SCALABLE COMPUTING: PRACTICE AND EXPERIENCE
Volume 8, Number 2, pp. 173–183. http://www.scpe.org
ISSN 1895-1767

APPLICATION OF BIOINFORMATICS AND SCALABLE COMPUTING TO PERFORM PROTEOMIC ANALYSIS OF STOMACH TISSUE FROM DIABETIC MICE

EDWARD O. LIST[†], DARLENE E. BERRYMAN[‡], AMANDA J. PALMER[†‡], ELAHU GOSNEY[†], SHIGERU OKADA[†§], BRUCE KELDER[†], JENS LICHTENBERG[¶], LONNIE R. WELCH[¶] AND JOHN J. KOPCHICK[†‖**]

Abstract. Modern molecular biology experiments generate large amounts of data, which the biologists then have to analyze these manually through slow and error-prone processes. Bioinformatics and scalable computing provide essential tools for a speedup of the proteomics analysis. An example for such a proteomic analysis and the influences of bioinformatics is presented in this paper.

The stomach is a versatile organ with a prominent role in digestion and with endocrine function. While stomach has been the target of several proteomic analyses interested in cancer or ulcer biology, no studies have analyzed the proteome of stomach with the onset of diabetes, despite the fact that diabetes has a significant impact on gastric function. In this study, proteomic analyses were performed on stomach samples collected from C57BL/6J mice with obesity and diabetes and compared with samples from non-diabetic, lean controls. Obesity and diabetes were induced in mice by placing 3 week old mice on a high fat diet for 16 weeks, while control mice remained on a low fat standard chow diet. Once diabetes was established in the high fat fed mice, 4 diabetic and 4 control mice were sacrificed and stomachs removed for proteomic analysis using 2 dimension gel electrophoresis (2-DE). The protein "spots" that made up the stomach proteomic profiles were quantified using PDQuest 7.0.0 software. Protein spots found to be increased or decreased in diabetic stomach as compared to control stomach were removed from the gel for identification by database searches from peak lists generated by both MALDI-TOF and MS/MS analyses. In conclusion, a total of 23 proteins are reported herein with 14 being increased and 9 being decreased in diabetic stomach as compared to control.

Key words. diabetes, obesity, proteomics, stomach, scalable computing, bioinformatics

1. Introduction.

Proteomics is a vastly growing research area that creates large amounts of data and information that needs to be analyzed in order to map proteomes, their changing expression levels and the interactions of involved proteins. The involved tasks and computational efforts are presented in this paper under the background of the proteomic analysis of stomach tissue from obese and diabetic mice. A scalable computing approach is presented that integrates the bioinformatics tools into a common platform.

The stomach is a highly acidic environment that serves to break down food and is the site where protein digestion begins. In addition to its role in digestion, the stomach communicates with the brain to regulate energy balance and metabolism. The gastrointestinal tract produces more than 20 hormones and is the largest endocrine organ in the body [1]. Gastrointestinal regulation of energy balance is achieved by direct communication with the brain via the vagus nerve and by various endocrine hormones that are produced in the stomach and intestine. Ghrelin, for instance, is produced by the stomach and travels through the blood to the hypothalamus where it plays an important role in the regulation of appetite, regulates glucose metabolism, reduces fat utilization and possibly regulates energy expenditure (reviewed by [2]). The role of the stomach in regulation of energy balance could have significant implications for treatment of obesity and type 2 diabetes.

Obesity and type 2 diabetes mellitus are two highly interrelated conditions that have approached epidemic proportions in recent decades. The most recent available data show that approximately two thirds of US adults are now either overweight or obese [3] with 10% of adults having diabetes. The onset of type 2 diabetes, which is characterized by hyperglycemia due to extreme insulin resistance, is highly correlated with obesity with as many as 90% of type 2 diabetics being overweight or obese [4]. Alarmingly, another 54 million people in the US are considered to be in a pre-diabetic state according to recent reports from the US Department of Health and Human Services [4], implying the health ramifications of diabetes are likely to challenge the healthcare system for years to come.

Gastrointestinal disturbances are common in diabetes. As many as 83% of patients with diabetes report significant upper gastrointestinal problems [5]. Symptoms vary but often include abdominal pain or discomfort, bloating, early satiety, nausea, and vomiting. Many of the gastrointestinal problems are thought to be due to altered motility in the gastrointestinal tract. A common disturbance, gastroparesis, or delayed stomach emptying, is a common complication of type 2 diabetes impacting approximately 30-50% of individuals with long-standing diabetes [6]. Delayed gastric emptying also represents an important clinical complication as the digestion and absorption rate of carbohydrates can

[†]Edison Biotechnology Institute, Ohio University
[‡]School of Human and Consumer Sciences, Ohio University
[§]Department of Pediatrics, College of Osteopathic Medicine, Ohio University
[¶]School of Electrical Engineering and Computer Science, Ohio University
[‖]Department of Biomedical Sciences, College of Osteopathic Medicine, Ohio University
[**]John J. Kopchick, Ph.D., Edison Biotechnology Institute, Ohio University, 101 Konneker Research Laboratories, The Ridges, Athens, Ohio 45701, Tel: (740) 593-4534 Fax: (740) 593-4975 E-mail: kopchick@ohio.edu

impact the hyperglycemic condition. The causes of gastroparesis are not fully understood but both autonomic neuropathy [7] and glycemic control [8, 9] appear to be major contributing factors. Animal models of diabetes have proven valuable for studying the morphological and physiological changes that occur in the stomach. Although not always consistent, studies have demonstrated several morphological distinctions in nerve supply to the stomach [10] and several differences in protein expression [11]. Further studies are required to better elucidate which pathways and proteins contribute to the altered physiology of the diabetic stomach.

While several mouse models of diabetes are available, most of these diabetic mouse models are monogenic, meaning that their diabetic phenotype results from a single mutated gene. Since in humans, type 2 diabetes is mainly a polygenic disease that often is triggered by a combination of poor lifestyle choices, the use of such monogenic mouse models is limiting. On the other hand, C57BL/6J mice develop obesity, hyperinsulinemia and hyperglycemia when fed diets that are high in fat content [13, 12, 14, 15]. Thus, these mice serve as a useful model for type 2 diabetes as they closely mimic disease progression of type 2 diabetes in humans.

Proteomics represents a powerful tool to globally assess and compare protein expression among tissues. While proteomics has been used repeatedly to study the protein expression changes that occur with helicobacter pylori and specific stomach cancers or cancer cells, no studies appear to address the protein expression changes in stomach that might occur as a result of diabetes. Therefore, in the study 2-DE proteomics was utilized to look at changes that occur in the diabetic stomach. Using this approach (reviewed in Kopchick et al 2001 [16]) combined with the use of the high-fat fed C57BL/6J mouse model of diet induced type 2 diabetes, protein changes in tissues such as pancreas [15] and skin [17] have previously been identified and have led to a better understanding of the molecular mechanisms of diabetes progression.

2. Proteomics Architecture. Proteomics is the study of the protein complement of the genome. Based on metabolic developments proteins can be up-regulated, down-regulated or subjected to various post-translational modifications. Symptoms of diabetes mellitus type 2 are unrecognizable in the initial years of the disease, however, preliminary studies have shown that certain proteins are up-regulated during this early phase. Detecting the complement of proteins that are regulated due to the diabetes would allow for measures to detect or prevent diabetes. Proteomics can be categorized into four main stages: Protein Mining, Protein-Expression Profiling, Protein Modification Mapping and Protein Network Mapping. Protein or proteome mining is the identification of proteins, their sequences and unique specifications. After the identification of a protein its expression is to be registered. The registration is based on the development of a protein over various stages of a certain development, in case of the proposed research the up- and down-regulation of protein during the course of diabetes. The mapping of post-translational modifications provides additional input to the expression of the protein. The modification of a protein can render it incapable of performing its intended functionality. After identifying the proteins and their profile based on a particular state of an organism, the interaction between the proteins are determined to link them together in signal-transduction and other complex biochemical pathways. The medical application of proteomics research is not limited to type 2 diabetes but can be extended to the broader spectrum of clinical diagnosis. Proteomics experiments can lead to certain problems when applied as high throughput protein identification approaches. Protein identification will lead to a high number of possible positive identification. To allow for high throughput proteomics, the protein identification stage of the experiment has to determine between false positives and false negatives, to present experiments based on the identification with the least error possible. But even with automated protein identification, data analysis is very time consuming. Biologists have to interpret and convert cryptic spreadsheet to determine the final result or prepare follow-up in silico experiments. While the automation of each stage of a proteomic experiment itself leads to great increase in productivity the elimination of intermediate steps necessary to convert and interpret data would yield an even greater increase and would allow fully automated pipelined proteomic experiments. This section identifies the most common proteomics workflow for biomarker discovery as shown in figure 2.1. This pipeline was modeled in silico, as the foundation of a proteomics software framework. The current implementation focuses on protein identification. The protein identification has two main objectives: The analysis of peak-lists or mass spectra for an identification based on mass spectrometry data, and the analysis of fragment peak-list detected by tandem mass spectrometry experiments. Each analysis is implemented as a framework that can contain n identification algorithms, which can be local version of the algorithms (with or without source control) or web-based versions. The current implementation supports Protein Prospector's MS-Fit (MS) and MS-Tag (MSMS), Matrixscience's Mascot (MS) and Mascot Ion Search (MSMS) and ProteoMetric's Profound (MS). The presented software provides a detailed report of all search results and their sources. Based on these search results additional metrics can be applied to make better predictions about the analyzed biomarker.

2.1. Image Analysis. In the image analysis a digital image is captured from the 2D Gel using a digital camera or a scanner. The captured imaged is cropped and resized to match previously captured images of different specimen. The

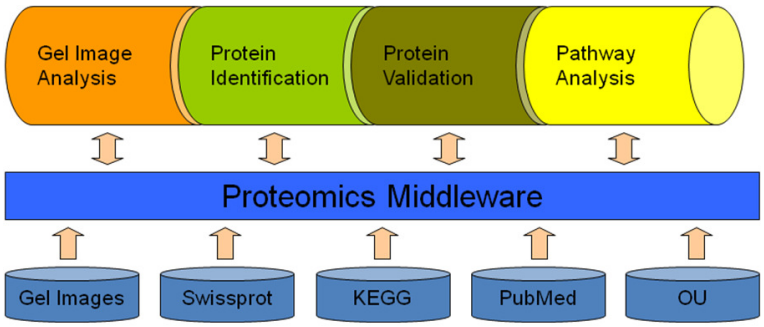

FIG. 2.1. *Proteomics Pipeline. This framework is modeled after the analysis of the biomarker discovery pathway employed in the analysis of type 2 diabetes.*

gamma rate of the picture is adjusted to allow for a better differentiation of protein spots in the image. The preprocessed image is then compared to images of specimen take at different time points. These comparisons allow the quantification of intensity difference of the shown proteins. With the intensity difference vector an intelligent selection of diabetes related proteins can be achieved. The selected spots are cut out of the 2D Gel and a Mass Spectrometry is conducted on them. Various software tools have been developed to handle this problem like BIORAD's PDQuest, Melanie, Image Master and Progenesis. Since the current tools can not identify interesting spots satisfactorily various improvements have been suggested in the literature.

2.2. Protein Identification. The most important data files produced are peak lists that provide information about the mass and intensity of the protein fragments based on the laser lights time of flight. In the protein identification process, this peak list data is submitted to a protein identification algorithm implemented in a web portal or a stand-alone proteomics system. The algorithm calculates the deviation of the provided mass information from the data stored in a database. As a result it produces a list which consists of the best matching proteins, their matching score and additional information about that protein (mass, isoelectric point, accession numbers, etc.). If the matched protein score are sufficient the results are verified by analyzing the peak lists of each fragment of the protein. If the results are verified the protein is analyzed whether or not it is previously linked to Type-2 Diabetes. In case the match produced by the algorithm has a poor score an analysis of the post-translational modification of this protein is conducted. The protein is submitted to the algorithm taking the various post-translational modification possibilities into account. The highest matched protein is set as the identified protein and its post-translational modification is being stored. These results are once again verified through an analysis of the fragment peak lists.

2.3. Protein Analysis. In the protein analysis the sequence of the identified protein is determined and human homologues are inquired. The human homologue inquiry is based on simple BLAST searches and lead, upon identification of a homologue, to a checking of the previous existence in the metabolic pathway of type 2 diabetes or in case of non-existence of a homologue to the discarding of the protein or the publication depending on the extend of the species type 2 diabetes research. If the identified protein is previously unknown the protein sequence has to be determined by additional mass spectrometric experiments. The derived sequence will be published and further analyzed for the homology to human proteins.

2.4. Protein-Protein Interaction. After deriving the homologue protein for the human species the existence in the type 2 diabetes related pathways will be determined. If the protein is existent in the diabetes pathways, the newly discovered connections will be linked into the pathway to extend the information content. If the protein is not existent yet, a new pathway will be created, which then will be linked to the existent pathways to extent the diabetes pathway knowledge.

3. Distributed Protein Identification. In order to illustrate the speedup gained by automatic analysis of MS and MS/MS generated peak lists it is necessary to understand the manual process of protein identification.

3.1. Manual Process. In previously proteomics experiments conducted in the analysis of mouse stomach tissue MS peaks were submitted to Matrix Science website where MASCOT peptide mass fingerprint searches were performed using the following search parameters (Database: NCBInr, Taxonomy: Mus musculus, Enzyme: Trypsin, missed cleavage: 1, Fixed Modifications: none specified "default", Variable modifications: Carbomidomethyl (C), Oxidation (M), Protein

Mass: not specified “default”, Peptide Tolerance: 1 Da “default”, Mass value: MH+ “default”, Monoisotopic: selected “default”).

To achieve a greater accuracy, MS/MS spectra were acquired in MS/MS 2kV Positive mode. Spectra were acquired for 6,000 laser shots or until 5 peptide fragment ions reached a S/N of 100, whichever was less. Fragmentation of the peptides was induced by the use of atmosphere as a collision gas with a pressure of 6 x10-7 torr and a collision energy of 2kV. The collected MS/MS peaks were once again submitted to Matrix Science website, where MASCOT MS/MS ion searches were performed and the following search parameters were used: (Database: NCBInr, Taxonomy: Mus musculus, Enzyme: Trypsin, missed cleavage: 1, Fixed Modifications: none specified-default, Variable modifications: Carbomidomethyl (C), Oxidation (M), Protein Mass: never specified-default, ICAT: not selected-default, Peptide tol. 2.0 Da-default, MS/MS tol. 0.8 Da-default, Data format: Mascot generic, Monoisotopic: selected-default, Precursor m/z : not specified, Instrument: Default).

3.2. Distributed Proteomics. In distributed proteomics a distribution of protein identification tasks is performed. Based on available resources peaks can be matched to peptides on different CPUs leading to a considerable speed-up of the identification process. The protein identification implementation presented in this pattern is easily distributable. To allow for multiple nodes to participate in a protein identification the information about the peptides has to be available and consistent on all nodes. Once the peptide information is available on all nodes, one main node distributes the matching processes for the various elements in the peaks array to the available nodes. Each node will execute the matching process and return the matching result array, which will be added to a main matching array. Once a node completes a matching process, it is available to receive another peak for matching until the main node has matching information for all elements of the peaks array.

The scoring and preparation of the final results has to be conducted on the main node, since information about the complement of peak matches for a protein is required. An illustration of this distributed protein identification under the control of one main node is depicted in figure 3.1. This distribution scheme is adjustable to any number of nodes and be multiplied to incorporate multiple main nodes at once, which can process multiple peak lists. While this might not be of interest for small local protein identification solutions it is of value for larger proteomics centers, which evaluate numerous peak lists at the same time for different laboratories.

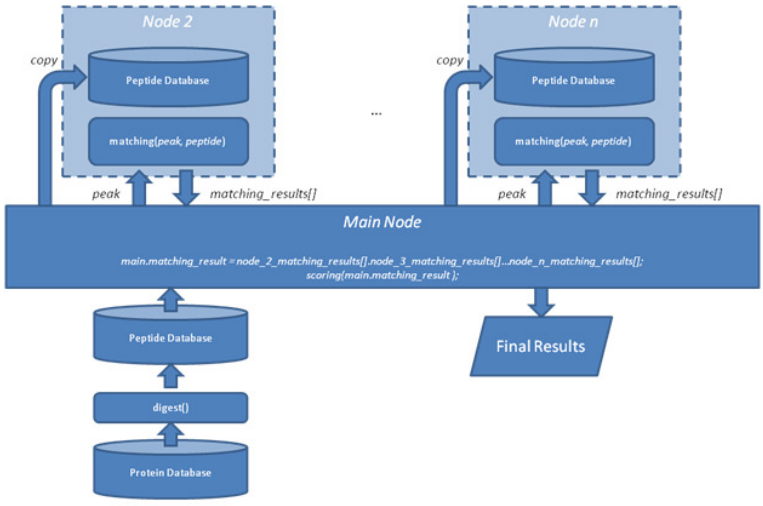

FIG. 3.1. *Distributed Proteomics Approach.*

4. Experimental Analysis. A distributed proteomics platform was implemented based on the architecture presented in figure 3.1. In order to test the framework an experiment was conducted that performs a proteomic analysis of stomach tissue from obese and diabetic mice.

A total of 70 three week old male C57BL/6J mice were purchased from Jackson Laboratory (Bar Harbor, ME). Obesity and type 2 diabetes were induced by placing 50 of the mice on a high-fat diet (#F1850, Bioserve, Frenchtown, NJ) for 16 weeks. The macronutrient content of the high fat diet was as follows: 17% of the calories were provided by protein, 27% by carbohydrates, and 56% by fat. The control mice (n=20) were placed on a standard rodent chow diet for

the same timeframe (Prolab RMH 3000, PMI Nutrition International, Inc., St. Louis, MO). The macronutrient content of the control diet was as follows: 26% of the calories were provided by protein, 60% by carbohydrates, and 14% by fat.

Mice were housed 2 per cage in a temperature controlled room (22°C) on a 14 hr light, 10 hr dark cycle. All mice were allowed ad libitum access to water and food. They were weighed once every two weeks and blood samples taken at 2, 4, 8 and 16 weeks on the diet to monitor diabetic progression. Once diabetes was established, 4 representative diabetic mice and 4 non-diabetic control mice were sacrificed by cervical dislocation and tissues collected. These protocols also have been approved by the Ohio University Institutional Animal Care and Use Committee and conform to local, state and federal laws.

Fasting blood glucose and plasma insulin were determined at four separate time points (2, 4, 8 and 16 weeks on the diets). For all plasma collection, mice were fasted for 8 hours starting at 7am and, subsequently, whole blood was obtained by tail bleeding. The first drop of blood was used for assessment of blood glucose using a OneTouch glucometer from Lifescan (Milpitas, CA). Approximately 250ul of blood were then collected using heparinized capillary tubes. Whole blood samples were centrifuged at 6000 x g and the plasma collected was used to determine insulin concentrations using the rat insulin ELISA kit and rat insulin standards (ALPCO: Windham, NH). As recommended by the manufacturer, the values were adjusted by a factor of 1.23 to correct for the species difference in cross-reactivity with the antibody.

During the course of the high fat and standard chow 16 week feeding, 4 representative diabetic mice from the high fat fed group with classical signs of type 2 diabetes (obesity, hyperinsulinemia and hyperglycemia), and 4 non-diabetic control mice from the low fat fed group were sacrificed via cervical dislocation. After the stomachs were removed, they were cut open in order to remove stomach contents and the stomach tissue washed with saline. The washed stomach tissue was then immediately frozen in liquid nitrogen, transferred to a freezer, and kept at -80°C until processing. The stomach samples were weighed in order to determine the amount of solubilization buffer to be used and then homogenized using a mechanical homogenizer in tubes containing solubilization buffer (7M urea; 2M thiourea; 3% 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS); 1% 3-(Decyldimethylammonio) propanesulfonate inner salt (SB3-10); 0.1% Bio-lytes 3-10 (Bio-Rad Laboratories Inc., Hercules, CA); 2mM tributylphosphine (TBP); 1.5% protease inhibitor cocktail (Sigma, St. Lewis, MO)). The solubilization buffer was made fresh just prior to homogenizing the tissues. The ratio of solubilization buffer used per mass of stomach was 4ml per gram stomach tissue. The samples were then homogenized a second time via sonication and centrifuged for 45 minutes at 150,000 x g. After centrifugation, the supernatant was collected and stored at -80°C.

The frozen solubilized protein samples were removed from -80°C storage and a small portion diluted so that protein concentrations could be determined using a Bradford assay [18]. One mg of solubilized stomach protein was diluted in freshly prepared solubilization buffer to a total volume of 400ul. To create a reduction environment for the proteins, 6ul of 200mM tributylphosphine (TBP) and 8ul of 1M Tris-HCl pH-8.8 were then added to the sample and samples were allowed to incubate at room temperature for 2 hours. After reduction, 6μl of freshly prepared 160 mg/ml iodoacetamide were added in order to alkylate the reduced disulfide bonds. Alkylations reactions were incubated at room temperature for 3 minutes and were repeated two additional times. After completion of the alkylation reactions, the samples were transferred to individual wells of disposable 17cm IPG trays (Bio-Rad Laboratories Inc., Hercules, CA) with 17cm IPG strips pH 3-10 (Bio-Rad). The trays containing IPG strips soaking in protein sample were then wrapped with plastic wrap and incubated at 20°C for 16 hours to allow for passive rehydration of the strips. The IPG strips containing the stomach protein samples were then removed from the disposable trays, briefly blotted onto filter paper in order to remove excess moisture, and placed into wells of an isoelectric focusing tray. The IPG strips were covered with mineral oil to prevent desiccation of the strip during the first dimension. For the first dimension, the isoelectric focusing tray was placed into a PROTEAN IEF cell (Bio-Rad) and the proteins separated via isoelectric focusing at 4,000 volts for 60,000 volt hours. Following isoelectric focusing, the IPG strips were removed from the focusing tray, briefly blotted onto filter paper to remove excess mineral oil and then placed into disposable IPG trays containing 1.5ml freshly prepared equilibration buffer (6M Urea; 2% SDS; 375mM Tris-HCl pH 8.8; 20% Glycerol). The samples were equilibrated at room temperature for 25 minutes.

The first dimension was performed using 17 cm IPG strips that were designed by the manufacturer to be resolved in the 2nd dimension by SDS-PAGE using the corresponding 17 cm large gel apparatus (Bio-Rad Laboratories Inc., Hercules, CA). For our experiments, 4.5 cm were cut from each end of the 17 cm IPG strips leaving the center 8cm, which was then placed on top of 15% polyacrylamide gel containing 4% stacking gels for the second dimension using the 8 cm small gel apparatus (Bio-Rad Laboratories Inc., Hercules, CA). The proteins were then separated via SDS-PAGE at 25mA per gel for 250 volt hours.

Following 2-DE, the gels were incubated with SYPRO Orange fluorescent stain (Molecular Probes, Inc., Eugene, OR). Since SYPRO Orange was intended by the manufacture for use with 1-D SDS-PAGE, a modified protocol by

Malone et al [19] was used. SYPRO Orange stained gels were then imaged using a VersaDoc 1000 Imaging System (Bio-Rad Laboratories Inc., Hercules, CA). Spot detection and spot densitometry were carried out using the Discovery Series PDQuest 2-DE analysis software package version 7.0.0 that came with the VersaDoc 1000 Imaging System. Protein spots to be identified were removed from the polyacrylamide gel using pipette tips that were cut with scissors so that the end of the tip corresponded to the size of the particular spot to be excised from the gel. The gel plugs containing the protein spots were placed in distilled water and frozen with dry ice and sent to the proteomics facility at the University of Michigan for identification by MS and MS/MS analyses.

For MS and MS/MS analyses, 5 uL of alpha-cyano-4-hydroxycinnaminic acid (5 mg/mL in 50% acetonitrile, 0.1% trifluoroacetic acid (TFA), 2mM ammonium citrate) matrix was added to the 30ul of digest extract for each well of the extraction plate. The samples were taken to dryness and 5ul of 50% acetonitrile/0.1% TFA was added back into the extraction well. This solution (0.5ul) was hand-spotted on a 192-well MALDI target and allowed to dry in atmosphere.

Mass spectra were acquired on an Applied Biosystems 4800 Proteomics Analyzer (TOF/TOF). MS spectra were acquired in Reflector Positive Ion mode. Peptide masses were acquired for the range from 800-3,500 Da. MS spectra were summed from 2,000 laser shots from an Nd-YAG laser operating at 355 nm and 200 Hz. Internal calibration was performed using a minimum of 3 trypsin autolysis peaks.

4.1. Results. High fat feeding in C57BL/6J mice resulted in a significant increase in weight gain (figure 4.1A). The increase in weight between high fat fed and low fat fed mice reached statistical significance ($P < 0.001$) by 4 weeks of feeding the respective diets and continued throughout the 16 week dieting period. By the end of the 16 week period, high fat fed mice weighed approximately 35% more than mice fed the low fat diet (figure 4.1A). The increased weight of high fat fed mice was accompanied by an apparent diabetic state. That is, the weight gain observed for the high fat fed mice was accompanied by an increase in both fasting blood glucose (figure 4.1B) and fasting plasma insulin (figure 4.1C) levels. Both insulin and glucose remained elevated throughout the 16 week period. This indicated that the high fat fed mice were suffering from a condition comparable to type 2 diabetes in humans.

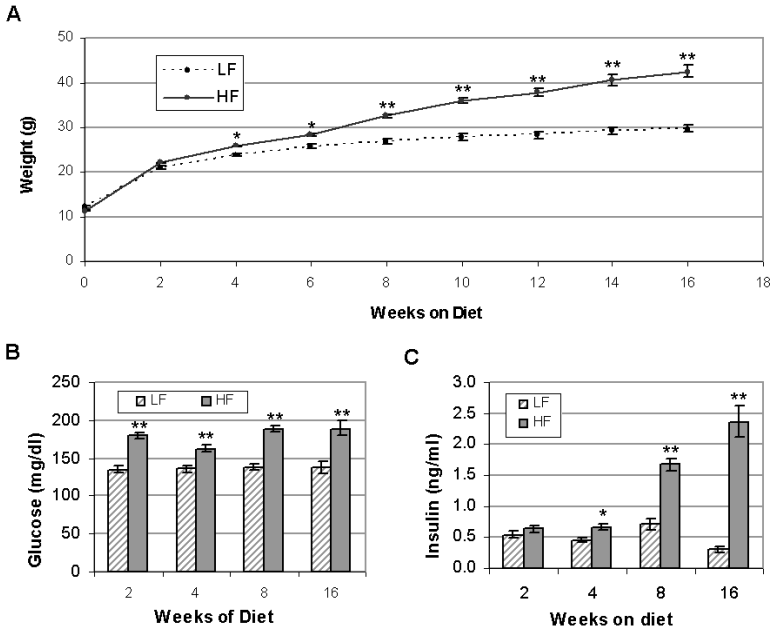

FIG. 4.1. *Progression of type 2 diabetes in C57BL/6J mice placed on a high fat diet compared to control mice fed standard chow. A-C) All measurements were taken at 2, 4, 8 and 16 weeks on the diet except for body weight, which was measured every two weeks. A) Body weights of mice fed a high fat diet began to separate from control mice after 2 weeks becoming statistically significant after only 4 weeks on the diet and continued throughout the 16 weeks of the diet phase of the study. Dashed lines (A) and striped bars (B&C) represent control mice fed a standard chow diet (n=20), while the dashed lines (A) and solid bars (B&C) represent mice fed a high-fat diet (n=50). B) Fasting blood glucose levels were significantly higher in the high fat fed mice at 2, 4, 8 and 16 weeks. C) Fasting plasma insulin levels increased at each of the time points measured and were statistically significant at 4, 8 and 16 weeks. Error bars represent the standard error of the means. Statistical analysis was performed using ANOVA. *($p<0.05$), **($p<0.01$).*

During the course of the high fat and standard chow 16 week feeding, 4 representative diabetic mice from the high fat fed group with classical signs of type 2 diabetes (figure 4.2) and 4 non-diabetic control mice from the low fat fed group were sacrificed. The 4 high fat fed mice selected had significantly higher body weights, circulating insulin and glucose when compared to the 4 controls and clearly showed two distinct groups, mice that are diabetic and mice that are non-diabetic (figure 4.2).

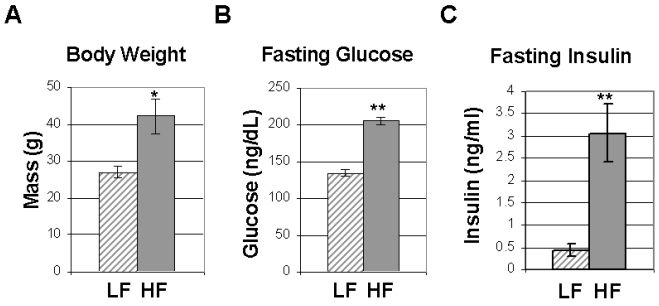

FIG. 4.2. *Physiological parameters of the eight C57BL/6J mice used for proteomic analysis just prior to sacrifice. A) Body weights, B) fasting blood glucose, and C) fasting plasma insulin levels of mice fed a high fat diet (solid grey bars; n=4) were all significantly greater than mice fed a low fat standard chow diet (stripped bars; n=4). Error bars represent the standard error of the means. Statistical analysis was performed using ANOVA. *($p<0.05$), **($p<0.01$).*

In order to identify stomach proteins that were altered in mice with type 2 diabetes, stomachs were harvested from these eight mice following cervical dislocation and proteins treated to maximize solubilization. Solubilized proteins from the 8 mouse stomach samples were separated using 2-DE on 17 cm IPG strips followed by 17 cm large format SDS-PAGE (figure 4.3, top gel). These initial gels contained a lot of unresolved space prompting the development of a slight variation in 2-DE technique (figure 4.3, bottom gel and figure 4.4). The simple removal of the outer 4.5 cm from the IPG strip followed by resolving the proteins using 8 cm small format SDS-PAGE (figure 4.3, bottom and figure 4.4) not only utilized less resources to obtain the second dimension, but roughly doubled the total number of detectable spots from approximately 300 to 600 (figures 4.3 and 4.4).

Following staining of the small format gels with SYPRO orange, the gels were used for imaging, quantification and spot selection. A typical 2-DE image obtained using this technique is shown in figure 4.4. From these images, PDQuest software version 7.0.0 was used to detect spots, and compare intensities between diabetic and control stomach proteins.

By comparing the combined proteomes of all 4 diabetic mouse stomachs to all 4 non-diabetic mouse stomachs, 23 protein spots were found to be altered (figure 4.4) with 14 being increased (Table 1) and 9 being decreased (Table 2). In order to identify these protein spots, all 23 were physically removed from the gels and analyzed by MS and MS/MS analyses. Mascot-MS and Mascot-MSMS database searches were performed on peak lists generated by MS and MS/MS analysis, respectively. Identities of the proteins are presented in tables 1 and 2.

Overall, a total of 14 protein spots later identified as 11 different proteins were found to be increased and 9 protein spots identified as 8 different proteins were found to decrease in the stomachs of high fat fed diabetic mice as compared to controls. The majority of the proteins identified are involved in energy metabolism and gastric motility or structure, with the remaining proteins playing roles in various pathways or processes including protein turnover, protection against reactive oxygen species and protection and overall integrity of the stomach mucosa.

Type 2 diabetes and obesity are overlapping conditions that are both characterized by marked alterations in the metabolism of energy nutrients, in particular carbohydrate and lipids. Therefore, it is not surprising that many of the proteins identified to be altered in the high fat fed mice are related to energy metabolism. Of the 18 distinct proteins found to be altered in these samples, 11 (61%) are known to participate at some level in energy metabolism. While some of the altered proteins, are involved in the energy production and metabolism involving multiple types of energy nutrients, like ATP synthase and cytosolic malate dehydrogenase, most of the proteins are specifically related to either lipid (11%) or carbohydrate metabolism (33%). As type 2 diabetes is ultimately a condition of impaired glucose tolerance with marked alterations in glucose metabolism, it was expected that proteins involved in glucose metabolism or homeostasis would be identified. These proteins included enolase, triosephosphate isomerase, galactose binding lectin (otherwise known as galectin-7), galectin-2, lactate dehydrogenase, and phosphoglycerate mutase.

Gastroparesis, a common disorder in patients with type I or type II diabetes, results from partial or complete paralysis of stomach muscles causing the movement of food into the intestines to slow or stop. Interestingly, 4 of the 23 protein spots

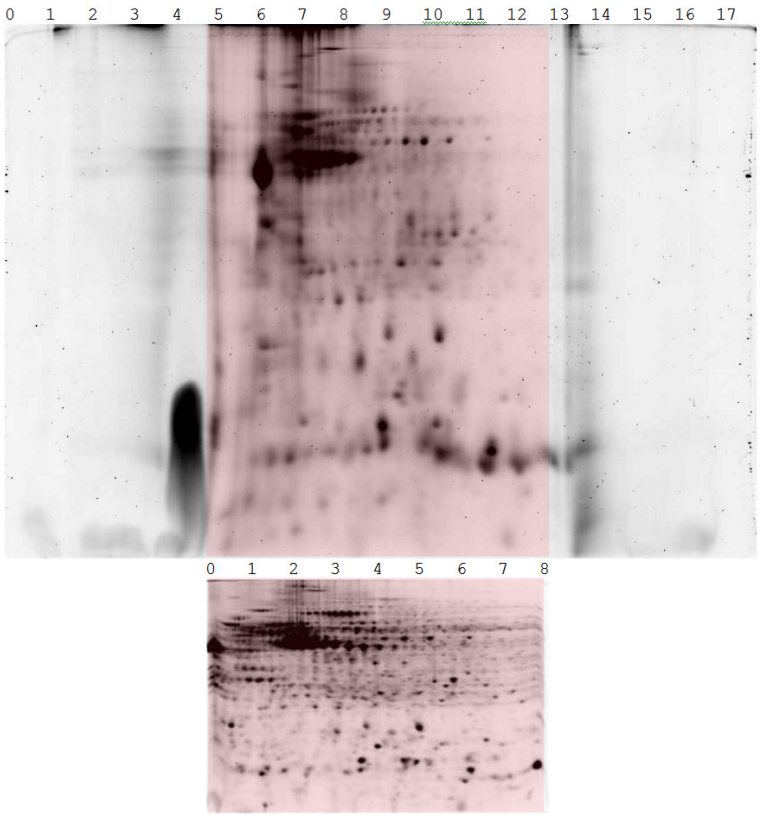

FIG. 4.3. *Removal of edges from IPG strips prior to the SDS-PAGE second dimension greatly enhances 2-DE image and protein separation. Identical stomach protein samples run in the first dimension on 17 cm IPG stips produce better images with more distinguishable protein spots when 4.5 cm from each end of the 17 cm IPG strips are removed and resolved on an 8 cm small format SDS-PAGE (bottom) as compared uncut 17 cm IPG stips resolved on 17 cm large format SDS-PAGE (top).*

(6%) that were found to be altered could potentially play a role in this disorder. Two of these proteins were identified as myosin light polypeptide 3 while the third was identified as myosin regulatory light chain 2. Myosin proteins are involved in muscle contraction, and therefore, any changes to this class of protein in diabetic stomach could be directly related to gastroparesis.

The fourth protein was identified as heat shock protein beta 1 (Hspb1), which is a protein that has been implicated in neuromuscular disorders. The exact function of Hspb1 is unknown; however, mutations in this gene have been associated with motor neuropathy [20, 21]. Kijima et al. hypothesize that a mutation in Hspb1 may cause the neurofilament network to become unstable, disrupting peripheral nerve stability [21]. It is conceivable that the decrease in Hspb1 expression observed in diabetic stomach in this study may also contribute to gastroparesis.

Only one protein identified, gastrokine-1, was found to be expressed exclusively by the stomach. Although the exact functions of this protein are not fully understood, several studies have attempted to further characterize gastrokine-1. Analysis of the human and mouse transcripts show high conservation of the sequence, suggesting an evolutionarily conserved and important function. Oien et al [22], reported finding gastrokine-1 expressed in all areas of the stomach including antrum, body and cardia; however, gastrokine-1 was absent in other gastrointestinal tissues, such as the colon, as well as gastric carcinomas of human patients. Gastrokine-1 has been reported to be a secreted protein produced at least by the secretory cells of the antral mucosa [23, 24]. Toback et al [23], suggests that it is a mitogenic protein that, as a component of the normal mucous layer, could also play a role in the protection of the stomach from various stresses or in maintaining mucosal integrity of the surface epithelial layer. More recent studies have shown a decrease in two isoforms of gastrokine-1 in Helicobacter pylori positive patients [25]. This subset of patients was also categorized as having moderate to severe gastritis. In these same patients, once H. pylori was eradicated, gastrokine-1 mRNA increase by 2.5-fold [26]. In our high fat fed diabetic mouse model, gastrokine-1 protein was found to increase in stomach, possibly in response to diabetic stress. Further studies are needed to reveal the relationship between gastrokine-1 and the diabetic stomach.

TABLE 4.1

Proteins found to be increased in stomach tissue of high fat fed diabetic mice as compared to non-diabetic control mice fed standard rodent chow. (a) Protein identified with 2 or more significant MS/MS peptide matches. (b) Protein identified with a significant MS match. (c) Indicates a weak identification with multiple extensive homology matches but do not satisfy the criteria for a or b.

Well #	Protein Name	Mascot Accession #	% Increase
H01 [c]	myosin, light polypeptide 3	gi—33563264	367%
B09 [c]	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit isoform 1	gi—6680748	260%
D23 [b]	triosephosphate isomerase 1	gi—6678413	89%
F21 [a]	myosin, light polypeptide 3	gi—33563264	88%
F07 [c]	phosphatidylethanolamine binding protein 1	gi—84794552	75%
B19 [a,b]	galectin-2	gi—13629138	48%
F09 [b]	Apoa1 protein	gi—61402210	46%
F20 [c]	ATP synthase, H+ transporting, mitochondrial F1 complex, beta subunit isoform 1	gi—31980648	41%
B20 [a,b]	fatty acid binding protein 3, muscle and heart	gi—6753810	37%
D11 [a]	gastrokine 1	gi—13384882	37%
F01 [a]	Peroxidase	gi—885932	36%
H09 [b]	lactate dehydrogenase 2, B chain	gi—6678674	34%
B24 [a,b]	lectin, galactose binding, soluble 7	gi—31543120	22%
F13 [c]	phosphatidylethanolamine binding protein 1	gi—84794552	19%

TABLE 4.2

Proteins found to be increased in stomach tissue of high fat fed diabetic mice as compared to non-diabetic control mice fed standard rodent chow. (a) Protein identified with 2 or more significant MS/MS peptide matches. (b) Protein identified with a significant MS match. (c) Indicates a weak identification with multiple extensive homology matches but do not satisfy the criteria for a or b.

Well #	Protein Name	Mascot Accession #	% Descrease
B16 [c]	enolase 1, alpha non-neuron	gi—12963491	57%
H15 [a]	cytosolic malate dehydrogenase	gi—387129	41%
B17 [c]	phosphoglycerate mutase 1	gi—10179944	40%
D21 [c]	myosin regulatory light chain 2, smooth muscle isoform	gi—38605043	40%
B23 [a,b]	triosephosphate isomerase 1	gi—6678413	26%
B04 [a]	enolase 1, alpha non-neuron	gi—12963491	23%
H19 [a]	heat-shock protein beta-1	gi—547679	16%
B10 [c]	aldehyde dehydrogenase family 3, subfamily A1	gi—56238010	16%
H02 [c]	carbonyl reductase 3	gi—27413160	16%

Several of the identified proteins have been linked to the diabetic condition in previous studies. Importantly, none of these proteins have previously been shown to be altered in the stomach in a diabetic state offering new areas of research. For example, apolipoprotein A1 (Apo A1) levels are decreased with diabetes, which is likely due to the lack of an insulin stimulatory effect on the Apo A1 promotor [27]. Diabetics have a well documented increase in cardiovascular complications with a characteristic decrease in high density lipoprotein levels (HDL) [28]. As the major HDL associated protein, Apo A1 has also been shown to be decreased in the plasma of type 2 diabetics [29]. However, Apo A-1 levels were elevated in stomach of the diabetic mice in this study. Although the relative contribution of the stomach in Apo A-1 production is not known, Apo A-1 gene expression was shown to be downregulated in mouse model of H. Pylori induced gastric cancer [30], suggesting that stomach Apo A-1 1 may be modified with other chronic conditions. Another example can be found in the two galectin proteins (galectin-2 and lectin, galactose binding-7, also called galectin-7), which are both members of the galectin protein family and which were both shown to be upregulated with diabetes in this study. While neither galectin-2 nor galectin-7 have been connected with diabetes, another family member, galectin-3 increases in diabetes [31], similar to the trend shown for both galectin-2 and 7, and has been implicated in β-cell destruction [32]. Collectively, these findings implicate the galectin family as critical to diabetes progression and also provide evidence that further studies are needed to examine other members of this family in the diabetic condition.

5. Conclusions. As shown through the analysis of the stomach tissue it is possible to conduct complex biological experiments in silico and derive significant results with a possible impact on medical applications. The presented scalable computing approach, though simple in its architecture provides a proof of concept to the area of distributed proteomics.

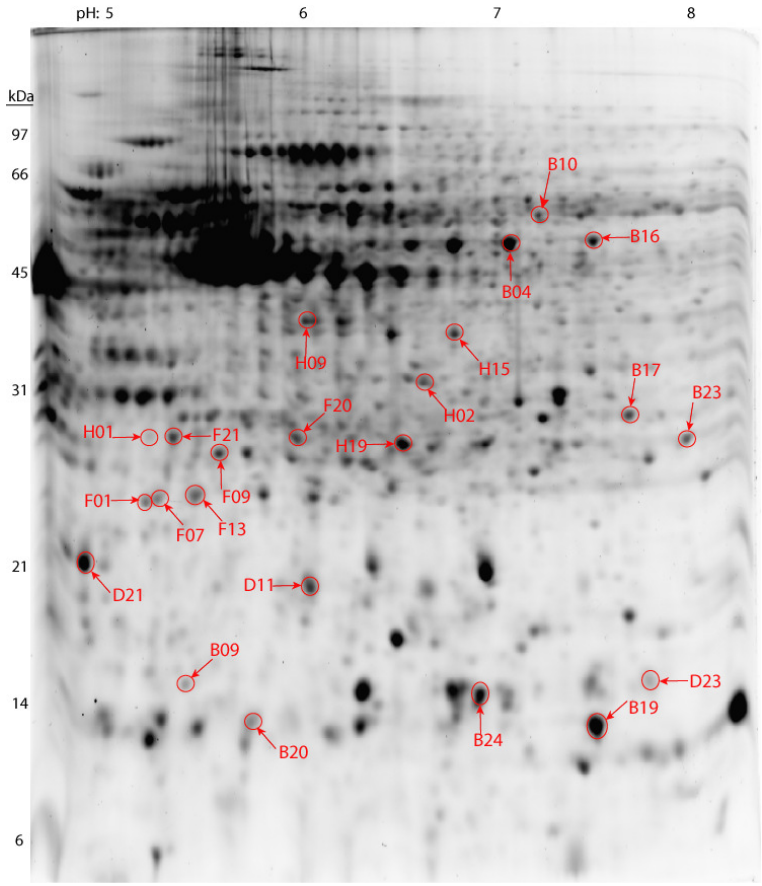

FIG. 4.4. *Two-dimensional gel from diabetic mouse stomach. This SYPRO Orange stained gel contains proteins isolated from the stomach of diabetic mice following 16 weeks on a high-fat diet. The gel was imaged using a VersaDoc 1000 Imaging System. The approximate pI and molecular weights are labeled along the top and left hand borders of the gel, respectively. Spot detection and densitometry were performed using the Discovery Series PDQuest 2-DE analysis software package version 7.0. Spots labeled in red represent proteins that were changed in the diabetic as compared to control stomach samples. All labeled spots were removed from the polyacrylamide gel and analyzed by both MALDI-TOF and MS/MS mass spectrometry at the Michigan Proteome Consortium. Protein identities and the percent increase or decrease are presented in tables 1 and 2, respectively.*

By providing first an automated proteomics approach and then a distributed protein identification the human bottleneck of manual identification can be eliminated in favor of a very efficient high performance computational analysis.

Concerning the presented biological experiment it can be said that while proteomic analysis has been used to elucidate proteins involved in stomach cancers and H. pylori infections, little or no attention has been given to diabetic stomach. The up and down regulated proteins reported in this study serve as targets for future studies that will allow better understanding of the molecular progression of diabetic gastroparesis as well as other diseases associated with gastric disturbances.

6. Acknowledgements. This work was supported in part by the State of Ohio's Eminent Scholar program, which includes a gift from the Milton and Lawence Goll family, by a grant from the AMVETS, by the WADA and by DiAthegen, LLC. D.B. is supported by a mentored career development award from NIH-NIDDK (DK064905). MS and MS/MS peak lists were generated by the Michigan Proteome Consortium (www.proteomeconsortium.org) which is supported in part by funds from The Michigan Life Sciences Corridor.

REFERENCES

[1] K. G. MURPHY AND S. R. BLOOM, *Gut hormones and the regulation of energy homeostasis*, Nature, 444(2006), pp. 854–859.
[2] M. GIL-CAMPOS, C. M. AGUILERA, R. CANETE AND A. GIL, *Ghrelin: a hormone regulating food intake and energy homeostasis.*, Br J Nutr, 96(2006), pp. 201–226.
[3] K. M. FLEGAL, M. D. CARROLL, C. L. OGDEN AND C. L. JOHNSON, *Prevalence and trends in obesity among US adults, 1999-2000.*, Jama, 288(2002), pp. 1723–1727.

[4] *Prevention CfDCa 2005 National diabetes fact sheet: general information and national estimates on diabetes in the United States. In: Services DoHaH (ed).*, Center for Disease Control and Prevention, 2005.
[5] C. H. KIM, F. P. KENNEDY, M. CAMILLERI, A. ZINSMEISTER, D. J. BALLARD, *The relationship between clinical factors and gastrointestinal dysmotility in diabetes mellitus.*, J Gastroint Motil, 3(1991), pp. 268–272.
[6] M. HOROWITZ, P. E. HARDING, A. F. MADDOX, *Gastric and oesophageal emptying in patients with type 2 (non-insulin-dependent) diabetes mellitus.*, Diabetologia, 32(1989), pp. 151–159.
[7] F. MEARIN, J. R. MALAGELADA, *Gastroparesis and dyspepsia in patients with diabetes mellitus.*, Eur J Gastroenterol Hepatol, 7(1995), pp. 717–23.
[8] R. FRASER, M. HOROWITZ, J. DENT, *Hyperglycaemia stimulates pyloric motility in normal subjects.*, Gut, 32(1991), pp. 475–478.
[9] M. SOGABE, T. OKAHISA, K. TSUJIGAMI, *Ultrasonographic assessment of gastric motility in diabetic gastroparesis before and after attaining glycemic control.*, J Gastroenterol, 40(2005), pp. 583–590.
[10] S. T. BRITLAND, R. J. YOUNG, A. K. SHARMA, D. LEE, A. K. AH-SEE, B. F. CLARKE, *Vagus nerve morphology in diabetic gastropathy.*, Diabet Med, 7(1990), pp. 780–787.
[11] T. KAWAGISHI, Y. NISHIZAWA, Y. OKUNO, K. SEKIYA, H. MORII, *Effect of cisapride on gastric emptying of indigestible solids and plasma motilin concentration in diabetic autonomic neuropathy.*, Am J Gastroenterol, 88(1993), pp. 933–938.
[12] R. S. SURWIT, M. N. FEINGLOS, J. RODIN, *Differential effects of fat and sucrose on the development of obesity and diabetes in C57BL/6J and A/J mice.*, Metabolism, 44(1995), pp. 645–651.
[13] R. S.SURWIT, C. M. KUHN, C. COCHRANE, J. A. MCCUBBIN, M. N. FEINGLOS, *Diet-induced type II diabetes in C57BL/6J mice.*, Diabetes, 37(1988), pp. 1163–1167.
[14] R. S. SURWIT, S. WANG, A. E. PETRO, *Diet-induced changes in uncoupling proteins in obesity-prone and obesity-resistant strains of mice.*, Proc Natl Acad Sci USA, 95(1998), pp. 4061–4065.
[15] L. QIU, E. O. LIST, J. J. KOPCHICK, *Differentially expressed proteins in the pancreas of diet-induced diabetic mice.*, Mol Cell Proteomics, 4(2005), pp. 1311–1318.
[16] J. J. KOPCHICK, E. O. LIST, D. T. KOHN, G. M. KEIDAN, L. QIU, S. OKADA, *Perspective: proteomics–see "spots" run.*, Endocrinology, 143(2002), pp. 1990–1994.
[17] E. O. LIST, D. E. BERRYMAN, A. J. PALMER, *Analysis of mouse skin reveals proteins that are altered in a diet-induced diabetic state: a potentially new method for detection of type 2 diabetes.*, Proteomics, in press.
[18] M. M. BRADFORD, *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.*, Anal Biochem, 72(1976), pp. 248–254.
[19] J. P. MALONE, M. R. RADABAUGH, R. M. LEIMGRUBER, G. S. GERSTENECKER, *Practical aspects of fluorescent staining for proteomic applications.*, Electrophoresis, 22(2001), pp. 919–932.
[20] O. V. EVGRAFOV, I. MERSIYANOVA, J. IROBI, *Mutant small heat-shock protein 27 causes axonal Charcot-Marie-Tooth disease and distal hereditary motor neuropathy.*, Nat Genet, 36(2004), pp. 602–606.
[21] K. KIJIMA, C. NUMAKURA, T. GOTO, *Small heat shock protein 27 mutation in a Japanese patient with distal hereditary motor neuropathy.*, J Hum Genet, 50(2005), pp. 473–476.
[22] K. A. OIEN, F. MCGREGOR, S. BUTLER, *Gastrokine 1 is abundantly and specifically expressed in superficial gastric epithelium, down-regulated in gastric carcinoma, and shows high evolutionary conservation.*, J Pathol, 203(2004), pp. 789–797.
[23] F. G. TOBACK, M. M. WALSH-REITZ, M. W. MUSCH, *Peptide fragments of AMP-18, a novel secreted gastric antrum mucosal protein, are mitogenic and motogenic.*, Am J Physiol Gastrointest Liver Physiol, 285(2003), pp. 344–353.
[24] T. E. MARTIN, C. T. POWELL, Z. WANG, *A novel mitogenic protein that is highly expressed in cells of the gastric antrum mucosa.*, Am J Physiol Gastrointest Liver Physiol, 285(2003), pp. 332–343.
[25] G. NARDONE, E. RIPPA, G. MARTIN, *Gastrokine 1 expression in patients with and without Helicobacter pylori infection.*, Dig Liver Dis, 39(2007), pp. 122–129.
[26] M. B.RESNICK, E. SABO, P. A. MEITNER, *Global analysis of the human gastric epithelial transcriptome altered by Helicobacter pylori eradication in vivo.*, Gut, 55(2006), pp. 1717–1724.
[27] A. D. MOORADIAN, M. J. HAAS, N. C. WONG, *Transcriptional control of apolipoprotein A-I gene expression in diabetes.*,Diabetes, 53(2004), pp. 513–520.
[28] G. YOSHINO, T. HIRANO, T. KAZUMI, *Atherogenic lipoproteins and diabetes mellitus.*, J Diabetes Complications, 16(2002), pp. 29–34.
[29] E. C. QUINTAO, W. L.MEDINA, M. PASSARELLI, *Reverse cholesterol transport in diabetes mellitus.*, Diabetes Metab Res Rev, 16(2000), pp. 237–250.
[30] S. TAKAISHI, T. C.WANG, *Gene expression profiling in a mouse model of Helicobacter-induced gastric cancer.*, Cancer Sci, 98(2007), pp. 284–293.
[31] C. IACOBINI, L. AMADIO, G. ODDI, *Role of galectin-3 in diabetic nephropathy.*, J Am Soc Nephrol, 14(2003), pp. 264–270.
[32] A. E. KARLSEN, Z. M. STORLING, T. SPARRE, *Immune-mediated beta-cell destruction in vitro and in vivo-A pivotal role for galectin-3.*, Biochem Biophys Res Commun, 344(2006), pp. 406–415.

Edited by: Dazhang Gu
Received: March 15, 2007
Accepted: May 17, 2007