (CM10 & IMAC-Cu) were tested with both U9 sample preparation and the original plasma sample diluted 1:40. iii) Appropriate buffers were applied to each array as per instructions (CM10 low stringency buffer kit, IMAC-Cu buffer kit, Bio-Rad, Laboratories) with 5µl or 10µl of sample added to wells containing buffer (a total of 200µl) and incubated for 40 minutes at the sample incubation step. iv) A human serum control sample was also included in the array. v) An acquisition protocol using two laser powers was generated on the instrument and the array data was collected.

d) Optimal dilutions

The viscosity of the plasma samples was a concern for this type of analysis. If the proteins in the plasma were to "clump" together then the adherence to the array surface would be compromised. A range of dilutions of the plasma (1:10, 1:15, 1:20, 1:25, 1:30, 1:35, and 1:40) was applied to a NP20 (this array captures the total protein in the sample) and the arrays read on the instrument at two laser settings.

e) Array chemistries

Aliquots of the pooled sample were run on several of the same type of array (CM10 or IMAC). From the resulting data a cross section of peak intensity was compared and the coefficient of variance between samples was calculated as an indication of accuracy of the methodology. The CV% for the analysis ranged from an average of 13 -21% over a total of 18 CM10 arrays for all laser conditions which was acceptable. The IMAC-Cu initially did not achieve similar results however addition of a low concentration of detergent to the buffer (0.1% Triton) enhanced the binding capacity of the array surface leading to improved CVs over 10 peaks of 9-30%.

f) Laser power

Proteins of different molecular weights reach the instrument detector after a period of time (time of flight (TOF)) relative to the mass/charge ratio(m/z). The larger proteins require more laser power to energise and fly through the time of flight tube. This "stretching of time'' leads to the proteins at high molecular weight to resolve less well than the smaller proteins which appear as distinct peaks in the spectra. However there is a trade off between low laser power and the resolution of large proteins and high laser power which will raise the background noise at the lower end of the spectrum thus skewing the low molecular weight proteins. For the best resolution of peaks therefore it is best practice to use two laser power settings, one for the lower molecular weight proteins and one for the larger proteins. To achieve the best results several laser power settings were tested on each array chemistry and protocols were developed from the results.

2.4 Analysis of samples - final

The arrays used in SELDI-TOF analysis consist of 8 specialised surface spots, designed to capture sub-sets of proteins from crude biological samples according to their chemical or biological properties similar to chromatographic methods. Initially three ProteinChip® array chemistries, anion, cation (Q10, CM10) and immobilized metal affinity capture coupled with copper (IMAC-Cu) were used to achieve the maximum possible number of protein peaks for profiling. The Q10 arrays did not give as good a range or resolution of peaks (signal to noise ratio) as the CM10 and IMAC-Cu arrays, therefore we decided to proceed with just two array chemistries.

Each array was placed in a specialised holder called a bio-processor (ProteinChip®) which accommodates the arrays in a 96 well format and allows for buffer and sample incubation. A plan of the bio-processors with array numbers, sample details and analysis protocols was entered in the ProteinChip® DataManager software and printed out as a guide to array preparation. A separate bio-processor for each BSE time point or scrapie year, plus individual array chemistry was assembled where samples were arranged on arrays in combinations of infected and non-infected control samples. A pooled sample was spotted on every array for the assessment of spectrum quality.

Initial pilot experiments had shown more consistent replication when the samples were diluted before addition to the bio-processor. Samples (15µl) were therefore pipetted into a 96 well mircotitre plate and 140µl buffer (CM10 or IMAC) + 0.1%triton added. The plate was covered and briefly vortexed.

IMAC arrays were first incubated with copper sulphate to bind copper ions to the surface as per instructions in the IMAC–Cu Buffer Kit (Bio-rad Laboratories Inc.) before the addition of buffers and samples. The appropriate buffer for each array chemistry was added as follows: - CM10:CM10 100mM sodium acetate pH 4 (Low stringency CM10 buffer Kit, Bio-rad Laboratories Inc.) and IMAC-Cu: 100mM sodium phosphate, 500mM sodium chloride pH7 (IMAC- Cu Buffer Kit, Bio-rad Laboratories Inc.). In the binding stages i.e. before addition of the sample all buffers had 0.1% Triton added which was excluded in the rinsing stages (after removal of the sample). The diluted sample (140µl) was added to the bio-proccessor wells and left to incubate on a shaking platform at room temperature for 40 minutes. After removal of the samples the arrays were washed three times with buffer (150 µl shaken on shaking platform for 5 mins each rinse) followed by a short water rinse. The arrays were then dried and two applications of 0.5µl matrix solution (sinapinic acid/50%acetonitrile/0.1% TFA) applied to each array spot.

SELDI-TOF MS analysis

Arrays were analysed on a ProteinChip® Enterprise system (Bio-rad Laboratories Inc.). Protocols were developed for all array types by adjusting laser settings to yield the optimal resolution of peaks. The arrays were read twice, once at a low laser power focused on the 0-50kDa spectral range and again at a high laser power focused on the 0-200kDa spectral range. Spectral data were collected using ProteinChip® DataManager software with an average of 583 laser shots per array spot. Calibration algorithms are necessary to convert the time of flight scale into a mass/charge (m/z) scale therefore all spectra were externally mass calibrated using equations constructed with an All in One Protein Standard (Bio-rad Laboratories Inc.) using 4 standards covering the focussed area in the low molecular range (7-30kDa) and 4 in the high molecular range (25-100kDa). All spectra peak intensities were normalised on total ion current. Protein peaks of similar mass found in spectra were clustered automatically using a cluster wizard function within the ProteinChip® DataManager. This function was set to define clusters of commonly found (across all the samples) peaks all of which had a height on the first pass of 5 times the signal to noise ratio and 5 times signal to noise valley depth with a minimum peak threshold of 20% i.e. the peak must occur in at least 20% of the spectra to be classified as a peak. Where the peak is not found in a spectrum the software estimates the missing peak using the cluster centroid peak height intensity.

All data were then exported to an excel file for further statistical analysis.

Statistical Analysis

All possible peaks/clusters were included in the analysis of both the scrapie and BSE array chemistries and laser power (CM10 and IMAC-Cu arrays at low and high laser power). The peak intensity values were log 10 transformed and averaged over replicate determinations for each sample as appropriate. As a preliminary analysis for each peak, the log-intensities were tested for differences between the two states (BSE /scrapie positive and BSE /scrapie negative controls) by Mann-Whitney U tests, and assessed collectively by the Benjamini and Hochberg false discovery rate (FDR) [11]. Prediction of status was in the range 0 (certain BSE/scrapie) and 1 (certain control), from a multiple regression on peak log–intensities fitted on the logistic scale. As there were many more peaks than samples, the number of predicting peaks was reduced by a penalty which excluded peaks with large regression coefficients (the least absolute shrinkage and selection operator (lasso) penalty, [8]). A two-level cross validation scheme was necessary, with partitions chosen at random. The outer level cross validation partitioned the samples into several unique groups. The number of partitioned groups was determined by the total number of samples i.e. in general 10% of the total number of samples is set aside for testing, and complementary training sets identify the penalty parameter giving the model with the best prediction. This used an inner cross-validation step again with 10% total number of samples in the testing with complimentary

training sets. The penalty parameter was chosen to maximise the likelihood of the predictions over all the sets of test samples in the inner partition. The model with this penalty parameter was then fitted to all observations, and used to predict the status of the “left out” test samples in the outer partition. This was repeated for all sets of test samples in the outer loop partition.

Peaks were considered to be “good’’ predictors if they were selected (i.e. have non-zero parameters) in a minimum over half of the testing sets in the outer partition, but were excluded if their non-zero parameters were not all of the same sign, indicating an influential and inconsistent outlier in a peak's data.

Ninety nine different random outer partitions were run, producing a range of predictions for each sample. All good peaks represented in > 80% of partitions were taken as the best predictor set. The quality of predictions over the 99 partitions was assessed by measurement of the area under the receiver-operating characteristic (ROC) curve, (AUC).

The robustness of the predictions to the scale on which the model was fitted, was tested by using leave-one-out cross validation with a linear discriminant analysis (LDA). LDA is equivalent to a multiple regression model fitted to the status (0 or 1) on the observed, rather than the logistic scale. This also allowed a rapid test of the classification power of randomly chosen groups of “Best Peaks”.

2.5 Protein identification

The peaks shown to be significantly differentiated between normal and diseased samples are identified by their mass (m/z) in the above methodology. Further steps are required to purify and identify the individual peaks based on the chemical properties indicated by the pH and type of array used in the SELDI-TOF process. Sequential fractionation of the plasma reduces the complexity of the sample revealing the protein of interest. A combinatory protocol (see Figure 2.1) was developed in this study based on previous studies [12]. Plasma samples of both control and diseased animals were diluted (100µl plasma plus 150 µl 9M urea diluted to 250 µl U1 buffer) and fractionated using mini-spin columns (SigmaPrep spin columns, Sigma-Aldrich) packed with 200µl Q Hyper DF or CM Hyper DF resin (Bioserpa). Columns were washed sequentially and fractions eluted with 500µl of buffers Tris-HCl pH9, HEPES pH7, NaAcetate, pH4 and 5 NaCitrate pH3 and finally an organic fraction with a solution of 33% isopropanol/0.1% TFA. Each fraction was spotted on a NP20, IMAC-Cu and CM10 array and analyzed on the ProteinChip® reader to assess the presence of the protein peak of interest. By examining the profiles on the arrays on which they were first observed, at each stage of the fractionation the protein of interest can be tracked and the fraction in which it appears further purified. The NP20 array surface captures all the protein present in the sample. This is important as an indicator of the complexity of the sample. Should the peak of

interest lie within an area of the spectrum which is highly complex then isolation of an individual protein would be extremely difficult leading to an aberrant identification. The fractions displaying the presence of the peak of interest were further fractionated on a hydrophobic column (Pierce Pepclean C18) by sequentially washing with 10%, 40% and 70% acetonitrile/0.1%TFA. Again these fractions were spotted on a NP20 array and analyzed on the ProteinChip® reader. The fractions displaying the presence of the peak of interest were then dried on a SpeedVac vacuum centrifuge. The final purification and isolation of the fractions was carried out using 1D SDS-PAGE electrophoresis (Novex mini-gel system, Invitrogen). Gels for this step were selected for the best separation and resolution of the low molecular weight proteins (Novex 10-20% Tricine with Tricine running buffer and 12% NuPage BisTris with MES running buffer (Invitrogen)). The samples were re-constituted in sample buffer according to the gel system used (Tricine SDS sample buffer for Tricine gel and LDS sample buffer for Bis-Tris gels):- 20µl of 1X sample buffer was added to the dried organic fractions and 40µl 1xsample buffer added to the dried aqueous fractions which appeared to be more precipitated. The samples were incubated for 30 minutes at room temperature briefly vortexed and heated on a hot block at 80°C for 10 minutes. 15µl of sample was applied to a 1D SDS-PAGE gel and the gel allowed to run until the stained standard dye front was near the bottom of the gel. The gel was stained with Coomassie blue (R-250 Sigma-Aldrich). Three standard protein ladders (Novex Sharp Pre-stained protein standard, MARK12 unstained standard (Invitrogen), Protein Precision Plus (Bio-Rad)) were used to indicate the approximate mass weight of the bands.

As a comparison to the above fractionation protocol we also used a bead technology to purify the samples (Proteominer bead kit, BioRad Laboratories) which is recommended for biomarker discovery. These fractionations were carried out as described in the kit.

The gels were sent to the Moredun Research Institute, Edinburgh, Proteomics Facility for the excision of bands and identification by peptide fragmentation using a tandem mass spectrometer as described below.

2.5.1 Mass spectrometer identification of excised bands

Protein bands were cut from SDS-PAGE gels under sterile conditions and place into sterile eppendorfs. Gel pieces were washed in 400µl 50% acetonitrile, in 25mM ammonium bicarbonate pH8.0 for 15 minutes 3 times. Gel pieces were washed in 100% acetonitrile for 5 minutes before being dried in a speed-vac for 30 minutes.

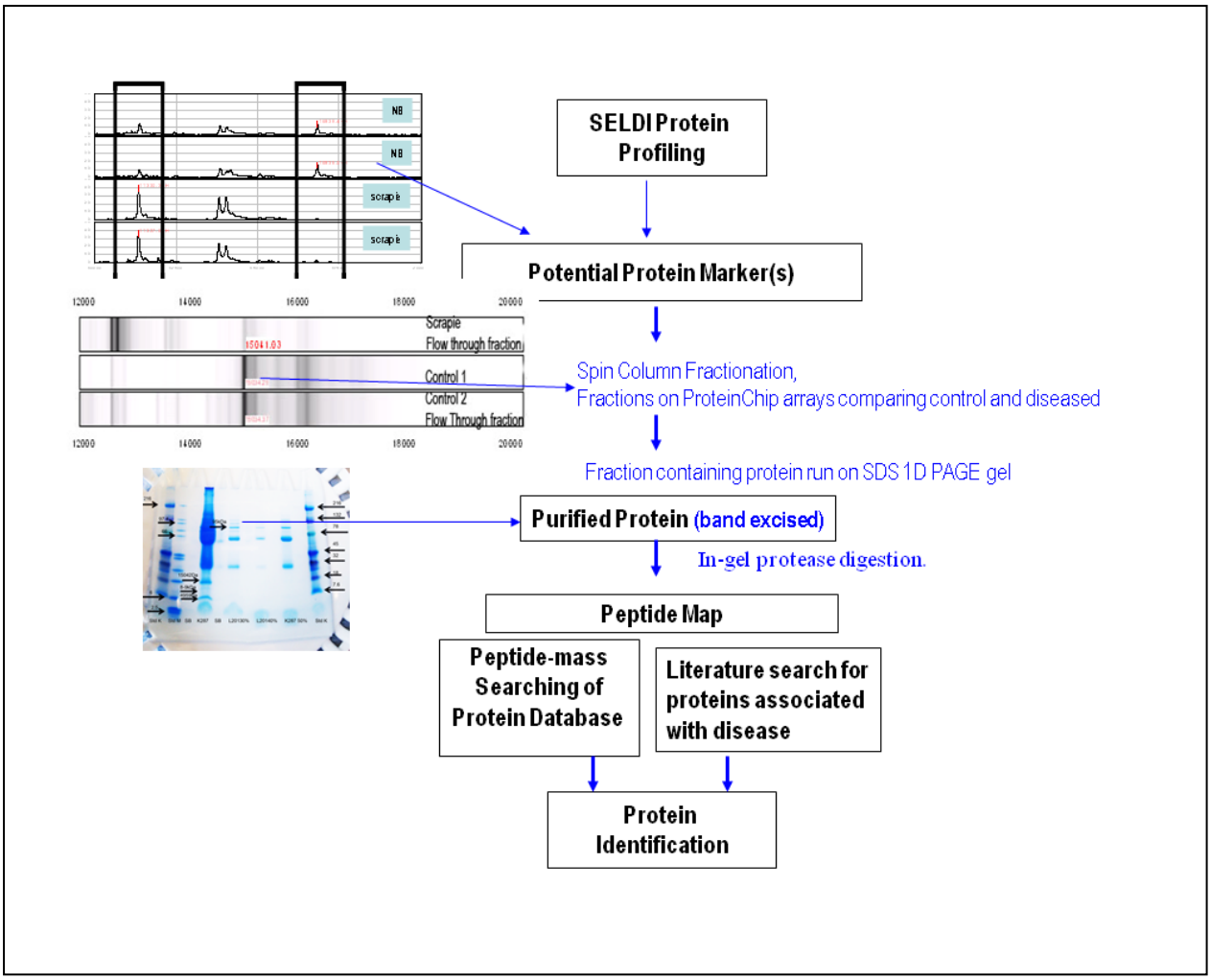
The gel pieces were covered with Trypsin solution (10µg/ml trypsin (Promega), 25mM ammonium bicarbonate pH8.0 and allowed to rehydrate for 10 mins. After 10mins if all the solution was absorbed a further 5 µl of trypsin was added. The samples were then incubated overnight at 37ºC

(16-24) hours. After incubation 10µl 50% Acetonitrile (Rathburn, 5% TFA (Fluka) was added and the sample was vortexed for 30 minutes.

For MALDI analysis, 0.5µl of the supernatant was mixed with the 0.5µl of matrix solution (10mg/ml α-cyano-4-hydroxycinnamic acid in 50% acetonitrile, 0.1% TFA) and spotted on the sample well of a MALDI sample plate. 10µl of Pepmix 1 standards (Bruker Daltonics) were mixed with 10µl matrix solution and 0.4µl was spotted adjacent to each sample spot. Spots were allowed to dry completely before being analysed using the Ultraflex II MALDI-ToF-ToF mass spectrometer. Data for Peptide Mass Fingerprint (PMF) analysis was accumulated from 10x100 shot batches. Each sample was calibrated using the adjacent standard. Where required, further analysis was carried out by obtaining ion fragmentation on individual peptides (MS/MS).

Data was processed using Flex analysis software and compiled into mass lists. Searches were performed using Mascot^{TM} software searching NCIB BLAST and SwissProt/EMBL databases. Mass tolerances were set at 50 ppm for PMF and 0.5 Da for MS/MS data.

Figure 2.1 Scheme of Protein Identification

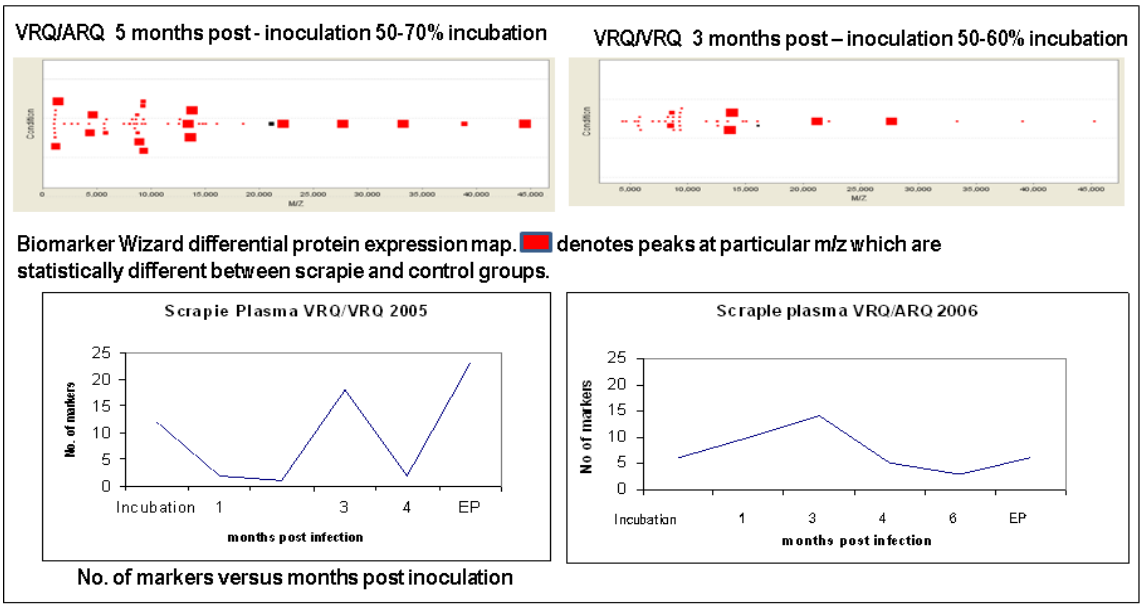


3. Results

3.1 Establishing scrapie markers

Initially the data for each genotype were analysed separately however there were insufficient samples particularly in control groups to achieve a statistically relevant outcome. Ideally age matched animals and samples taken at the same time would be preferred however we were limited to the archived samples available to us. To assess the possibility of combining the samples, we utilised the Biomarker Wizard software to visualise expression differences at each time point and in each year sample batch. A graph of the number of expression differences versus the months at which the samples were taken after inoculation indicated that the trends were similar with a peak around 50% of the total incubation time (Figure 3.1).

Figure 3.1 Biomarker Wizard map and graphs of the number of markers visualised at each sample time point.



No. of markers versus months post inoculation

As the m/z of the peaks found in each year sample batch and genotype (2005 VRQ/VRQ & 2006 ARQ/VRQ) were also similar we decided to pool the data. A summary of the groups and results can be seen in Table 3.1.

Table3. 1 Summary of time course groups

Group	% Inc	Number of samples (Scrapie/Control)			No. Best Peaks	AUC range	AUC median
		VRQ/VRQ	VRQ/ARQ	Total			
1	Uninfected	5/3	5/3	10/6	*ANP	-	-
2	1-29	5/0	5/2	10/2	*ANP	-	-
3	30-49	4/3	2/0	6/3	*ANP	-	-
4	50-59	4/3	5/0	9/3	*ANP	-	-
5	60-69	5/3	5/3	10/6	*ANP	-	-
6	70-99	4/0	5/3	9/3	*ANP	-	-
7	100	13/3	5/3	18/6	10	0.6-0.97	0.92

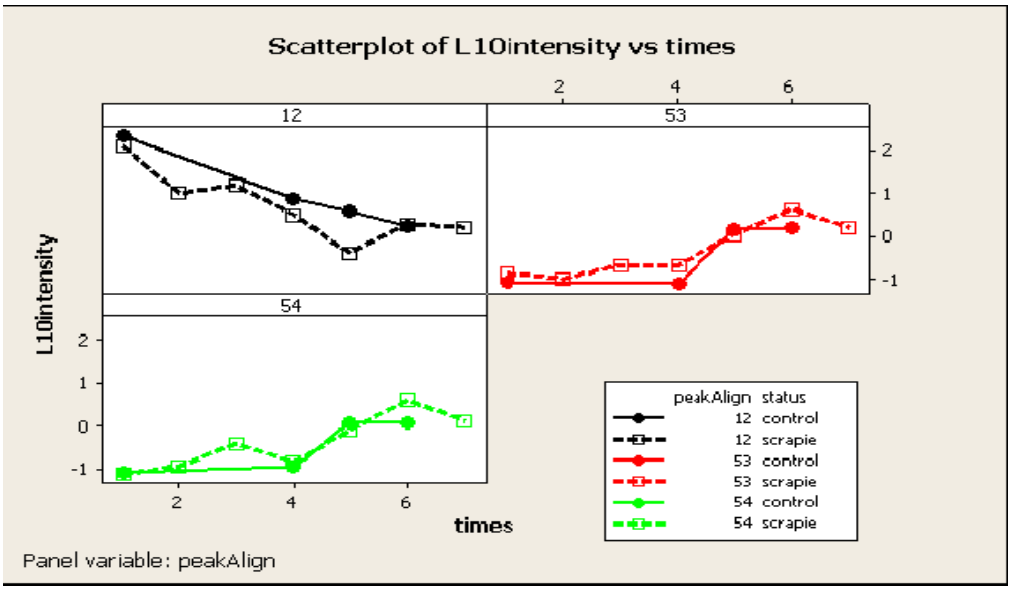
*analysis not possible due to lack of data

We therefore combined the two sample sets (VRQ/VRQ and VRQ/ARQ) using the incubation time as an indicator for the formation of relevant groups. This resulted in a successful analysis of the clinical group however analysis of the other groups was not possible.

Natural Variation

We also investigated the natural variation in protein expression differences throughout time. Taking three peaks (peak no. 12, 53 and 54) at particular m/z mass present in all the samples we compared expression differences over all time points.

Figure 3.2 Variation of protein expression over time

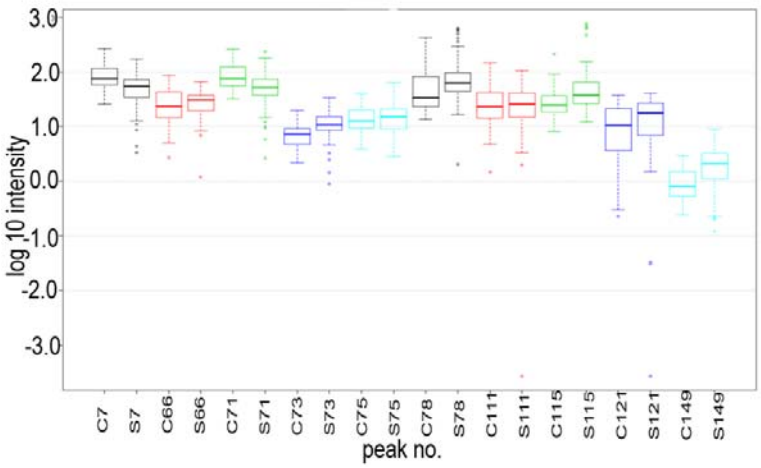


As can be seen in Figure 3.2, although there are some differences between the infected (scrapie) group and the normal control group the overall trend between groups is the same. This would imply that the contribution of natural protein expression differences would be negligible when comparing the control and infected animals at each time point.

Statistical Analysis

Plasma samples were analysed on two array types (CM10 & IMAC-Cu) at two laser settings (high and low power). This generated 150 clusters which after omitting peaks with log intensities <0 and cross validation analysis, 10 clusters ("Best Peaks'') were found to be differentially expressed at the clinical stage of disease (Figure 3.3). The area under the curve (AUC), of the receiver operator curve (ROC) values, range between 0.6 – 0.97 with a median of 0.92 suggesting that the 10 "Best Peaks'' (Figure 3.4A) have some predictive value. In a randomised analysis where the peaks and status are not linked the AUC ranges between 0.25 – 0.61 with a median of 0.44 (Figure 3.5A). The outcome of this analysis would suggest that these peaks could have arisen by chance giving further proof that the 10 Best Peaks (Figures 3.3, 3.4D and Table 3.2) have information specific to scrapie. The distributions of the 99 predictions of disease status for each sample are displayed in Figure 3.4B. If prediction status is taken as a threshold of 0.5, in the scrapie group 17 sheep were predicted to be positive for scrapie with one false negative and 4 sheep in the control group were predicted negative for scrapie with 2 false positive sheep. Figure 3.4C is the same information as in 3.4B, the average mis-classification rate shown in a different format. The distribution of the 99 partitions is shown for the 10 "Best Peaks'' in Figure 3.4D. None of the proteins were FDR significant for the Mann - Whitney test with little difference in log- intensities between status showing the importance of joint distributions (Figure 3.4A-D and Table 3.2)

Figure 3.3 Scrapie clinical group "Best Peaks''



Box plots of log 10 intensities for 18 scrapie and 6 controls ordered by status (C – control, S - scrapie) and peak number (as in Table 3.3)

Table 3.2 Protein "Best Peaks'' descriptions as visualised in Figures 3.2-3.3.

Peak	Peak No.	No of X Vals.	Fig 3.1 colour.	m/z(Da)	median co-efficient	M-W *p* value	FDR<0.05	Array Type /laser power
1	149	99	Black	95250	-2.573	0.066	FALSE	IMAC.H
2	121	99	Black	9668	-1.377	0.137	FALSE	IMAC.H
3	115	99	Black	8568	-1.015	0.018	FALSE	IMAC.H
4	78	99	Black	8601	-4.257	0.104	FALSE	CM10.L
5	75	99	Black	8034	-1.382	0.673	FALSE	CM10.L
6	73	99	Black	6945	-0.883	0.137	FALSE	CM10.L
7	71	99	Black	5812	4.657	0.001	FALSE	CM10.L
8	66	99	Black	2691	3.44	0.077	FALSE	CM10.L
9	111	97	Red	7474	1.09	0.09	FALSE	IMAC.H
10	7	90	Green	5716	0.629	0.006	FALSE	CM10.H

Figure 3.4 Summaries of parameters and predictions from lasso model analysis (see also Table 3.2)

A. Receiver Operator Characteristics Curves

Plots of proportions of true positives (sensitivity), against false positives (1-specificity) from 99 random partitions. Area under the curve (AUC) ranges from 0.60 to 0.97, with a median of 0.92.

B. Predicted animal Disease Status

Box-plots of sample predictions from 99 random partitions, range from 0=scrapie positive to 1=scrapie negative. Animals are ordered (from the left , scrapie 1-16 and controls 1-6) by increasing median prediction within scrapie and control status.

C. Prediction classification

Plots of the numbers of status predictions, from the median animal predictions, at 9 prediction thresholds. Colours: black=scrapie, red=controls and blue=average of scrapie and control numbers. Solid lines denote correct predictions and dotted lines denote wrong predictions.

D. Most predictive protein peaks

Box-plots of parameter values for each peak (+ down regulation and – up regulation of scrapie samples) from 99 random partitions. Colour represents the number of partitions in which the peak appeared (black =99, red=97 & green=90) and peak labels are as in Table 3.1

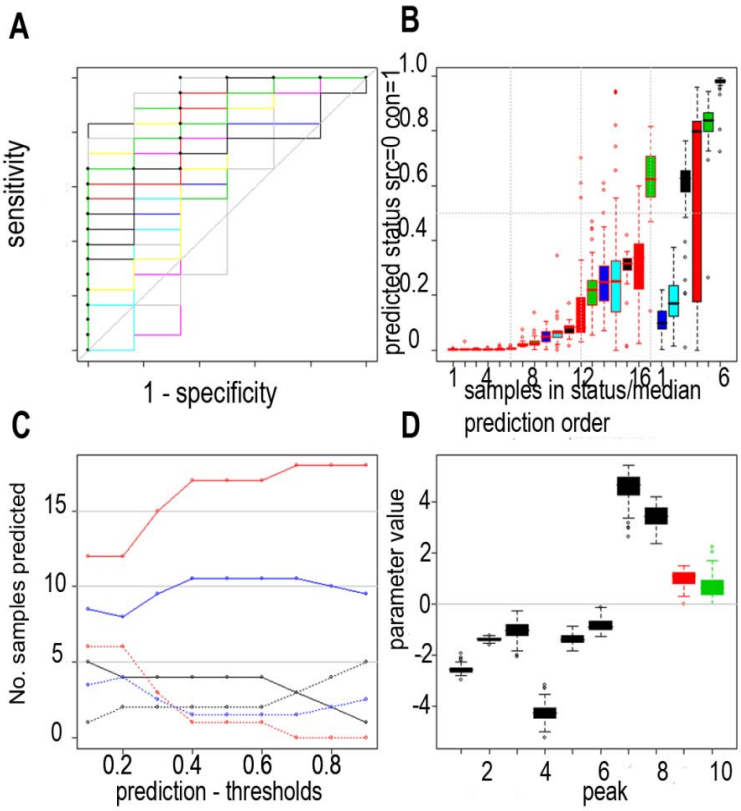
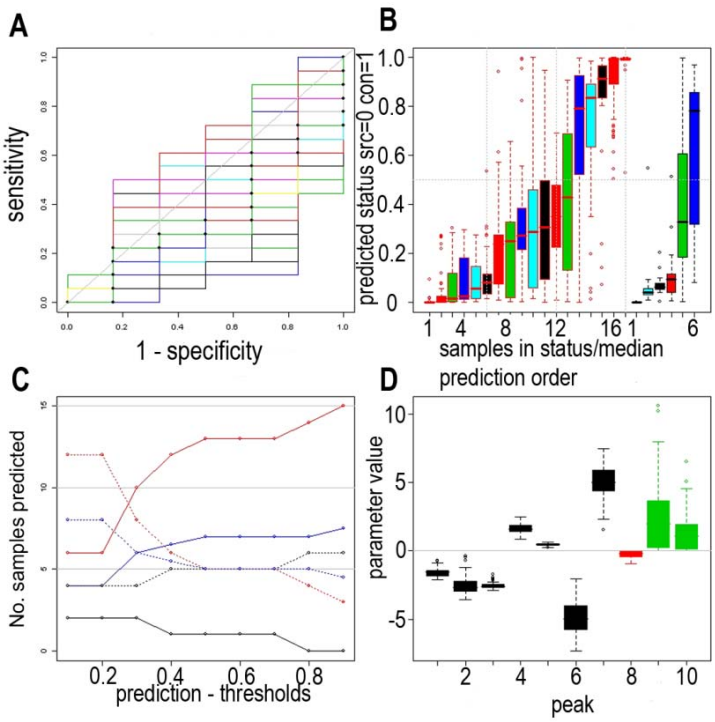


Figure 3.5 Randomised plots

A. Receiver Operator Characteristics Curves

Plots of proportions of true positives (sensitivity) against false positives (1-specificity) from 99 random partitions. Area under the curve (AUC) ranges from 0.25 to 0.61, with a median of 0.44

B - D descriptions as in Figure 1.



3.2 Establishing BSE markers

3.2.1. Predictive Models and Best Peaks for all time points

As is the nature of large animal studies predicting the length of disease incubation is not absolute and the project from which the plasma was obtained experienced many drawbacks one of which was that the animals were alive longer than anticipated. We were therefore unable to receive the full complement of samples at the clinical stage of disease as this would have been out with the duration of the current funding. A decision was made to use data from the animals for which we had clinical samples and that had been confirmed as being infected (BSEp), by clinical signs and the deposition of PrP^{Sc} in brain tissue visualised by western blotting. A predictive classification model was constructed using these samples and non-infected age matched control animals (CONco). Data from the remaining animals was classified into two groups, those which were still alive (BSEal) at a cut off date of 10th February 2010, and those which had died and were non positive for BSE (BSEnp). These data and data from a non scrapie group of plasma samples (*toxoplasma gondi*) were analysed by cross – validation to test the prediction potential of each of the time points "best peaks'' panels of protein markers. A summary of the samples can be seen in Table 3.3.

Table 3.3. Summary of sample numbers and groups

Group	Months Post innoc.	BSEal	BSEnp	BSEp	CONco
1	Not innoculated	12	7	19	10
2	2	11	6	21	9
3	4	11	6	18	10
4	6	12	7	17	10
5	8	11	7	20	10
6	10-12	12	4	21	9
7	18-28	12	3	18	8
8	clinical	0	3	10	0

Data from groups BSEp and CONco were statistically analysed using linear regression methodology to establish a predictive classification model at each time point as described in the Methods section. Sample data were partitioned into outer and inner training/testing sets according to the number of samples in each time point group (Table 3.4) in a two level cross validation scheme to identify the penalty parameter giving the model the best prediction (see Methods section).

Table 3.4 Cross validation scheme training/test sets

Group	BSEp	CONco	outer N fold	inner N fold
1	19	10	14	13
2	21	9	10	9
3	18	10	14	13
4	17	10	9	8
5	20	10	10	9
6	21	9	10	9
7	18	8	13	12
8	10	0	9	8

Results are visualised in a set of graphs for each time point, similar to Figure 3.4, which are shown in a supplementary file “BSE Prediction Models.pdf’’; (A) the area under the curves (AUCs) of receiver operator curves (ROCs) gave a measure of true positives (sensitivity) against false positives (specificity) from the 99n-fold partitions, (B) box plots of sample disease status predictions from the 99 partitions of the nested cross validation analysis, (C) a prediction classification graph, using the “Best Peaks’’ at each time point where the median animal predictions are shown at nine threshold values and (D) the most predictive markers are shown as box plots of the parameter values for each peak. The prediction error, AUCs for all time points and the ranking indicative of the most informative set of markers (ranked 1-7), are summarised in Table

3.4. A more detailed description and figures of box plots of the marker sets is set out under a heading for each time point. Protein markers which occur in more than one of the “Best Peaks’’ sets appear in related colours in Tables 3.6-13 e.g. Peak No. 65 coloured brown at 4973m/z appears in T3,T5 & T8.

A total of 201 peaks were observed in spectra over all conditions (array type laser condition and time point groups), 44 IMAC-Cu peaks at low laser, 46 at IMAC-Cu high laser, 60 CM10 low laser and 51 at CM10 high laser. Duplication of peaks occurred between the two different array types (CM10 and IMAC-Cu). Whilst we cannot be certain they are the same proteins, the masses are extremely close. 20 protein peaks at low laser shared similar masses in both IMAC-Cu and CM10 arrays and 6 at high laser were also shared across arrays. This implies that we are indeed picking up definite peaks and not spurious artefacts.

Peaks which were present in >80% of the cross- validation analysis were considered predictive peaks. Box plots (Figures 3.6 - 3.14) of the proteins differentially expressed at each time point (“Best Peaks’’) showing the log 10 intensity of each peak (BSE (B) and Control(C)), as numbered in the accompanying tables are shown under each heading Group T1 - 8 below. Peaks found to be differentially expressed at one or more time points were found over all conditions, 9 IMAC-Cu peaks at low laser, 9 IMAC-Cu high laser, 19 CM10 low laser and 13 at CM10 high laser. Peaks are displayed in tables 3.5 - 3.12 those which appear in more than one time point are coloured according to similar mass/charge /ratio (m/z).

Table 3.5 Summary of all time points, area under the curve (AUCs), prediction error and ranking

In this table we are interested in determining the set of “Best Peaks’’ which has the highest prediction potential. T1 is excluded from this table as the animals were not inoculated at this point and therefore could not be predictive for the presence of BSE.

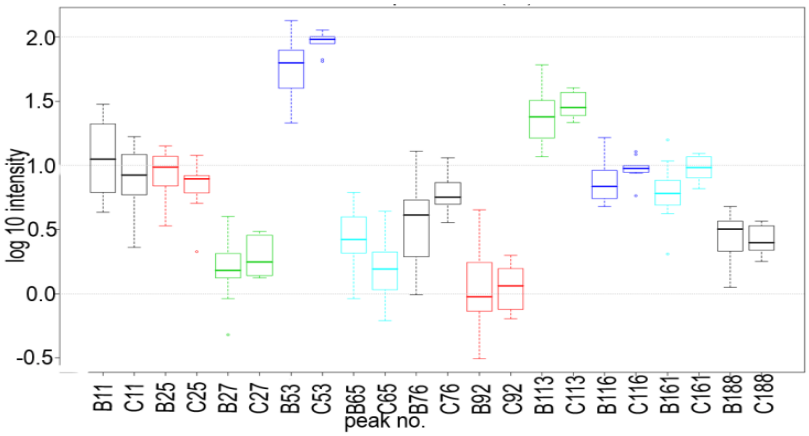
Time point	prediction error						AUC X.validation						Rank
	Min.	1st Qu.	Median	Mea	3rd Qu.	Max.	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
2	4.00	7.00	8.00	7.92	9.00	11.00	0.66	0.78	0.80	0.80	0.83	0.88	5
3	5.00	7.00	8.00	8.33	9.00	11.00	0.76	0.81	0.82	0.82	0.85	0.90	3
4	4.00	6.00	7.00	7.37	8.00	11.00	0.73	0.82	0.84	0.84	0.86	0.91	2
5	7.00	9.00	10.00	10.27	11.00	13.00	0.67	0.73	0.76	0.76	0.80	0.86	6
6	9.00	11.00	12.00	12.48	14.00	17.00	0.51	0.64	0.67	0.67	0.72	0.78	7
7	6.00	7.00	8.00	8.02	9.00	12.00	0.71	0.78	0.81	0.81	0.83	0.88	4
8 clinical	1.00	3.00	4.00	4.11	5.00	7.00	0.70	0.84	0.87	0.87	0.90	0.94	1

Group T1 – Natural variance

The sheep were recruited to the transfusion project at five separate time intervals, September 2006 – January 2007, therefore rather than being predictive; this set of markers displays the natural difference between animals at the start of the study. Blood samples were taken from all sheep before the BSE group were inoculated.

Assigned groups (control and BSE) before dozing with BSE did show protein differences in the cross validation analysis (Figure 3.6). This would be expected due to the wide variation in the environment from which the animals originated and the time between groups before recruitment. The linear discriminant analysis which was carried out for all the "Best Peaks'' sets on animals not included in the building of the classification model and the toxoplasma infected animals, indicates in this initial grouping (Figure 3.7) that these markers do not have any predictive value as would be expected. From this point on the BSE and control groups were sampled at similar intervals from the point of inoculation.

Figure 3.6 Box plots of T1 "Best Peaks'' (log 10 peak intensities v peaks BSE (B) and Control(C))

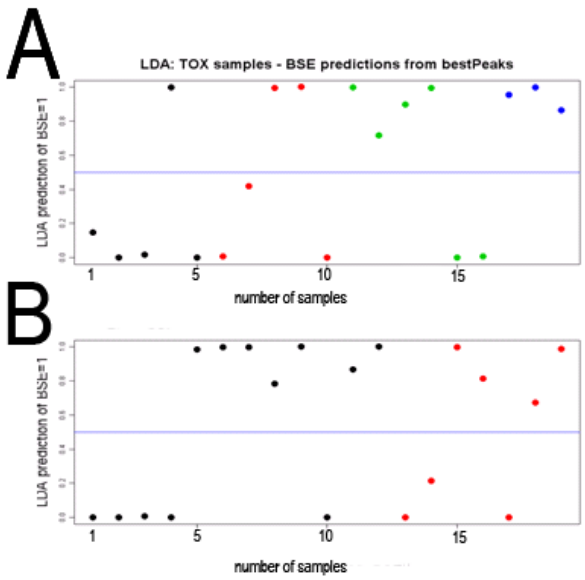


Box plots of log 10 intensities for 19 BSE and 10 controls ordered by status (B – BSE, C – control) and peak number (as in Table 3.6).

Table 3.6 T1 Best Peaks

Peak no.	Array/laser condition	No.X Validation	median Coef.	M-W *p* value	med. Intensity	BSE.med intensity	CON.med intensity	M/Z. median	M/Z.min	M/Z.max
188	IMAC.L	99	9.804	0.484	0.447	0.504	0.397	14760	14744	14780
161	IMAC.L	99	-8.29	0.008	0.857	0.781	0.983	4076	3976	4082
115	IMAC.H	99	-2.583	0.094	0.944	0.835	0.976	3938	3847	3993
113	IMAC.H	99	-2.798	0.228	1.433	1.377	1.451	1199	1195	1226
65	CM10.L	99	4.013	0.024	0.324	0.422	0.192	4963	4959	4973
53	CM10.L	99	-6.905	0.003	1.869	1.798	1.981	1464	1436	1476
27	CM10.H	99	-10.255	0.211	0.181	0.181	0.249	16090	16025	16129
25	CM10.H	99	3.423	0.228	0.91	0.987	0.896	14770	14757	14785
11	CM10.H	99	1.778	0.308	1.011	1.046	0.925	5706	5668	5798
76	CM10.L	98	-0.816	0.062	0.698	0.614	0.752	8266	8260	8294
92	CM10.L	92	0.576	1	0.035	-0.022	0.06	12606	12549	12686

Figure 3.7 Linear Discriminant analysis with T1 "Best Peaks''

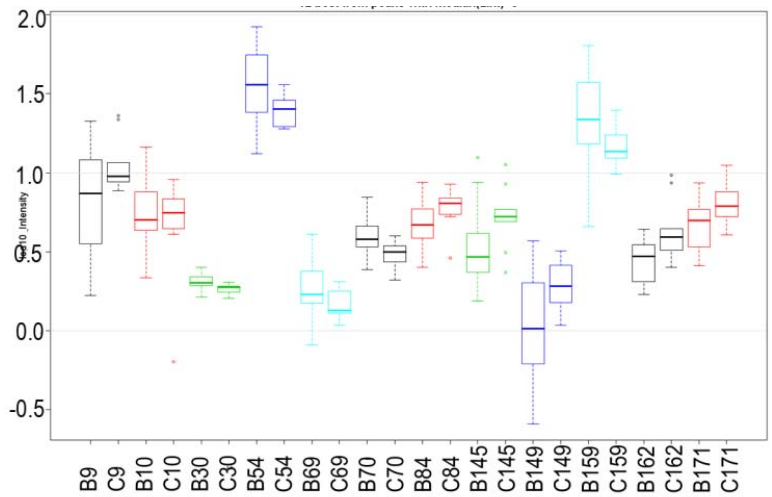


A – *T. gondi* infected and related controls
Black – *T.gondi* infected(ToxH)
Red – *T.gondi* infected (ToxL)
Green – non infected (ToxM)
Blue – non infected (Tox S)

B – BSE groups not included in classification model.
Black – BSE infected (BSEal)
Red – non-positive **(**BSEnp)

Group T2 – 2 months

Figure 3.8 Box plots of T2 "Best Peaks'' (log 10 peak intensities v peaks BSE (B) and Control(C))



Box plots of log 10 intensities for 21 BSE and 9 controls ordered by status (B – BSE, C – control) and peak number (as in Table 3.7)

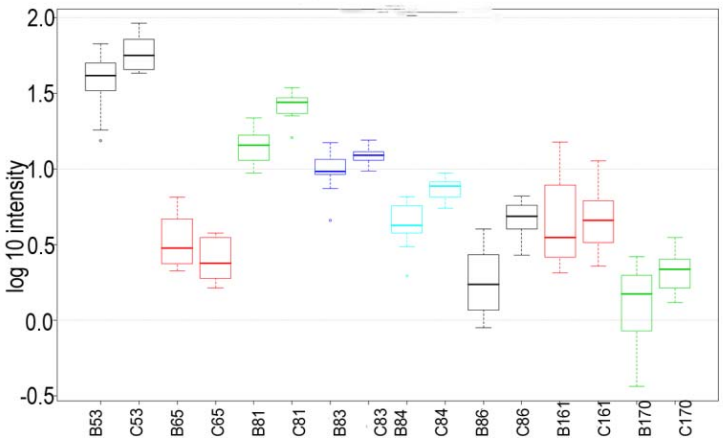
A panel of 12 protein peaks (Fig 3.8) were found to be predictive for BSE with AUCs in the range of 0.66 - 0.88 with a median of 0.82 (Table 3.7). Six protein peaks were up-regulated in the BSE samples. Peak masses (m/z ratio) ranged between 2kDa – 96kDa. Predictions of disease status were shown to be 3 false negatives in the BSE positive group (n= 21) and 4 false positives in the control group (n=9) resulting in a measure of Sensitivity of 86% and a specificity of 56%

Table 3.7 T2 Best Peaks

Peak no.	Array/laser condition	noX Validation	median Coef.	M-W *p* value	med. Intesity	BSE.med intensity	CON.med intensity	M/Z.median	M/Z.min	M/Z.max
171	IMAC.L	99	-6.172	0.063	0.722	0.699	0.79	7497	7491	7503
162	IMAC.L	99	-3.713	0.014	0.506	0.47	0.594	4382	4356	4409
159	IMAC.L	99	0.867	0.032	1.241	1.338	1.134	2011	2000	2024
145	IMAC.H	99	-2.742	0.028	0.53	0.466	0.721	66269	66168	66874
84	CM10.L	99	-2.231	0.114	0.704	0.67	0.805	9199	9195	9204
70	CM10.L	99	3.906	0.012	0.557	0.578	0.5	5961	5940	5967
69	CM10.L	99	4.086	0.147	0.225	0.229	0.129	5894	5888	5900
54	CM10.L	99	4.684	0.028	1.529	1.556	1.404	1889	1887	1919
30	CM10.H	99	13.535	0.015	0.291	0.302	0.276	22882	22809	23007
10	CM10.H	99	2.942	0.983	0.71	0.702	0.745	5590	5581	5617
9	CM10.H	99	-2.592	0.072	0.931	0.869	0.977	4478	4316	4509
149	IMAC.H	95	-0.405	0.04	0.175	0.015	0.281	92325	90196	96777

Group T3 – 4 months

Figure 3.9 Box plots of T3 "Best Peaks'' (log 10 peak intensities v peaks BSE (B) and Control(C))



Box plots of log 10 intensities for 18 BSE and 10 controls ordered by status (B – BSE, C – control) and peak number (as in Table 3.8)

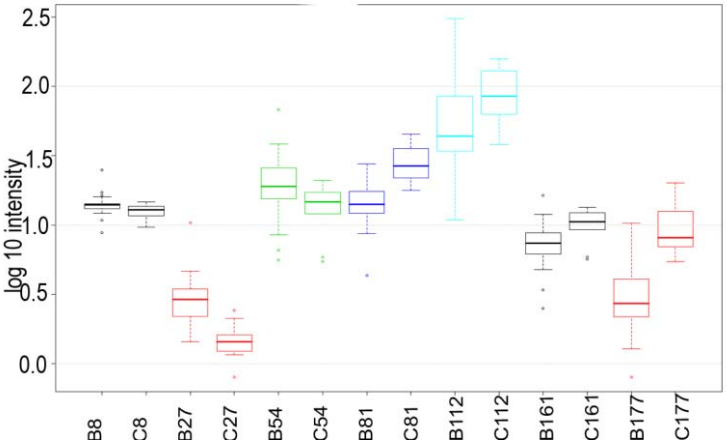
A panel of 8 protein peaks (Figure 3.9) were found to be predictive for BSE with AUCs in the range of 0.76 - 0.90 with a median of 0.84 (Table 3.8). Two protein peaks were up-regulated in the BSE samples. Peak masses (m/z ratio) ranged between 2kDa – 9.5kDa. Predictions of disease status were shown to be 4 false negatives in the BSE positive group (n= 18) and 3 false positives in the control group (n=10) resulting in a measure of sensitivity of 78% and a specificity of 70%

Table 3.8 T3 "Best Peaks''

Peak no.	Array/laser condition	nXV	median Coef.	M-W *p* value	med. Intesity	BSE.med intensity	CON.med intensity	M/Z.median	M/Z.min	M/Z.max
170	IMAC.L	99	-0.795	0.007	0.221	0.173	0.337	7332	7327	7387
86	CM10.L	99	-3.833	0	0.43	0.238	0.686	9465	9459	9472
84	CM10.L	99	-4.585	0	0.749	0.625	0.886	9199	9195	9204
83	CM10.L	99	-3.453	0.012	1.011	0.981	1.087	9169	9163	9177
81	CM10.L	99	-7.964	0	1.221	1.157	1.439	8725	8719	8731
53	CM10.L	99	-3.339	0.01	1.647	1.617	1.75	1464	1436	1476
65	CM10.L	94	0.383	0.135	0.465	0.476	0.377	4963	4959	4973
161	IMAC.L	91	0.807	0.287	0.566	0.544	0.659	4076	3976	4082

Group T4 – 6 months

Figure 3.10 Box plots of T4 "Best Peaks'' (log 10 peak intensities v peaks BSE(B) and Control(C))



Box plots of log 10 intensities for 17 BSE and 10 control ordered by status (B – BSE, C – control) and peak number (as in Table 3.9)

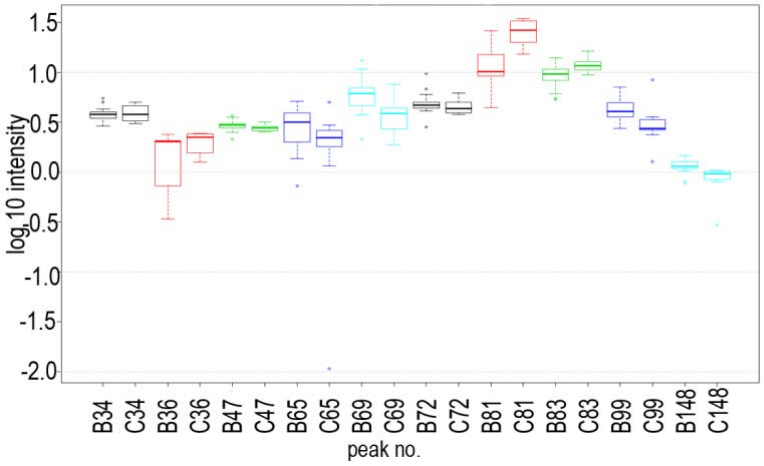
A panel of 7 protein peaks (Figure 3.10) were found to be predictive for BSE with AUCs in the range of 0.73 - 0.91 with a median of 0.86 (Table 3.9). Three protein peaks were up-regulated in the BSE samples. Peak masses (m/z ratio) ranged between 2kDa – 16kDa. Predictions of disease status were shown to be 4 false negatives in the BSE positive group (n= 17) and 3 false positives in the control group (n=10) resulting in a measure of sensitivity of 76% and a specificity of 70%

Table 3.9 T4 Best Peaks

Peak no.	Array/laser condition	No. X.Validation	median Coef.	M-W *p* value	med. Intensity	BSE.med intensity	CON.med intensity	M/Z.median	M/Z.min	M/Z.max
161	IMAC.L	99	-3.177	0.019	FALSE	0.935	0.868	4076	4082	3976
112	IMAC.H	99	-2.763	0.085	FALSE	1.801	1.643	1155	1174	1149
81	CM10.L	99	-7.25	0	TRUE	1.244	1.148	8725	8731	8719
8	CM10.H	99	6.041	0.027	FALSE	1.135	1.147	4023	4082	3999
27	CM10.H	98	2.182	0	TRUE	0.341	0.464	16090	16129	16025
54	CM10.L	94	0.754	0.019	FALSE	1.205	1.278	1889	1919	1887
177	IMAC.L	80	-0.81	0	TRUE	0.634	0.435	8720	8724	8665

Group T 5 – 8 months

Figure 3.11 Box plots of T5 "Best Peaks'' (log 10 peak intensities v peaks BSE(B) and Control(C))



Box plots of log 10 intensities for 20 BSE and 10 control ordered by status (B – BSE, C – control) and peak number (as in Table 3.10)

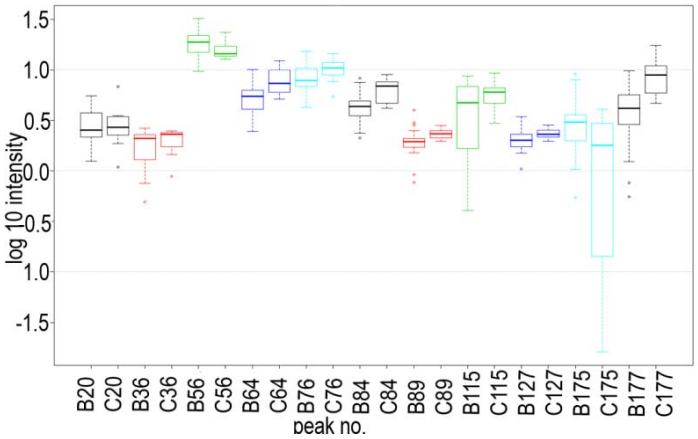
A panel of 10 protein peaks (Figure 3.11) were found to be predictive for BSE with AUCs in the range of 0.67 - 0.86 with a median of 0.76 (Table 3.10). Four protein peaks were up-regulated in the BSE samples. Peak masses (m/z ratio) ranged between 5kDa – 144kDa. Predictions of disease status were shown to be 5false negatives in the BSE positive group (n= 20) and 5 false positives in the control group (n=10) resulting in a measure of sensitivity of 75% and a specificity of 50%

Table 3.10 T5 "Best Peaks''

Peak no.	Array/laser condition	No. XVal.	median Coef.	M-W *p* value	med. Intesity	BSE.med intensity	CON.med intensity	M/Z.median	M/Z.min	M/Z.max
148	IMAC.H	99	3.328	0.001	0.039	0.058	-0.016	81298	79753	82361
99	CM10.L	99	3.698	0.001	0.561	0.606	0.44	15051	15039	15071
83	CM10.L	99	-3.317	0.005	1.013	0.985	1.065	9169	9163	9177
81	CM10.L	99	-4.697	0	1.172	1.007	1.423	8725	8719	8731
69	CM10.L	99	1.836	0.005	0.698	0.787	0.587	5894	5888	5900
47	CM10.H	99	19.557	0.109	0.455	0.474	0.441	143497	143170	145008
34	CM10.H	99	-6.328	0.619	0.577	0.577	0.577	38749	38607	38829
65	CM10.L	97	0.233	0.155	0.415	0.5	0.346	4963	4959	4973
36	CM10.H	89	-0.412	0.074	0.307	0.305	0.35	53280	52899	54352
72	CM10.L	86	0.891	0.328	0.67	0.671	0.637	6905	6849	6930

Group T 6 – 10 -12 months

Figure 3.12 Box plots of T6 "Best Peaks'' (log 10 peak intensities v peaks BSE (B) and Control(C))



Box plots of log 10 intensities for 21 BSE and 9 control ordered by status (B – BSE, C – control) and peak number (as in Table 3.11)

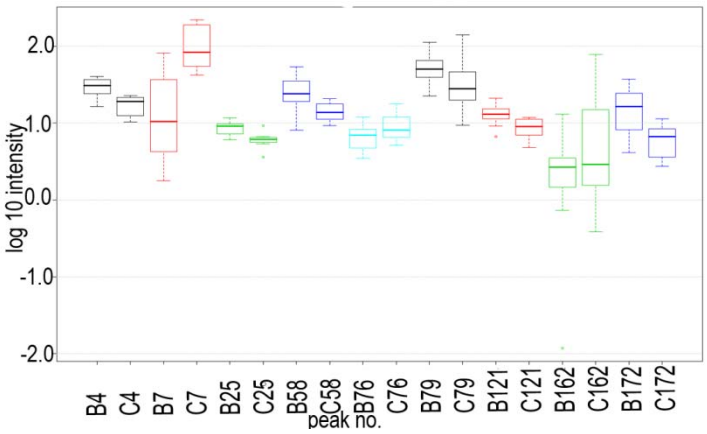
A panel of 11 protein peaks (Figure 3.12) were found to be predictive for BSE with AUCs in the range of 0.51 - 0.78 with a median of 0.66 (Table 3.11). Two protein peaks were up-regulated in the BSE samples. Peak masses (m/z ratio) ranged between 2kDa – 53kDa.Predictions of disease status were shown to be 6 false negatives in the BSE positive group (n= 21) and 6 false positives in the control group (n=9) resulting in a measure of sensitivity of 71% and a specificity of 34%

Table 3.11 T6 Best Peaks

Peak no.	Array/laser condition	No.X Validation	median Coef.	M-W *p* value	med. Intesity	BSE.med intensity	CON.med intensity	M/Z.median	M/Z.min	M/Z.max
177	IMAC.L	99	-0.942	0.001	0.714	0.619	0.949	8720	8665	8724
175	IMAC.L	99	2.215	0.104	0.462	0.483	0.252	8132	8068	8211
115	IMAC.H	99	-1.845	0.226	0.693	0.673	0.779	3938	3847	3993
89	CM10.L	99	-3.65	0.022	0.303	0.288	0.366	11334	11327	11353
84	CM10.L	99	-7.167	0.006	0.65	0.637	0.839	9199	9195	9204
64	CM10.L	99	-3.393	0.019	0.765	0.74	0.865	4628	4584	4633
56	CM10.L	99	4.863	0.077	1.229	1.276	1.159	2031	2025	2038
36	CM10.H	99	-4.229	0.178	0.334	0.323	0.361	53280	52899	54352
127	IMAC.H	93	-2.93	0.086	0.335	0.303	0.365	13425	13372	13498
20	CM10.H	81	0.41	0.859	0.42	0.403	0.429	12598	12464	12689
76	CM10.L	80	-0.555	0.094	0.952	0.895	1.016	8266	8260	8294

GroupT7 – 18 – 28 months

Figure 3.13 Box plots T7 "Best Peaks'' (log 10 peak intensities v peaks BSE (B) and Control(C))



Box plots of log 10 intensities for 18 BSE and 8 controls ordered by status (B – BSE, C – control) and peak number (as in Table 3.12)

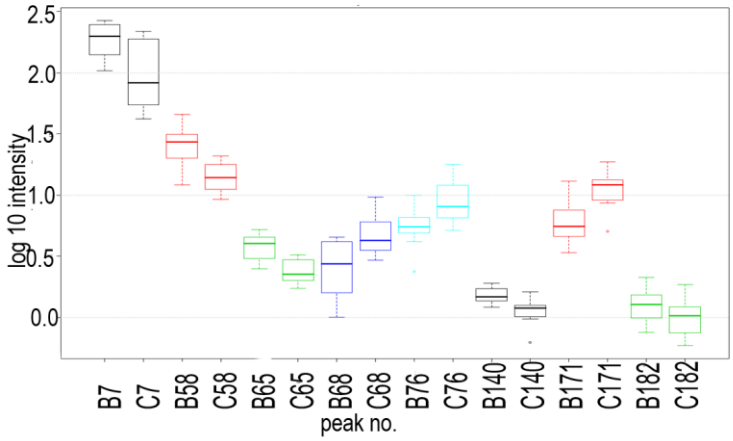
A panel of 9 protein peaks (Figure 3.13) were found to be predictive for BSE with AUCs in the range of 0.72- 0.88 with a median of 0.83(Table 3.12). Six protein peaks were up-regulated in the BSE samples. Peak masses (m/z ratio) ranged between 2kDa – 15kDa. Predictions of disease status were shown to be 4 false negatives in the BSE positive group (n= 18) and 4 false positives in the control group (n=8) resulting in a measure of sensitivity of 78% and a specificity of 50%

Table 3.12 T7 "Best Peaks''

Peak no.	Array/laser condition	No. X.Val	median Coef.	M-W *p* value	med. Intensity	BSE.med intensity	CON.med intensity	M/Z.median	M/Z.min	M/Z.max
172	IMAC.L	99	1.797	0.005	1.027	1.214	0.823	7526	7519	7580
162	IMAC.L	99	-0.552	0.531	0.429	0.429	0.463	4382	4356	4409
121	IMAC.H	99	6.205	0.003	1.064	1.118	0.958	7567	7524	7589
79	CM10.L	99	2.982	0.047	1.667	1.703	1.448	8605	8594	8616
25	CM10.H	99	5.223	0.001	0.877	0.962	0.787	14770	14757	14785
7	CM10.H	99	-1.492	0	1.495	1.021	1.92	3280	3252	3325
4	CM10.H	99	7.724	0.001	1.385	1.486	1.281	2507	2495	2573
76	CM10.L	98	-1.234	0.115	0.858	0.842	0.905	8266	8260	8294
58	CM10.L	89	0.598	0.006	1.312	1.382	1.142	2694	2690	2697

Group T8 – clinical stage of disease

Figure 3.14 Box plots of T8 "Best Peaks''



Box plots of log 10 intensities for 10 BSE and 8 control ordered by status (B – BSE, C – control) and peak
as in Table 3.13.

A panel of 8 protein peaks (Figure 3.14) were found to be predictive for BSE with AUCs in the range of 0.70 - 0.94 with a median of 0.86(Table 3.13). Five protein peaks were up-regulated in the BSE samples. Peak masses (m/z ratio) ranged between 2kDa – 31kDa. Predictions of disease status were shown to be 1 false negatives in the BSE positive group (n= 10) and 2 false positives in the control group (n=8) resulting in a measure of sensitivity of 90% and a specificity of 75%.

Table 3.13 T8 "Best Peaks''

Peak no.	Array/laser condition	No. X.Val	median Coef.	M-W *p* value	med. Intesity	BSE.med intensity	CON.med intensity	M/Z.median	M/Z.min	M/Z.max
182	IMAC.L	99	1.545	0.129	0.041	0.104	0.013	10771	10760	10854
171	IMAC.L	99	-7.367	0.054	0.902	0.745	1.083	7497	7491	7503
140	IMAC.H	99	6.662	0.144	0.129	0.168	0.074	31155	31048	31570
76	CM10.L	99	-1.194	0.115	0.793	0.74	0.905	8266	8260	8294
65	CM10.L	99	4.859	0.238	0.49	0.604	0.352	4963	4959	4973
7	CM10.H	99	2.44	0	2.239	2.298	1.92	3280	3252	3325
58	CM10.L	96	1.019	0.006	1.302	1.434	1.142	2694	2690	2697
68	CM10.L	93	-0.891	0.003	0.563	0.436	0.628	5815	5800	5844

3.3 Testing the model

For the purposes of testing the resultant best peaks sets against samples of unknown TSE disease status, plasma samples from an independent source were obtained (Moredun Research Institute, Edinburgh). Groups of plasma samples from sheep infected with *Toxiplasma gondi* and corresponding control animals were analysed by the same protocol as the BSE and scrapie plasma samples. Plasma from the BSE groups (BSEal, BSEnp) not included in building the models as detailed in Table 3.3 were also similarly analysed. The resultant data were tested by linear discriminant analysis(Figure 3.15) using each of the Best Peaks sets (T1-T8) from each time point. The BSE group of samples were matched by time i.e. T1 samples were tested with T1 panel of markers found in the classification model. The clinical panel of markers were tested against the last time point available which was 18-28 months.

Figure 3.15 Linear Discriminant Analysis

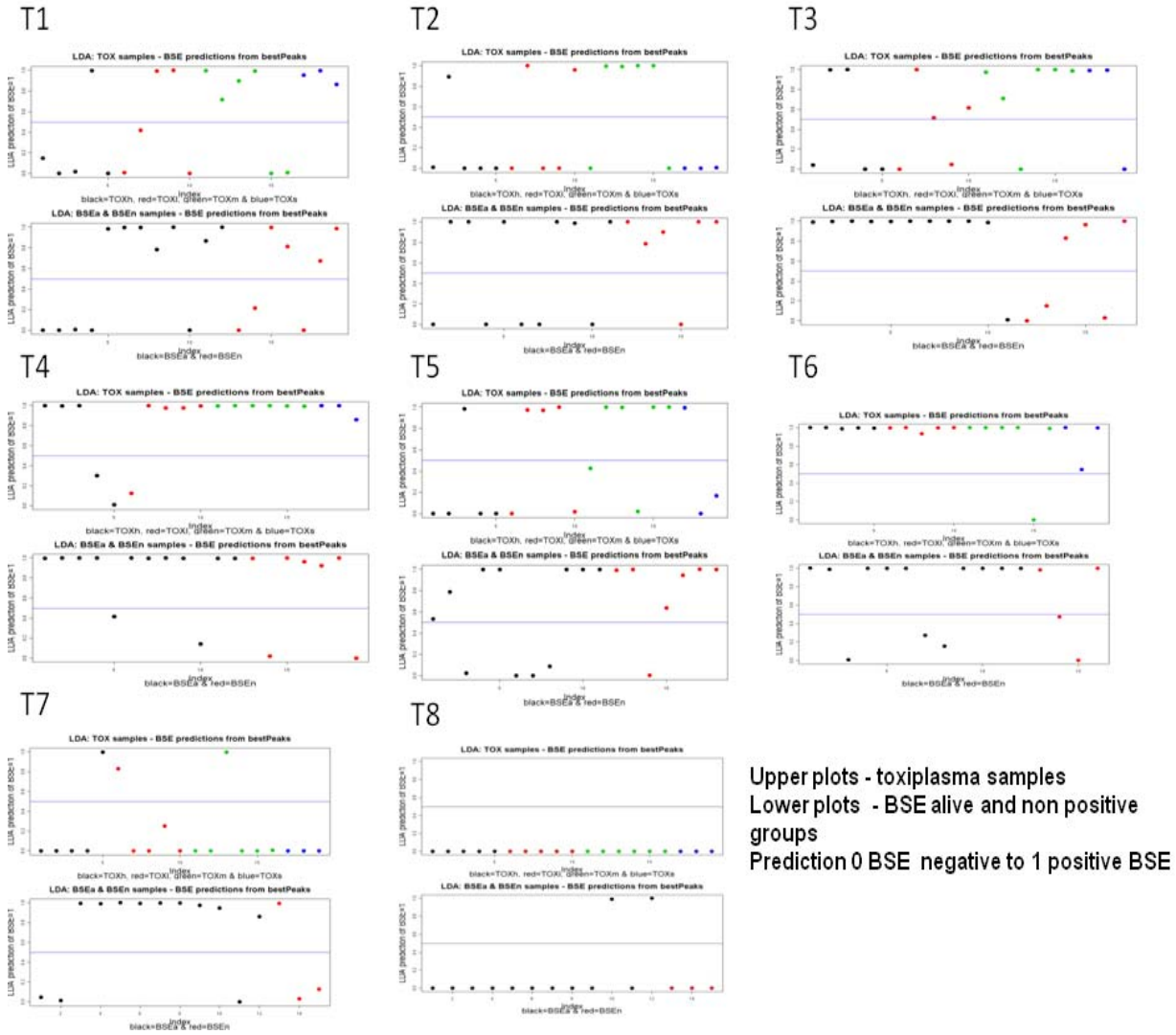


Figure 3.15 Each pair (Toxiplasma (upper) and BSE (lower)) of LDA graphs are shown for each time point. Prediction of BSE disease status 0 – BSE negative (lower) to 1 – BSE positive (top). In the toxoplasma plots (upper) coloured dots denote groups :- black - TOX H, red TOX, green – TOX med and blue - TOX s. In the BSE groups: black denotes BSE alive and red, BSE non positive.

Pre incubation – T1

Both toxoplasma and the BSE plasma samples did not separate in to either BSE or control. This would be expected as there is no difference at this stage between the BSE group and the control group i.e. no disease

2months – 18/28months T2-T7

None of the panels of markers at each time point successfully separated the toxoplasma samples as non-BSE. False positive ranged from 18 of 19 (T6) to 3 of 19 (T7). Of the known i.e PrP^d positive animals (n = 7) there were an average of 2 false negatives in groups T2-T6 and 4 in T7.

Clinical T8

The "Best Peaks'' panel at the clinical stage of disease predicted all the toxoplasma samples and all but two animals in the BSE al group as negative for BSE. Of the known i.e PrP^d positive animals (n = 7) there was only 1 animal predicted positive.

BSE samples not included in the training set

A summary of the findings from the LDA analysis can be seen in Table 3.14. Each time point T1-8 is scored positive (P) or negative (N) for BSE. Animals still alive at the time of building the classification model which have since died and post mortem diagnosis carried out on tissue by ICC and western methodologies shown to be PrP^d positive are indicated in blue. Animals which were still alive at the time of analysis are indicated in green and animals non positive for PrP^d indicated in orange.

Table 3.14 LDA analysis summaries

Time group Animal No	1	2	3	4	5	6	7	8	Prediction correct /no. samples
N159	N		N	P	P	P	P	N	3/7
N161	N	N	P	P	P	P	N	N	4/8
N164	N	P	N	N	N	N	N	N	1/8
N168	P	P	P	P	N				died non BSE symptoms
N169	N	P	P	P	P	P	N	N	
N175	N	P	P	P	N	N	P	N	
N181	N	N	P	P		P	P	N	
N201	P	P		P	P	N			1/5
N202	N	N	P	P	P				2/5
N204	P	P	N	P	P	P	N	N	5/8
N208	P	P	P	N	P	P	P	N	6/8
N219	P	N		P	P	P	P	N	
N223	P	N	P	P	N	N	P	N	4/8
N241	P		P	P	N	N	P	N	
N246	P	P	P	P	P	P	P	N	
N248	N	P	P	N	P	P	P	P	6/8
N250	P	P	P	N	P				4/5
N259	P	N	P	P	P	P	N	N	5/8
N264	P	P	N	P	P	P	P	P	

Alive at time of analysis, now dead and +ve for PrPd

Still alive

Non positive for TSE

P – LDA positive for BSE

N – LDA negative for BSE

3.4 Summary of differentially expressed proteins in scrapie and BSE

Table 3.15 summarises the "Best Peaks'' sets at each time point in the scrapie and BSE (T1 – pre-inoculation, T2 – 2 months post inoculation (m.p.i.) T3 – 4 m.p.i., T4 – 6 m.p.i., T5 – 8 m.p.i., T6 – 10/12m.p.i., T7 – 18-28 m.p.i., T 8 – clinical stage of disease) sheep models.

Table 3.15 Differentially expressed protein peaks

The peak number from the original data numbering (1- 201) appears in the first column followed by the type of array and laser condition with which the peak was found e.g. CM10 array at high or low laser setting, the mass to charge ratio (m/z) expressed as a median, minimum and maximum, each time point T1-T8 and scrapie clinical (orange boxed area) peaks similar to BSE groups. The yellow banding denotes candidate proteins for identification.

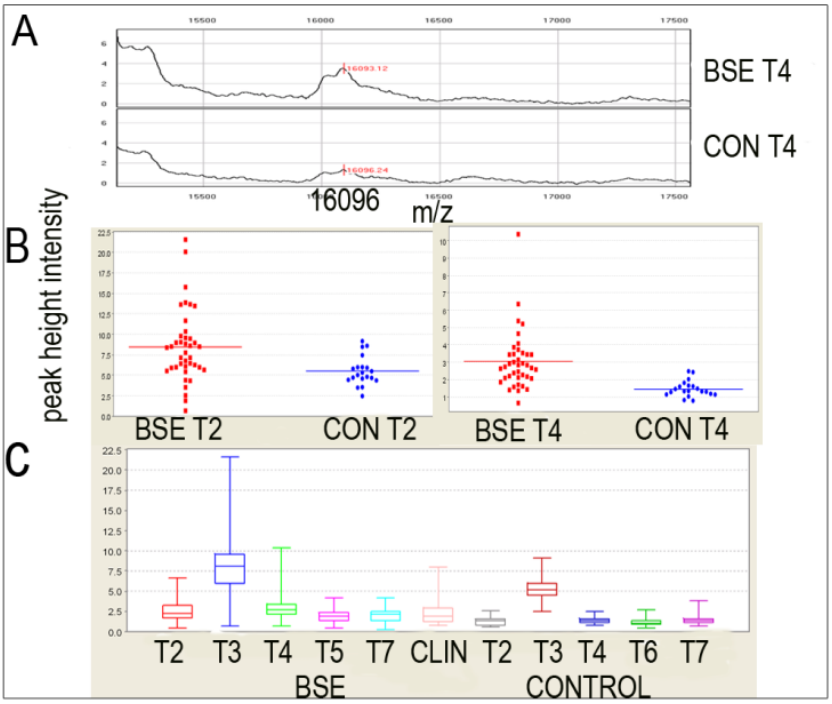
Peak No.	Array/Las	MZ.median	MZ.min	MZ.max	T1	T2	T3	T4	T5	T6	T7	T8	scrapie	Peak No.	m/z med	m/z min.
4	CM10.H4	2507	2495	2573							↑					
7	CM10.H7	3280	3252	3325							↓	↑				
8	CM10.H8	4023	3999	4082				↑								
9	CM10.H9	4478	4316	4509		↓										
10	CM10.H10	5590	5581	5617		↓										
11	CM10.H11	5706	5668	5798	↑								↓	7	5716	5703
20	CM10.H20	12598	12464	12689						↓						
25	CM10.H25	14770	14757	14785	↑						↑					
27	CM10.H27	16090	16025	16129	↓			↑								
30	CM10.H30	22882	22809	23007		↑										
34	CM10.H34	38749	38607	38829					↑							
36	CM10.H36	53280	52899	54352					↓	↓						
47	CM10.H47	143497	143170	145008					↑							
53	CM10.L2	1464	1436	1476	↓		↓									
54	CM10.L3	1889	1887	1919		↑		↑								
56	CM10.L5	2031	2025	2038						↑						
58	CM10.L7	2694	2690	2697							↑	↑	↓	66	2691	2687
64	CM10.L13	4628	4584	4633						↓						
65	CM10.L14	4963	4959	4973	↑		↑		↑			↑				
68	CM10.L17	5815	5800	5844								↓	↓	71	5812	5809
69	CM10.L18	5894	5888	5900		↑			↑							
70	CM10.L19	5961	5940	5967		↑										
72	CM10.L21	6905	6849	6930					↑							
75	CM10.L24	8037	8031	8049									↑	75	8034	8025
76	CM10.L25	8266	8260	8294	↓					↓	↓	↓				
79	CM10.L28	8605	8594	8616							↑		↑	78	8601	8590

Peak No.	Array/Las	MZ.median	MZ.min	MZ.max	T1	T2	T3	T4	T5	T6	T7	T8	scrapie	Peak No.	m/z med	m/z min.
81	CM10.L30	8725	8719	8731			↓	↓	↓							
83	CM10.L32	9169	9163	9177			↓		↓							
84	CM10.L33	9199	9195	9204		↓	↓			↓						
86	CM10.L35	9465	9459	9472			↓									
89	CM10.L38	11334	11327	11353						↓						
92	CM10.L41	12606	12549	12686	↓											
99	CM10.L48	15051	15039	15071					↑							
112	IMAC.H1	1155	1149	1174				↓								
113	IMAC.H2	1199	1195	1226	↓											
115	IMAC.H4	3938	3847	3993	↓					↓						
120	IMAC.H9	7466	7429	7516									↓	111	7474	7441
123	IMAC.H12	8550	8482	8576									↑	115	8568	8563
127	IMAC.H16	13425	13372	13498						↓						
140	IMAC.H29	31155	31048	31570								↑				
145	IMAC.H34	66269	66168	66874		↓										
148	IMAC.H37	81298	79753	82361					↑							
149	IMAC.H38	92325	90196	96777		↓							↑	149	95250	93749
159	IMAC.L2	2011	2000	2024		↑										
160	IMAC.L3	3276	3264	3283												
161	IMAC.L4	4076	3976	4082	↓		↓	↓								
162	IMAC.L5	4382	4356	4409		↓					↑					
170	IMAC.L13	7332	7327	7387			↓									
171	IMAC.L14	7497	7491	7503		↓						↓				
172	IMAC.L15	7526	7519	7580							↑					
175	IMAC.L18	8132	8068	8211						↑						
177	IMAC.L20	8720	8665	8724				↓		↓						
182	IMAC.L25	10771	10760	10854								↑				
187	IMAC.L30	14536	14523	14586	↑											

3.5. Protein Identification

Several candidate proteins were selected for identification. This selection was based on the spectral peak profile, presence in more than one of the BSE time point groups and /or the scrapie end point “Best Peaks’’ sets and previous studies in other groups (personal communication). The presence of a particular protein peak in the “Best Peaks’’ sets is the product of the statistical classification model built as described in the Methods section. These proteins represent markers with the most information about the TSE disease state within each time point group. Whilst the rigorous statistical analysis generates a protein peak in one time point group but not in others, this does not necessarily mean that at other stages of disease it is not differentially expressed, rather, that there are other proteins which rank higher at the different time points. Taking this into consideration when choosing candidate proteins for identification, the relationship of the proteins appearing in the “Best Peaks’’ sets, to other time points of disease, was also examined using the ProteinChip Database Manager software, to assess p-values by Mann-Whitney or Kruskal-Wallis non-parametric tests. The software also generates a visual plot of these analyses in the form of box or scatter plots. Whilst some peaks are statistically significant at some time points and not at others the general trend (up/down regulated) in the course of disease is an important indicator of the protein’s potential as marker of the presence of TSE disease. As an example, a peak at a median of 16090 (Figure 3.16) is up-regulated in BSE disease in the T4 “Best Peaks’’ set at the six months stage of disease. Comparing the *p*-values for each time point in the ProteinChip DataManager software, it was found that at most time points this protein is up-regulated during disease and is statistically significant (T2 - p =0.0001,T3 - p = 0.001, T4 - p = 0000007, T5 - p =0.001, T6 - p = 0.1, T7 - p = 0.005, T8 - Clinical – p = 0.02). This protein would be a candidate for identification and a possible candidate for inclusion in a final panel of markers if the levels can be verified by other technologies.

Figure 3.15 Example of protein peak (peak no. 27 in Table 4.1 16096 m/z) up-regulated in BSE T4 "Best Peaks'' group and also differentially expressed at other time points.



A – BSE and corresponding control sample as seen on CM10 array.

B – Scatter plots of T2 BSE vs T2 control (p =0.001) and T4 BSE vs T4 control (p = 0.0000075). **C** –Boxplots of time points throughout disease.

For the identification of candidate markers a more powerful mass spectrometer than the SELDI must be used to maximise the search for a candidate protein from the peptide mass fingerprint. We utilised the proteomics unit at the nearby Moredun Institute where staff specifically trained in this approach analysed the resultant gels and fragments on an Ultraflex II MALDI-ToF-ToF mass spectrometer. We were limited however by health and safety considerations which prohibited the use of material (gels) generated from the BSE plasma (CAT 3), in the mass spectrometer and therefore isolated proteins of interest from scrapie infected and control plasma samples. These plasma samples were fractionated and fractions analysed on ProteinChip arrays which indicated the presence of the peak of interest and other surrounding peaks before finally isolating and further purifying the protein of interest on a 1D-SDS PAGE gel. A scheme of identification steps can be seen in Figure 2.1. The resultant 1D SDS PAGE gels can be seen in Figure 3.17 where the bands indicated were excised, tryptically digested and the peptides identified by MALDI- TOF mass spectrometry and MASCOT data mining. In Table 3.16 the proteins identified, number of peptides and the MASCOT score are shown for each protein identification.

Figure 3.17

1D SDS-PAGE gels from which proteins were identified

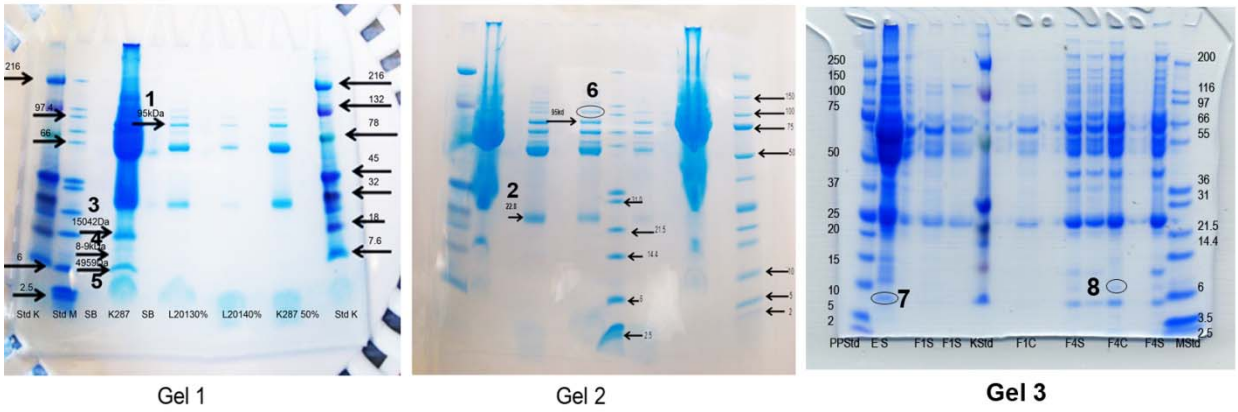


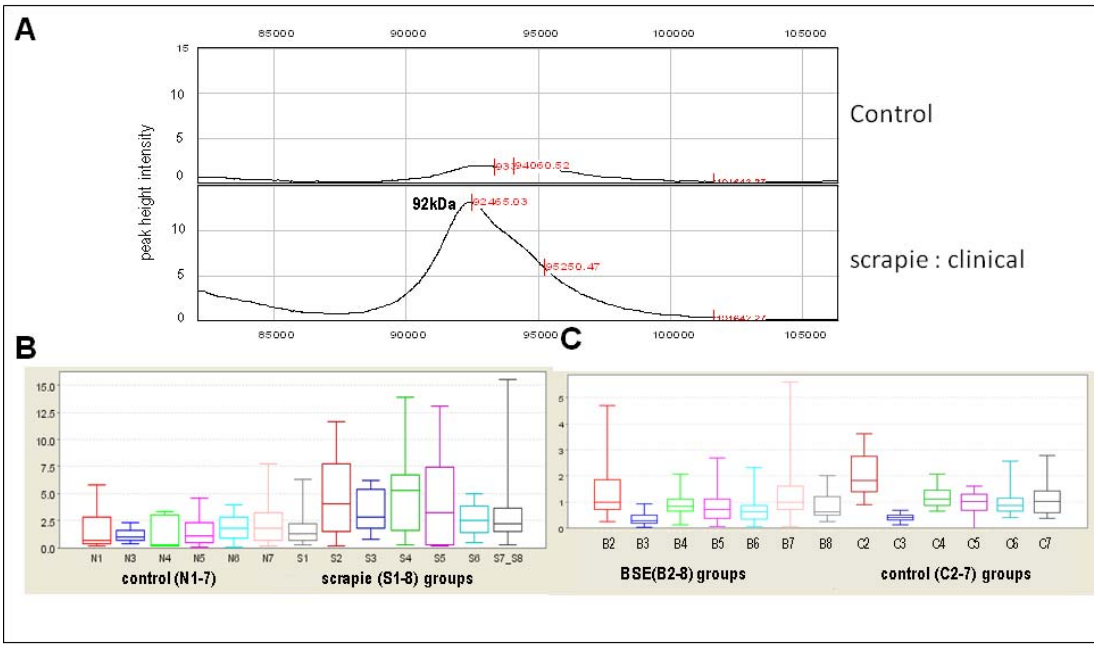
Table 3.16 Identification of protein bands excised from gels in Figure 3.17

Gel No.	Band No.	Sample/ Fraction	Protein identified	mass weight (kDa)	MASCOT score	Sequence coverage %
1	1	scrapie / flow through & 30% organic fraction	TRFE_BOVIN Serotransferrin	77703	96	42
2	2	scrapie / flow through & 50% organic	gi52366986 (ovis aries) Immunoglobulin lambda light chain	11322	78	76
1	3	scrapie / MWCO fraction	HBB sheep (ovis aries) Haemoglobin subunit beta	16063	115	69
1	4	scrapie / MWCO fraction	No result	8-9		
1	5	scrapie / MWCO fraction	No result	4		
2	6	Scrapie/ flow through & 30% organic fraction	PLMN_SHEEP Plasminogen gi|881556 Fibrinogen alpha chain	37638 36053	67 115	42 10.1
3	7	Scrapie/ Proteominer bead elution(HEPES)	No result	8		
3	8	Scrapie/Proteominer bead fraction 4(organic)	Gi 86438511(Bos Taurus) ApoN protein	28517	68	20

Several proteins excised from the gels for the identification of particular individual peaks did not result in a similar molecular mass weight when identified by peptide mass fingerprinting.

A peak at 92kDa (IMAC peak no.149 Table 3.15) up-regulated in scrapie and down-regulated in BSE at 2 months post-inoculation (see Figure 3.18) was excised from Gel 2 band 6 (Figure 3.17). This was identified as a complex of plasminogen (ovis aries) and fibrinogen. This extract was further analysed on a Bruker Daltonics Ultraflex II ToF-ToF further verifying the identification.

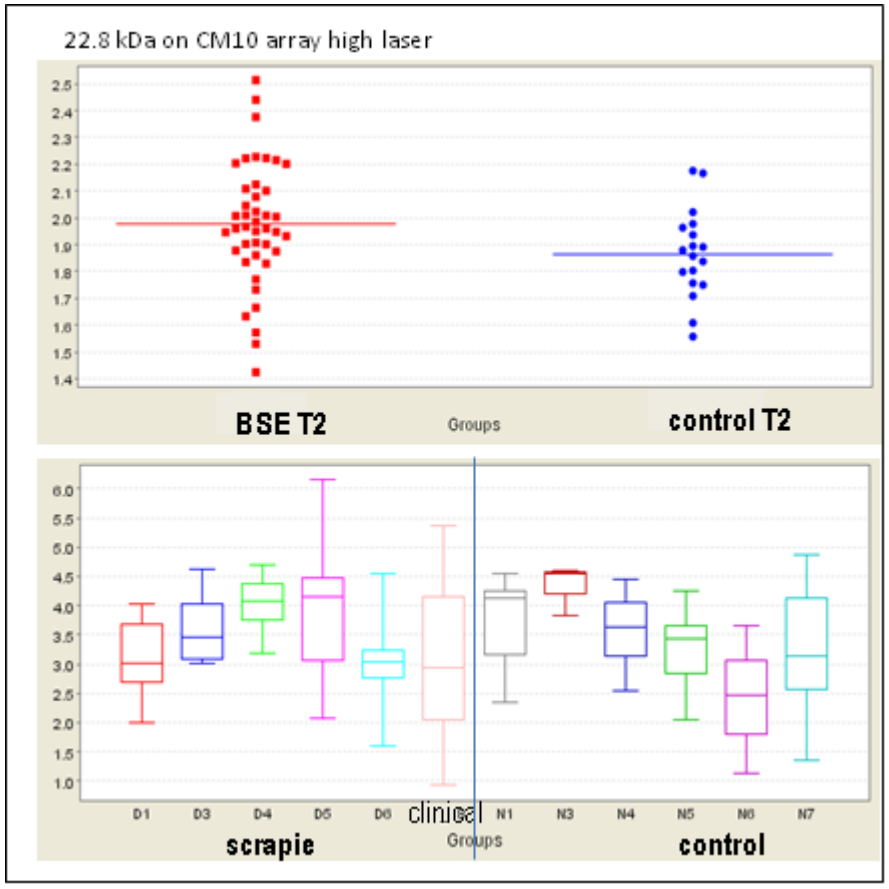
Figure 3.18 Identification of 92kDa peak



A. Protein peak (No.149 in Table 4.1) at 92kDa visualised on DataManager software.
B. 92kDa is up - regulated in scrapie time points N1 - 7(normal) S1 - 8 (scrapie).
C. Protein peak at 92kDa down - regulated at time point 2 in BSE. BSE02-Clin (BSE time points) CON02 – 27(control time points)

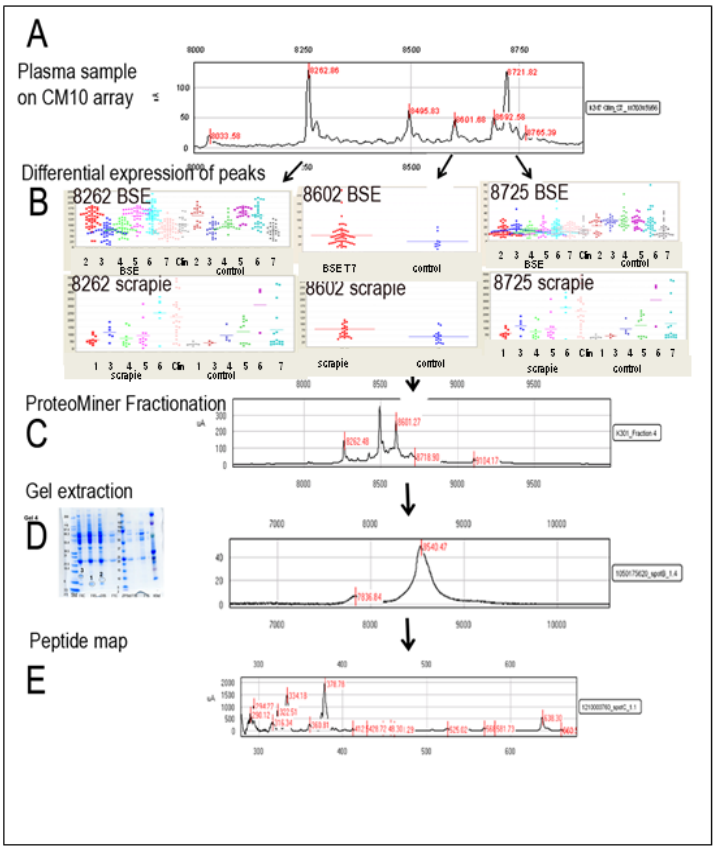
We were interested in isolating proteins with similar molecular weights to proteins differentially expressed in this study and by other groups (personal communication). A peak at 22kDa up–regulated in BSE at 2 months post inoculation (Table 3.15 Peak No.30) was excised from Gel 2 band 2. This was identified by peptide mass fingerprinting as Immunoglobulin lambda light chain protein (Ovis aries) with a mass weight of 11321daltons. As this is half the mass weight of the band excised it is reasonable to assume that the band at 22.8kDa is a dimer. This peak was not significantly up-regulated in the scrapie samples (Figure 3.19).

Figure 3.19 Differential expression of peak no. 30 at 22kDa (Table 4.1)



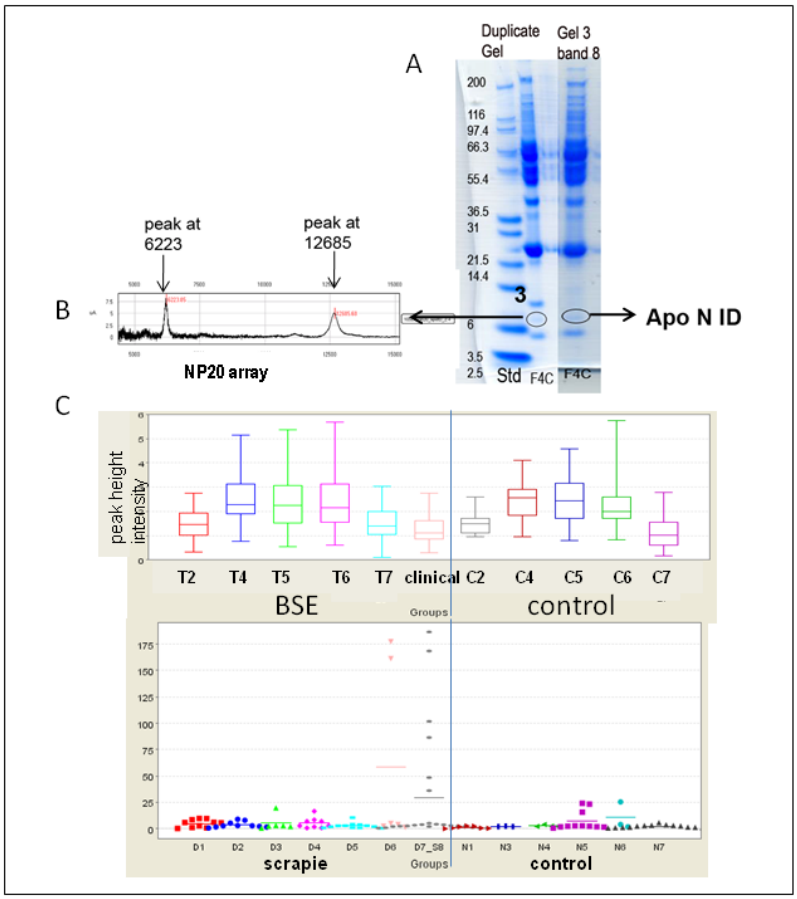
There were many proteins with a molecular mass between 8-9kDa which appeared in various time points of the "Best Peaks'' sets (see Table 3.15). For this reason we were interested to isolate proteins at this mass weight. As in the other identifications a fractionation protocol was followed to eliminate some of the proteins surrounding the differentially expressed proteins. An example of these closely associated peaks can be seen at 8262, 8602, 8725m/z (Figure 3.20A) which are differentially expressed in both BSE and scrapie (Figure 3.20B). The flow through fraction of a plasma sample (see Methods) was further fractionated using a molecular weight cut-off tube and run on a 1D SDS- PAGE gel and the band excised tryptically digested and analysed on the MALDI-TOF (Band 4 in Gel 1). No proteins could be identified from this which was unexpected. To investigate this further and determine if we could in fact isolate a protein of this molecular weight we extracted a plasma sample using the ProteoMiner kit. As can be seen in Figure 3.20C some peaks were preferentially expressed. This extract was run as before on a gel, the band eluted and spotted on an NP20 array to look for the 8kDa protein. Figure 3.20D visualises a peak at 8540 which is extremely broad covering a large area. The extract was tryptically digested and the peptide mass fingerprint submitted to MASCOT which again returned no identified protein. Our conclusion to this is that there is a mixture of all the proteins present at this molecular weight and at low concentration making identification of individual proteins very difficult.

Figure 3.20 Scheme for the identification of 8kDa peak



A Several protein peaks from 8-9kDa visualised on CM10 array
B Proteins within this range differentially expressed (statistically significant at some time points throughout disease
C Fractionation of neat plasma using ProteoMiner bead kit and visualisation of protein peaks in extract on NP20
D Excision and elution of gel band extract visualised on NP20
E Tryptic digest peptide map visualised on SEND array

Figure 3.21 Identification of Band 8 as Apo N



A – SDS – PAGE Novex Bis –Tris 12% gel stained with coomassie blue
B – SELDI NP20 array showing a peak at 6223 and 12685 daltons
C_Box plots (upper panel) of BSE infected plasma groups (T2 –T7 & clinical) and control groups (C2-C7). Scatter plots of scrapie infected plasma (D1-D7&8clinical) and control (N1-N7) groups.

A protein peak at 12598 was down-regulated in the T6 “Best Peaks’’ set. We were interested in this peak as a protein of this molecular mass had been shown to be up - regulated in scrapie by another TSE laboratory (Dr S. Booth, National Microbiology Laboratory, Canada, personal communication). Although we have few samples in the scrapie infected groups, an up-regulation can be seen in the late stage samples which would be in agreement with the Canadian group. We extracted this protein from Gel 3 as described in the protein identification protocol (see Methods section). A similar band was extracted from a duplicate gel, passively eluted and spotted on a NP20 array to check the molecular weight. As can be seen in Figure 3.21 this was within an acceptable range of the molecular weight of the protein to be identified. Peptide mass fingerprinting identified this protein as Apolipoprotein N (Bos Taurus). This protein was first identified in bovine

tissue [13] as a 12kDa protein processed from the C terminus of a 29kDa precursor protein and is thought to be similar to ApoF which is a cholesterol transport regulatory protein. There is sequence homology with murine tissues but not human. ApoN may play a role in steroidogensis and immunoregulation

4. Discussion

In previous studies we had established the use of SELDI - TOF technology as a potential methodology for the detection of TSE disease in murine and bovine brain tissue. Given the success using brain tissue, we sought in the present study to apply this methodology to blood samples

The complexity of whole blood limits its use in biomarker discovery therefore it is necessary to fractionate the blood into components. Plasma is a rich source of proteins both originating in plasma and from tissues throughout the body. Natural cellular processes release proteins into the plasma in small concentrations during the course of life. When these processes are disrupted, for example in disease, the natural turnover of proteins present and character of the plasma proteome will change and result in a different profile or “fingerprint’’. Evidence from other studies indicated that the plasma proteome is a rich source in determining disease status and a current international study, carried out by the Human Proteome Organisation (HUPO) the “Plasma Proteome Project’’, is now reporting characterisation of the plasma proteome as a reference of proteins found normally in plasma [10].

We sought to establish a methodology using plasma samples to determine the presence of TSE disease by a data driven algorithm. As there are many variables in the analysis of blood we deliberately kept the sample preparation to a minimum. Preliminary studies further fractionating the plasma using albumin depletion columns and denaturants in our hands did not improve the number or quality of the protein spectra obtained from the SELDI-TOF analysis and indeed increased the variability between samples. Simple dilution of the plasma in buffer and pre- mixing of the samples in a microplate, before addition to the arrays for binding resulted in higher reproducibility. We also randomised the position of samples on the arrays, always including an equal mix of disease and normal samples plus pooled samples. This meant that any spurious results arising from an aberrant array surface or binding artefact would affect both disease status groups. Any total failures of array analysis were apparent both in the peak intensity of the pooled sample and in the total ion count normalisation. Thus at each stage of analysis we endeavoured to maximise the quality of the data entering the statistical model.

In a recent study in which we applied differential protein expression profiling for the detection of BSE in cattle brain tissue [7] we found that a successful outcome is reliant on analysing a large number of samples and in particular those of “normal’’ status. In TSE studies, the numbers required are particularly problematic as there are limited resources for controlled experimental samples from large animals. The scarcity of such samples limits projects which rely on genetic or proteomic differential expression data. The BSE plasma samples in this study originated from a

unique experiment using groups of sheep which were either infected with BSE or were non-infected controls. Blood samples were taken from each animal at several time points from inoculation through to clinical disease. This experiment generated the largest resource of TSE infected blood samples taken under controlled experimental conditions likely to become available and as such, we endeavoured to utilise this resource (despite the relatively low numbers) for the application of this technology, in the search for a much needed non invasive pre-mortem test for the detection of TSE disease.

The statistical methodology which we applied in building the classification model to these data was particularly and necessarily stringent as was previously demonstrated in the cattle brain study. This resulted in data from the scrapie study being insufficient to build a classification model for each group of samples at time points during the course of disease. Only the terminal group of samples had sufficient information to build a predictive model. In this case the predictive model looks particularly impressive with the measurement of sensitivity and specificity expressed as the area under the ROC curve (AUC) as 0.92. This however is biased towards the infected disease status due to the low number of controls (18 scrapie and 6 controls), therefore we must be cautious in applying this "Best Peaks'' set to a large set of unknown samples. Similarly although higher numbers of individual animals were available for the BSE study, the control numbers were still relatively low, limiting the specificity and sensitivity of the prediction model. Despite this, protein peaks of similar molecular mass (m/z) are present in more than one of the "Best Peaks'' sets for individual time points and are duplicated in two different array chemistries, suggesting that these proteins may give an indication of the mechanisms involved in TSE disease pathogenesis. Results of the linear discriminant analysis for the clinical stage of disease at first glance look encouraging with all the toxoplasma samples predicted as non-BSE implying that there is BSE specific information in the protein marker combination. However there are only two sheep predicted as BSE in the samples sheep of unknown disease status (BSE alive and BSE non-positive). Some of these sheep which have had extremely long incubation times, have since been culled and have been found to be positive for BSE by the detection of PrP^{d}. This raises the question that the markers at the clinical stage of disease are indicative of the physiological state of the sheep at the end stage of disease rather than being specific to BSE disease or were the sheep, with longer incubations at too early a stage for protein expression differences to be detected? It will be interesting to correlate these results with the final outcome of the project of which these sheep were the donors **(**DOH project –"The effect of Leucodepletion on transmission of BSE by transfusion of sheep blood components", PI.,Prof. Jean Manson). Blood and blood products from the infected donor sheep was transfused to other groups of uninfected sheep at certain time points after inoculation well before clinical signs are present. The point at which BSE infectivity is demonstrably transmitted from donor to recipient animals will perhaps elucidate the results obtained in this study. All the

profiling data from our study is stored on a virtual database and could be re-analysed in the light of more information or samples becoming available which may strengthen the statistical model.

We have identified some of the proteins however these identifications will have to be further validated and repeated to ensure accuracy.

A peak up-regulated in scrapie at 92-96kDa (Figure 3.18A-B) was identified as a plasminogen and fibrinogen complex. A down regulated peak of similar molecular mass is present in the BSE "Best Peaks'' set of data at two months post inoculation and was found to be present in all the time points using the ProteinChip DataManager software (Figure 3.18C). Fibrinogen has been shown to carry out chaperone like activity with an inhibitory role in the formation of spontaneous protein misfolding and plasminogen has been shown to bind the disease associated form of the prion protein (PrP^{d}) [14]. Zerr *et al* [15] observed a raised plasminogen concentration but paradoxically lower plasminogen specific activity in the plasma of CJD patients compared to other dementias. Complexes of fragments of fibrinogen and plasminogen are known to occur which may have a direct effect on the activation of plasminogen [16]. Thus this finding may indicate some involvement of these proteins in TSE pathogenesis. The concentrations of these proteins are abundant in plasma so it is not perhaps surprising to isolate these individually; however the differential expression of a complex between scrapie and BSE is interesting and may represent differences in the transmission mechanisms i.e. fibrinogen is also an acute phase protein which may be more likely to be up-regulated in the subcutaneously inoculated ovine scrapie model in contrast to the orally inoculated ovine BSE model.

We did attempt to isolate low molecular weight proteins however these did not produce identifications by peptide mapping (Figure 3.20). Apolipoprotein N was identified from Gel 3 band 8 (Figure 3.17) which corresponded to a molecular weight in the BSE "Best peaks'' time point 6. ApoN was first identified by O'Bryan et al in ovarian follicular fluid and is thought to be present in other tissues and have a similar function to ApoF which is a cholesterol transport regulatory protein.

Immunoglobulin lambda light chain was identified from Gel 2 band 2 (Figure 3.17) and correlated with a protein peak at BSE 2 months as being up regulated and down regulated at the clinical stage of scrapie. This protein has been implicated in scrapie pathogenesis [17] in previous ovine studies where an up-regulation in IgG concentrations were observed in some sheep displaying clinical signs of disease. This variability may explain our results where we find no statistical difference between groups. Qualitative differences in subtypes of IgG were also observed however the large variations in IgG concentration between individuals prevented this protein to be considered diagnostically. The mechanisms of this switching of isotypes were considered to be

important for disease pathogenesis and the formation of amyloid however the underlying process leading to this were not investigated. Simon *et al* [18] conclude that the presence of this protein in urine is associated with the immunoglobulin concentration in blood plasma.

5. Conclusions

Differential protein expression profiling gives us a snap-shot of the proteome at a given time during disease. We have attempted to use the profiling information to create predictive models of disease. Due to the scarcity of samples we have not produced robust models suitable for a pre-mortem test however the methodology could be extended to larger population studies in the future. Certainly our predictive models at both the clinical stage in scrapie and BSE have produced reasonable sensitivity and specificity levels (scrapie AUCs 0.95 and BSE 0.87) which taken together with the previous study in cattle brain tissue suggest that given a larger number of both infected and control samples this methodology could produce a data driven methodology for the prediction of TSE disease at the end stage of disease.

The study has also highlighted the practical difficulty in finding a point at which a pre-mortem test would be usefully carried out with so many natural variations in protein expression. We grouped samples in the scrapie model by incubation length to pull the different genotypes together to form a relevant group on which to perform statistical analyses.

We considered also grouping the BSE samples in relevant incubation length groups but this would have seriously diminished the numbers in each group and it was thought that months post inoculation, by the grouping in which the samples were taken, would be more meaningful as points of intervention, should a pre-mortem test be possible.

The “Best Peaks’’ sets of proteins for each time point have given us other avenues of investigation. The proteins identified in this study all have some resonance in processes leading to TSE disease and require more verification. This methodology produces profiles of proteins which will not be full length proteins but rather fractions or modifications of the whole protein. We know from many studies that the quantification of the full length proteins do not always correlate with their association with disease as it is portions such as side chains interacting with other proteins which are responsible for mechanistic processes. We have demonstrated that complexes can be visualised as peaks in the disease profile and some such as the low molecular weight proteins although visualised cannot be easily isolated. Therefore this methodology is providing a mechanistic insight into cellular process which have been difficult to study in traditional methodologies.

References

1. Wells GA, Scott AC, Johnson CT, Gunning RF, Hancock RD, Jeffrey M, Dawson M, Bradley R: **A novel progressive spongiform encephalopathy in cattle**. *Vet Rec* 1987, **121**(18):419-420.
2. Ironside JW, Sutherland K, Bell JE, McCardle L, Barrie C, Estebeiro K, Zeidler M, Will RG: **A new variant of Creutzfeldt-Jakob disease: neuropathological and clinical features**. *Cold Spring Harb Symp Quant Biol* 1996, **61**:523-530.
3. Grassi J, Comoy E, Simon S, Creminon C, Frobert Y, Trapmann S, Schimmel H, Hawkins SA, Moynagh J, Deslys JP *et al*: **Rapid test for the preclinical postmortem diagnosis of BSE in central nervous system tissue**. *Vet Rec* 2001, **149**(19):577-582.
4. Barron RM, Campbell SL, King D, Bellon A, Chapman KE, Williamson RA, Manson JC: **High titers of transmissible spongiform encephalopathy infectivity associated with extremely low levels of PrPSc in vivo**. *Journal of Biological Chemistry* 2007, **282**(49):35878-35886.
5. Lasmezas CI, Deslys JP, Robain O, Jaegly A, Beringue V, Peyrin JM, Fournier JG, Hauw JJ, Rossier J, Dormont D: **Transmission of the BSE agent to mice in the absence of detectable abnormal prion protein**. *Science* 1997, **275**(5298):402-405.
6. Barr J, Waddington D, Barron RM: **Differential protein expression profiling in BSE disease**. *Arch Physiol Biochem* 2010, **10**:10.
7. Barr JB, Watson M, Head MW, Ironside JW, Harris N, Hogarth C, Fraser JR, Barron R: **Differential protein profiling as a potential multi-marker approach for TSE diagnosis**. *BMC Infect Dis* 2009, **9**:188.
8. Hastie T, Tibshirani R., Friedman J. (ed.): **The Elements of Statistical Learning**: Springer Science and business Media 2009.
9. Linke T, Doraiswamy S, Harrison EH: **Rat plasma proteomics: effects of abundant protein depletion on proteomic analysis**. *J Chromatogr B Analyt Technol Biomed Life Sci* 2007, **849**(1-2):273-281.
10. Rai AJ, Stemmer PM, Zhang Z, Adam BL, Morgan WT, Caffrey RE, Podust VN, Patel M, Lim LY, Shipulina NV *et al*: **Analysis of Human Proteome Organization Plasma Proteome Project (HUPO PPP) reference specimens using surface enhanced laser desorption/ionization-time of flight (SELDI-TOF) mass spectrometry: multi-institution correlation of spectra and identification of biomarkers**. *Proteomics* 2005, **5**(13):3467-3474.
11. Benjamini H: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *The Royal Statistical Society, Series B, Methodological* 1995, **57**(11):289 -300.
12. Le Bihan MC, Tarelli E, Coulton GR, Noble JL, Dua RS, Isacke CM, Gui GP, Hou Y, Harris N, Coulton G: **Evaluation of an integrated strategy for proteomic profiling of skeletal muscle**
13. O'Bryan MK, Foulds LM, Cannon JF, Winnall WR, Muir JA, Sebire K, Smith AI, Keah HH, Hearn MT, de Kretser DM *et al*: **Identification of a novel apolipoprotein, ApoN, in ovarian follicular fluid**. *Endocrinology* 2004, **145**(11):5231-5242.
14. Fischer MB, Roeckl C, Parizek P, Schwarz HP, Aguzzi A: **Binding of disease-associated prion protein to plasminogen**. *Nature* 2000, **408**(6811):479-483.
15. Zerr I, Bodemer M, Kaboth U, Kretzschmar H, Oellerich M, Armstrong VW: **Plasminogen activities and concentrations in patients with sporadic Creutzfeldt-Jakob disease**. *Neurosci Lett* 2004, **371**(2-3):163-166.
16. Lucas MA, Fretto LJ, McKee PA: **The relationship of fibrinogen structure to plasminogen activation and plasmin activity during fibrinolysis**. *Ann N Y Acad Sci* 1983, **408**:71-91.
17. Collis SC, Kimberlin RH: **Further studies on changes in immunoglobulin G in the sera and CSF of Herdwick sheep with natural and experimental scrapie**. *J Comp Pathol* 1983, **93**(2):331-338.
18. Halimi M, Dayan-Amouyal Y, Kariv-Inbal Z, Friedman-Levi Y, Mayer-Sonnenfeld T, Gabizon R, Dayan Y, Engelstein R, Caspi S, Yanai A *et al*: **Prion urine comprises a glycosaminoglycan-light chain IgG complex that can be stained by Congo red**