

Identification and Differential Expression Analysis of Putative and Known Plasma Biomarkers for Human Hypertrophic Cardiomyopathy

Tori Richmond¹, Manor Askenazi¹, Leo Bonilla¹, Robert E. Gerszten², Jennifer Sutton¹, Emerson Liu², David Sarracino³

¹Biomarker Research Initiatives in Mass Spectrometry (BRIMS) Center, Thermo Fisher Scientific, Boston, MA; ²Cardiology Division and Center for Immunology & Inflammatory Diseases, Massachusetts General Hospital and Harvard Medical School, Boston, MA; ³Harvard-Partners Center for Genetics and Genomics, Cambridge, MA

Introduction

Myocardial infarction (MI) is a major cause of morbidity and mortality worldwide, with approximately 500,000-700,000 deaths caused by ischemic heart disease in the United States each year. The causes of the disease have been well-established, but the build-up to a massive heart attack is difficult to predict, particularly during early stages of disease progression when prevention is still possible. Proteomics – both large-scale protein identification and quantitative differential expression – may be used to identify improved diagnostic markers and contribute to our understanding of biochemical pathways triggered during myocardial injury that might one day serve as targets for drug therapy.

In this study, samples were obtained in a time-course fashion from patients undergoing alcohol septal ablation for hypertrophic cardiomyopathy (a “planned” myocardial infarction) and analyzed in triplicate to track changes in their protein profile during the course of treatment. Bottom-up proteomics analysis was accomplished in two steps, namely (1) off-line fractionation by strong anion exchange chromatography (SAX) of plasma proteins in order to circumvent the need for depletion of high-abundance proteins (e.g., albumin, IgG), followed by (2) replicate bottom-up analysis of digested fractions, by nanospray LC/MSⁿ with a Thermo Scientific LTQ linear ion trap mass spectrometer, to assess semi-quantitative changes in the protein profile over the course of the planned MI.

Goal

Demonstrate the use of a specific sample handling and pre-fractionation protocol combined with on-line nanoflow LC/MSⁿ with a high-performance linear ion trap mass spectrometer (LTQ) for in-depth identification and semi-quantitative analysis of both putative and known plasma protein biomarkers of human MI.

Experimental

Sample Collection

Samples were obtained from patients undergoing alcohol septal ablation for hypertrophic cardiomyopathy, a “planned” myocardial infarction. Blood was drawn at indicated time points from central catheters into citrated tubes (Becton Dickinson, 0.105 M buffered sodium citrate, whole blood ratio 1:9) and centrifuged at 2000×g for 10 min. Plasma was aliquoted and immediately frozen and preserved at –80 °C until use.

Pre-fractionation

Samples of plasma (1.6 mL) were thawed on ice and diluted (1:10) 1 hour prior to injection in 80 mL of 15% n-propanol, 25 mM ammonium acetate pH 8.8. Prior to loading, diluted samples were spun at 3000×g for 10 min to remove any particulates. The samples were loaded in a 150 mL sample loop, and an additional 36 mL of Solvent A (see below) was used to rinse out the sample loop and rinse off unbound proteins from the column prior to starting the gradient. A polymeric strong anion exchange column was used. A gradient was formed with (Solvent A) 25 mM ammonium acetate 15% n-propanol(v/v) pH 8.8, and (Solvent B) 2 M ammonium acetate 15% n-propanol(v/v) pH 8.8, at a flow rate of 0.8 mL/min.

The pre-fractionation gradient is listed in the following table:

Time (Minutes)	%B
0	0
650	17.5
800	50
900	100
1000	100
1000.1	0
1100.1	0

Column Regeneration

The column was regenerated at a flow rate of 1 mL/min with 10 column volumes of each of the following: acetonitrile, 0.1% formic acid; 40% isopropanol, 15% acetic acid; 40% isopropanol, 0.1 N NaOH; 2 M ammonium acetate, 15% n-propanol pH 8.8; 80 mL 25 mM ammonium acetate, pH 8.8 15% n-propanol.

Fraction Collection

Eight mL fractions were collected into 15 mL polypropylene tubes, frozen to –80 °C and lyophilized to dryness using a Savant SSC250 SpeedVac[™] with a –110 °C trap and high vacuum pump.

Digestion

Lyophilized fractions were reconstituted in 500 µL of 6 M urea, 100 mM TRIS (pH 7.8). A 100 µL aliquot was taken for enzyme digestion according to published procedures (Kinter et al)¹. Following digestion, samples were reconstituted to a final sample volume of 500 µL prior to injection and analysis by LC-MS/MS.

Key Words

- LTQ[™]
- Cardiac Disease
- Differential Expression
- Human Plasma
- Protein Fractionation

HPLC Separation

Ten μL fractions were loaded and desalted for 10 min on a trapping column at 10 $\mu\text{L}/\text{min}$. Peptides were separated using a handpacked, 75 $\mu\text{m} \times 100 \text{ mm}$ column, 200 \AA , 5 μm particle size, at a flow rate of 100 $\mu\text{L}/\text{min}$ pre-split; 350 nL/min post-split.

The gradient was formed with the following mobile phases: A – Water w/0.1% Formic Acid; B – Acetonitrile w/0.1% Formic Acid.

The chromatographic gradient is listed in the following table:

Time (Minutes)	%B
0	5
240	60
241	80
265	80
266	5
285	5

Mass Spectrometry

An LTQ linear ion trap was operated in positive ion nanospray mode, using a Data Dependent™ Triple Play method as follows:

- 1: Full scan MS: 400–1800 m/z (survey scan)
- 2: ZoomScan™ MS of most intense ion from survey scan (high-resolution scan)
- 3: MS/MS of most intense precursor ion from survey scan

Dynamic Exclusion Settings

Repeat Count: 2

Repeat Duration: 30s

Exclusion List Size: 200

Exclusion Duration: 120s

Data Dependent Settings

Min. Signal Required: 1000

MS² Isolation Width: 2 m/z

MS² Normalized Collision Energy™: 30%

Data Analysis

All time points and fractions were analyzed in triplicate and their semi-quantitative changes determined using prototype differential expression software, including a pipeline of automated data analysis with chromatographic alignment and relative quantification of LC/MS features without the use of stable-isotope labels. In this pipeline, identification via database (IPI human, indexed for tryptic cleavages, M+16) searching using SEQUEST® with aggressive filtering of the results preceded alignment and calculation of peak areas from raw MS signals. Relative quantification was performed only on those precursor ions (MS) which produced ZoomScans and MS/MS spectra of sufficient quality to yield positive protein IDs.

Results

Figure 1 shows the local chromatographic environment of fractions G4 and G5 from the SAX profile across the different time points (0, 1, and 24 hours). Note the time-dependent gross changes in profile in and around fractions G4 and G5, especially 24 hours after the initiation of the septal ablation procedure.

A meta-consensus summary report was generated to tabulate SEQUEST scores and the qualities of related ZoomScans for identified peptides and proteins. In nine analyses of fraction G4 (3 time points \times 3 replicates per time point), approximately 32 proteins were confidently identified using a criteria of two or more distinct peptide sequences with an Xcorr > 2.5. Among the proteins identified were two splice isoforms of C-reactive protein (shown in Figure 2), a known marker of myocardial injury. Figure 3 shows the sequence alignment of both CRP isoforms.

In the data analysis pipeline used here, several internal consistency checks were applied both to the raw data, as well as to the results from the database search. One of the QC tests applied to the raw data prior to database searching involves a scoring of the goodness of fit between the theoretical and the experimental isotope distribution of the precursor ion using the ZoomScan profiles from each triple play event. This score, labeled zQual in Figure 2, is bound between the values of 0 (no match) and 1 (perfect match). For the purpose of our experiments, only MS/MS spectra with an associated zQual value >0.65 were used for database searching. As a result, of the approximately 13,000 triple plays generated for each sample analysis, only half of the ZoomScan data and their corresponding MS/MS spectra were searched.

Two CRP peptides, (K)RQDNEILIFWSK(D) and (R)GYSIFS YATK(R), were monitored to examine differential expression levels between the time points. Automated chromatographic alignment of the RICs was performed within a 3 m/z unit-wide window around the doubly-charged precursor ion signal in the MS scan. Following alignment, peak areas were calculated and averaged for the three technical replicates. These average areas were then normalized to the T_0 average value and their ratios computed between each time point.

Figure 4 shows the change in CRP expression across the three time points, T_0 , T_1 , and T_{24} as calculated by the response in the raw MS signal from the (K)RQDNEILIFWSK(D) peptide. The expression ratio $T_0:T_{24}$ is 1:2.2 (t-test value of 0.001 for a statistically significant difference). This ratio is in good agreement with recent reports in the literature for the kinetics of release of CRP following myocardial injury²⁻⁵.

In order to arrive at a global view of the changes in expression profiles (i.e., up- and down-regulation of protein ‘families’), we devised an alternative procedure that extracts and groups differentially expressed gene ontology (GO) annotations for all the proteins identified



Figure 1: SAX plasma fractionation, showing significant change in the 24 hour timepoint sample.

g|IPI00022389.1|rs|sp|P02741-1|Splice (71 Peptides, 7 distinct) avgXC_{corr}=3.30 [9 of 1767]
 g|IPI00022389.1|rs|sp|P02741-1|Splice_isoform_1_of_P02741_C-reactive_protein_precursor|mass:25039|Human [IPI] [SwissProt]

Scan	RT	M+H	Pep_M+H	Δ M+H	z	zQual	Xo	Δ Xo	Sp	Ion%	Other	Rank	NumTryp	ParDemt	Δ ACN	Peptide
14514	93.60	2781.37	2780.46	0.91	3	0.73	5.07	0.00	3388	34	2	1	2	1	4.44	K_APLTKPLKAFVLCVLFHYTELSSTR. G [8]
12657	83.23	1549.19	1548.82	0.37	2	0.94	4.99	0.00	2132	82	0	1	2	0	5.67	K_RQDNELIFWSK. D [29]
14643	92.07	1393.63	1392.72	0.91	2	0.91	3.94	0.00	665	57	0	1	2	0	7.80	R_QDNELIFWSK. D [5]
14130	87.67	1821.49	1820.92	0.57	2	0.96	3.23	0.00	550	50	2	1	1	0	8.00	K_YEVQGEVFTKPLQWLP. - [12]
09993	67.92	1136.68	1136.56	0.11	2	0.90	3.13	0.00	483	59	0	1	2	0	9.28	R_GYSIFSYATK. R [15]
13518	86.04	1457.15	1456.77	0.38	2	0.88	3.01	0.00	381	44	1	1	1	0	9.39	Y_LGGPFSPNLNWR. A [1]
13653	88.83	1932.79	1931.91	0.88	3	0.67	2.31	0.00	1069	40	2	1	2	0	8.69	K_AFTVLCVLFHYTELSSTR. G [3]

g|IPI00181794.1|rs|NP_000558|sp|C-reactive (27 Peptides, 5 distinct) avgXC_{corr}=2.85 [15 of 1767]
 g|IPI00181794.1|rs|NP_000558|sp|C-reactive_protein_pentaxin-related|mass:25107|Human [IPI] [RefSeq]

Scan	RT	M+H	Pep_M+H	Δ M+H	z	zQual	Xo	Δ Xo	Sp	Ion%	Other	Rank	NumTryp	ParDemt	Δ ACN	Peptide
14514	93.60	2781.37	2780.46	0.91	3	0.73	5.07	0.00	3388	34	2	1	2	1	4.44	K_APLTKPLKAFVLCVLFHYTELSSTR. G [8]
14130	87.67	1821.49	1820.92	0.57	2	0.96	3.23	0.00	550	50	2	1	1	0	8.00	K_YEVQGEVFTKPLQWLP. - [12]
13518	86.04	1457.15	1456.77	0.38	2	0.88	3.01	0.00	381	44	1	1	1	0	9.39	Y_LGGPFSPNLNWR. A [1]
13653	88.83	1932.79	1931.91	0.88	3	0.67	2.31	0.00	1069	40	2	1	2	0	8.69	K_AFTVLCVLFHYTELSSTR. G [3]
14157	86.90	2101.40	2103.07	-1.67	3	0.78	2.19	0.00	489	31	0	1	2	2.3	3.25	R_MPPPRDKTMRFFIFWSK. D [3]
14157	86.90	2101.40	2103.07	-1.67	3	0.78	2.12	0.08	510	32	0	2	2	2.3	3.25	R_MPPPRDKTMRFFIFWSK. D [2]

Figure 2: Metaconsensus report showing database identification of two CRP splice isoforms in fractions G4 and G5.

```

MEKLLCFLVL TSLSHAFGQT DMSRKAFVFP KESDTSYVSL KAPLTKPLKA FTVCLHFYTE 60
MEKLLCFLVL TSLSHAFGQT DMSRKAFVFP KESDTSYVSL KAPLTKPLKA FTVCLHFYTE 60

LSSTRGYSIF SYATKRQDNE ILIFWSKDIG YSFTVGGSEI LFEVPEVTVV PVHICTSWES 120
LSSTRGTVFS RMPPRDKTMR FFIWFSKDIG YSFTVGGSEI LFEVPEVTVV PVHICTSWES 120

ASGIVEFWVD GKPRVRKSLK KGYTVGAEAS IILGQEQDSF GGNFEQSLSL VGDIGNVMNW 180
ASGIVEFWVD GKPRVRKSLK KGYTVGAEAS IILGQEQDSF GGNFEQSLSL VGDIGNVMNW 180

DFVLSPEIN TIYLGGPFSP NVLNWRALKY EVQGEVFTKP QLWP 224
DFVLSPEIN TIYLGGPFSP NVLNWRALKY EVQGEVFTKP QLWP 224

SPLICE ISOFORM 1 OF P02741 C-REACTIVE PROTEIN PRECURSOR
C-REACTIVE PROTEIN, PENTRAXIN-RELATED
  
```

Figure 3: Splice isoforms of C-Reactive Protein identified in fractions G4 and G5. Amino acid coverage is shown in red (33% for splice isoform and 30% for CRP)

in this experiment (Figure 5). [Only MS/MS spectra with an associated zQual >0.65 were used for database searching against the human IPI database. Furthermore, only protein IDs derived from proteolytic peptides which exceeded set charge state vs. Xcorr thresholds (+1, >1.8; +2, >2.5; +3, >3.0) were used to extract differential GO annotations.]

Differential GO annotations for individual proteins were calculated by counting the frequency of successful sequencing events for all its peptide sequences per time point, and comparing this frequency across all three time points in the experiment (T₀, T₁ and T₂₄ hours).

For a given GO annotation (e.g., “acute phase response”), Fisher’s exact test was performed to determine whether there was a significant imbalance in the number of fragment detections having this annotation. The negative base10 log of the test’s p-value was taken to generate the graph in Figure 5. In this figure, each bar represents a broad (involving at least ten proteins) and differentially expressed GO annotation category. The length of each GO bar represents the negative log of the test p-value, and the direction of the bar represents increase (red) vs. decrease (blue) at T₂₄ hr vs T₀ hr.

If the imbalance in detection leaned towards the T₂₄ time point (i.e., an “upregulation”) the number remained positive; otherwise, the value was multiplied by -1. In this way, the magnitude of the number represents improbability, but the sign of the number represents the direction in which the change occurred.

Conclusion

Preliminary results from this model allowed the quantification and identification of known biomarkers of cardiac disease (e.g., C-reactive protein, CRP) as well as several biologically relevant low-level soluble factors with potential applications as biomarkers. Relative quantification was accomplished using label-free approaches. No isotope tags or derivatization chemistry was used as quantification was based on relative MS peak area of peptides identified by their CID spectra.

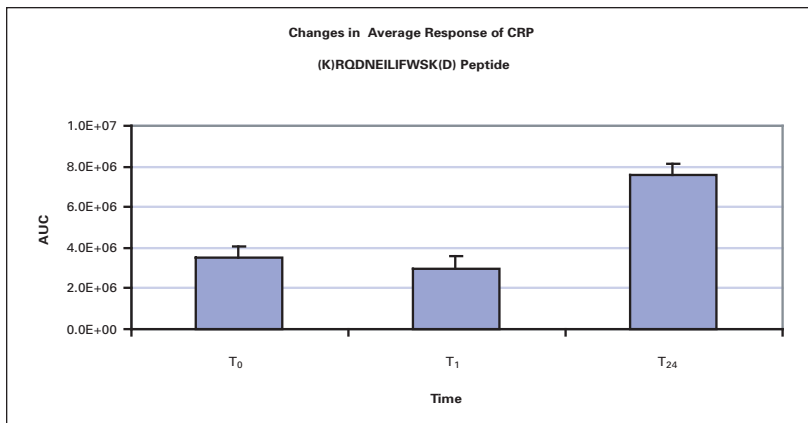


Figure 4: Time-course variation in the expression levels of CRP during septal ablation procedure as measured by the raw MS signal of the (K)RQDNEILIFWSK(D) peptide.

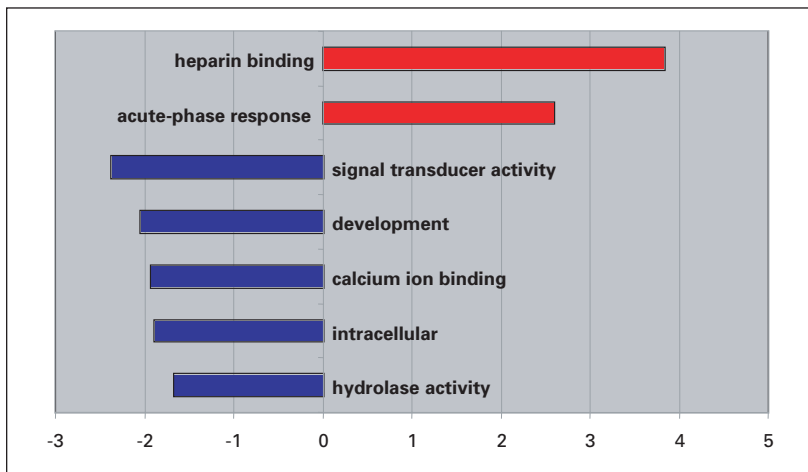


Figure 5: Differential Gene Ontology Expression Plot Between T₀ and T₂₄.

This workflow involved the analysis of a highly complex protein digest and was dependent upon an LC/MS methodology that achieves the best possible peptide detection and analysis, in order to detect important, low-level protein components. In addition to this, many of the peaks used for quantification were small, but accurate comparisons were possible due to the large number of data points defining these peaks. The LTQ acquires five MS or MS/MS spectra/second, ensuring the best possible data for accurate analytical determinations such as these.

Up-Regulation of CRP, as measured by a 2.2-fold increase in its raw MS signal (concentration) at T₂₄ vs T₀, is consistent with the observed general elevation of acute-phase response factors of which it is a part.

Future work will involve validation of unique putative markers that were also identified in this analysis.

References

1. Kinter, M. and Sherman, N. *Protein Sequencing and Identification Using Tandem Mass Spectrometry*. Wiley-Interscience Series on Mass Spectrometry, Series Editors Dominic Desiderio and Nico M.M. Nibbering, New York, (2000), pp 162-163
2. *Am J Cardiol*. 2004; 93: 822-825
3. *Chest* 2004; 125: 384-389
4. *Act Cardiol*. 2004 Jun; 59 (3): 291-296
5. *Cerebrovasc Dis*. 2004 Jul 23; 18 (3): 214-219

Legal Notices

©2007 Thermo Fisher Scientific Inc. All rights reserved. SEQUEST is a registered trademark of the University of Washington. All other trademarks are the property of Thermo Fisher Scientific Inc. and its subsidiaries. This information is presented as an example of the capabilities of Thermo Fisher Scientific Inc. products. It is not intended to encourage use of these products in any manners that might infringe the intellectual property rights of others. Specifications, terms and pricing are subject to change. Not all products are available in all countries. Please consult your local sales representative for details.

View additional Thermo Scientific LC/MS application notes at: www.thermo.com/appnotes

In addition to these offices, Thermo Fisher Scientific maintains a network of representative organizations throughout the world.

Africa
+43 1 333 5034 127

Australia
+61 2 8844 9500

Austria
+43 1 333 50340

Belgium
+32 2 482 30 30

Canada
+1 800 530 8447

China
+86 10 5850 3588

Denmark
+45 70 23 62 60

Europe-Other
+43 1 333 5034 127

France
+33 1 60 92 48 00

Germany
+49 6103 408 1014

India
+91 22 6742 9434

Italy
+39 02 950 591

Japan
+81 45 453 9100

Latin America
+1 608 276 5659

Middle East
+43 1 333 5034 127

Netherlands
+31 76 587 98 88

South Africa
+27 11 570 1840

Spain
+34 914 845 965

Sweden/Norway/Finland
+46 8 556 468 00

Switzerland
+41 61 48784 00

UK
+44 1442 233555

USA
+1 800 532 4752

www.thermo.com



Thermo Fisher Scientific, San Jose, CA USA is ISO Certified.

AN62500_E 11/07S