Serum profiling and biomarker discover of rat mammary tumors using mass-coded abundance tags (MCAT)

Craig Rowell[1], Glen Puckett[1], Kevin Raorty[1], Marion Kirk[2], Landon Wilson[2], Sam Wang[2], Mark Carpenter[3] and Coral A. Lamartiniere[1]

[1]Department of Pharmacology and Toxicology and [2]Cancer Center Mass Spectrometry Shared Facility, University of Alabama at Birmingham, Birmingham, AL 35294

[3]Department of Discrete and Statistical Sciences, Auburn University, Auburn AL 36849

ABSTRACT

Advances in Mass-spectrometry techniques allow for the rapid processing and evaluation of complex biological mixtures such as blood/serum. These samples represent a protein rich environment as well as a sentinel monitoring system of the entire organism. The central tenet of these studies is that changes in the microenvironment of a tissue, brought about by a disease process, will lead to sufficient changes in the protein and peptide pattern of the serum, such that the differences can be accurately detected and correctly associated with a particular disease state. Using mass-spectrometry approaches we have developed techniques that allow us to compare samples from tumor-free and tumor present serum samples simultaneously to find biomarkers that indicate the presence of cancer. To examine potentially important but less abundant proteins, ultrafiltration (UF) was used to eliminate the more abundant proteins and combine this with the non-isotopic peptide tags (*S*-methylthioacetimidate and *S*-methyl thiopropionimidate) described by Beardsley and Reilley (J. Proteome Res. 2: 15-21, 2003) to differentiate our samples. Use of these mass-coded abundance tags (MCAT) allows for simultaneous evaluation of serum samples from tumor present, and tumor free animals. Using an oa time-of-flight mass-spectrometer (Q-tof) with electrospray ionization we produce high quality spectrums to screen for peptides that have only one tag. Specificity of tagging increases the likelihood that the peptide resulted from a protein unique to either the control or conditioned state. Using the ms/ms function of the Q-tof we sequence the peptide and identify the parent protein. Specifically, our lab is using UF, MCAT and the Q-tof to evaluate rat models of chemically-induced tumors. By using animal models we overcome much of the variability that may exist in human serum samples due to differences in gender, diet and cancer initiation. We have shown that these systems allow for the identification of both small molecules such as Alpha S1 casein precursor (24 kDa) as well as proteins greater than the MCO such as Fibrinogen alpha/alpha E precursor and Coagulation factor 2 (86 and 70 kDa, respectively). With positive sequence identification we can now evaluate the tumors themselves to determine if the proteins are over-expressed in the tumor vs. normal tissues. Using this method of " bottom-up" analysis provides information on the nature and composition of our samples to more rapidly identify those proteins that are unique to the tumor state of the animals.

INTRODUCTION

Serum represents a rich repository of information about the body in terms of biomarkers and a readily accessible fluid for collection. The difficulty of using serum for biomarker analysis lies in the high abundance of relatively few proteins. Anderson has published that 99% of the protein concentration in serum is derived from 22 proteins[1] (Figure 1). These proteins cover a large dynamic range. Therefore, in order to screen the serum for unique proteins related to a developing biological alteration (such as tumor growth) we must have mechanisms to enrich for the low abundance proteins, quickly associate their presence with the disease state, and identify them to determine their role in the disease process.

To date, the majority of serum studies to date have used human samples. The reasons are that these samples are characterized (type and stage of cancer as well as information about the patient) and relatively large data sets can be acquired. The underlying problem with these projects is the inherent complexity introduced by looking at human populations. Therefore, the development of analytical technologies is confounded by having to evaluate both the inherent complexity of the analyte and the extrinsic factors associated with clinical samples.

Our goal is to create a system that reduces the complexity of the analyte. We are accomplishing this in two ways 1) using a rat model, we can control the day to day variances of diet, light cycle, temperature etc., as well knowing the cause of the tumorogenesis and the date of induction, and 2) simplify the analyte by eliminating most of the high abundant proteins using ultrafiltration (UF). For this, we are applying some of the fractionation procedures used in MUD-PIT studies (i.e tryptic digestion and multi-phasic HPLC). The result is a sample that is a mixture of all the peptides produced by both endogenous cleavage events as well as the *in vitro* tryptic fragmentation, but will lack an abundance of peptides from the most prevalent proteins in the serum.

The resulting tryptic fragmentation pattern (an N-terminus and a C-terminus with either a lysine group or an arginine group) is then used to enhance the analysis by treating the sample with one of the two MCAT reagents (*S*-methylthioacetimidate (S-MTA): +41Da or *S*-methyl thiopropionimidate (S-MTP): +55Da)[2]. These reagents undergo amindination reactions with all lysine residues and the N-terminus. Therefore, each condition receives one tag (e.g. tumor samples are tagged with S-MTA and control are tagged with S-MTP). Hence, we can then evaluate both the tumor and tumor-free serum simultaneously. Particularly, for MALDI-TOF analysis this allows us to overcome several of the variables involved with comparisons of different spots.

We propose that by using a controlled environment as well as micro-scale fractionation procedures we can readily apply the use of mass-coded abundance tags (MCAT' s) coupled with mass-spectrometry to rapidly assess biomarkers and discover circulating proteins that indicate the presence of tumors.

METHODS AND MATERIALS:

Female Sprague Dawley rats were treated with NMU (50 mg/kg BW) at 50 days of age then palpated twice a week to monitor tumor development. After tumors had developed the animals were killed, whole trunk blood was collected and serum was separated and stored at -80C. Serum samples were then treated similarly to the protocol described by Radhakrishna et al[3] (Figure 2). Briefly, samples were processed by ultrafiltration using a 30 kDA molecular weigh cut-off. The collected fractions were then denatured, reduced and alkylated prior to digestion with trypsin. Following digestion samples were checked on the MALDI-TOF to ensure complete digestion and evaluate sample recovery (Figure 3).

Micro-column construction

Each sample required its own micro-column (Figure 4) to ensure that there were no carry-over peptides between samples. Each column consisted of fused silica capillary of approximately 11 cm (180 um i.d.) attached via a PEEK sleeve to a 4 cm piece of fused silica (250 um i.d.). The 11 cm sections were packed with Jupiter RP material to a length of 1 cm. The second section consisted of 2 cm of PolyA (SCX) and the third section was packed with 2 cm of Jupiter RP beads.

HPLC

Using a Waters 660E pump and switching system the following conditions were used to for peptide separation. All fractions were collected into cells of a 96 well plate with conical bottom to ensure greater peptide recovery.

Fraction segments were:
1) 100% ACN
2) 25% KCL followed by 100% ACN
3) 50% KCL followed by 100% ACN
4) 75% KCL followed by 100% ACN
5) 100% KCL followed by 100% ACN

After collected fractions were lyophilized, they were rehydrated in a solution containing either SMTA or SMTP (Figure 6) and allowed to sit at room temperature overnight to allow complete tagging. Tagged samples were then lyophilized and resuspended in 0.5% FA and applied to a 96 well Montage plate (C18 Zip Plate, Millipore). Samples were eluted from the plate and collected in a 96 well plate and once again lyophilized. Then these samples are ready for mass-spectrometry.

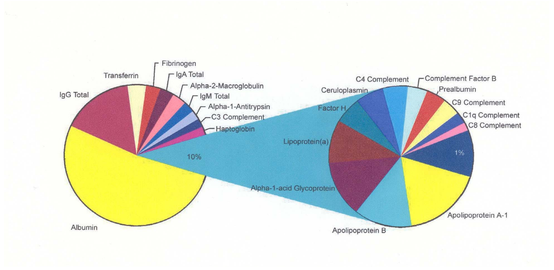
MALDI-TOF

Using a Voyager Elite the samples were analyzed as described by Beardsley & Reilly[2]. Briefly, samples were resuspended in 1% FA and mixed 1:1 with DHB matrix. Each spotted sample consisted of 1 ul of S-MTA and 1 ul S-MTP labeled fractions. Using the peptide reflector in the positive mode, 100 shots per sample were collected.

Analysis

Spectrums from each collected spot were exported to Excel (Table 1). These excel sheets were then imported into SAS. Presently, we are designing a search algorithm to bin all peaks + or – 1da, then to search for differences in peak of +14, +28, +42, +55, +85 and +110 daltons to locate peaks of interest, including those that are similar to each other.

Predominant Serum Proteins



90% of serum protein concentration is made by 10 proteins
9% is made of 12 proteins
1% is made of all other proteins and peptides.

Figure 1. This chart shows the 22 most abundant proteins found in serum. Taken together they make up 99% of the serum protein concentration.

Serum separation and peptide treatment protocol

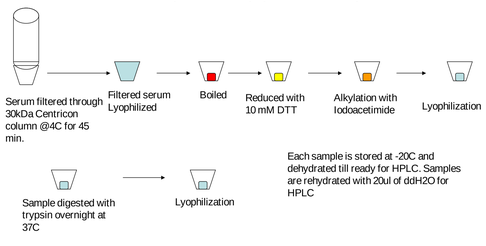


Figure 2.Workflow of the sample preparation steps as described by Radhakrishna et al. Ultrafiltration of the serum reduces the complexity of the sample by removing predominant proteins, primarily albumin. Tryptic fragmentation allows for MCAT additions and the fragment sequence information can be used for protein identification.

MALDI-TOF spectrum of UF samples from control and tumor serum before HPLC and tagging

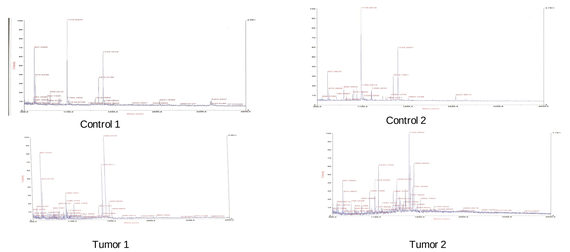


Figure 3. These MALDI-TOF spectra illustrate the need for separation methods prior to mass spectrometry. For the various sample types there are a finite number of peaks that can be visualized at any time. To increase the number of peaks available for evaluation we are developing HPLC fraction methods.

Micro-scale tri-phasic column for serum peptide separation

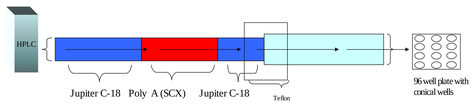


Figure 4. This is a schematic of the micro-scale HPLC columns constructed for peptide separation. The nature of the column allows us to use the inherent chemistry of the peptides to get better separation and reduce handling, therefore, our overall recovery should improve.

Mass-coded Abundance Tags (MCAT)

- Peptides produced from tryptic digestion are tagged with one of two reagent tags.
- Each tag will attach to all lysine amino acids or to any free N-terminus
- QUEST (Quantitation Using Enhanced Signal Tags) allows peptides from different sample treatments or conditions to be evaluated simultaneously.
- Simultaneous evaluation allows for rapid identification of unique peptides.

1) S-methylthioacetimidate (S-MTA): +41Da; *2) S*-methyl thiopropionimidate (S-MTP): +55Da

Post-HPLC processing for MCAT addition and preparation for mass-spectrometry

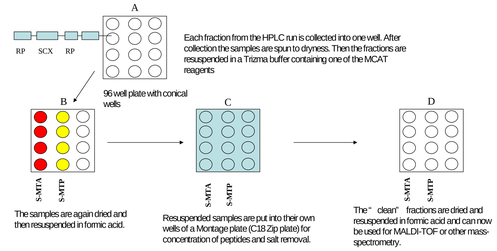


Figure 6. Outline of the treatment of the samples after they have been fractionated by HPLC. These steps add the MCAT reagents as well as prepare the samples for mass-spectrometry.

MALDI-TOF spectrum of tagged peptides from serum

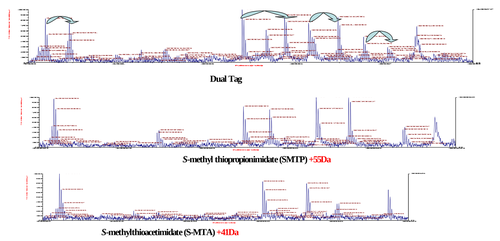


Figure 7. Appearance of MALDI-TOF spectrum during a run. The dual sample spectrum is the result of a spot containing samples of each tag. Arrows indicate peaks that are the same peptide in each sample, but vary by mass due to the tag addition.

Peak list from individual samples.

55 tag			41 tag				
Peak Number	Mass	RI	Mass	RI	Theoretical Mass additions (Da)		
					14 Da	28 Da	55 Da
1	959.78	11.26	945.65	14.22	959.65	973.65	1000.65
2	1032.86	17.52	1018.72	19.26	1032.72	1046.72	1073.72
3	1033.85	9.54	1019.73	14.33	1033.73	1047.73	1074.73
4	1072.91	22.16	1058.78	23.41	1072.78	1086.78	1113.78
5	1073.90	14.05	1059.77	14.60	1073.77	1087.77	1114.77
6	1115.93	8.26	1101.81	9.26	1115.81	1129.81	1156.81
7	1161.95	9.06	1102.82	9.12	1116.82	1130.82	1157.82
8	1235.96	8.66	1147.77	11.85	1161.77	1175.77	1202.77
9	1236.97	9.26	1148.78	8.38	1162.78	1176.78	1203.78
10	1268.96	8.58	1221.81	10.77	1235.81	1249.81	1276.81
11	1274.99	9.86	1222.81	8.41	1236.81	1250.81	1277.81
12	1358.88	18.53	1239.83	9.25	1253.83	1267.83	1294.83
13	1359.87	14.76	1240.78	8.23	1254.78	1268.78	1295.78
14	1375.04	9.51	1246.79	9.27	1260.79	1274.79	1301.79

Table 1. Peaks generated from samples tagged with either S-MTA or S-MTP. Calculations are made to look for specific mass differences between peaks based on multiples of the mass difference of the tags as well as to look for missed tags. Peak matching allows for rapid localization of peaks unique to one of the conditions. Peaks are color coded and three of the theoretical additions to the peptides are shown. A shift of 14 Da indicates that the peptide only had one molecule of each tag added and a shift of 28 Da would indicate that two of the molecules were added to the peptide.

MS/MS spectrum from a Micromass Q-tof.

Peptide	Parent Protein
FPNAEFAEITK	Rat Albumin
IMSMIDEAK	Apolipoprotein A-1 precursor
ATIDQNLEDLR	Apolipoprotein A-IV precursor
LGPLVEQGR	Apolipoprotein E precursor
AFSPVASVESASGEVLHSPK	Rat Fetuin precursor
LTAQPAPSPEDLALSMGTIK	Alpha I inhibitor III precursor

Table 1. Peptides produced from whole serum digested with trypsin and incubated with S-methylthiacetimidate. Red letters indicate the sites of tag addition. Tags add only on the N-terminus and on all Lysine residues (K). Peptides were then sequenced using a Q-TOF mass spectrometer (Micromass). The N-terminus labeling is non-amino acid specific. However, the labeling will not add to the C-terminus when Arginine (R) is present as a result of the tryptic digestion.

From Ultrafiltration trial (this sample was not modified)

Peptide	Parent Protein
ADTGTTSEFIEAGGDI	Fibrinogen alpha/alpha E precursor*
YQNFDPEVKLV	Coagulation factor 2 Rat*
FPVAPFPEVEGK	Alpha S1 casein precursor bovine**
YLGYLEQLLR	Alpha S1 casein precursor bovine**

*The average mass of these intact proteins is 86Kda and 70Kda, respectively. This demonstrates that we can detect the presence of proteins outside the MW cut-off of the ultrafiltration probe.**The average mass of this intact protein is 24Kda. This demonstrates that proteins of the appropriate MW may enter the probe.

Figure 8. Unique MALDI-TOF peaks can be evaluated by CID ms/ms to provide information on the amino acid sequence of the peptide as well as detect the number of sites of tag addition. The peptide sequence is then used to get protein identification as shown in the lists on the right.

DISCUSSION

Advances in mass spectrometry are allowing the evaluation of more complex samples. Therefore, several labs have undertaken the goal of screening serum for biomarkers of disease states/processes. Current literature points to the fact that only about 1500 proteins have been identified in serum by different technologies. This seemingly small number demonstrates the need for improved separation technologies in order to ensure that the broadest number of proteins is studied.

Many studies (including our initial attempts here) have focused on looking at presence/absence of proteins for an indication of biological significance, purposefully avoiding quantitative analysis. This avoidance may, impart, be due to mass spectrometry not being used accurately for quantitation. However, advances such as the development of MCAT reagents should allow for this type of analysis to be done reliably.

Finally, it is important to remember that the data provided by these experiments is a direct reflection of the study design. Therefore, high dimensional project development will be open to several sources of variation specifically because the instrumentation is so sensitive. So, we suggest that the developmental projects should be conducted on more rigorously controlled samples (i.e. animal models) until the technical variability can be established and corrected or modeled.

REFERENCES

1. N. L. Anderson and N. G. Anderson. The human plasma proteome: History, character, and diagnostic prospects. *Mol Cell Proteomics* 1, 845-67 (2002).
2. Beardsley, R.L. & Reilly, J.P. Quantitation using enhanced signal tags: a technique for comparative proteomics. *J Proteome Res* 2, 15-21 (2003).
3. Radhakrishna S. Tirumalai, King C. Chan, DaRue A. Prieto, Haleem J. Issaq, Thomas P. Conrads, and Timothy D. Veenstra. Characterization of the Low Molecular Weight Human Serum Proteome. *Mol. Cell. Proteomics* 2, 1096 – 1103 (2003).

ACKNOWLEDGEMENTS

This research was supported by DOD BC 17-00-1-0119, Susan Komen DISS0201242, NIH 1 R01 CA61742-07 and DOD BC 021371. The mass spectrometers were purchased with NIH/NCRR Shared Instrumentation grants (S10 RR11329 and S10 RRRR13795) and support from the UAB Health Services Foundation-General Endowment Fund. The operation of the UAB Mass Spectrometry Shared Facility comes from the NCI Core Grant (P30 CA13148) to the UAB Comprehensive Cancer Center.