Appendix E - UNIVERSITY HONORS PROGRAM
SENIOR PROJECT - APPROVAL

Name: Graham Beattie

College: Arts & Sciences Department: BCMB

Faculty Mentor: Dr. Kennedy

PROJECT TITLE: Metabolomics: The Newest 'Omic Field of Research

I have reviewed this completed senior honors thesis with this student and certify that it is a project commensurate with honors level undergraduate research in this field.

Signed: John R Kennedy, Faculty Mentor

Date: 5/3/04

General Assessment - please provide a short paragraph that highlights the most significant features of the project.

Comments (Optional):

Metabolomics

The Newest 'Omic Field of Research

Graham Beattie

BCMB 409

April 28, 2004

Introduction:

Metabolomics is the latest addition to a recent collection of new scientific fields that are categorized as 'omic fields of study. Its appearance follows those of genomics and then proteomics. Genomics is the field that studies the genome, or the entire set of genes that an organism possesses. Genes are sequences of DNA, and their expression causes transcription to make mRNA, which codes for proteins. Gene expression studies involve the profiling of the transcriptome, which is the entire set of mRNA that is present at any given time in an organism. The mRNA transcripts of the transcriptome are translated to form proteins. Proteomics deals with studies of the proteome, or the entire set of proteins that are present in an organism. Metabolites are byproducts of metabolic reactions that are carried out by proteins. Metabolomics is the study devoted to the metabolome, which is the entire collection of small molecule metabolites that are present in an organism at any given time. All of the 'omic fields of study are linked together in that expression of the genome causes the formation of the transcriptome, which codes for the proteome, which in turn carries out metabolic activities to cause fluctuations in the metabolome.

Since metabolites are the end products of gene expression, they are a reflection of all the collective biochemical processes that are occurring simultaneously in an organism, and their measurement can give a "snapshot" of the metabolic state and overall health of an organism. The purpose of this paper is to discuss the technologies and data analysis techniques that are employed by the field of metabolomics, as well as its application to multiple scientific disciplines including disease diagnosis, drug discovery, drug toxicology, functional genomics, systems biology, and nutrition.

Technologies:

Nuclear magnetic resonance (NMR) spectroscopy has been used since the late 1950's by chemists to determine the molecular structure of known and unknown compounds. NMR utilizes a nuclear magnetic resonance spectrometer, which detects energy absorption by the nuclei of compounds in the sample, and generates a spectrum with peaks that correlate to absorptions, or resonances. Traditionally, NMR has been used to detect hydrogen molecules in compounds. The nucleus of hydrogen is a proton, and it has a nuclear spin of either +1/2 or –1/2. These two spin states normally have the same energy. When these nuclei are placed in a magnetic field, the energy levels of the spin states change in such a way that the –1/2 spin state has higher energy than the +1/2 spin state. The lower energy of the +1/2 spin state is favored, and thus more nuclei are in the +1/2 spin state than the –1/2 spin state under an electrical field. There is now an energy difference of ΔE between the two states. After samples have been placed under a magnetic field, they are exposed to electromagnetic radiation, typically with a frequency (v) of 60-500 MHz. Energy is calculated as the product of magnetic field intensity (h) and frequency (v); $E_o=hv$. The magnetic field in the NMR spectrometer is increased slowly over time, and the radiation frequency remains constant, causing the energy to slowly raise over time. When the energy (E_o) is equal to the ΔE between the spin states of the nuclei in the compound, the nuclei in the +1/2 state absorb this energy, causing them to reverse spin and assume the higher energy state of the –1/2 spin state. When this occurs, absorption or resonance peaks appear on the spectrum, which is a graph of energy vs. increasing magnetic field (ppm). Multiple resonance peaks on a spectrum indicate different groups of chemically inequivalent protons in different parts of the compound.

The observed chemical shifts of the different peaks can be compared to tables which identify the groups making up the compound, and this information can then be used to determine the molecular structure. This is how ^{1}H NMR is used to determine the molecular structure of compounds. NMR spectroscopy can be used to detect any nucleus with a spin, including the nuclei of ^{19}F, ^{31}P, ^{13}C, ^{14}N, and ^{23}Na. Metabolomicists use NMR, not only to find molecular structure of compounds, but also to differentiate between small molecule metabolites in a sample. In this instance resonance peaks can correspond to entire compounds, not just different groups on one compound (Loudon, 1998).

Ultraviolet (UV) spectroscopy is used to quantify and identify compounds containing conjugated double bonds. Conjugated double bonds are two double bonds with only one single bond between them. NMR is not very effective for differentiating between conjugated double bonds and normal double bonds, so UV spectroscopy must be used to help determine the structure of compounds containing conjugated double bond structures. The UV spectrometer generates a UV spectrum of absorbance vs. wavelength (λ). Peaks on the UV spectrum indicate the λ_{max} of the compound, and this value can be looked up to find what type of diene the compound contains. The absorbance at λ_{max} can be used for quantification of the compound because the absorbance is related to, and increases with, the number of molecules in the sample (Loudon, 1998).

Mass spectrometry (MS) is generally used to determine the molecular mass of a compound, and also to shed light on the molecular structure. The instrument used is a mass spectrometer, and compounds are vaporized in a vacuum within the device. The vaporized compounds are then bombarded with a high energy electron beam, causing the

ejection of an electron from the compound, and the formation of a cation with a + charge. These cations then undergo a series of fragmentation reactions in each of which one atom is removed from the compound until it is completely degraded to an individual cation. This degradation process yields many fragmentation cations along the way, which are separated by their mass-to-charge ratios (m/z) by the mass spectrometer. Since the charge is generally +1, the m/z ratio correlates to the mass of each cation. A mass spectrum is then generated, which is a graph of the relative abundance of each cation vs. their m/z ratios. The ion on the spectrum with the highest m/z value is called the molecular ion, and corresponds to the cation initially formed by the ejection of an electron form the complete compound. The mass of the molecular ion is equal to the mass of the original compound. The masses of the other cations on the mass spectrum can then be used to determine the identity and quantity of the atoms making up the compound by looking at the mass difference between the cations to see what type of atom was removed at each subsequent fragmentation reaction. This gives enough information to determine the molecular formula and a partial structure of the compound. Metabolomicists use separation techniques to purify and separate small molecule metabolites in a sample from each other. Once separated, the individual metabolites can be analyzed by mass spectrometry to give their molecular mass, molecular formula, and partial molecular structure (Loudon, 1998).

High performance liquid chromatography (HPLC) is a separation technique that can be used to separate and purify compounds in a mixture based on charge, hydrophobicity, size, or affinity for certain molecules. HPLC utilizes a uniform column bed that is packed with tiny spheres designed to separate mixtures based on one of these

characteristics. The sample is placed in a solvent, and then run through the machine at a rate of one column volume per minute. The nature of the material in the column slows some compounds passing through, while allowing others to pass through at faster rates or freely. The HPLC machine automatically places eluted solvent in different tubes based on time of elution. A readout from the machine indicates which tubes have accumulation of compounds in them. These tubes should contain purified collection of compounds because different compounds will be eluted into different tubes due to their differential rate of flow through the column. Gas chromatography (GC) is another separation technique, which separates mixtures based on the boiling points of the components. Mixtures are injected into an injection port, which is maintained at a temperature higher than the boiling point of the compounds in the sample. The components of the sample vaporize differentially based on their boiling points, and are then carried through a column by an inert gas. When compounds pass through the column, they are detected by a detector, which relays this information to a recorder. This is similar to the detection of compound elution in HPLC. The two types of columns utilized in GC include packed columns and capillary columns. Packed columns are 1-5 m in length and 5 mm in diameter. Capillary columns are 10-100 m long and have an inner diameter of 250 μm. Capillary columns have a higher separation efficiency, and are thus more suited for metabolomic studies (Alberts *et al.*, 2002). The purified samples eluted from the separation techniques of HPLC and GC can then be analyzed by MS for metabolomic studies.

NMR and MS are the two platforms for metabolite analysis that are used most frequently in the field of metabolomics. UV is used in conjunction for structural

determination only when the metabolite compounds of interest contain conjugated double bonds, causing the failure of NMR to reveal accurate structures. There are several advantages to the use of NMR over MS. These include identification and quantification of metabolites in complex mixtures without the necessity for sample preprocessing by separation techniques, and its relatively low time consumption (Adams, 2003). NMR is also advantageous in its ability to specifically detect and identify compounds containing nuclei with proton spin, which is a characteristic of the vast majority of metabolites (Sumner *et al.*, 2003). Additionally, NMR has the capacity to detect hundreds of small molecule metabolites simultaneously, and its non-destructive nature allows metabolite detection and localization of the individual metabolites within cells, tissues, and organs. The drawback to NMR is its requirement for relatively high analyte concentrations for detection, meaning that it is difficult to identify the metabolites that are naturally low in concentration (Watkins and German, 2002). Alternatively, MS is capable of identifying and quantifying metabolites in samples regardless of how low their concentrations may be. The drawback of MS is the fact that it cannot identify or quantify individual metabolites within a mixed sample, and samples must first be processed by separation techniques to give purified metabolite samples for MS analysis (Adams, 2003). Neither NMR nor MS offer good means of metabolite identification and quantification on their own, and there is no other single technique that is capable of effective metabolomic analyses. For successful identification and quantification of all metabolite classes, modifications on existing technologies or hyphenated techniques must be utilized (NIH, 2003).

Several modifications on NMR techniques have been established for better metabolite detection. One of these is the use of ^{13}C NMR for tracking ^{13}C-labeled metabolites through metabolic networks to reveal network topologies (Grivet *et al.*, 2003). Another example of an NMR modification that was found to increase resolution for accurate metabolite identification was two-dimensional (2D) J-resolved (JRES) NMR. 2D JRES NMR can offer higher resolution by analyzing the skyline projection of the JRES spectrum, which has separate axes for chemical shift and spin-spin coupling data for the reduction of background noise on the spectrum. This technology was used to determine the relative levels of citrate, taurine, and alanine throughout the different stages of medaka embryogenesis (Viant, 2003). Other modifications of NMR techniques that have been utilized for better metabolite detection include 2D ^{1}H NMR, in situ ^{1}H NMR, and 2D ^{1}H-^{13}C NMR (Grivet *et al.*, 2003).

Many combinatory or hyphenated techniques have been used to increase the usefulness of MS in metabolomic studies as well. GC-MS has been around for a long time, but it is not capable of high throughput, meaning that it is time consuming. The incorporation of a time of flight (TOF) mass spectrometer into the GC-MS method has been shown to greatly increase the throughput of metabolite detection because the TOF mass spectrometer has the capacity to generate more than 50 spectra per second (Guttman *et al.*, 2004). A study that was aimed at the detection of all detectable metabolites in the model plant *Arabidopsis thaliana* compared the success of detection using different hyphenated MS techniques. These techniques included GC-MS, capillary electrophoresis (CE)-MS, and HPLC-MS. Using simple GC-MS, 300 of the estimated 5000 metabolites in *Arabidopsis thaliana* were detected. Three different methods of CE-MS were studied,

and they collectively detected approximately 1500 metabolites. The unavailability of CE technology in normal labs lead to the overall preference of metabolite detection in the plant extracts by HPLC-MS, a technique that is more sensitive than GC-MS, and more reasonable than CE-MS. Experimentation with different types of HPLC columns were undertaken, and they revealed that hydrophilic interaction columns were best suited for detection of polar and water-soluble metabolites, and reversed-phase columns were the most suitable for detection of metabolites rich in lipophilic component content. The study discovered that the use of monolithic silica C18 capillary columns could decrease time consumption relative to normal particle-packed columns due to the increased porous nature of the capillary column material, while simultaneously increasing resolution of peaks for metabolite detection. Experiments with monolithic silica C18 capillary columns that were 30 cm, 60 cm, and 90 cm long showed that increased column length in monolithic C18 reversed-phase HPLC-MS slightly decreased the time efficiency and throughput, but greatly increased the resolution of peaks for successful identification of lipophilic metabolite classes in crude plant extracts (Tolstikov *et al.,* 2003). Other hyphenated techniques involving MS that have shown improvement in metabolite detection experiments include liquid chromatography tandem mass spectrometry (LC-MS-MS), Fourier transform ion cyclotron mass spectrometry (FTMS), HPLC-NMR-MS, and HPLC-NMR-MS-MS (Sumner *et al.*, 2003).

The national institute of health has recognized the potential application of metabolomics to numerous scientific fields of study, as well as the inherent problems with existing technology used for metabolomic studies in relation to their quantification, scope, and throughput capacities. They have published a solicitation that offers

researchers grant money for the development of an integrated set of technologies well suited for metabolomic studies (NIH, 2003).

Data Analysis:

NMR and MS spectral data can be analyzed by computers that perform principle component analysis (PCA) to create PCA plots. The computer analyzes the data and assigns principle component (PC) values to the variables, in this case the metabolites, that describe the greatest variance between data sets. When the PCs are plotted on separate axes of a graph, the resulting scores plot can reveal patterns within a group that are useful to researchers for later comparison in future studies. A similar technique is that of Soft Independent Modeling of Class Analogy (SIMCA), which is used to create computer generated models with similar results to those of PCA (Reo, 2002).

Another data analysis method is that of Heirarchical cluster analysis (HCA). In this type of analysis, computers generate dendograms illustrating the groups of similar metabolites, and their distance from each other based on a number of distance measures. Self organizing maps (SOMs) are an ideal means of spectral data analysis in that they utilize artificial intelligence in computers for better group differentiation based on similarities of data sets within groups, and superior visualization when compared to PCA, SIMCA, or HCA (Sumner *et al.*, 2003).

Disease Study and Diagnosis:

There are many diseases in all different types of organisms that are metabolic in origin, meaning that they are not caused by infectious pathogens, such as different kinds of viral, fungal, or bacterial agents. The origins of metabolic disease lie in the environmental influences on metabolism. Organisms that are developing metabolic

diseases will have altered metabolism compared to those deemed healthy. The metabolome will shift in the concentration of certain metabolites during the development of metabolic disease in a way that is indicative of what type of metabolic disease is affecting an organism. Metabolomic profiling can indicate differences in the metabolomes of healthy and diseased individuals, therefore identifying biomarkers for the disease, and creating a basis of comparison that can be used to diagnose others similarly (Viant *et al.*, 2003).

An example of a metabolomic study on disease is illustrated by a study that focused on withering syndrome in the red abalone, a Pacific shellfish. Withering syndrome is caused by a pathogen called Rickettsiales-like prokaryote (RLP), which infects the digestive epithelial cells, leading to death from subsequent onset of withering syndrome. Withering syndrome is unlike normal metabolic diseases because it requires the infection with a pathogen before disease can occur. It does, however, bear some similarity with environmentally induced metabolic diseases, since there is a much higher rate of withering syndrome in red abalone that are in the presence of environmental attenuators of stress. The environmental factors that can affect the onset of withering syndrome include seawater temperature and quality. When the combination of infection by an RLP and environmental stress occur simultaneously, it is very likely that withering syndrome will set in as the disease state, in which it exerts its effects on abalone's metabolism. For this study, samples of foot muscle, digestive gland, and hemolymph samples were taken from diseased, stunted, and healthy abalone for metabolic profiling purposes. All of the samples were analyzed by ^{1}H NMR, yielding hundreds of peaks on the NMR spectra that corresponded to specific metabolites. Principle component

analyses were conducted on the NMR spectra in order to reduce the data set to those metabolites that were significant due to either similarities or differences between spectra. PCA score plots of the principle components were then generated by a computer, and clustering of data points lead to three visible clusters that represented the principle components of the data collected via NMR from diseased, stunted, and healthy abalone. This illustrates that the health of an abalone could be determined by NMR and PCA score plots without ever actually identifying the principle components. This held true for all tissue and hemolymph samples. Although it was not necessary to identify any of the metabolites that were represented by peaks on the NMR spectra for the diagnosis of withering syndrome, the identity of the metabolite that had the greatest variability between the healthy and diseased states was determined. This metabolite was homarine, which was shown to increase in the diseased and stunted abalone relative to its lowest concentration in the healthy abalone. Homarine can be used as a biomarker for the progression towards withering syndrome in red abalone. This study provided disease diagnosis methods for withering syndrome in red abalone. These include NMR spectra and subsequent PCA plots for comparison in future studies. Another method that can be utilized now, after establishing that homarine is a biomarker for disease and noting that glycine levels dropped in all of diseased and stunted samples, is HPLC for the determination of glycine to homarine ratios, which would become smaller as the disease state increases (Viant *et al.*, 2003). This method of profiling in diseased and healthy organisms could be extended to study other metabolic diseases in the future, and lead to the discovery of new biomarkers of disease.

Metabolomic mapping is another method that can be used to study the differences in biochemical processes that occur in diseased vs. healthy organisms. Metabolomic mapping can follow the individual components a known initial amount of a specific metabolite through all of the different biochemical pathways they are involved in throughout the body to determine the final destinations of all the metabolites in the initial pool. This can be achieved by injecting a known concentration of a single metabolite into healthy and diseased individuals, and then tracking their differential movement through organs, tissues, cell types, and subcellular compartments in the healthy and diseases individuals. This can reveal the biochemical processes that have taken place, and identify metabolite distributions that are correlated with healthy and diseased states (Nishikawa *et al.*, 2003).

An example of metabolomic mapping for disease study in mammals can be seen in a study that mapped vitamin C in diabetic and normal rats. $_{L}$-ascorbic acid (ASA) is a form of vitamin C, and it is known to occur at lower concentrations in the plasma and tissues of diabetic individuals in comparison to its higher abundance in plasma and tissue of healthy individuals. This holds true for diabetic rats as well as diabetic humans. Lower ASA levels in diabetics suggest that ASA is degraded more rapidly in diabetics than it is in healthy individuals. In the presence of free radicals, ASA is normally oxidized to form dehydro-$_{L}$-ascorbic acid (DHA). This is the mechanism by which ASA participates in oxidant defense, and its oxidized form of DHA is normally reduced to its original form. In situations of high free radical abundance, accumulated DHA can be converted into diketo-$_{L}$-gulonic acid (DKG). This is a delactonization reaction and is irreversible. Numerous subsequent DKG degradation products are formed after

delactonization of DHA has occurred. This study utilized NMR to follow the distribution and degradation of ASA that was labeled with a ^{19}F isotope. Labeling was necessary to discriminate between injected ASA and naturally occurring ASA. Labeling the ASA with the ^{19}F isotope also made for easy detection, quantification, and identification by the highly sensitive method of ^{19}F NMR. The labeled form of ASA used was 6-deoxy-fluoro-ascorbic acid (F-ASA), and it was shown to have the same structural conformation and exhibited similar degradation behavior to natural ASA. It is hypothesized that increased oxidative stress in diabetic individuals is the cause of lower ASA levels. If oxidative stress is the cause of ASA loss in diabetics, then the final destination of the ASA metabolites that are lost should be the multiple DKG degradation products. ASA can alternatively be degraded by a metal catalyzed oxidation reaction. This study mapped F-ASA degradation in diabetic and healthy rats to determine the mechanisms of F-ASA degradation that occurred in these individuals based on the degradation end products. The metal catalyzed degradation products of F-ASA were analyzed by ^{19}F NMR after incubation of F-ASA in PBS with Cu^{2+}. Four stable degradation compounds were detected for Cu^{2+} degradation of F-ASA, with the major degradation compound producing a resonance peak at –212.45 ppm. When F-DHA was incubated in similar conditions, the major initial degradation product was found at –213.4 ppm, and was converted later to the –212.45 ppm degradation product. The results of these two degradation experiments show that metal catalyzed degradation of F-ASA does not rely on F-DHA as an intermediate like it does under conditions of oxidative stress. One of the major differences in the F-ASA distribution between diabetic and healthy rats that was uncovered in this study was that F-ASA concentrations were much lower in the plasma,

kidney, and spleen of the diabetic rats. Metabolomic mapping revealed that the lowered level of F-ASA in these areas was due to urinary loss of F-ASA. It was determined that 40% of the F-ASA injected into diabetic rats was lost in the urine within one day. This can be compared to 18% F-ASA lost in the urine of healthy rats at 24 hours. ^{19}F NMR was used to analyze the urine of the healthy and diabetic rats. The healthy rat urine contained F-DKG degradation products, F-1 through F-7, but the diabetic rat urine was missing some of the F-DKG degradation compounds between F-4 and F-7. This evidence suggests that diabetics have not only accelerated F-ASA degradation and loss, but also less complex degradation processes that is similar, but not idendical, to degradation from oxidative stress. To determine whether the degradation products were of biologic origin or a result of in situ degradation in the urine, F-ASA was incubated in normal and diabetic rat urine for one day, and the degradation products were observed. The degradation products that resulted from F-ASA incubation in rat urine were similar to those of metal catalyzed degradation experiments, but not similar to the products found in the rat urine 24 hours after injection. This study shows that accelerated nonmetal catalyzed ASA degradation of biological origin in diabetics is the cause of lower plasma, kidney, and spleen levels of ASA. It also demonstrated the differences in the urinary degradation compounds of ASA in healthy vs. diabetic rats, which suggests that urine analysis could be used for diagnosis of diabetes in the future (Nishikawa *et al.*, 2003). Metabolomic mapping could be utilized to study the effects of numerous other metabolic diseases, and has the potential to increase knowledge of disease mechanisms, and identify new methods of diagnosis.

Drug Discovery:

Metabolomics has the potential to be useful in the pharmaceutical arena by aiding in the process of drug discovery. One way in which metabolomics could be useful for drug discovery is by conducting metabolomic studies on disease mechanisms to elucidate new drug targets. An example of this type of metabolomic study can be seen in the study of hypoxia-inducible factor 1 (HIF-1), which is a transcription factor involved in the progression of many cancers. The heterodimer HIF-1, which is over-expressed in cancerous cells, is made up of α and β subunits. This study compared the metabolomes of normal Hepatoma cells with those deficient in their HIF-1β subunits (Hepa c4 cells). Hepa c4 tumors have slowed rates of growth when compared to normal Hepatoma tumors. NMR analysis of tumor extracts revealed that Hepa c4 cells have lower concentrations of glycine, betaine, and cholines. These are results of the failure to upregulate glycolysis in the Hepa c4 cells, while it is upregulated in normal Hepatoma cells and numerous other types of cancerous cells. Glycine is required for purine ring synthesis, and purines are required for ATP synthesis. The ATP content of the Hepa c4 cells was 20% of the ATP content in normal Hepatoma cells. ATP is necessary for cell proliferation and tumor growth, and the lack of ATP in the Hepa c4 cells can account for the slowed growth rate in Hepa c4 tumors. This study, which illustrated that HIF-1β deficient Hepatoma tumors have slowed rates of growth due to their failure to upregulate glycolysis for ATP production, shows that the HIF-1 transcription factor could be a good drug target to slow the progression of cancer and decrease the rate of tumor growth (Griffiths and Stubbs, 2003).

Metabolomics could also be useful in drug discovery by offering a method of screening the plant kingdom for naturally occurring bioactive compounds. Natural products of plants and their derivatives currently make up over 40% of the drugs designed in the pharmaceutical arena, yet only a miniscule portion of the 350,000 known plant species have been researched for therapeutic activity. LC-UV-MS and LC-NMR profiling techniques can be used along with LC-bioactivity assays to identify bioactive compounds in crude plant extracts. This can be seen in a study that screened crude plant extracts from an African plant, *Erythrina vogelii*, for antifungal agents with bioactivity against *Cladosporium cucumerinum.* These techniques were used to identify nine compounds that exhibited antifungal activity in LC-bioactivity assays with *Cladosporium cucumerinum.* The dereplication process of comparing these compounds to libraries of known antifungal agents revealed that two of these compounds were previously unknown. This study successfully identified an isoflavone and an isoflavanone as previously unknown antifungal agents (Wolfender *et al.*, 2003). This type of analysis of crude plant extracts could be replicated in additional plant species for antifungal activity against additional fungal pathogens, as well as other types of pathogens. This could lead to a future advance in the rate of discovery of naturally occurring therapeutic agents that can be used as templates for new drug designs.

Drug Toxicology:

Drugs are not perfect in the sense that they may affect numerous biochemical pathways, in addition to the ones they are designed to alter. Even when a drug can successfully cause desired changes in metabolism, it can alter other aspects of

metabolism that eventually lead to detrimental effects. Drug induced toxicity can occur when long term drug therapies lead to chronic metabolite imbalances and cause negative side-effects such as disease. By monitoring changes in metabolite concentrations, it is possible to determine which metabolic pathways are unintentionally affected by drug therapies and indicate when drugs have to potential to cause toxic effects. This is not a new concept, and in fact, succinate, glycine, and dimethylamine levels in blood are already measured in their response to drug therapies because it is known that their presence in blood can predict kidney damage. It is important that other metabolites correlated with toxic effects be identified so that potentially toxic drugs never reach the public. Paradigm Genetics is a company that has created a predictive toxicology program for just that purpose, and they hope to identify shifts in the metabolome content that are characteristic of ill effects (Adams, 2003).

Metabonomics is a subdivision of metabolomics that is dedicated to the study of drug toxicity through the utilization of NMR for analysis of the metabolome and subsequent pattern recognition methods to identify toxic effects. The pattern recognition methods, like PCA, are useful for the determination of the toxin type, since different types of toxins cause different types of shifts in metabolite concentrations within the metabolome. Metabolite patterns that are indicative of toxic effects can be identified by using NMR to determine the metabolite concentration changes that occur after exposing an organism to multiple known toxins, followed by principle component analysis. These PCA plots can be used as a reference for comparison when analyzing the effects of drug treatments to determine whether they are potentially toxic, and to determine the type of toxic effect that may occur. These types of studies have already identified some

metabolite patterns that are indicative of toxicity such as the correlation between toxic damage to the proximal tubules in the renal cortex with appearance of glucose, amino acids, and some organic acids in urine samples, as well as the correlation between renal or hepato-toxins with lowered concentrations of urinary citrate. Metabonomics can also be used to identify individuals that are more likely to have negative side-effects to drug therapies which may not cause harm to some people. This can be done by analyzing the metabolites in urine samples of different individuals to determine whether the potentially toxic effects of a drug therapy are more or less likely to cause harmful imbalances in their metabolomes. For example, a study of the urinary metabolomes in two different mouse strains, Alpk:ApfCK and C57BL10J, was able to identify differential characteristics of the metabolomes, which may make them more or less at risk for drug induced toxicities. NMR analysis and PCA can be very effective for identifying toxic effects of drug treatments, and offer a superior method to previously used clinical chemistry methods. This is due to their increased sensitivity and ability to observe changes indicative of toxicity when clinical chemistry cannot (Reo, 2002).

A specific example of a drug that has been shown to have toxic effects through metabolomic studies is that of rosiglitazone, a type II diabetes drug. This drug was designed with the purpose of lowering the levels of serum glucose, insulin, and triacylglycerides, which are all unusually high in the serum of individuals with type II diabetes. Rosiglitazone can successfully lower these levels, but metabolomic studies have revealed that the drug treatment does not lower them in a beneficial way, and can actually cause more harm than good. Metabolomic analyses revealed the mechanism of action which rosiglitazone uses to lower plasma triacylglyceride levels, which was the

blocking of lipid export into blood. Although plasma analysis would indicate positive effects of the drug, it actually causes an increase in fatty acid synthesis, leading to higher fat content in liver and adipose in comparison to their fat contents in type II diabetics prior to drug treatment. Rosiglitazone is an example of a drug that has been identified through metabolomics as one that not only does not achieve its desired effects in an appropriate way, but as one that causes more harm than good by increasing overall fat content to cause negative side effects such as hepatic lipidosis (German *et al.*, 2003a). Future metabolomic studies will be able to identify additional drug duds that do not lead to their intended effects through appropriate means of intervention.

Functional Genomics:

Genomics is the field of study dealing with studies of the genome, which is the entire set of genes that an organism possesses. Functional Genomics the subdivision of the genomics field in which specific unidentified genes are studied with the purpose of determining their functional role in an organism. DNA sequencing techniques sparked the beginning of the genomics era, which has had a huge impact on science and advanced numerous disciplines in the scientific arena. Functional Genomics relies on gene expression studies that analyze the mRNA content of the transcriptome in order to construct an mRNA profile for the organism under study (Sumner *et al.*, 2003). This is done by utilizing DNA microarray technology, which has revolutionized gene expression studies, and allowed the simultaneous detection of the expression or repression of thousands of genes simultaneously on a tiny specialized gene chip. By studying the expression of specific gene sets in organisms that have been exposed to different conditions or modified genetically, microarrays provide functional genomicists a

powerful tool for predicting the functions of individual genes with known sequences (Foos *et al.*, 2003).

Arabidopsis thaliana is a model plant and is one of the twenty plus organisms with genomes that have been completely sequenced since the invention of automated DNA sequencing. By strategically perturbing genes of interest, and studying the resulting changes in the various gene expression products, functional genomicists have been successful in identifying the functions of vast amounts of genes. Several characteristics of plants make them ideal organisms for gene expression studies. Chemical exposure, ionizing radiation exposure, and foreign DNA incorporation are all methods used to mutate genes of interest for gene expression studies in plants. It is much easier to cause these mutations in plants than in animals, and it also does not upset anyone over the use of animals for experimental research. The mRNA making up the transcriptome, proteins making up the proteome, and metabolites making up the metabolome represent a cascade of gene expression products in an organism. Studying changes in these products relative to the types of mutations that were induced can help researchers elucidate genes that are related to each other, as well as their function. Another approach to determining gene functions is by assigning genes predicted functions based on sequence similarities with other genes that have already been functionally identified. Although these methods have been successful in identifying functions of a portion of the genes in *Arabidopsis thaliana*, not very many of the genes assigned functions have been confirmed to have their predicted functions through experimentation, and many more have not been assigned predicted functions at all (Sumner *et al.*, 2003).

Phenotypes can be determined by measuring the changes in metabolite concentrations that occur as a result of gene expression and protein activities. Metabolites are a good measure of what type of post-transcriptional response to gene expression takes place in a cell or tissue. Metabolites can alter in concentration as a result of a biochemical pathway, which was modulated by gene expression, and identification of the pathway responsible for altering the metabolome can be very useful for assigning functions to genes that directly affect metabolism (Sumner *et al.*, 2003). An example of this type of gene activity can bee seen by the changes in concentrations of several nitrogenous metabolites during diurnal changes in plants, which are the result of altered nitrogen metabolism pathways (Stitt and Fernie, 2003). When gene expression does not dictate the activity of proteins, or cause changes in metabolite concentrations, there is a good chance that the protein must be activated at the post-transcriptional level (Sumner *et al.*, 2003). This can be seen in nitrate reductase, which is regulated post transcriptionally by various methods involving sugar, light, or carbon dioxide, as well as members of the 14-3-3 protein family. In another example, glutamine synthetase is regulated post-transcriptionally by shifts in the concentration of ammonium metabolites. Some genes that exert effects on metabolism do so in such a way that it is impossible to detect any changes in the proteome. Because of their superior sensitivity, metabolomic detections of metabolite concentration changes can be very useful in these circumstances to reveal what has occurred, which can then indicate the specific gene function (Stitt and Fernie, 2003). So, metabolomics can be very useful for determining gene functions when the gene directly effects metabolism, the gene produces proteins that are regulated post-

transcriptionally, and when the activity of the protein is so acute that it can only be detected at the metabolite level.

Metabolomic strategies for functional genomics are based on metabolic profiling, and have the potential to accelerate the rate of gene function discovery in plants by determining gene function based solely on the detection and analysis of the metabolome without conducting studies on the transcriptome and proteome. This is partly due to the metabolite richness in plants that is acquired through the numerous different metabolic reactions that can be carried out by an individual protein in different subcellular compartments. Plants are said to have between twenty and fifty thousand genes in their genomes, and are estimated to have tens of thousands of metabolites, with an estimated 200,000 different metabolites throughout the entire plant kingdom (Hall *et al.*, 2002). The metabolomic functional genomic process should begin with the identification and quantification of the metabolites that shift in concentration due to a series of genetic perturbations. The metabolites that should be targeted for detection in these experiments should include carbohydrates, amino acids, organic acids, lipids, fatty acid constituents, vitamins, phenylpropanoids, terpenoids, alkaloids, and glucosinolates. After metabolite detection, researchers should look for groups of metabolites that seem to vary in their concentrations after similar perturbations, and decide which metabolic reactions they are involved in to assign gene functions (Sumner *et al.*, 2003).

Data acquisition methods for metabolite profiling in plants need to include means to collect accurate samples from different subcellular compartments because there are differential reactions displayed in the different subcellular compartments, causing each to have its own unique metabolome. Methods include traditional microdissection, and a

biochemical process that causes subcellular components to break away from each other, and uses a device to detect markers in each so their origin can be identified after completion. NMR spectroscopy can be used to detect metabolites in some of the different subcellular compartments without harming the cells or altering their composition (Stitt and Fernie, 2003). Most metabolite detection techniques related to functional genomics are conducted on plant extracts. The two most commonly used methods for this detection are GC-MS and HPLC-photodiode array-MS because of their reliability in identification of abundance and metabolite class for metabolic profiling. By incorporating TOF technology and peak deconvolution software into GC-MS measurements, increased resolution can was achieved, which doubled the quantity of metabolites that could be detected in crude plant extracts (Hall *et al.*, 2002).

FANCY was coined by Steve Oliver, and represents a metabolomic approach for assessing gene functions in yeast. FANCY stands for functional analysis by co-response in yeast, and assigns gene functions similar to the functions of known genes when they exhibit a co-response. A co-response is indicated by similar changes in metabolite concentrations in an organism when unknown and known genes are perturbed at different times. FANCY assigns functions as to which biochemical pathways the gene is involved in (Sumner *et al.*, 2003). It seems likely that FANCY techniques could be implemented for gene function determination in plants in the same way that it has bee used in yeast. A study that was presented by Denise Jacobs at the First International Congress on Plant Metabolomics showed evidence that FANCY for plants is a feasible approach. She conducted genetic pertubations in 2000 various genes of *Catharanthus roseus* to alter the proteome, and saw the similar co-response of alkaloid accumulation for approximately

one third of the trials, which shows that genes with co-responses to perturbations do not just reside in yeast (Hall *et al.*, 2002).

Transgenic plants, in which foreign DNA has been integrated with the native genomic DNA, are used for metabolomic studies directed towards the elucidation of gene functions through the measurement and analysis of changes in the metabolome that result from the incorporation of a transgene. A functional genomics study, which was conducted at Plant Research International in the Netherlands, utilized transgenic tomato plants that were designed to overexpress the Lc/C1 transcription factor. The researchers constructed metabolic profiles of normal tomatoes and transgenic tomatoes by analyzing crude plant extracts with four different hyphenated techniques for metabolite detection, quantification, and identification. The techniques used in this study included the combinatory technologies of LC/photodiode array detection, LC-TOF-MS, direct infusion-TOF-MS, and GC-MS. The data that was obtained from this study correlated overexpression of the Lc/C1 transcription factor to extremely high levels of accumulation of flavonoids, as well as an accumulation of methylsalicylate (Hall *et al.*, 2002). These findings give the researchers a basis for assigning gene function to Lc/C1 through determination of the biochemical pathways that were activated to cause these metabolite accumulations.

Knockout mutations are another type of genetic modification that can be utilized in plants for functional analysis of genes by metabolomic study. The term knockout in this sense generally characterizes the underexpression of a specific gene within the genome in order to ascertain the effects of its activity absence on the phenotype of the organism. Knockouts will help elucidate gene function if the gene of study cannot be

deleted because that gene is one that the organism cannot live without. RNA interference can successfully cause knockout effects by antisense inhibition of gene expression, where a molecule binds specifically to the mRNA transcript of the gene under study and blocks translation, preventing protein assembly (Stitt and Fernie, 2003). Knockout mutant plants provide the metabolomic analysis of gene function a contrast to the transgenic plant mutations, because the certain gene of interest is not expressed due to antisense inhibition, rather than inserted as a duplicate and overexpressed. In these types of experiments, researchers should generally see a drop in the concentration of the metabolites that the knocked out gene is normally responsible for forming as a result of its expression. An example of a metabolomic knockout experiment determining gene function can be seen above in the study that implicated hundreds of genes in the accumulation of alkaloids in *Catharanthus roseus* (Hall *et al.*, 2002). Metabolomic profiling in knockout mutants has lead to the discovery of additional functions of some genes that were already functionally annotated. Transketolase has been characterized as an important enzyme in the Calvin cycle for a long time, but metabolic profiling of Transketolase knockouts have shown that it also exerts effects on the regulation of the shikimic acid pathway, as well as phenylpropanoid metabolism (Stitt and Fernie, 2003).

Functional genomics is not only capable of defining the functions of individual genes, but also will be able to interrelate different genes that work together, and characterize the connectivity between multiple genes, biochemical pathways, and metabolites for a better understanding of the features of entire biological systems. This type of functional genomics is called integrated functional genomics, and will be largely based on metabolomic studies. In response to the metabolomic advancements of

functional genomics, The Samuel Roberts Noble Foundation has started an integrated functional genomics program, and other integrated functional genomics programs are sprouting up at Iowa State University, the University of Massachusetts, the Max Plank Institute, and the Plant Research Institute (Hall *et al.*, 2002).

Systems Biology:

The goal of systems biology is to develop enough knowledge about all of the different biochemical processes and networks that exist within cells, tissues, and organisms to get a good understanding of all the aspects of an entire integrated biological system. The goals of systems biology are quite similar to those of integrated functional genomics, which are described above. The methods for data collection in systems biology bear great similarity to those employed by functional genomics in that both study the effects of genetic perturbations in plant genomes through quantitative measurement and identification of the response in the metabolome (Weckwerth, 2003). While the goal of functional genomics, to identify individual gene function, can be attained without measuring changes in the transcriptome and proteome, these measurements must be taken for information to be accumulated regarding all biological processes encompassed by systems biology (Sumner *et al.*, 2003). The metabolomic data acquisition processes employed by functional genomics and systems biology differ in one fundamental way. Functional genomics relies on metabolic profiling in which there is a known group of targets for analysis, while the metabolomic aspect of systems biology is based on the global analysis of all metabolites present in an unbiased way (Weckwerth, 2003).

Due to the heterogeneous tissue makeup of plants and the existence of subcellular compartments within plant cells, there are numerous different concentration gradients of

metabolites within cells and tissues (Weckwerth, 2003). This makes analysis of metabolites in unbiased ways difficult, and it is necessary to analyze the individual metabolomes of different subcellular compartments, cell types, and tissue types for accurate data that describes the variability of enzyme activity for creation of a picture of the processes that take place throughout the entire system (Sumner *et al.*, 2003). An experiment in which plastid phosphoglucomutase was under-expressed illustrates the differential activity of single enzymes in different subcellular organelles. In addition to various other metabolic changes, the decresed expression of phosphoglucomutase increased amino acid synthesis in the plastid, while simultaneously decreasing amino acid synthesis in all the other subcellular organelles (Stitt and Fernie, 2003).

In some respects, this non-targeted analysis of metabolite components in systems biology is a superior technique to the existing techniques of transcriptome and proteome profiling that have been used for systems biology in the past. The technology available for metabolite profiling allows for correct metabolite quantification and identification with relative ease, while methods of mRNA profiling and protein profiling used in the past have been far more involved and difficult (Weckwerth, 2003). mRNA profiling has, however, advanced tremendously since the invention of DNA microarrays, and companies are beginning to offer protein arrays as well (Foos *et al.*, 2003). These new platforms for studying the transcriptome, proteome, and metabolome stand to make significant advances in the quantity of complex biological processes that are discovered in a system-wide manner.

Identifying network connectivity between multiple pathways is crucial to the advancement of systems biology. It is a common misconception that genes drive entire

biological systems (Weckwerth, 2003). Changes in the expression of genes or the transcriptome usually lead to changes in the proteome and metabolome as well, but there are many cases where this is not true due to post-transcriptional regulation or cross-talk between pathways (Stitt and Fernie, 2003). The multiple data sets that are required for analysis of entire systems include measurements of the transcriptome, proteome, and multiple metabolomes of tissues, cells, and organelles. These, combined with knowledge of growth stages, environment, and diurnal/circadian rhythmicity must be compiled and integrated to understand all biological processes that occur in an organism simultaneously (Weckwerth, 2003).

Although the preferred techniques for metabolite identification and quantification are usually GC-MS and LC-MS, the holistic approach of systems biology requires implementation of multiple techniques that are suited for certain types of metabolites. This is because systems biology requires an unbiased global analysis of metabolites, and the two most popular techniques, although very effective usually, lack in their ability to detect and analyze certain metabolite classes. Ionic compounds have been identified and quantified by CE-MS, which has much higher resolution for these than either GC-MS or LC-MS. One study showed that normally undetected saponins can be quantified and identified by using electrospray ionization (ESI)-LC-MS-MS to reveal their variance in different species of *Medicago*. By sampling from the multiple sites that vary in their metabolomes, it is possible to track metabolites through organisms to reveal network topologies. After initial sampling and metabolomic mapping have revealed a network topology, it is critical that its response to altered gene expression be measured in order to reveal any connectivity between networks, and also to discover completely new pathways

that may exist. Profiling of the other gene products upstream of metabolites, including proteins and mRNA, can aide in these discoveries (Weckwerth, 2003). Regulatory sites in networks can be seen as those at which metabolite flux occurs. This means that an enzyme or pair of enzymes with opposite activities work at a common site in a pathway, and catalyze forward and reverse reactions, causing metabolite concentrations to shift back and forth between the substrates for the two opposite reactions in a reciprocal fashion. After identifying these regulatory sites according to metabolite flux, researchers can use these sites to design genetic perturbation experiments targeting the genes that code for the forward and reverse enzymes controlling flux. These types of experiments can reveal a great deal of information about the regulatory mechanisms in pathways, and throughout the entire system. For example, the identification of the site of glutamine flux has provided researchers with the perturbation targets of enzymatic catalysts of the forward and reverse reactions, GOGAT and Glutamate dehydrogenase. The perturbation of the genes responsible for these, and the analysis of the response will provide information about the regulatory networks controlling the flow of reduced nitrogen into metabolism (Stitt and Fernie, 2003).

While linking experimental data to regulatory networks is essential in systems biology, determining the connectivity between multiple metabolic pathways is also a required step in the development of a knowledge base encompassing all processes within a system. Snapshot analysis involves multiple sequential analyses of the metabolites in wild type and genetically altered or perturbed plants simultaneously in an attempt to find metabolites that display Pearson's correlations. After the presence of Pearson's correlations has been established, distance maps are generated by using Pearson's

correlation coefficient to calculate and determine connectivity distances between the many metabolites analyzed in the snapshot analysis (Weckwerth, 2003).

Computer modeling can be utilized in systems biology to compile and integrate the determinations of processes in various regulatory networks along with the connectivity maps from snapshot analysis for Pearson's correlations. The generated models visualize the connectivity between metabolites, pathways, and networks, all in relation to each other. The computer models also have predictive abilities for the modes of action of unknown genes based on the information that has already been input. This aspect of computer models makes them very useful for researchers as a guide to help them design perturbation experiments, because the computer models can predict when the proteome and metabolome are going to act independently of the transcriptome (Sumner *et al.*, 2003).

Metabolomic approaches to systems biology can define the metabolic variance between different subcellular organelles, cells, tissues, and even entire organisms. An example of metabolomic assessment of variance between organisms can be seen in a study that determined the metabolic variance between *Arabidopsis* ecotypes and between mutants of the same line. Interestingly, the study showed that there is greater metabolic variance between different ecotypes of *Arabidopsis* than there is between mutants of the same line (Weckwerth, 2003). The incorporation of metabolomics into systems biology has already begun to advance the amount of knowledge about the integration of complex biological systems and what factors cause them to function. As more research is undertaken by systems biologists, this knowledge base of system interactions will continue to grow, along with its applications to other fields.

Diet and Nutrition:

Metabolomics offers an approach to the field of diet and nutrition that is far more sensitive to changes in organismal health than previously used and current methods for the detection of changes in metabolic status and health due to the components of an individual's diet. The experimental and analytical techniques utilized by metabolomics have the potential to greatly advance our understanding of how diet affects metabolism, and subsequently, the overall status of health in humans. The main focus of most nutritional research to date has been on identifying the aspects of specific biochemical pathways that certain dietary nutrients affect. The limited scope of these studies has prevented nutritionists from discerning the larger connectivity of the multiple pathways that a food may affect due to its complex composition of multiple compounds (German *et al.*, 2003a).

The multiple elements making up the foods that humans ingest exert their effects by up-regulating or down-regulating the expression of numerous genes that are responsible for metabolic activity. In fact, the field of nutrigenomics is dedicated to studying the relationship between the molecular makeup of foods and the transcriptome, resulting from gene expression (German *et al.*, 2003b). Since metabolites are the end products of gene expression, the concentrations of various metabolites and metabolite classes are modulated by the genes that are expressed or repressed, and metabolite concentration measurements can give valuable information about the current metabolic status or phenotype (Watkins and German, 2002). This metabolomic style approach to studying the mechanisms of metabolism in relation to dietary components offers a much wider dynamic range for identifying aspects of human metabolism than the traditional

approach, and when data in the form of metabolite concentrations is combined with the genomics technology of DNA microarrays that detect expression of thousands of genes simultaneously, information about biochemical pathways involved and gene functions can be discerned (German *et al.*, 2003b).

Human populations range across a broad spectrum in respect to the characteristics of their physical appearances, like skin color, height, weight, hair color, and many other readily viewable traits, indicating that humans are a very diverse species. Human metabolism, although not easily visualized like external phenotypes, is no different, in that humans have great variability in their metabolic response to foods. The source of diversity across human populations lies within the genomes of different individuals, where genetic polymorphisms alter the activities of homologous genes in different people. Nutrigenetics is dedicated to studying the causes of differential nutritional requirements and metabolic responses to similar foods across human populations at the genetic level (German *et al.*, 2003a). Metabolomic studies offer a good platform for the analysis of different individuals to assess their different nutritional needs (German *et al.*, 2002). This is because it offers a method to eventually identify the significance of different concentrations of certain metabolites in tissues and blood. Once enough metabolomic studies have been conducted to give a good understanding of the overall metabolic system, metabolite levels will be indicative of metabolic pathway activities and genes expressed, which in turn will give nutritionists enough information for dietary recommendations. Metabolomics could effectively end the era of nutrition in which dietary recommendations are standardized for entire populations, and begin another in which people will be able to attain greater overall health through access to

information about their own bodies and how to manipulate their metabolisms (German *et al.*, 2003b).

Scientists have identified many biomarkers associated with various metabolic diseases, and they are the current means of diagnosis, which indicate whether an individual is or is not in a diseased state. Generally, it has been common practice to deem individuals "healthy" when diagnostics are unable to detect any of the known biomarkers for metabolic diseases. Metabolic diseases do not appear spontaneously, but are caused by long-term metabolite imbalances. Metabolic diseases include artherosclerosis, obesity, hypertension, type II diabetes, osteoporosis, and some inflammatory diseases. While biomarkers for metabolic disease do not appear until late in the disease state for many diseases, individual metabolite concentrations shift out of their healthy range long before the establishment of the disease state. In this sense, metabolomics offers a preventative technology with its capability to predict disease before onset. The success of this technique has already been proven through the measurement of blood cholesterol levels to determine the likelihood that an individual will be affected by the metabolic disease artherosclerosis. Artheroscleosis is caused by chronic imbalances in cholesterol metabolism (German *et al.*, 2003a). Although the use of cholesterol as a biomarker for artherosclerosis was successful in many aspects, this case also shows the need for metabolomic assessments of multiple metabolites, rather than a single biomarker. There are numerous mechanisms by which cholesterol accumulation can occur, and different treatments are preferential for each. While cholesterol level measurement can indicate that an individual is at risk for artherosclerosis, it does not specify why this is true, or what to do about it. The reasons

for cholesterol accumulation can be determined by measuring multiple other sterol metabolite levels, and the identity of the sterols that have accumulated can in turn indicate the treatment of choice (German *et al.*, 2003b).

According to German *et al.* (2002), "metabolomics should be led by the field of nutrition." Modern nutrition will focus on studies of the genome, proteome, and metabolome, in respect to changes that occur in each as a response to foods (German *et al.*, 2002). For these studies to be productive, specialized sampling techniques will be required, as it is difficult to effectively quantify many metabolites in blood. Lipids are an example of a metabolite class that is difficult to measure in blood, where results are skewed in samples smaller than 50μL (Watkins and German, 2002). The main barrier standing in the way of metabolomic implementations in nutrition is the lack of complete understanding of metabolism. There is not yet a reference source containing information on all metabolites in the human body, but an integrated database with information linking metabolite structures, metabolite concentrations, protein levels, mRNA levels, and genes to biochemical pathways would complete the data set required for analysis to completely understand the metabolic processes in humans (German *et al.*, 2003a). In assembling this database, it is preferable to organize metabolites in classes for analysis due to their involvement in similar metabolic events. This has already been done with lipids, allowing them to be quantified accurately in blood using the size of each fatty acid chain as a reference (German *et al.*, 2003b). Scientists should agree on a standard for data storage so that their data can be compiled easily into one central database, and interpreted after data sets are complete. Once complete, these databases will enable nutritionists to identify metabolic differences among populations, determine the risk of specific

metabolic diseases, and construct individual diets suited for each person with the goal of improving the overall health of humans. Greater understanding of human metabolism will be beneficial to entire populations, not just to individuals with metabolic diseases (German *et al.*, 2002).

Conclusion:

The study of the metabolome and the fluctuations that occur in its metabolite content can provide massive amounts of data, which can be analyzed to produce great amounts of information about the biochemical processes that occur within organisms, organs, tissues, cells, and subcellular organelles. The interpretation of data collected in metabolomic studies has the potential to advance a wide variety of scientific research objectives through the discovery of new biomarkers for disease diagnosis, identification of new targets for drug therapy in disease, discovery of naturally occurring therapeutic agents, determination of the possible toxic effects of drug treatments, assignment of functional annotations to unknown genes, understanding of all processes within entire biological systems, and implementation of individual diets most suitable for particular nutritional needs. In order for these possible advancements to be made through metabolomics, a couple of things need to happen first. An integrated set of technologies that is well suited for the detection, quantification, and identification of all metabolite abundances and classes must be established for complete metabolomic analysis. Reference sources that link metabolites to all of the states that they are associated with and clearly define the meaning of their relative abundances must be constructed as well. The best way to achieve this is through the creation of a database that links different concentrations of each individual metabolite to the different healthy or diseased states,

gene expressions, biochemical pathways, and integrated networks that they are associated with throughout each organism under study. The diverse set of potential scientific advancements based on metabolomics will be realized only after the invention of ideal assessment technologies and establishment of good reference sources for all metabolites.

References:

1. Adams, A. Metabolomics: Small-Molecule 'Omics. The Scientist. 17, 38-45 (2003).

2. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and P. Walter. Molecular Biology of the Cell Fourth Edition, Garland Science, New York, 2002.

3. Anonymous. Department of Health and Human Services. Roadmap Initiative: Metabolomics Technology Development. National Institute of Health, Sep. 29, 2003. http://grants.nih.gov/grants/guide/rfa-files/RFA-RM-04-002.html

4. Foos, L., Eryuksel, E., and M.H. Brutsche. Functional genomics and gene microarrays - the use in research and clinical medicine. Swiss Med. Wkly. 133, 31-38 (2003).

5. German, J.B., Roberts, M.A., Fay, L., and S.M. Watkins. Metabolomics and Individual Metabolic Assessment: The Next Great Challenge for Nutrition. J. Nutr. 132, 2486-2487 (2002).

6. German, J.B., Roberts, M.A., and S.M. Watkins. Genomics and Metabolomics as Markers for the Interaction of Kiet and Health: Lessons from Lipids. J. Nutr. 133, 2078-2083 (2003a).

7. German, J.B., Roberts, M.A., and S.M. Watkins. Personal Metabolomics as a Next Generation Nutritional Assessment. J. Nutr. 133, 4260-4266 (2003b).

8. Griffiths, J.R., and M. Stubbs. Opportunities for studying cancer by metabolomics: preliminary observations on tumors deficient in hypoxia-inducible factor 1. Advan. Enzyme Regul. 43, 67-76 (2003).

9. Grivet, J., Delort, A., and J. Portais. NMR and microbiology: from physiology to metabolomics. Biochimie. 85, 823-840 (2003).

10. Guttman, A., Varoglu, M., and J. Khandurina. Multidimensional separations in the pharmaceutical arena. Drug Discovery Today. 9, 136-144 (2004).

11. Hall, R., Beale, M., Fiehn, O., Hardy, N., Sumner, L., and R. Bino. Plant Metabolomics The Missing Link in Functional Genomics Strategies. The Plant Cell. 14, 1437-1440 (2002).

12. Loudon, G.M. Organic Chemistry Third Edition, Benjamin/Cummings Publishing, Redwood City, 1998.

13. Nishikawa, Y., Dmochowska, B., Madaj, J., Xue, J., Guo, Z., Satake, M., Reddy, V., Rinaldi, P.L., and V.M.Monnier. Vitamin C metabolomic mapping in experimental diabetes with 6-deoxy-6-fluoro-ascorbic acid and high resolution ^{19}F-nuclear magnetic resonance spectroscopy. Metabolism. 52, 760-770 (2003).

14. Reo, N.V. NMR-Based Metabolomics. Drug and Chemical Toxicology. 25, 375-382 (2002).

15. Stitt, M., and A.R. Fernie. From measurements of metabolites to metabolomics: an 'on the fly' perspective illustrated by recent studies of carbon-nitrogen interactions. Current Opinion in Biotechnology. 14, 136-144 (2003).

16. Sumner, L.W., Mendes, P., and R.A. Dixon. Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry. 62, 817-836 (2003).

17. Tolstikov, V.V., Lommen, A., Nakanishi, K., Tanaka, N., and O. Fiehn. Monolithic Silica-Based Capillary Reversed-Phase Liquid Chromatography/Electrospray Mass Spectrometry for Plant Metabolomics. Anal. Chem. 75, 6737-6740 (2003).

18. Viant, M.R. Improved methods for the acquisition and interpretation of NMR metabolomic data. Biochemical and Biophysical Research Communications. 310, 943-948 (2003).

19. Viant, M.R., Rosenblum, E.S., and R.S. Tjeerdema. NMR-Based Metabolomics: A Powerful Approach for Characterizing the Effects of Environmentas Stressors on Organism Health. Environ. Sci. Technol. 37, 4982-4989 (2003).

20. Watkins, S.M., and J.B. German. Toward the implementation of metabolomic assessments of human health and nutrition. Curr. Opin. Biotech. 13, 512-516 (2002).

21. Weckwerth, W. Metabolomics in Systems Biology. Annu. Rev. Plant Biol. 54, 669-689 (2003).

22. Wolfender, J., Ndjoko, K., and K. Hostettmann. Liquid chromatography with ultraviolet absorbance-mass spectrometric detection and with nuclear magnetic resonance spectroscopy: a powerful combination for the on-line structural investigation of plant metabolites. J. Chrom. A. 1000, 437-455 (2003).