The use of proteome similarity for the qualitative and quantitative profiling of reperfused myocardium.

Johannes P. C. Vissers, Sandrine Pons, Anne Hulin, Renaud Tissier, Alain Berdeaux, Joanne B. Connolly, James I. Langridge, Scott J. Geromanos, Bijan Ghaleh

The use of proteome similarity for the qualitative and quantitative profiling of reperfused myocardium

Johannes P. C. Vissers [1] [*] , Sandrine Pons [2] , Anne Hulin [3] , Renaud Tissier [2] , Alain Berdeaux [2] , Joanne B. Connolly [1] , James I. Langridge [1] , Scott J. Geromanos [4] , Bijan Ghaleh [2]

[1] Waters Corporation Waters Corporation , Atlas Park, Simonsway, M22 5PP, Manchester,GB

[2] IMRB, Institut Mondor de recherche biomédicale INSERM : U841 , Université Paris XII Val de Marne , Hôpital Henri Mondor 51, av du mal de lattre de tassigny 94010 CRETEIL CEDEX,FR

[3] Service de pharmacologie clinique AP-HP , Hôpital Henri Mondor , Université Paris XII Val de Marne , 51, av du Maréchal de Tassigny, Créteil,FR

[4] Waters Corporation Waters Corporation , 34 Maple Street, Milford, MA 01757,US

* Correspondence should be adressed to: Johannes Vissers <hans_vissers@waters.com >

Abstract

An LC-MS based approach is presented for the identification and quantification of proteins from unsequenced organisms. The method relies on the preservation of homology across species and the similarity in detection characteristics of proteomes in general. Species related proteomes share similarity that progresses from the amino acid frequency distribution to the complete amino sequence of matured proteins. Moreover, the comparative analysis between theoretical and experimental proteome distributions can be used as a measure for the correctness of detection and identification obtained through LC-MS based schemes.

Presented are means to the identification and quantification of rabbit myocardium proteins, immediately after inducing cardiac arrest, using a data-independent LC-MS acquisition strategy. The employed method of acquisition affords accurate mass information on both the precursor and associated product ions, whilst preserving and recording the intensities of the ions. The latter facilitates label-free quantification. The experimental ion density observations obtained for the rabbit sub proteome were found to share great similarity with five other mammalian samples, including human heart, human breast tissue, human plasma, rat liver and a mouse cell line. Redundant, species-homologues peptide identifications from other mammalian organisms were used for initial protein identification, which were complemented with peptide identifications of translated gene sequences. The feasibility and accuracy of label-free quantification of the identified peptides and proteins utilizing above mentioned strategy is demonstrated for selected cardiac rabbit proteins.

INTRODUCTION

The rapid increase in genome sequencing has led to the decoding of several complete genomes, including human [1]. Since the decoding of entire genomes, significant emphasis has been placed on the identification of the functionally translated gene products that comprise the proteome. Mass spectrometry (MS) has firmly been established as the primary technique for identifying and characterizing proteins, and to probe the structure and function of proteins. Consequently, mass spectrometry is now commonplace in the biological laboratory. The two most widely applied techniques are matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF) [2] and electrospray (ESI) [3]/nanoelectrospray ionization [4]. MALDI-TOF is often used in conjunction with some type of gel-isolation of proteins [5 ,6], whereas ESI is typically used in combination with liquid chromatography (LC) [7]. Recent MALDI-TOF research improvements to overcome limitations associated with the identification of mixtures or the presence of contaminates include in post source decay and MALDI MS/MS [8 ,9]. LC column and interface miniaturization [10 –12], the utility of alternative fragmentation techniques to collision induced dissociation such as electron transfer dissociation or electron capture dissociation [13 ,14] and alternative scanning approaches to data-dependent based acquisitions [15] have been described and are active research areas in the LC-ESI-MS based identification of peptides and proteins.

The relative quantification of protein expression changes is important in understanding protein function. For example, subtle changes in the expression level of a protein, in response to stimuli, may result in a significant biological effect. A traditional approach to the relative quantification and identification of proteins couples the separation and visualization power of two-dimensional (2D) gel electrophoresis. Spots are subsequently excised, digested and analyzed by either MALDI-TOF PMF or LC-MS to identify protein spots of interest. This hyphenated approach relies upon optical density measurements of protein spots from a 2D-gel to provide relative quantification between samples. There are however several disadvantages associated with such an approach. Firstly, the isoelectric

dimension of 2D-gels is not amenable to proteins with extreme isoelectric points, such as very acidic or basic proteins, and are therefore often under represented. Secondly, the mass range of proteins that can be separated and analyzed is also compromised by the use of a 2D gel strategy itself. The final disadvantage is that, despite significant effort on improving sample preparation procedures prior to gel electrophoresis, the ability to solubilize hydrophobic membrane proteins is limited [16].

Several non-gel based approaches have been developed that utilize various stable isotope labeling strategies in combination with separation and subsequent analysis by mass spectrometry [17 –26]. Despite the ability of these techniques to quantitatively compare protein levels between samples and across conditions, there is additional complexity inherent in such approaches. In addition, the decision to perform a quantitative study must be decided upon early in the experimental strategy. As an alternative, data independent exact mass LC-MS strategies have recently described [27 –29], where quantification is achieved via normalization of the LC-MS datasets and comparison of the observed tryptic peptide intensities across samples. A further development of this approach uses the same data to provide absolute quantification values for each protein present in a complex mixture, using a response factor, calculated from the intensity ratio of the average or sum of the three best ionizing peptides from the protein of interest and an internal standard protein [30]. Other non-labeled techniques include peptide match score summation [31] and spectrum sampling [32] for relative protein quantification and spectral counting based techniques such as protein abundance index (PAI) [33] and exponentially modified PAI (emPAI) for absolute protein quantitation [34].

The most common approach to protein identification, searching of tandem LC-MS data against a non-redundant protein database, relies upon the availability of a protein or translated nucleotide databases to match the experimental precursor and fragment ion data. It is widely believed that this approach cannot be used for the identification of proteins from unknown or partially sequenced genomes, due to the lack of available sequence information. The most common approach to working with proteins from partially sequenced organisms is to use de novo sequencing, to generate long stretches of primary amino acid sequence for error tolerant databank or BLAST searching [35 ,36]. Subsequently the amino acid sequence of the protein can be assembled from the homologous peptides, or the gene can be cloned and sequenced.

It is, however, possible to utilize tandem MS data, or data independent, alternate scanning data, as applied in this study, to search against related species databases to identify proteins by the presence of homologous peptides. Due to the large amount of sequence redundancy between proteomes [37], this makes a viable approach for the identification of proteins from poorly sequenced organisms. For example, rat, E. coli and human exhibit a similar distribution of protein and tryptic peptide molecular weight. The amino acid compositions for each proteome are very similar as well (half of the commonly occurring amino acids account for 70% of the total proteome in each case). This homology can be used to assemble the sequence of a gene product from an unsequenced organism by piecing together contiguous tryptic peptides which have homology to a sequenced genome. This approach has been used in this study for the qualitative and quantitative characterization of proteins from the rabbit myocardium.

The development of cardioprotective strategies is a major area of research in the cardiovascular field as illustrated by the tremendously vast number of publications reporting strategies aimed at reducing myocardial infarct size [38 –40]. Although this issue is of major importance, it is however not sufficient to induce a complete protection of the heart against the different deleterious consequences of myocardial ischemia, i.e. , infarction, stunning, arrhythmias and endothelial dysfunction [41 ,42]. Among them, improving the functional recovery of the reperfused myocardial tissue remains one the main goals to achieve as it is well known that the reduction of myocardial infarct size does not end up with a concomitant improvement of regional myocardial function [43 ,44]. Therefore, in order to gain new insight into the physiopathology of myocardial ischemia and reperfusion, it was investigated whether a particular protein profile could be obtained from the comparison of the proteomic analysis of both the 'non-ischemic' zone and the area 'at risk' of hearts that were submitted to ischemia and prolonged duration of reperfusion.

EXPERIMENTAL

The animal experiments were performed in accordance with the official regulations of the French Ministry of Agriculture.

Experimental protocol

Two male New Zealand rabbits (2–2.5 kg) were anesthetized with pentobarbital sodium (20–30 mg/kg, i.v.). They were intubated and mechanically ventilated with 100% oxygen (ventilation rate: 25 breaths/min; tidal volume: 25 ml). A catheter was positioned in the rabbit's ear marginal artery for arterial pressure measurement (Statham P23ID strain gauge, Statham Instruments, Oxnard, CA, USA). An external electrocardiogram (ECG) was also recorded. A left thoracotomy was performed at the fourth intercostal space. The pericardium was opened and a 4/0 Prolene suture was passed beneath a major branch of the left coronary artery. The ends of the ligature were passed through a short segment of propylene tubing to form a snare. Regional myocardial ischemia was induced by pulling the snare through the tubing. Ischemia was confirmed by the presence of regional modifications of the myocardial surface and by the occurrence of ST segment

deviation of the ECG. A 30 min period of coronary artery occlusion was performed and the snare was released to induce reperfusion. The chest was then closed in layers and a small tube was left in the thorax to evacuate air and fluids after surgery. Rabbits underwent 5 days of coronary artery reperfusion.

Delineation of the area 'at risk'

After completion of reperfusion, animals received heparin and sodium pentobarbital (50 mg/kg, i.v.). Potassium chloride was then administered intravenously to induce cardiac arrest and the hearts were excised. Rapidly, the ascending aorta was cannulated and perfused (120 mmHg) retrogradely with saline followed by Evans blue (1%) after ligation of the previously occluded artery. The area 'at risk', i.e. , the area previously submitted to coronary artery occlusion and reperfusion, was identified as the non blue area. The non-ischemic zone was the blue area. Myocardial samples of both areas were immediately frozen into liquid nitrogen.

Protein extraction

Myocardial samples were scissor minced, washed with ice-cold NaCl and homogenized using a Teflon Potter homogenizer in ice-cold buffer containing 50 mM Tris-HCl (pH 7.4 at 4°C), 150 mM NaCl, 1 mM EDTA, 5 µl/ml protease inhibitor cocktail, 1 mM sodium orthovanadate, 5 mM sodium fluoride and 1 mM sodium Na_2 β-glycerol phosphate (Sigma Aldrich, St. Louis, MO, USA). The homogenates were centrifuged at 40,000g for 30 min at 4°C to collect the cytosol. The cytosolic protein concentration was determined by the BCA Protein Assay Kit (Pierce, Rockford, IL, USA).

Protein digestion

250 µg of each myocardial protein sample was taken and washed three times with 500 µL of 0.1% RapiGest (Waters Corporation, Milford, USA) in 50 mM ammonium bicarbonate (Sigma Aldrich) using a 5000 Da Mw cut off spin column (Amicon Millipore, Billerica, MS, USA). The protein solution was concentrated to approximately 100 µL. After incubation of the solution at 80°C for 15 min, a 5 µL aliquot of 100 mM DTT (Sigma Aldrich) in 50 mM ammonium bicarbonate was added and the solution heated at 60°C for 15 min. Following the reduction of disulphide bonds, alkylation was facilitated by the addition of a 5 µL aliquot of 200 mM iodoacetamide (Sigma Aldrich) in 50 mM ammonium bicarbonate. The solution was incubated in the dark at room temperature for 30 min. Sequencing grade TMPK treated trypsin (Promega, Madison, MI, USA), was added at a 1:25 w/w ratio and the solution was left overnight at 37°C. TFA was added to a final concentration of 0.5% (v/v) to hydrolyze RapiGest and the solution incubated at 37°C for 20 min before being vortexed and centrifuged. The supernatant was collected and transferred to a clean microcentrifuge tube. The resulting tryptic digest solutions were diluted to a concentration of 0.18 µg/µL before being mixed 1:1 (v/v) with a tryptic digest of Saccharomyces cerevisiae Enolase, at a concentration of 50 fmol/µL. Reference proteome digests from human heart, human breast tissue, human plasma, rat liver lysate and a mouse cell line were treated identically. The LC-MS analyses were performed using 2 µL of the final protein digest mixtures.

LC-MS configuration

Nanoscale LC separation of tryptic peptides was performed with a nanoACQUITY system (Waters Corporation), equipped with a Symmetry C_{18} 5 µm, 5 mm × 300 µm precolumn and an Atlantis C18 3 µm, 25 cm × 75 µm or a Atlantis C18 3 µm, 15 cm × 75 µm analytical reversed phase column (Waters Corporation). The samples, 2 µL full loop injection, were initially transferred with an aqueous 0.1% formic acid solution to the precolumn at a flow rate of 4 µL/min for 3 min. Mobile phase A was water with 0.1% formic acid whilst mobile phase B was 0.1% formic acid in acetonitrile. After desalting and preconcentration, the peptides were eluted from the precolumn to the analytical column and separated with a gradient of 3% to 40% mobile phase B over 90 min at a flow rate of 300 nL/min, followed by a 10 min rinse with 90% of mobile phase B. The column was reequilibrated at initial conditions for 20 min. The column temperature was maintained at 35°C. The lock mass compound, [Glu[1]]-Fibrinopeptide B, was delivered by the auxiliary pump of the LC system at 250 nL/min at a concentration of 100 fmol/µL to the reference sprayer of the NanoLockSpray source of the mass spectrometer. All samples were analyzed in triplicate.

Mass spectrometric analysis of tryptic peptides was performed using a Q-Tof Premier mass spectrometer (Waters Corporation, Manchester, UK). For all measurements, the mass spectrometer was operated in v-mode with a typical resolution of at least 10,000 FWHM. All analyses were performed in positive mode ESI. The time-of-flight analyzer of the mass spectrometer was externally calibrated with a NaI mixture from m/z 50 to 1990. The data were post-acquisition lock mass corrected using the doubly charged monoisotopic ion of [Glu[1]]-Fibrinopeptide B. The reference sprayer was sampled with a frequency of 30 s. Accurate mass LC-MS data was collected in an alternating, low energy and elevated energy mode of acquisition [29]. The spectral acquisition time in each mode was 1.5 s with an 0.1 s interscan delay. In low energy MS mode, data was collected at constant collision energy of 4 eV. In elevated energy MS mode, the collision energy was ramped from 15 eV to 40 eV during each 1.5 s integration. One cycle of low and elevated energy data was acquired every 3.2 s. The RF amplitude applied to the quadrupole mass analyzer was adjusted such that ions from m/z 300 to 2000 were efficiently transmitted, ensuring that any ions observed in the LC-MS data less than m/z 300 were known to arise from dissociations in the collision cell.

Data processing and protein identification

Continuum LC-MS data was processed and searched using ProteinLynx GlobalSERVER version 2.3 (Waters Corporation). Protein identifications were obtained by searching either superorder Euarchontoglires (supraprimates), infraclass Eutheria (placental mammals), or species specific Homo sapiens (human), Rattus norvegicus (rat), Mus musculus (mouse) UniProtKB/Swiss-Prot (http://www.uniprot.org/) databases. An Oryctolagus cuniculus (rabbit) database was constructed from all reported NCBI protein sequences of Oryctolagus cuniculus to which protein amino acid sequence information from an Ensemble translated genome Oryctolagus cuniculus database (http://www.ensembl.org/Oryctolagus_cuniculus/index.html) was appended. Protein amino acid redundancy was cleared from all utilized databases with dbtoolkit [45] prior to use. Sequence information of Enolase Saccharomyces cerevisiae was added to the databases to afford the ability to normalize the data sets or to conduct absolute quantification [30].

Data independent scanning protein identifications were accepted when more than three fragment ions per peptide, seven fragment ions per protein and more than 2 peptides per protein were identified. Furthermore, the identification of the protein had to occur in at least 2 out of 3 replicate injections. By using replication as a filter, the false positive rate is minimized as false positive identifications have a random nature and as such do not tend to replicate across injections [46]. Typical search criteria used for protein identification included automatic peptide and fragment ion tolerance settings (approximately 10 and 25 ppm, respectively), 1 allowed missed cleavage, fixed carbamidomethyl-cysteine modification and variable methionine oxidation, N/Q deamidation, N-terminal acetylation modifications.

RESULTS AND DISCUSSION

Theoretical proteome distributions

The identification and quantification of proteins from an organism whose proteome is not well represented is challenging. However, to a large extent, homology between related species is highly preserved at the protein level and can be potentially used for protein identification purposes of organism of which the proteome is not well characterized. Moreover, the enzymatically derived peptides generated from these proteins are constructed from a relatively small set of 20 commonly occurring amino acids. Combined, these similarities can be used to extract valuable information from LC-MS data sets in order to obtain both qualitative and quantitative information. To illustrate commonality of related species at the proteome level, theoretical protein and peptide molecular weight (Mw) distributions for the species investigated were calculated. Figure 1a shows the protein Mw distributions for rat, mouse, human and a combined empirical and theoretically predicted rabbit database. The average Mw values of the proteins in the four databases are 55.1, 54.1, 52.1 and 53.2 kDa. The median Mw values equal 41.9, 36.4, 39.8 and 37.7 kDa, respectively. The estimated average standard deviation based on a normal distribution is approximately 30–35 kDa for all databases. The results illustrate very similar distributions of intact protein molecular weights across species. As will be demonstrated in a subsequent paragraph, the Mw distribution of correctly identified proteins from complex biological sample matrices should reflect the theoretical distribution of the proteome under investigation. Distinct variation between the two might be indicative of either an alternative sample preparation fractionation strategy or of incorrect, database protein Mw biased search identifications.

Figure 1b illustrates the theoretical Mw protonated precursor mass frequency distribution of proteolytic peptides for the earlier mentioned and discussed four databases. Trypsin was used as the protease of interest, a single, miss-cleavage was allowed only in instances where the lysine or arginine was followed by an aspartic or glutamic acid, the mass of cysteine residues were modified to the carbamidomethyl derivative. The results shown in Figure 1b illustrate that the proteome distributions at the peptide level follow the same trend as the theoretical protein Mw distributions. The average and median protonated Mw of the peptides of the four databases of interest are 1776.7 and 1442.7 Da, respectively. These trends – in line with the protein identification - should also be followed by the database search results at the peptide level. This will also be demonstrated in subsequent paragraphs.

Given that protein Mw and peptide Mw distributions of mammalian proteome databases are similar, it is to be expected that the same similarity is retained at the amino acid composition level. Figure 1c shows the amino acid composition of the four mammalian species specific databases. The high degree of similarity at the protein and peptide Mw level is also reflected at the amino acid composition level. As stated, this is to be expected since the number of naturally according amino acids is relatively small and the species of interest all belong to the same class. Interestingly, half of the most frequently occurring amino acids represent approximately 70% of all the amino acids in each proteome. Due to this high degree of similarity, the number of highly ionizing peptides to a given protein is directly proportional to its intact Mw. As such, it is easier to identify medium to higher molecular weight proteins at or near the limit of detection than those of lower molecular weight. The relatively high frequency of certain amino acids also restricts the number of possible linear sequences that can be generated. In addition, the encoding of the amino acid sequence of polypeptides and proteins is not a random biosynthetic process, which also further limits the number of peptide and protein sequences that can be produced.

Experimental proteome distributions

In addition to illustrating the theoretical similarity between proteomes at the protein, peptide and amino acid levels, a high degree of uniformity at the ion detection level of proteolytically digested samples from the four different species under study will be demonstrated.

Using a similar starting amount of protein, each sample was digested and approximately the same amount of material analyzed in triplicate. With equal on-column sample loads and a high degree of similarity between proteomes, the expected number of ion detections across a gradient elution are expected to be similar for every LC-MS experiment. Figures 2a–c illustrate the number of ions detected per unit time, the detected intensity per unit time and the detected ion intensity distribution, respectively. Figure 2a shows the average number of detected ions for three replicate injections per unit time for the 'at risk' zone and the non-ischemic zone investigated proteomic rabbit samples. The distributions show a high degree of similarity. Supplemental Figure 1 illustrates similar distributions for (sub) proteome samples of human heart, human breast tissue, human plasma, rat liver lysate and a mouse cell line. Here, the average number of detected ions for three replicate injections per unit time is shown as well. As can be observed, a very similar number of ions per unit time were detected for all investigated samples, which suggests that from an LC-MS perspective, the samples are similar. The detected ion intensity per unit time for the two rabbit sample types is shown Figure 2b and for the five reference (sub) proteomes in Supplemental Figure 2 . As previously observed for the ion intensity distributions shown in Figure 2a , the number of detected ions are also very analogous.

It should also be noted that the majority of the number of observed detections occur in the lower two orders of magnitude of the detection dynamic range (approximately 80%), which is illustrated in Figure 2c and Supplementary Figure 3 . However, these detections only constitute for about 30 to 35% of the total detected ion intensity, which is in agreement with earlier publications [46 ,47]. In other words, the more abundant proteins for the investigated samples comprised about 65 to 70% of the total mass, i.e. total detected intensity, loaded onto an LC column that can be detected by a mass spectrometer. Moreover, the obtained peptide detection distributions shown in Figure 2a and Supplementary Figure 1 are in very good agreement with the theoretical peptide Mw distributions shown in Figure 1b . This suggests that the applied data-independent LC-MS detection measurements are not biased and that the (sub) proteome of interest – or at least the part that can be detected – has been adequately sampled. The latter should ideally also hold for the identified peptides since this will provide additional confidence in correctness of identification. An overview of the detection characteristics of the various studied samples is provided in Table 1 . Shown are the average values of three technical replicates. A summary of the number of detected ions per unit time and the total detected intensity per unit time for the investigated and reference (sub) proteomics is provided in Table 2 .

The results shown in Table 1 indicate that on average 80,000 – 100,000 ions were detected above threshold with the applied analytical LC-MS conditions. At this stage, all isotopic and charge stage components are still considered. An ion, however, should have sufficient intensity to be able to generate collision-induced fragment ions of sufficient intensity for reliable identification. Approximately 70% of the detected ions are above an intensity threshold which may contribute to possible identification. Interestingly, they encompass on average 90 % of the total detected intensity. Dependent on the size and sequence of a peptide, a certain number of isotopes at one or more charge states, will contribute to the overall number of detected features. Deisotoping and charge stage reducing the data set reveals that on average 17,000 – 20,000 peptides (accurate mass-retention time clusters) were detected for all investigated (sub) proteomes. In other words, on average, approximately five isotopic features per peptide were detected. Furthermore, the maximum cluster intensity was found to be similar and in the same order of magnitude for all samples. The detection and identification dynamic ranges were 3.5 and 2.7 orders, respectively, which is in agreement with a previous study [46]. The results provided in Table 2 show that the majority of the ions are detected in the two lowest orders, approximately 80%. Although large in number, they only contribute to 30 – 35% of the total detected intensity, which again holds for all investigated sample types. Lastly, the number of clusters in mass sufficient regions is typically greater than 95% of the total number of detected, deconvoluted monoisotopic components.

Qualitative analysis

To further demonstrate the un-biased and reproducible nature of the applied data-independent LC-MS scanning method, replication of the deconvoluted mono-isotopic masses at a given retention time was determined for the four rabbit sub proteome heart samples. The clustering mass precision and retention time tolerances are automatically determined by the detection processing software. For these particular samples, considering 12 LC-MS experiments in total, the average and median mass precision equaled 2.41 and 1.84 ppm, respectively. The average and median retention time reproducibility were 0.42 and 0.31%. The results shown in table 3 illustrate that detection replication at the deconvoluted mass level is approximately 65 to 70% for all investigated rabbit samples. It should be noted that the loaded on-column amounts were balanced based on colorimetric detection results and as such may vary. However, the fraction of replicating, deconvoluted mono-isotopic components is constant and hardly affected by variations in the analyzed amounts. The peptide and protein replication details follow replication at the deconvoluted accurate mass-retention time cluster level and are often even higher since only the identified components have to be considered. Typical values are around 80%. The latter discussion is beyond the scope of this of the current manuscript, but has been demonstrated in previous studies [46 ,47]

An example of the identification of a protein by data independent, continuous ion scanning is shown in Figure 3 , whereby the collision energy applied to the collision cell of the mass spectrometers is varied between a normal and elevated energy state. Precursor ions are not isolated with the first analyzer of the mass spectrometer, providing a parallel approach to ion detection and sequencing. The low energy segment of the obtained data sets is typically used for quantification of the proteins, whereas the combined low and elevated energy

information are utilized for qualitative, identification purposes [29]. The bottom pane of Figure 3 shows the low energy spectrum and the top pane the elevated energy one of one of the detergent phase samples. Shown inset is the identification of a peptide by means of the applied search algorithm that is dedicated for searching multiplexed fragmentation data.

The frequency distribution of the masses of the identified rabbit peptides and proteins vs. the respective theoretically predicted distributions are shown in Figure 4 . In general, the identified and theoretical distributions show very similar trends. A slight bias towards higher mass for both the identified peptides and proteins is observed. This is to be expected since smaller proteins typically generate a smaller number of proteolytic peptides that can be detected and identified. Hence, the opportunity or chance of detecting a peptide of a small molecular weight protein is smaller than that of an average sized or high molecular weight protein. In practice this means that LC-MS based identification strategies slightly oversample the mass domain of a (sub) proteome at both the peptide and protein level. The identification of polypeptides and small proteins is therefore challenged. The latter can be observed in the results presented in Figure 4 . However, a more severe oversampling of higher molecular weight proteins could be indicative of an error in the protein identification process. Namely, not all peptides to a protein ionize with the same efficiency. The ionization rates of peptides to a protein of a fixed concentration and length can vary by up to 3 orders of magnitude. Furthermore, the number of peptides in any given ionization efficiency region is directly proportional to the Mw of the intact protein. The larger a protein, the more highly ionizing peptides there will be available for detection and identification. As such, if the bias to higher molecular weight is correct, the number of peptides identified to each of these proteins should be proportional to the Mw and concentration of the proteins. However, in most instances, these higher Mw proteins are reported with one or two peptides to match, which affects the credibility of the identification.

Preservation of homology

It is to be expected that homology across species is preserved [29]. To illustrate this, the data independent LC-MS data of the rabbit heart samples were first searched against mammalian databases (Euarchontoglires (supraprimates) or Eutheria (placental mammals)). The results of two of the identified proteins, namely isocitrate dehydrogenase 2 and cysteine and glycine rich protein 3 (cardiac LIM protein), in terms of identified peptides with a minimum identification probability of 95% and associated intensities, are summarized in Table 4 . As can be seen from the results presented in table 4 , the two rabbit proteins share certain homology with human, rat, and mouse, which allows for protein identification. The latter stands to reason since these species are part of the same mammalian superorder based on molecular genetic sequence. In other words, preserved homology between and amongst species can be utilized for the initial identification of species redundant peptides, which can be subsequently re-aligned to assemble the protein amino acid sequence. The identified peptides were mapped against the predicted protein amino acid rabbit sequences of the two example proteins based on the translation of the genomic sequence of rabbit, which is shown in Figures 5a and 5b . In all instances, the same homologous peptides were identified in human, rat and mouse.

Next, the data was searched against a predicted rabbit protein amino acid database. The identified peptides are presented in the second column of Table 4 and underlined in Figures 5a and 5b . Regardless of the database, i.e. mammalian or rabbit, the same peptides were identified, where the theoretical rabbit database afforded the identification of additional peptides. The results in Table 4 are presented in descending order of peptide intensity. To illustrate data and ionization distribution consistency, a human heart data set was searched against the rabbit database. The intensities of the identified homologous human peptides are provided in parentheses in Table 4 . The applied data independent LC-MS scanning technique acquires both precursor ion and fragment ion data throughout the entire chromatographic peak width. This enables the use of integrated peak areas as an additional element for the characterization of the identified proteins. The relative relationship of the intensity measurements of the identified tryptic peptides to the same protein, along with the associated accurate mass measurements and retention times, provide an additional dimension of specificity for a given proteolytic peptide map of a protein. It can be seen that the homologous rabbit and human peptides identified to isocitrate dehydrogenase 2 and cysteine and glycine rich protein 3 (cardiac LIM protein) exhibit very similar peptide intensity ratios, which strengthens the identification of the proteins of species poorly represented in amino acid sequences databases. Per sample, approximately 350 – 400 proteins were identified from a mere estimated 0.5 μg protein digest on column load using the above mentioned scanning and identification approaches. Further validation of these qualitative identifications is currently on-going.

Quantitative analysis

As mentioned, the data independent LC method of acquisition also affords for label free quantification of the identified proteins. The latter is demonstrated for filamin C, gamma and annexin 2, which were both found to be differently expressed in terms of their relative abundance. Shown in Figure 6 is the variation in average intensities of accurate mass-retention time clusters of the two investigated rabbit samples, i.e. 'at risk' vs. 'non- ischemic', for one of the biological samples. A perfect 45-degree diagonal distribution intersecting through zero and displaying a minimum degree of deviation throughout the detected range with little intensity variation would be indicative of the absence of differential expression between the samples of interest. As can be seen, the intensity variation of the majority of the commonly detected peptide components in both investigated conditions is mild. Hence, the proteins amounts did not very greatly as a result of the induced perturbation. However, the peptides identified to filamin C, gamma and annexin 2 for the same relative precursor intensity

variation to their parent protein. The average fold changes ('area at risk' vs. 'non ischemic zone') were 2.61 and 0.75, respectively. Moreover, the reported likelihood of relative quantification for both proteins reported by the applied quantification software was found to be greater than 95%. In addition, an average root-mean-square probability value was expressed based on a tailed, two-sample unequal variance Student t-test and found to be equal to 0.0092 for filamin C, gamma and 0.012 for annexin 2. The technical variation of the applied scanning technique is approximately 10 – 15% after normalization using an internal standard [46], which allows for the detection of mild changes in protein abundance in systems and data sets with little biological variation.

Filamins (A, B and C) were first described as non-muscle actin-binding proteins. They are large cytoplasmic proteins consisting of an amino-terminal actin-binding domain and a rod-like domain of 24 repeated anti-parallel beta-sheets interrupted by 2 flexible loops that form hinge structures. More than 30 different proteins have been reported to interact with filamins, suggesting that filamins have a wide role as structural components of the cytoskeleton [48 –51]. Through interactions with actin and membrane proteins, filamins link the cytoskeleton to the plasma membrane and are believed to be essential in cell motility and membrane stability. Locally expanded swellings of cardiac jelly and mesenchymal cells undergo extensive remodeling to form the heart valves. Several ligands and signaling pathways are implicated: VEGF, NFATc, Notch, Wnt/beta catenin, BMP, TGFb, ErbB and NF1/Ras. Filamin A is required for cell-cell contact in vascular development and cardiac morphogenesis. Annexin 2 is not described in the context of cardiac research. It should be noted that the observed increase in protein abundance of annexin 2, although accurate, is very mild.

CONCLUSION

Proteome similarity at the amino acid, peptide and protein levels has been described from both theoretical and practical perspectives for four mammalian species. Upon the use of unbiased analytical sampling and identification methods, as demonstrated in this manuscript, the obtained detection and identification results typically follow theoretically predicted trends. Moreover, ion density measurements in terms of intensity and number, as measured by means of LC-MS, were comparable for all investigated (sub) proteomes.

The preserved homology between and amongst species can be utilized for the identification of proteins of so-called 'unsequenced organisms' such as rabbit. Translated DNA sequences can be queried to identify the correct open reading frame. The identified open reading frame can be subsequently processed in-silico to generate all theoretical peptides and their associated product ions. The in-silico accurate mass-retention time components can then be queried against the experimental data to identify non-homologous peptides. This increases the number of peptides that can be identified and utilized across technical replicates and conditions. Cross-species identified and theoretically predicted peptides can be subsequently mapped back to the parent protein to obtain more comprehensive protein amino acid sequence coverage.

The near 100% duty cycle of the utilized data independent, alternate scanning method of acquisition also records the intensity and chromatographic profile of the precursor ion and allows for the label-free quantification of the identified peptides and proteins. Despite mild biological protein expression between the samples of interest, several proteins were quantified to illustrate that the relative peptide and protein abundances are consistent with label-free LC-MS quantification approaches.

References:
- 1. Bentley DR. Nature. 409 : 2001 ; 942 -
- 2. Karas M, Bachmann D, Hillenkamp F. Anal Chem. 57 : 1985 ; 2935 -
- 3. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Science. 246 : 1989 ; 64 -
- 4. Wilm M, Mann M. Int J Mass Spectrom Ion Processes. 136 : 1994 ; 167 -
- 5. Yates JR III, Speicher S, Griffin PR, Hunkapiller T. Anal Biochem. 214 : 1993 ; 397 -
- 6. James P, Quadroni M, Carafoli E, Gonnet G. Biochem Biophys Res Commun. 195 : 1993 ; 58 -
- 7. Fournier ML, Gilmore JM, Martin-Brown SA, Washburn MP. Chem Rev. 107 : 2007 ; 3654 -
- 8. Medzihradszky KF, Campbell JM, Baldwin MA, Falick AM, Juhasz P, Vestal ML. AL Burlingame Anal Chem. 72 : 2000 ; 552 -
- 9. Suckau D, Resemann A, Schurenberg M, Hufnagel P, Franzen J, Holle A. Anal Bioanal Chem. 376 : 2003 ; 952 -
- 10. Vissers JPC. J Chromatogr A. 856 : 1999 ; 117 -
- 11. Valaskovic GA, McLafferty FW. J Am Soc Mass Spectrom. 7 : 1996 ; 1270 -
- 12. Zhang S, Van Pelt CK. Expert Rev Proteomics. 1 : 2004 ; 449 -
- 13. Mikesh LM, Ueberheide B, Chi A, Coon JJ, Syka JE, Shabanowitz J, Hunt DF. Biochim Biophys Acta. 1764 : 2006 ; 1811 -
- 14. Bakhtiar R. Z Guan Biotechnol Lett. 28 : 2006 ; 1047 -
- 15. Liebler DC. Nat Methods. 1 : 2004 ; 16 -
- 16. Gorg A, Weiss W, Dunn MJ. Proteomics. 4 : 2004 ; 3665 -
- 17. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Nat Biotechnol. 17 : 1999 ; 994 -
- 18. Zhou HL, Ranish JA, Watts JD, Aebersold R. Nat Biotechnol. 19 : 2002 ; 512 -

- 19 . Griffin TJ , Gygi SP , Rist B , Aebersold R . Anal Chem . 73 : 2001 ; 978 -
- 20 . Chakraborty A , Regnier F . J Chromatogr A . 949 : 2002 ; 173 -
- 21 . Smith RD , Anderson GA , Lipton MS , Pasa-Tolic L , Shen Y , Conrads TP , Veenstra TD , Udseth HR . Proteomics . 2 : 2002 ; 513 -
- 22 . Ong SE , Kratchmarova I , Mann M . J Proteome Res . 2 : 2003 ; 173 -
- 23 . Krijgsveld J , Ketting RF , Mahmoudi T , Johansen J , Artal-Sanz M , Verrijzer CP , Plasterk RHA , Heck AJR . Nat Biotechnol . 21 : 2003 ; 927 -
- 24 . Oda Y , Huang K , Cross FR , Cowburn D , Chait BT . Proc Natl Acad Sci . 96 : 1999 ; 6591 -
- 25 . Yao XD , Freas A , Ramirez J , Demirev PA , Fenselau C . Anal Chem . 73 : 2001 ; 2836 -
- 26 . Stewart II , Thomson T , Figeys D . Rapid Commun Mass Spectrom . 15 : 2001 ; 2456 -
- 27 . Wang W , Zhou H , Lin H , Roy S , Shaler TA , Hill LR , Norton S , Kumar P , Anderle M , Becker CH . Anal Chem . 75 : 2003 ; 4818 -
- 28 . Radulovic D , Jelveh S , Ryu S , Hamilton TG , Foss E , Mao Y , Emili A . Mol Cell Proteomics . 3 : 2004 ; 984 -
- 29 . Silva JC , Denny R , Dorschel CA , Gorenstein MV , Kass IJ , Li GZ , McKenna T , Nold MJ , Richardson K , Young P , Geromanos SJ . Anal Chem . 77 : 2005 ; 2187 -
- 30 . Silva JC , Gorenstein MV , Li GZ , Vissers JPC , Geromanos SJ . Mol Cell Proteomics . 5 : 2006 ; 144 -
- 31 . Allet N , Barrillat N , Baussant T , Boiteau C , Botti P , Bougueleret L , Budin N , Canet D , Carraud S , Chiappe D , Christmann N , Colinge J , Cusin I , Dafflon N , Depresle B , Fasso I , Frauchiger P , Gaertner H , Gleizes A , Gonzalez-Couto E , Jeandenans C , Karmime A , Kowall T , Lagache S , Mahé E , Masselot A , Mattou H , Moniatte M , Niknejad A , Paolini M , Perret F , Pinaud N , Ranno F , Raimondi S , Reffas S , Regamey PO , Rey PA , Rodriguez-Tomé P , Rose K , Rossellat G , Saudrais C , Schmidt C , Villain M , Zwahlen C . Proteomics . 4 : 2004 ; 2333 -
- 32 . Liu H , Sadygov RG , Yates JR . Anal Chem . 76 : 2004 ; 4193 -
- 33 . Rappsilber J , Ryder U , Lamond AI , Mann M . Genome Res . 12 : 2002 ; 1231 -
- 34 . Ishihama Y , Oda Y , Tabata T , Sato T , Nagasu T , Rappsilber J , Mann M . Mol Cell Proteomics . 4 : 2005 ; 1265 -
- 35 . Shevchenko A , Sunyaev S , Loboda A , Bork P , Ens W , Standing KG . Anal Chem . 73 : 2001 ; 1917 -
- 36 . Sunyaev S , Liska AJ , Golod A , Shevchenko A , Shevchenko A . Anal Chem . 75 : 2003 ; 1307 -
- 37 . Patthy L . Protein Evolution . Blackwell Publishing ; Oxford 1999 ; 75 -
- 38 . Aouam K , Tissier R , Bruneval P , Mandet C , Berdeaux A , Ghaleh B . Am J Physiol Heart Circ Physiol . 288 : 2005 ; H2763 -
- 39 . Cohen MV , Yang XM , Downey JM . Am J Physiol . 277 : 1999 ; H1754 -
- 40 . Couvreur N , Lucats L , Tissier R , Bize A , Berdeaux A , Ghaleh B . Am J Physiol Heart Circ Physiol . 291 : 2006 ; H1345 -
- 41 . Murry CE , Jennings RB , Reimer KA . 74 : 1986 ; 1124 -
- 42 . Ovize M , Przyklenk K , Hale SL , Kloner RA . Circulation . 85 : 1992 ; 2247 -
- 43 . Tissier R , Berdeaux A , Ghaleh B , Couvreur N , Krieg T , Cohen MV , Downey JM . Front Biosci . 13 : 2008 ; 284 -
- 44 . Zhao ZQ , Corvera JS , Halkos ME , Kerendi F , Wang NP , Guyton RA , Vinten-Johansen J . Am J Physiol Heart Circ Physiol . 285 : 2003 ; H579 -
- 45 . Martens L , Vandekerckhove J , Gevaert K . Bioinformatics . 21 : 2005 ; 3584 -
- 46 . Vissers JPC , Langridge JI , Aerts JMFG . Mol Cell Proteomics . 6 : 2007 ; 755 -
- 47 . Silva JC , Denny R , Dorschel CA , Gorenstein MV , Li GZ , Richardson K , Wall D , Geromanos SJ . Mol Cell Proteomics . 5 : 2006 ; 589 -
- 48 . Kyndt F , Le Scouarnec S , Jaafar P , Gueffet JP , Legendre A , Trochu JN , Jousseaume V , Chaventré A , Schott JJ , Le Marec H , Probst V . Arch Mal Coeur Vaiss . 100 : 2007 ; 1013 -
- 49 . Charitakis K , Basson CT . Circulation . 115 : 2007 ; 2 -
- 50 . Armstrong EJ , Bischoff J . Circ Res . 95 : 2004 ; 459 -
- 51 . Feng Y , Chen HH , Moskowitz IP , Mendonza AM , Vidali L , Nakamura F , Kwiatkowski DJ , Walsh CA . Proc Natl Acad Sci . 103 : 2006 ; 19836 -

Figure 1

Figure 1a. Theoretical protein Mw distribution of the proteomes of a human, mouse, rat and theoretically constructed rabbit database. Leading signal peptides and initiating amino acids were truncated from the protein amino acid sequences if annotated. Figure 1b. Theoretical protonated peptide Mw distribution of a human, mouse, rat and theoretically constructed rabbit database. See section Results and Discussion for details. Figure 1c. Distribution of the amino acid composition of a human, mouse, rat and theoretically constructed rabbit database.
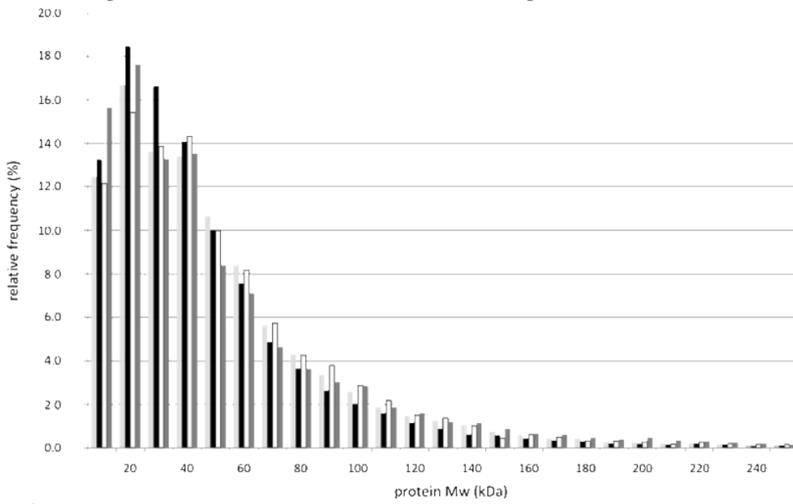
Figure 1a
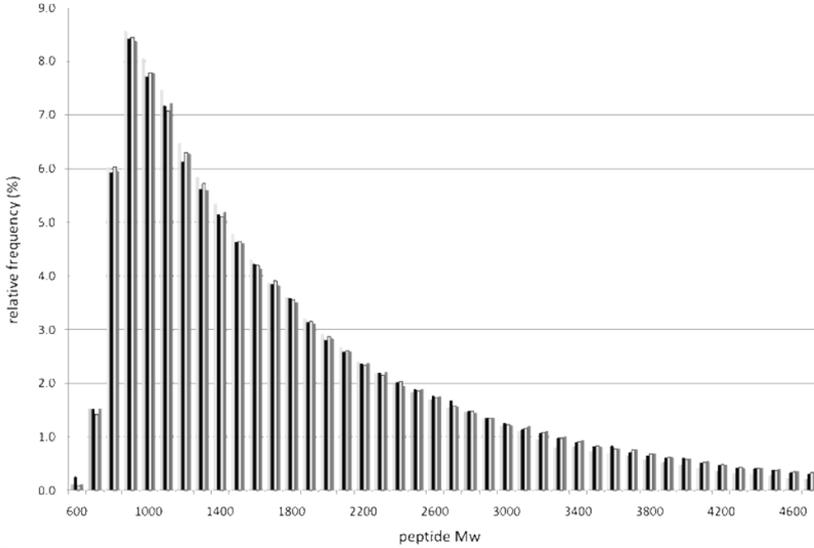
Figure 1b
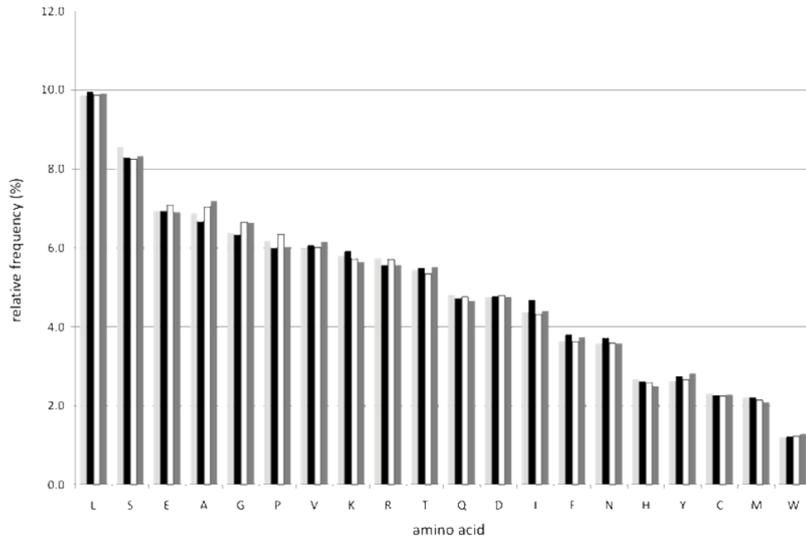
Figure 1c

Figure 2

Figure 2a. Number of detected ions/unit time for 'at risk zone' (black) and 'non ischemic' (grey) rabbit heart samples. See section Results and Discussion for details. Figure 2b. Ion intensity/unit time for 'at risk zone' (black) and 'non ischemic' (grey) rabbit heart samples. See section Results and Discussion for details. Figure 2c. Number of clusters/intensity bin for 'at risk zone' (black) and 'non ischemic' (grey) rabbit heart samples. See section Results and Discussion for details.
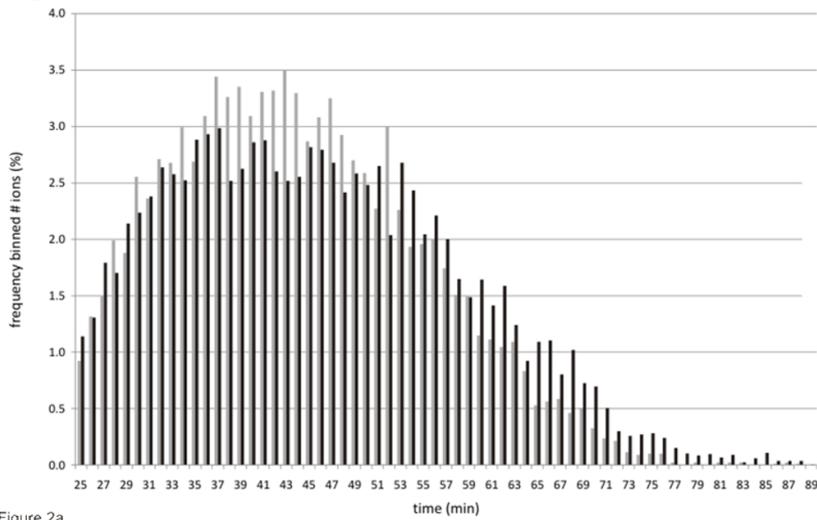
Figure 2a
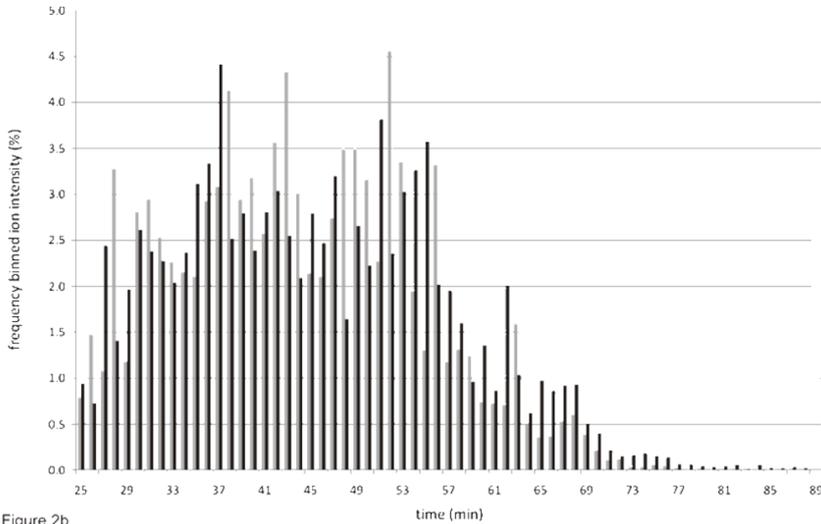
Figure 2b
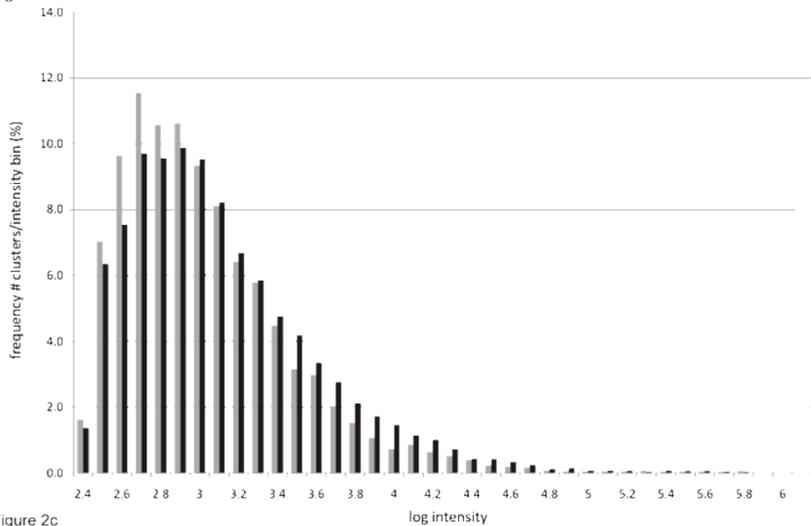
Figure 2c

Figure 3

Low (bottom) and elevated (top) energy spectra of isocitrate dehydrogenase 2 from rabbit heart tissue. Protein identification and quantitation was performed by data independent, alternate scanning experiments. Shown inset is the deconvoluted and annotated spectrum after database searching and identification.
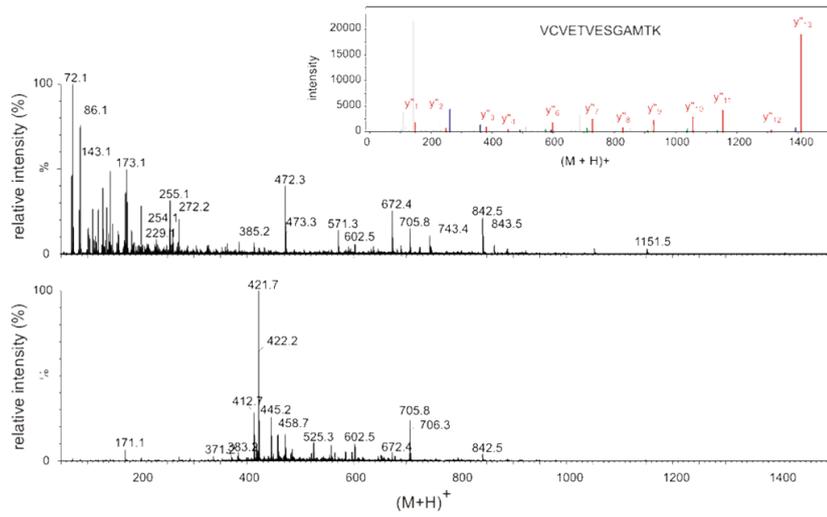
Figure 4

Identification of homologues human, mouse and rat peptide sequences mapped against the predicted protein amino acid of rabbit isocitrate dehydrogenase 2 (top) and rabbit glycine rich protein 3 (cardiac LIM protein) (bottom) based on the translation of the genomic sequence of rabbit. The identified rabbit peptides are underlined.
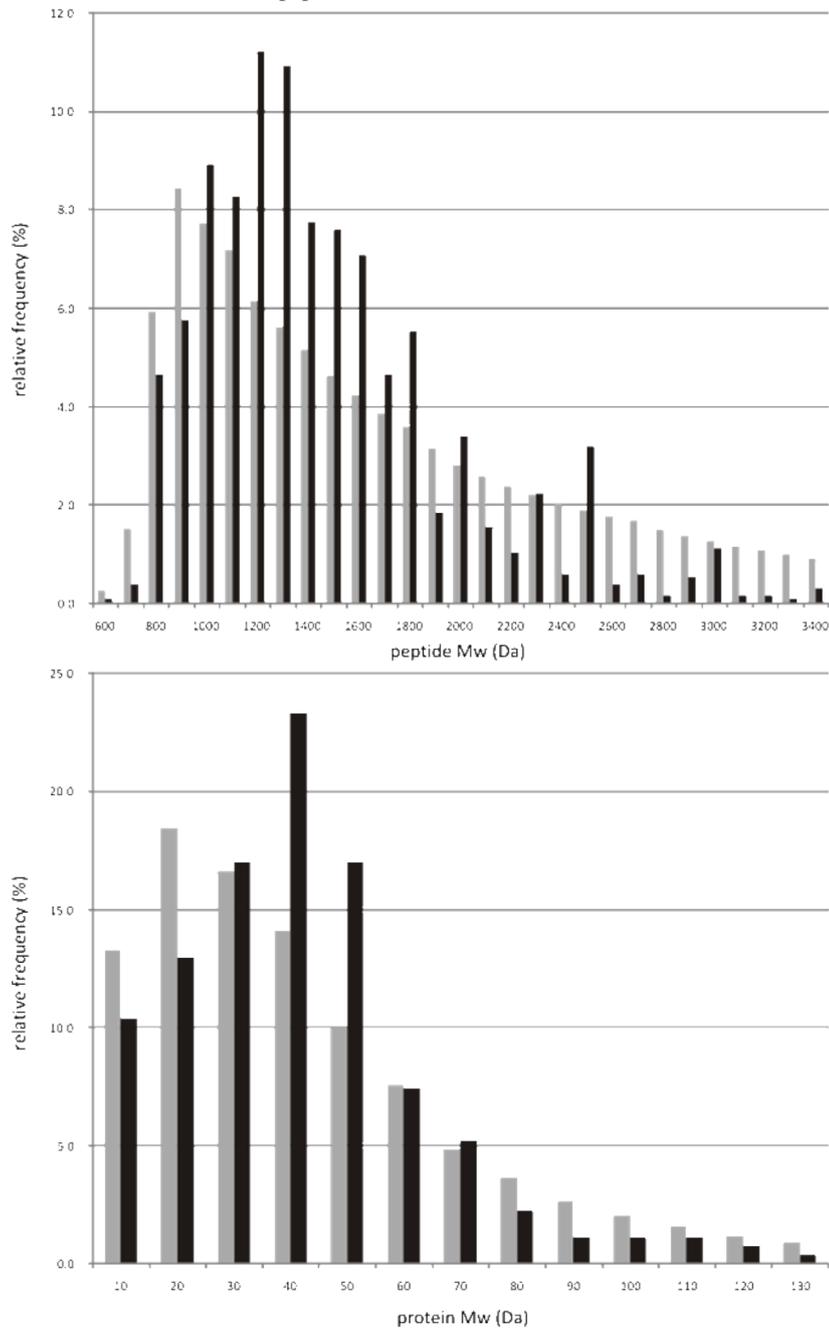
Figure 5

Empirically identified and theoretical database peptide Mw (a) and protein Mw (b) distributions of one of the investigated 'at risk zone' rabbit heart samples. The grey bars represent the theoretical distributions and the black bars the empirical database search results.

```
MAGYLRVVRSLCRAQGSGPTWAPAALATPTLQEQPRRHYADKRIKVSKPVVEMDGDEM    translated
                                                             human
                                                             mouse
                                                             rat

TRIIWQFIKEKLILPHVDVQLKYFDLGLPNRDKTDDQVTIDSALATQKYSVAVKCATI    translated
  IIWQFIK                YFDLGLPNR                            human
  IIWQFIK                YFDLGLPNR                            mouse
  IIWQFIK                YFDLGLPNR                            rat

TPDEARVEEFKLKKMWKSPNGTIRNILGGTVFREPIICKNIPRLVPGWTKPITIGRHA    translated
                    NILGGTVFR            LVPGWTKPITIGR        human
                    NILGGTVFR            LVPGWTKPITIGR        mouse
                    NILGGTVFR            LVPGWTKPITIGR        rat

HGDQYKATDFVVDRAGTFKLVFSPKDGSSAKEWEVYNFPGGGVGMGMYNTDESISGFA    translated
                                                             human
       ATDFVVDR                                              mouse
       ATDFVVDR                                              rat

HSCFQYAIQKKWPLYMSTKNTILKAYDGRFKDIFQEIFDKHYKTDFDKNKIWYEHRLI    translated
                      DIFQEIFDK                  LI          human
                      DIFQEIFDK                  LI          mouse
                      DIFQEIFDK                  LI          rat

DDMVAQVLKSSGGFVWACKNYDGDVQSDILAQGVGSLGLMTSVLVCPDGKTIEAEAAH    translated
DDMVAQVLK                                                    human
DDMVAQVLK                                                    mouse
DDMVAQVLK                                                    rat

GTVTRHYREHQKGRPTSTNPIASIFAWTRGLEHRGKLDGNQDLIRFAQTLEKVCVETV    translated
                      LDGNQDLIR          VCVETV              human
                      LDGNQDLIR                              mouse
                                                             rat

ESGAMTKDLAGCIHGLSNVKLNEHFLNTTDFLDTIKSNLDRALGKQ               translated
ESGAMTKDLAGCIHGLSNVKLNEHFLNTTDFLDTIK                         human
      DLAGCIHGLSNVKLNEHFLNTTDFLDTIK                           mouse
      DLAGCIHGLSNVKLNEHFLNTTDFLDTIK                           rat
```

Figure 5a

```
    MPNWGGGAKCGACEKTVYHAEEIQCNGRSFHKTCFHCMACRKALDSTTVAAHESEIYC    translated
                                    ALDSTTVAAHESEIYC            human
                                    ALDSTTVAAHESEIYC            mouse
                                    ALDSTTVAAHESEIYC            rat

    KVCYGRRYGPKGIGYGQGAGCLSTDTGEHLGLQFQQSPKPARSATTSNPSKFTAKFGE    translated
    K                                                           human
    K                                                           mouse
    K                                                           rat

    SEKCPRCGKSVYAAEKVMGGGKPWHKTCFRCAICGKSLESTNVTDKDGELYCKVCYAK    translated
                                    SLESTNVTDKDGELYCK           human
                                    SLESTNVTDKDGELYCK           mouse
                                    SLESTNVTDKDGELYCK           rat

    NFGPTGIGFGGLTQQVEKKE                                        translated
    NFGPTGIGFGGLTQQVEK                                          human
    NFGPTGIGFGGLTQQVEK                                          mouse
    NFGPTGIGFGGLTQQVEK                                          rat
```

Figure 5b

Figure 6

Binary comparison of the log intensity measurements obtained from the accurate mass–retention time components clusters for 'at risk zone' vs. 'non-ischemic zone'. Deviation from the main 45-degree diagonal line reflects fold changes of peptide cluster intensities, that is, differential protein expression. Highlighted are the identified peptides to filamin C, gamma (white) and annexin 2 (black).
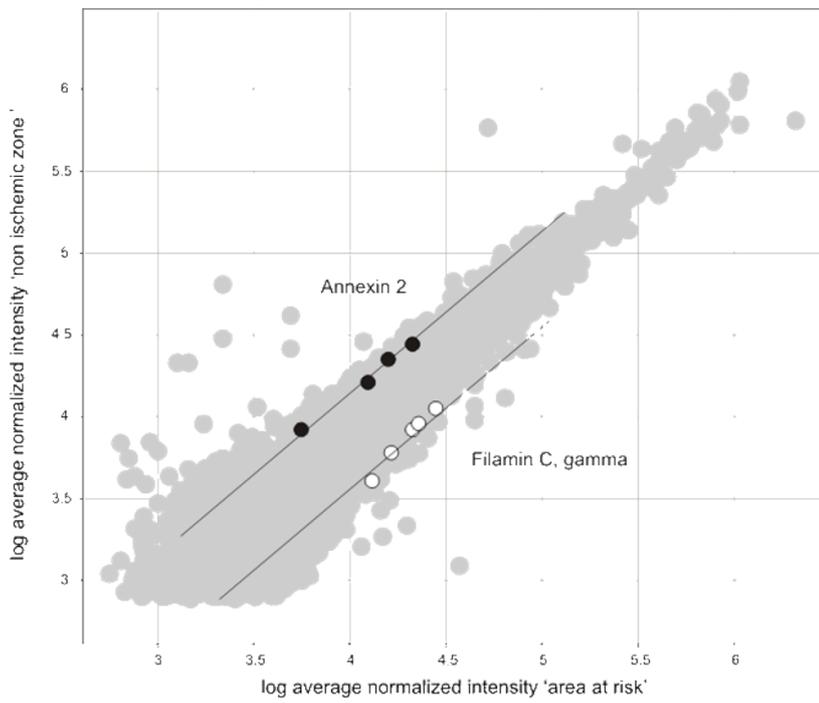
Table 1

Average ion detection statistics for three technical replicate experiments per sample type.

	rabbit		human			mouse	rat
	at risk	non ischemic	heart tissue	breast tissue	plasma		
# ion detections above detection threshold[1]	94,470	86,661	82,680	81,289	70,654	109,990	97,538
# ion detections above identification threshold[2]	66,129	61,901	61,310	59,028	56,588	74,032	66,938
% total number	70	71	74	73	80	67	69
% total detected intensity	89	92	90	92	89	90	92
# accurate mass - retention time clusters[3]	18,894	17,686	17,517	16,865	16,168	21,152	19,125
maximum cluster intensity	9.50E+05	1.12E+06	2.20E+06	1.80E+06	4.21E+06	1.82E+06	1.75E+06
ion detection dynamic range	3.4	3.4	3.4	3.6	3.6	3.6	3.6
effective dynamic range[4]	2.6	2.6	2.7	2.8	2.7	2.8	2.8
# clusters in mass sufficient regions	17,912	17,085	16,203	16,359	15,360	20,771	18,494
% clusters in mass sufficient regions	95	97	93	97	95	98	97

[1] all ion detections above processing threshold prior to deconvolution

[2] ions illustrating an intensity value above a minimum required for deconvolution and charge state reduction

[3] number of deconvoluted de-isotoped and charge state reduced precursor ion

[4] effective dynamic range of the identified peptide precursors

Table 2

Comparison of the distribution of replicating ion detections vs . peptide precursor identifications throughout the detection dynamic range of an LC-MS experiment.

	rabbit		human			mouse	rat
	at risk	non ischemic	heart tissue	breast tissue	plasma		
# ions/unit time[1]							
fraction (contribution) 2 lowest orders (%)	81	83	81	75	81	72	82
fraction (contribution) 2 highest orders (%)	19	17	19	25	19	28	18
detected intensity/unit time[1]							
fraction (contribution) 2 lowest orders (%)	32	39	34	33	35	31	32
fraction (contribution) 2 highest orders (%)	68	61	66	67	65	69	68

[1] only ion detections replicating 3 out of 3 technical replicates were considered.

Table 3

Replication rates of deconvoluted monoisotopic precursor masses of the four investigated rabbit heart samples.

	at risk 1	at risk 2	non-ischemic 1	non-ischemic 2
3 out of 3 replicate experiments	9032	8958	8828	7971
2 out of 3 replicate experiments	3778	3274	3443	3355
1 out of 3 replicate experiments	6084	5454	5854	5921
replicating fraction	0.67	0.69	0.68	0.66

Table 4

Overview of the peptides identified[1] and deconvoluted (charge state reduced and de-isotoped) intensities to rabbit and human isocitrate dehydrogenase 2 and cysteine and glycine rich protein 3 (cardiac LIM protein) searched against a placental mammalian database and a translated genomic rabbit database[3] .

	isocitrate dehydrogenase 2			
	rabbit	human[2]	mouse	rat
LNEHFLNTTDFLDTIK	214,409	214,409 (20,020)	214,409	214,409
LVPGWTKPITIGR	208,809	208,809 (19,416)	208,809	208,809
DLAGCIHGLSNVK	143,269	143,269 (15,502)	143,269	143,269
YFDLGLPNR	142,912	142,912 (15,124)	142,912	142,912
VCVETVESGAMTK	137,794	137,794 (6737)		
NILGGTVFR	133,898	133,898 (16,340)	133,898	133,898
LDGNQDLIR	121,766	121,766 (4911)	121,766	
LIDDMVAQVLK	119,704	119,704 (5983)	119,704	119,704
ATDFVVDR	46,521		46,521	46,521
VSKPVVEMDGDEMTR[3]	23,278			
IIWQFIK	22,821	22,821	22,821	22,821
DIFQEIFDK	22,505	22,505	22,505	22,505
TDDQVTIDSALATQK[3]	19,209			
AQGSGPTWAPAALATPTLQEQPR[3]	17,161			

	cysteine and glycine rich protein 3 (cardiac LIM protein)			
	rabbit	human	mouse	rat
NFGPTGIGFGGLTQQVEK	296,219	296,219 (25,994)	296,219	296,219
ALDSTTVAAHESEIYCK	201,394	201,394 (9914)	201,394	201,394
SLESTNVTDKDGELYCK	192,561	192,561 (35,844)	192,561	192,561
KALDSTTVAAHESEIYCK[3]	84,251			
TVYHAEEIQCNGR[3]	78,985			

[1] minimum peptide identification probability of 95% and replication of 2 out of 3 technical replicates

[2] the values provided in parentheses are the intensities of the peptides identified to a human heart sample. Note the similarity in relative peptide precursor intensity order between the rabbit and human identifications

[3] complementary peptide identification results by searching the data against a translated genomic rabbit database