Dynamic Metabolomics for Exploring a Novel View of Metabolic Network:from Development of High-throughput Techniques to Correlation Network Analysis

行平, 大地

https://doi.org/10.15017/1441324

出版情報：九州大学, 2013, 博士（農学）, 課程博士
バージョン：

Dynamic Metabolomics for Exploring a Novel View of Metabolic Network:

from Development of High-throughput Techniques to Correlation Network Analysis

Yukihira Daichi

Kyushu University

2014

Acknowledgements

This study was promoted under enthusiastic encouragements and valuable advices from my leaders. I would like to express my gratitude to Professor Hiroyuki Wariishi, Assistant Professor Hirofumi Ichinose in Metabolic Architecture Design Laboratory, and Professor Takuya Kitaoka in Bioresources Chemistry Laboratory, Kyushu Univsersity. I am also deeply grateful to have a lot of suggestions from Associate Professor Daisuke Miura, Associate Professor Yoshinori Fujimura and Dr. Daiki Setoyama in Innovation Center for Medical Redox Navigation, Kyushu University. Their assistance was indispensable throughout the study for my study. I also enjoyed good associations with students and staffs in my laboratory. My peaceful life during the PhD course was due to those in close and warm contacts with me. Sincere gratitude is expressed to my family for all the supports I have had. The financial support from Japan Society for the Promotion of Science (JSPS) is acknowledged.

Abbreviation

MS	Mass spectrometry
m/z	Mass to charge
GC	Gas chromatography
LC	Liquid chromatography
CE	Capillary electrophoresis
RT	Retention time
EI	Electron impact (ionization)
ESI	Electrospray ionization
MALDI	Matrix-assisted laser/desorption ionization
Q	Quadrupole
IT	Ion trap
TOF	Time of flight
FT-ICR	Fourier transform-ion cyclotron resonance
CID	Collision-induced dissociation
PTS	Phosphotransferase sugar uptake system
9-AA	9-Aminoacridine
F6P	Fructose 6-phosphate (hexose phosphate)
6PG	6-Phosphogluconate
GSH	Glutathione (reduced form)
dTMP	Thymidine monophosphate

CMP	Cytidine monophosphate
UMP	Uridine monophosphate
F16P	Fructose 1,6-bisphosphate
AMP	Adenosine monophosphate
GMP	Guanosine monophsphate
dTDP	Thymidine diphosphate
CDP	Cytidine diphosphate
UDP	Uridine diphosphate
ADP	Adenosine diphosphate
GDP	Guanosine diphsphate
dTTP	Thymidine triphosphate
CTP	Cytidine triphosphate
UTP	Uridine triphosphate
ATP	Adenosine triphosphate
GTP	Guanosine triphsphate
NADH	Nicotinamide adenine dinucleotide
dTDPg	Thymidine diphosphate 4-oxo-6-deoxy-glucose
UDPG	Uridine diphosphate glucose
dTDPa	Thymidine diphosphate 3-Acetamido-3,6-dideoxy-galactose
UDPGN	Uridine diphosphate N-acetylglucosamine
GSSH	Glutathione (oxydated form)
NADPH	Nicotinamide adenine dinucleotide phosphate
CoA	Coenzyme A

AcCoA	Acetyl coenzyme A
PCA	Principal component analysis
CRA	Centering resonance analysis
CA	Correspondance analysis
GGM	Gaussian graphical model

Table of Contents

List of Tables

List of Figures

Chapter 1.

Introduction for Dynamic Metabolomics

1.1 Metabolites in dynamics of cellular system

This study aimed at cultivating a systematic view for the metabolic network of microorganisms, as a model system. Systematic understanding of metabolism would allow to control, enhance and expand the metabolic system as desired. In addition to bio production, the benefit of systematic view would further influence food, material and medical science.

Experimentally, we exploited a time-course profile of metabolite abundances with an untargeted method, which could be called *dynamic metabolomics*. This section aims to provide a basic knowledge and a general background of metabolomic study, and the relevance of analyzing the dynamics of microbial metabolome. Although there are numerous types of applications with a partial to full use of metabolomics, we mention these facts with only a limited extent.

1.1.1 What is metabolite?

Metabolite is a class of organic low-molecular-weight compounds that are formed in the metabolism of a living organism and found in biofluids or cells (including tissues and organs). Small peptides such as the tripeptide glutathione are also metabolites, while polymerized compounds including proteins or nucleic acids, direct products of gene expression, are usually not considered as metabolites (Smolinska *et al.* 2012). There are two classes of metabolites: primary and secondary metabolites. The first can be found in living species with a broad distribution, and is essential component of central metabolism involved with energy production, growth, or development process. Secondary metabolites, such as

alkaloids or hormones, involve with non-essential but important biological functions specific to species (Herbert 1989). Since secondary metabolites are synthesized from a part of primary metabolites, it is still important to investigate the dynamics of primary metabolites in the study of the secondary metabolite synthesis with a systematic view.

1.1.2 Metabolome and metabolomics

Metabolome is defined as the comprehensive set of metabolites produced or present in a biosystem (Dunn *et al.* 2005). Metabolomics, a study of metabolome, is an emerging discipline following to proteomics, transcriptomics and genomics, which represent the 'omics' study of protein, mRNA and DNA, respectively. Such high-throughput and quantitative technologies are key progresses that have been leading an innovation in the life sciences, as the systems biology. (Hood 2003, van der Greef *et al.* 2007). In human, the number of primary metabolites is no more than a few thousands. However, it is still much lower than the number of genes (>30,000), RNA transcripts (>100,000) or proteins (>1,000,000) that exist in the body, while a comprehensive metabolome analysis is currently difficult. In addition, primary metabolites are largely identical across species as well as different cells in a human body. Hence, metabolomics is expected to lead to a simpler but inclusive view of biological system in organisms. This fact is one of major factors boosting the metabolomic studies. Metabolite analysis itself is of course not a new research field. The notion of metabolomics has formed on the basis of the accumulating knowledge of metabolism, advances of analytical tools and bioinformatics.

1.1.3 Metabolite as context-dependent phenotype

Metabolites are often referred to as the compound-level phenotype or the functional

endpoint, reflecting genetic traits of the organism and environmental context. A quantitative snapshot of metabolome, metabolic profile, serves as "fingerprints" that should correlate with the phenotypes (Fiehn 2001, Bernini *et al.* 2009). Metabolomics is also an important topic in quantitative biology (Griffin 2006, Oldiges *et al.* 2007, Mapelli *et al.* 2008). Due to its closeness to the phenotype, metabolic profile is presumably affected by environmental/genetic alteration, e.g. disease or gene modification, in an immediate manner. Metabolite profiling has thus received much attention as a promising tool for clinical systems biology to detect early metabolic perturbations, even before the appearance of disease symptoms (van der Greef *et al.* 2007). Recent advances in metabolomic study were encouraged by complementation of other omics study, *i.e.* genomics, transcriptomics and proteomics (Trauger *et al.* 2008, Suhre *et al.* 2011, Nicholson *et al.* 2011).

In contrast to DNA sequence as a basic and static blueprint of an organism, gene expression and following biosyntheses and bioconversions involve with dynamics. Consequently, depending on the time-scale of interest, the implication of metabolome data will undergo distinct interpretation. For example, Dunn *et al.* traced the the flux of metabolites in seconds with comparison to turnover in proteome which is measured in minutes to hours (Dunn *et al.* 2011). It has been demonstrated that metabolome dynamics can serve as great source for inferring dynamics of metabolic network both theoretically (Cakır *et al.* 2009) and experimentally (Sriyudthsak *et al.* 2013). Inferring biological network using dynamic 'omics' data can expand our view of systematic structure of metabolism. Metabolic network inference using dynamic metabolomic data has long been discussed (Sontag *et al.* 2004).

1.2 Technical basics for high-throughput metabolomics

In this section, the current state of analytical techniques for metabolomics is reviewed to clarify their advances and limitations with a perspective for high-throughput metabolomics. These aspects are particularly relevant in Chapter 2, where we conducted a development and validation of a high-throughput method for metabolite analysis using matrix-assisted laser desorption/ionization-mass spectrometry (MALDI-MS). An overview of MALDI-MS in terms of metabolite analysis is also provided in the last section.

1.2.1 How to measure metabolite abundances

Global studies of biomolecules were once limited to genes, transcripts and proteins, but technological evolutions in the past decade allowed for the untargeted profiling of metabolites as a new 'omics' study. Unlike mRNAs and proteins, which are constituted of 4 and 20 chemical building blocks, *i.e.* nucleic acids and amino acids, respectively, metabolites exhibit a high chemical diversity (*e.g.*, in molecular weight, polarity, solubility, *etc.*) because they includes various structural class of compounds ranging from amino acids to alkaloids. This fact makes it almost impossible to analyze a universe of metabolites by a single platform (Geier *et al.* 2011). Additionally, the reliable quantification of metabolites is hindered by the greatly wide range of intracellular metabolite concentrations (sub-nM to an order of 100 mM). Therefore, multimodal analytical techniques have been developed to analyze metabolites both quantitatively and qualitatively. Whilest the profiling study of metabolites was first reported 1950s (Goldsmith *et al.* 2010), it is only a decade since metabolomics was considered as a

separated research field (Fiehn 2001). Traditionally, quantitative analysis of metabolites in interest has been carried out using enzyme-based assays (Cook *et al.* 1978, Bergmeyer *et al.* 1985, Hajjaj *et al.* 1998). However, such assays require rather a large volume of sample despite the small amount of metabolites (less than 5%) in a cell. Furthermore, the scope of analyses is limited to single or a few metabolites per sample.

In modern metabolomics, there are two major analytical platforms: nuclear magnetic resonance (NMR) or MS. These tools can provide both quantitative and structural information in the scope of metabolome, though the quality of information differs. They work complementary if both available, but it can be redundant (Lindon and Nicholson 2008) Because the development of MS greatly contributed to the progress of metabolomics (As is the case of GC (Koek *et al.* 2011), we focus on MS-based metabolomic techniques. Basically, MS is an analytical system to measure the mass of ion molecule by electric or magnetic field. In addition, hyphenation (sequential connection of instruments for online analysis) of chromatographic separation is frequently employed. Since each part of equipment is customized to the posterior part, an overview of metabolomic analysis is reviewed backward, namely MS analysis, ionization and separation. Experimental aspects are also reviewed in the next section.

Regardless of the sort of MS system employed, workflow of hyphenated-MS-based metabolome analysis is typically constituted of following steps; sample preparation, online separation, quantitation by MS, raw data processing and data analysis. The sample preparation step usually includes quenching, extraction, partition and concentration. However, it should be noted that, if desired target metabolites are quantitatively detected, sample preparation steps are fundamentally not necessary. Since these steps require appreciable costs

including materials, times and labor, extensive optimization of experimental design is relevant in high-throughput metabolomics. Importantly, most of steps are inevitable for the column chromatography. This fact is encouraging to omit the chromatographic pre-separation for the higher throughput. In this section, a brief introduction of sample preparation is presented to explain the necessity of each step for conventional methods.

1.2.2 MS

MS enables analysis of diverse chemicals including bio-molecules on the basis of ion molecular mass-to-charge ratio (*m/z*). The critical properties of MS are mass resolution and quantitative dynamic range. Supporting properties are scan speed and *m/z* range, *etc*.. Practical properties include the cost of instrument and the difficulty of operation and maintenance. Accurate mass analysis is usually desired to perform putative annotation of a signal, comparing the measured *m/z* and the expected *m/z* of a specific metabolite. However, this approach may be too optimistic in the sense of metabolome because the experimental principle of metabolomics maximizes the possibility of detecting any unknown metabolites (See *Identification and database* for further discussion).

As a rough interpretation, mass resolution is a reciprocal number with which a peak is distinguishable from another peak multiplied. An insufficient resolution thus results in two or more peaks being fused, where peak quantification is hindered. The most common definition of resolution is given by $M/\Delta M$, where M corresponds to *m/z* and ΔM represents the full width at half maximum (FWHM). When a peak has an *m/z* 500 and a FWHM of 0.05, resolution is $M/\Delta M$ = 500/0.05 = 10,000. Mass accuracy is a relative difference between theoretical *m/z* and one that a MS provides, usually characterized by part per million (ppm)

unit. An instrument with 100 ppm accuracy can provide information on an ion of *m/z* 500 within ±0.05 error. It is affected by the MS's stability and resolution.

Supported by a long history of its use, single quadrupole (Q) offers a robust MS analysis with a nominal unit resolution. With a similar mechanism, quadropole ion trap (QIT) system allows capturing and fragmentation of selected ions, and provides MSn mass spectra, which offers additional structural information of ion species. The use of nominal mass spectrum could be limited to 'fingerprint' analysis, its data source is at the same time replaceable with any similar low-resolution data that offered by other instruments. Considering its lower cost and maintenance ease, QMS still possesses considerable potential for being exploited for fingerprint metabolomic study in any laboratory (Enot *et al.* 2008, Beckmann *et al.* 2008). Nevertheless, direct connections between mass spectrum and chemical entity is often crucial for rational understanding of biological systems based on the MS (Junot *et al.* 2010). Triple-Q (QqQ) MS is a powerful tandem MS system for targeted metabolomics. QqQ MS avoids the problem of low mass resolution in QMS analysis by exploiting the structural specificity of analyte, even with those effectively isobaric. Each quadrupole has a separate function: the first quadrupole (Q1) filters ions with pre-defined *m/z* (precursor). The second quadrupole (Q2) transfers the ions while introducing a collision gas for collision-induced dissociation (CID). The third quadrupole (Q3) analyze the known fragment ions (product ions) generated in Q2. Multiple pairs of precursor and product ion (transitions) per compound is desirable for more confident identification of detected peaks (Tsugawa *et al.* 2013).

A reflectron time-of-flight (TOF) MS offers higher mass resolution (3,000~15,000) and accuracy (10 ppm) and a hybrid Q-TOF MS could potentially achieve a resolving power

as high as 50,000. Arguably TOF is the most popular MS system because of its simple instrumentation as well as satisfactory performance in ordinary situations. In the context of metabolomics, however, the resolving power of TOF is still insufficient to distinguish effectively isobaric compounds (Dunn *et al.* 2005, Castrillo *et al.* 2003, Davey *et al.* 2008, Yang *et al.* 2009, Sun and Chen 2011). Fourier transform-ion cyclotron resonance-MS (FT-ICR-MS) is at the endpoint of the MS evolutions in terms of mass resolving power, mass accuracy and sensitivity (Brown *et al.* 2005).

Orbitrap, a new analytical tool in which detections occurs by measuring the frequency of the oscillating ions in electrostatic field, followed by fast Fourier transform to determine their *m/z*. Despite its simpler architecture compared to FT-ICR-MS, sensitivity, resolution and accuracy can follow FT-ICR, and are much better than Q or TOF (Makarov *et al.* 2006).

Currently, the most successful balance of performance and costs may be achieved by a quadrupole time-of-flight (QTOF)-MS or an Orbitrap MS, expanding the application field of metabolomics (Bino *et al.* 2004, Scheltema *et al.* 2008, Viant and Sommer 2012, Lu *et al.* 2010). At *m/z* 500 a current performance would be 10,000~100,000 (FWHM) with mass errors ranging from sub-ppm to 10 ppm. Such hybrid instruments are generally constructed with quadrupole and TOF mass analysers separated by a higher pressure collision cell which can be used to perform CID of selected ions. These systems have been utilized for untargeted metabolomics studies to lead novel biological insights (Wikoff *et al.* 2009, Yanes *et al.* 2010, Jain *et al.* 2012, Zhang *et al.* 2012).

Instrumental performance of MS, *e.g.* sensitivity, mass resolution and accuracy, is still making rapid progress. It is naturally assumed that the better performance leads to the better results. However, it has not been well documented which level of the performance is

required for which level of scientific readout. The upcoming generation of MS might provide an unexplored approach for metabolomics and related research fields rather than just an extension of existing methodology. Therefore, researchers and manufacturers should cooperatively share strategic aims for an advanced metabolomic study.

1.2.3 Ionization

Ionization is the most critical step in MS analysis because neutral (not ionized) compounds are fundamentally not detected in MS, and quantitative performance is strongly dependent on ionization stability. This is the reason why MS instrument itself is not well suited to quantitative analysis. Ionization yield is dependent on both the molecular species and ionization method. One metabolite may be ionized by one ionization method, but not by another, or at a totally different yield. The experimental aim in global metabolomics studies is to obtain a comprehensive, quantitative, and unbiased view of the metabolome, and a key to this goal is the ionization event.

In 1980s, two soft ionization techniques, *i.e.* electrospray ionization (ESI) (Tolstikov and Fiehn 2002, Want *et al.* 2006, Waybright *et al.* 2006, Nordström *et al.* 2006) and MALDI (Vaidyanathan *et al.* 2006, Vaidyanathan and Goodacre 2007b), opened the door for MS-based comprehensive analysis of biomolecules (Domon and Aebersold 2006). However, the history of hard ionization is much longer. Electron impact ionization (EI) (Fiehn 2001, Jellum 1977, Jonsson *et al.* 2004, Kopka 2006) has long been utilized in GC-MS analysis, and is still one of the most popular and robust ionization systems in metabolomics. Basically, EI is applicable to gas phase sample, while ESI or MALDI to liquid or solid, respectively. These characteristics diversify the utilization of MS in biomolecule analysis. Other different

ionization methods have been also probed in a metabolomics context such as atmospheric pressure chemical ionization (APCI) (Aharoni *et al.* 2002), desorption electrospray ionization (DESI) (Chen *et al.* 2006), and desorption/ionization on silicon (DIOS) (Vaidyanathan *et al.* 2005). All the above-mentioned ionization techniques discriminate differently and specifically/uniquely against certain physicochemical properties of analytes.

1.2.4 Separation methods

Chromatographic separation is employed mainly because of two distinct purposes. One is to identify the metabolites by retention time (RT) and the other is to avoid ion suppression, a phenomenon where an ionization yield is affected by the co-eluted matrix.Especially, ESI is known to suffer from ion suppression (Ikonomou *et al.* 1990). Therefore, although an extracted aliquot can be directly inject into MS, chromatographic separation is employed prior to MS analysis to enhance quantitative and sensitive measurement of ionized metabolites. Detailed description about separation techniques is out of the scope of this paper, because a high-thoughput metabolomics is the main focus, where separation step is often omitted to reduce analysis time. An overview of chromatographic techniques is presented in the following section to clarify the points that high-throughput techniques must overcome. Gas chromatography (GC) and QMS first served as a robust analytical system for metabolomics. In recent studies, liquid chromatography-MS (LC-MS) become one of the most popular analytical systems (Lisec *et al.* 2006, De Vos *et al.* 2007).

Prior to MS analysis, isolated metabolites are separated chromatographically by using relatively short solvent gradients (on the order of minutes) that allow for high-throughput analysis of large numbers of samples. The physiochemical landscape of the metabolome is

highly heterogeneous, so to increase the number of compounds detected, multiplexed methods for the extraction and separation of metabolites are used (Patti 2011). Similarly, reversed-phase chromatography is better suited for the separation of hydrophobic metabolites, whereas hydrophilic-interaction chromatography generally separates hydrophilic compounds more effectively.

GC and EI-MS is a great combination in terms of reproducibility. Thanks to the excellent peak capacity and reproducible RT of the silica column, and reproducible mass fragment pattern produced by EI, GC-MS can benefit by robust library-based identification of detected peaks, *i.e.* no reference compounds is usually not required for known metabolites. Analytes must be volatile, or otherwise trimethylsilyl derivatization is performed. Low molecular weight hydrophilic metabolites including sugar and amino acids can be analyzed once derivatized. (Jiye *et al.* 2005). Compounds with higher molecular weight cannot be volatilized even after derivatization because their original boiling points are too high. In addition, thermal decomposition also hinders the analysis. For such compounds, alternative analytical system such as LC is required.

LC-MS is capable of analyzing a wide range of metabolites by exploiting appropriate combinations of column and mobile phases. The hyphenation of LC and MS was realized by the invention of ESI, enabling ionization of the liquid eluent (Whitehouse *et al.* 1985). Although LC columns have been designed to hydrophobic compounds, recent technological advances allowed for the separation of hydrophilic metabolites. Reversed-phase ion-pair LC is also a useful method for metabolomics (Luo *et al.* 2007). In addition, ultra-performance liquid chromatography (UPLC) emerged as a significant boost for LC-MS-based metabolomics (Bruce *et al.* 2013). Employing pumps with a maximum pressure of more than

100 MPa and pressure-torelant columns, shorter measurement time, *e.g.* a few minutes, was realized. Alternatively, capillary electrophoresis mass spectrometry (CE/MS) also provides excellent separation of hydrophilic metabolite such as amino acids or intermediates of the glycolytic system enabling also excellent quantification of them (Soga and Heiger 2000, Soga *et al.* 2002, Soga *et al.* 2003, Soga *et al.* 2009, Sugimoto *et al.* 2010, Ito *et al.* 2013, Cai and Henion 1995). Although ion-pair LC-MS system is able to measure these metabolites, separation performance of CE-MS is generally superior to it. The current situation for peak annotation is, however, much similar or worse than that of LC-MS due to poor reproducibility and stability.

These metabolite 'profiling' methods demand careful control over the chromatographic process to ensure reproducibility and require significant time, effort and expertise for data pre-processing in order to deconvolve, align and annotate peaks correctly. Unfortunately, any chromatography column matrix will undergo gradual detoriation with repetative use, resulting in significant changes in data characteristics after a period of constant operation in larger (>200 samples) profiling experiments. Moreover, reproducibility of column separation is still dependent on the condition of pumps and columns used, including vendor lot. This makes it extremely difficult to develop a generally useful database of metabolite detection in LC-MS analysis. Users therefore annotate the chromatographic peaks using standard compounds. Whilst a careful operation is still necessary, in-house libraries of RT using standard compounds are also frequently developed. Although separation of chemicals is the heart of analytical chemistry, chromatographic techniques inevitably pose daunting drawbacks to the researchers, especially in non-targeted analyses. Hence, non-chromatographic technique is desirable for high-throughput metabolomics if possible.

1.2.5 Sampling and preparation

Quenching is an immediate arrest of cellular metabolism, namely enzymatic reaction, through inactivation and denaturalization of enzymes without loss of metabolites. Cold methanol is routinely employed because of its excellent inactivation effect of enzymes. Another choice is rapid and ambient washing followed by freezing by liquid nitrogen, which is suitable for samples intolerant to organic solvents. There are many variants of methods customized to sample of interest (Faijes *et al.* 2007, Canelas *et al.* 2008). The subsequent steps are performed strictly at a low temperature to avoid any metabolic reaction by remaining enzymatic activity. Additionally, rapid sampling techniques have been developed with consideration for the high turnover rates of intracellular metabolites such as G6P, ATP or citrate, which are usually in the order of 1–2 s (Weibel *et al.* 1974, de and van 1992). Such methods aims at collecting biological samples represent *in vivo* conditions as closely as possible.

The principal aim of extraction is to isolate metabolite from biological samples with a maximum recovery and without chemical alteration or degradation. Extraction is frequently accompanied with cell disruption, especially for cell wall. The reproducibility of extraction could be enhanced by extensive inactivation of enzymatic activity as the quenching did (Villas-Boas *et al.* 2005, Wittmann *et al.* 2004, Winder *et al.* 2008). Biphasic extraction is frequently performed to precipitate protein and cell debris, and partitioning the metabolite based on their polarity (Want *et al.* 2006, Yanes *et al.* 2011). As one ionization method may be suited to the compounds with a certain degree of polarity, appropriate partitioning of metabolites is desirable to avoid precipitation of non-ionized substances on the ion source. In another word, such laborious conditioning could be omitted if the instrument is tolerant to the

crudeness of samples. Whilst chromatographic techniques are vulnerable to the crudeness, off-line techniques such as LDIs could overcome the situation.

The importance of rapid sampling techniques has long been discussed in tems of the metabolite dynamics (Cole *et al.* 1967). Integration and automation contributed to decreasing operation cost and manual-handling errors. Weibel et al. introduced a method combining rapid sampling and automated analysis was introduced (Weibel *et al.* 1974) for metabolic profiling of yeast cells. The system was readily refined for minute-scale (Gonzalez *et al.* 1997) or sub-minute scale metabolite analysis (Sáez and Lagunas 1976, Theobald *et al.* 1993, Theobald *et al.* 1997, Visser *et al.* 2002, Mashego *et al.* 2007).

Metabolomic study of sub-second time scale has been achieved by Buchholz et al. (Buchholz *et al.* 2002), They performed metabolome analysis and kinetics modeling of metabolic perturbation against nutritional pulse in *E. coli* cells. The analytical system is equipped with automated sampling device introduced by Schaefer et al., enabling harvesting with intervals of 0.22 s per sample of culture media (Schaefer *et al.* 1999). Although their analytical method itself was not metabolomic, *in vivo* kinetics of glycolysis was investigated. A metabolomic study with a time scale of 100-millisecond was also reported by using a stopped-flow sampling technique (Buziol *et al.* 2002, Chassagnole *et al.* 2002).

1.3 Treatment of a numerous series of MS data

In this section, an overview of raw MS data treatment is presented. It is characteristic to dynamic metabolomics that hundreds of MS spectra are generated per every single experiment. Since a high-throughput metabolome analysis employs no pre-separation techniques, further consideration is required to annotate the MS signals: although chromatogram alignment is thus avoided, greater dependence on MS data quality poses more severe technical difficulties compared to an ordinary workflow of metabolomic data processing. In addition, identification process is actually the bottleneck for further progress of metabolomics. High-throughput analysis would succeed only when efficient identification was achieved.

1.3.1 Raw data processing

The chromatogram of MS analysis is subjected to elaborate processing workflows including peak-picking (Dixon *et al.* 2006), deconvolution (Jonsson *et al.* 2004, Jonsson *et al.* 2005) and alignment.

An important precaution is that one metabolite compound could lead to more than one peak derived from monoisotopic ion, isotope, adduct or ion product generated during the ionization process. The mono isotopic peak is generally the interest, and systematic methods to distinguish mono isotopic peaks in a mass spectrum have been reported (Werner *et al.* 2008, Matsuda *et al.* 2009). In contrast, isotopic peaks also convey indispensable information for identification of unknown compounds (Miura *et al.* 2010b), which is discussed in the

following session.

Once the signal intensities are aligned into a tabulated table, subsequent steps include scaling and transformation of the variables The processed variable table is then applied to some statistical methods that meet the purpose of the study. It has been indicated that variable processing may have more profound effects on the results of analysis and conclusive interpretation than the employed statistical methods (van den Berg *et al.* 2006). This technical aspect encourages a standardized data processing and reporting in metabolomics, even though it is still arguable whether a consensus will be established.

1.3.2 Identification and database

Identity of detected ion species can be roughly classified into five levels: ion formation, presence of certain elements, elemental composition, topological chemical structure and stereochemical structure. Unambiguous determination of stereochemical structure can be done only by NMR. In hyphenated-MS analysis, chemical structures are estimated based on the probability of RT, MS and MS^n data (Rojas-Chertó *et al.* 2011). Even if such comparative information can provide true-positive estimation, it should be noted that false-positives cannot be completely excluded, resulting in numbers of candidates. In metabolome studies, MS-based metabolite identification may thus frequently be mentioned in a wide sense because empirical identity is often satisfactory for systematic interpretation of the acquired data. Due to this ambiguity, there are currently few software programs available that can exactly tell metabolite identifications. Nevertheless, exploiting accurate mass information, peak annotation methods and databases have experienced continuous progresses (Kind and Fiehn 2010, Iijima *et al.* 2008, Brown *et al.* 2011, Wishart 2009, Wishart 2011).

Given a monoisotopic peak derived from an organic compound is observed, its elemental composition can be deduced, though more than one answers may be listed. For example, using currently available MS instruments with the highest mass accuracy (< 1 ppm of mass error), elemental compositions are uniquely determined for most of metabolites of < 300 Da. Inversely, however, identifying metabolites > 300 Da requires exponentially higher mass resolution if single MS spectra are used, which is probably not feasible (Kind and Fiehn 2006). Therefore, CID spectra are often acquired to obtain information about the chemical structure. When the metabolite was characterized to exhibit the same RT and fragmentation pattern than those of reference compound, identification is considered formally done in most of metabolomic studies, though a strict chemical identification remains pending. In terms of elemental composition, empirical rules were also effective to select chemically plausible estimation (Kind and Fiehn 2007).

It is also important to distinguish known-unknown and unknown-unknown metabolites in the identification process. The former is a group of metabolites that are publicly known but not yet annotated on the MS data in hand. This class of metabolites can be retrieved from public metabolite databases such as the Human Metabolome Database, METLIN, KNApSAcK or MassBank (Wishart *et al.* 2009, Forsythe and Wishart 2009, Smith *et al.* 2005). Although single match of molecular mass only allows putative assignments, METLIN and MassBank contain MS/MS data that were acquired by CID experiments with different collision energies or different platforms/sites, which can enhance the confidence of identification. Unfortunately, it is known that fragmentation patterns of low-energy CID (< 1 keV) can significantly differ, affected by the instrumental configurations, even if the same parameter set is applied for the same MS product. In addition, since there are no standard

pre-processing methods for MS spectra, comparison of data derived from different research site is hindered. Furthermore, as mentioned in previous section, RT in LC-MS or CE-MS is even less reproducible. These facts may be the reason why metabolome researchers develop an in-house database of MS experiments and hesitate to publish their data. GC-MS, as the only analytical platform that can overcome these problems, would remain the best practice in metabolomics.

The latter class, unknown-unknown, is the fundamental problem in metabolomics. They are not registered in any databases, and thus cannot be identified directly by comparative searches. Although metabolite databases have grown considerably over the past decade, a substantial number of query derived from biological samples do not return any matches from any databases (Patti *et al.* 2012). Complementary experiments (*i.e.*, other sequential MS^n experiments or H/D exchanges) may help identifying the chemical structure, but these additional analyses requires intensive efforts other than those for metabolome analysis itself. Presuming that unknown-unknown metabolites still possess some similarity with known metabolites, machine-learning methods for estimating chemical substructure were attempted using fragmentation patterns of known metabolites as training data (Rojas-Cherto *et al.* 2012).

1.4 Data analysis of dynamic metabolome

Temporal dynamics can be yet another dimension of metabolomic data. This section focuses on strategies for analyzing metabolome dynamics in an unbiased manner. In Chapter 3 and 4, a network analysis of dynamic metabolome data was demonstrated.

1.4.1 Purpose of analysis

As the compound-level phenotype of genetic traits, metabolome provides a variety of ways to investigate the biological system. Whereas the fingerprinting method is a kind of pattern recognition, integrative approach has been proposed for better interpretation of metabolome (Weckwerth and Morgenthal 2005, Cakir *et al.* 2006). , Metabolome data also served as indispensable information for flux analysis or kinetic modeling of cellular metabolism (Mendes *et al.* 2005, Nikerel *et al.* 2009). In contrast, sophisticated simplification approach of metabolome data was also reported (Hageman *et al.* 2008). Since a single metabolite can be a substrate for a number of different enzymes, metabolite can serve as connective information for various metabolic pathways. This concept may turn down the top-down understanding of metabolome alterations, *i.e.* as the consequence of changes in mRNA or protein level. Inversely, metabolite should be regarded as end-point evidences for the changes in both mRNA and proteins (Ellis *et al.* 2007, Alm and Arkin 2003). Although this situation is much alike to other 'omics' studies such as transcriptomics or proteomics, several exceptions should be noted. Firstly, the metabolome data we currently face are absolutely not a comprehensive view of the true metabolome entity. We only have limited

access to some clippings of metabolome, even when detected but unknown metabolites are taken into accounted. Secondly, as mentioned just before, a considerably large part of detected peaks remains unknown in spite of tremendous efforts including instrumentation and informatics (Patti 2011). Lastly, the variances in abundance of metabolites often require context-dependent interpretation because the abundance of metabolite itself may convey a limited insight for metabolic state. For example, when a metabolite accumulates in a sample, there are *a priori* two reasons: increased in-flux or decreased out-flux. In addition, there may be multiple fluxes for the metabolite, and the regulation mechanism of the fluxes is often elusive. Specific approaches to decipher these questions are, however, mostly out of scope in the context of metabolomics, even though metabolite analysis might help.

Taken all, the anonymousness of metabolome discouraged the efforts to treat metabolomics efficiently as an extension of traditional 'omics' sciences. Targeted metabolomics is a way to largely neglect these drawbacks, and is also consequently compatible with the conventional studies involved with metabolite analysis. Once targets are posed, a number of experimental optimizations become available. However, an important paradigm of omics study is discovery of hypothesis through analyzing the biological system as a whole (Goodacre *et al.* 2004). This implies that importance of developing metabolomic data mining techniques.

1.4.2 Network thinking

Multivariate statistical modeling is essential for 'omics' studies. Starting from univariate testing (Box *et al.* 1978) and validation (Broadhurst and Kell 2006), metabolome studies employed a variety of supervised methods including partial least squares (PLS) (Wold

et al. 2001), orthogonal projections to latent structures (OPLS) (Trygg and Wold 2002), and Random Forest (Breiman 2001), and unsupervised methods including principal components analysis (PCA) (Jackson 2005), (fuzzy) cluster analysis (Li *et al.* 2009), neural networks (NN) (Taylor *et al.* 2002), support vector machines (SVM) (Mahadevan *et al.* 2008), while some methods could work in both mode of supervise. Unsupervised methods are popular pattern-recognition strategies for metabolomics, which roughly examine how similar a set of samples are to one another on the basis of their metabolite profiles. This approach was simple and has prospered because it was naturally obey 'guilt and association', where the provenance could be ultimately attributed to one gene knockout (Altshuler *et al.* 2000). However, considering the currently relevant cases where a huge number of single origin or a mixture of factors forms sample classes, alternative methods such as machine learning is indispensable (Kell and King 2000).

Barabási et al. brought about a concept of metabolic networking to the construct of the metabolic pathway (Barabasi and Oltvai 2004). As well as mRNAs and proteins, the structural properties of the metabolite network have been investigated (Wagner and Fell 2001). Metabolomic network analysis was thought to be relevant because the distribution of metabolites through the metabolic pathway (or flux) was not accessible by analyses at mRNA or protein level. Correlation analysis of metabolites is one approach, which was utilized for discovering novel pathways (Weckwerth and Fiehn 2002) and inferring unknown metabolic networks (Steuer *et al.* 2003). It has been suggested that such correlation networks were occasionally related to the known biochemical network, but sometimes not.

Again, the metabolic pathway known today is the reconstruction of knowledge in biochemistry. Genome sequencing has recently been integrated into the reconstruction,

predominantly for microorganisms yet, leading to genome-scale models (GEMs). Systematic description of metabolic pathway opened the door for model-based systems biology of metabolism (Frazier *et al.* 2003, Price *et al.* 2004). Especially, such systematic representation of the whole metabolic pathway realized also systematic analysis of biomolecular networks. The first GEM is a reconstructed model for *Haemophilus influenza* published in 1999 (Edwards 1999). Since then, a number of GEMs have increasingly been reported (Vidal *et al.* 2011, Zhuang *et al.* 2011, Chang *et al.* 2011). Of all organisms that have been analyzed through a constraint-based metabolic reconstruction, *E. coli* has gained the most attention as a model organism (Feist and Palsson 2008). In the present study, a recently reported GEM of *E. coli* (Orth *et al.* 2011) was utilized to characterize the metabolite correlation network constructed by using dynamic metabolome data.

1.5 MALDI-MS

Overall, pre-separation of metabolome sample contributes to reproducible ionization and identification. Identification is further supported by MS^n analysis. From the view of high-throughput, major drawback of chromatographic separation is laborious sample pre-processing and analysis time itself. Technical disadvantage includes that maintaining separation quality requires highly skilled manpower, and that subsequent data processing becomes rather complicated. Furthermore, since simple and comprehensive software for MS data processing is currently absent, appropriate peak alignment and annotation require considerable time, labor and expertise. However, high-throughput methods should naturally be fast and easy. At the same time, such method is compatible to identification process and less affected by ion suppression.

MALDI-MS is the key technology for the future high-throughput metabolomics, as well as for the present study. Considering its distinct properties compared to other popular analytical techniques for metabolomics, this section dedicates to present a background for quantitative analysis of low-molecular-weight compounds by MALDI-MS.

1.5.1 Why not for low-molecular-weight compounds?

MALDI advanced the LDI techniques toward a MS-based biomolecular analysis using matrix compounds that mediated the energy transfer, which circumvented stark fragmentation of even low molecular weight organic molecules observed in LDI analysis (Glish *et al.* 1989, Schlag *et al.* 1992, Alexander *et al.* 1993). In typical MALDI analysis, the

sample is co-crystalized into a dried-droplet spot with an excess amount of solid *matrix* compound, which facilitate the soft ionization, embedding the cellular biomolecules. A rapid desorption/ionization is induced by using a pulse of radiation from an ultraviolet, visible or infrared laser, generating positively or negatively charged ions.

MALDI is applicable to various fragile and non-volatile samples including biomolecules (Castro *et al.* 1992) such as protein/peptide (Mustafa *et al.* 2007, Strupat *et al.* 1991, Karas and Hillenkamp 1988, Karas *et al.* 1989, Karas *et al.* 1990), oligosaccharides (Stahl *et al.* 1991, Franz *et al.* 2001) and synthetic polymers (Bahr *et al.* 1992, Danis *et al.* 1992, Dey *et al.* 1995) such as dendrimers (Li *et al.* 2006) and other macromolecules (Senko and McLafferty 1994). Unlike ESI, MALDI produces predominantly simple-charged ions, which allows an MS analysis with rather moderate mass resolution (like TOF) as well as more straightforward interpretation (Keller and Li 2001). MALDI thus enjoyed its Renaissance with the combination of TOF MS (Schriemer and Li 1996). Other advantages of MALDI are very high absolute sensitivity and tolerance for contamination and buffers (Keller and Li 2001), which enable an analysis of rather crude samples, *e.g.* direct whole-cell analysis (Lay 2001, Welker and Moore 2011).

Despite a number of advantages, MALDI-MS analysis for low-molecular-weight compounds has been hindered by as well a number of disadvantages. Firstly, the mass resolution of linear TOF-MS instruments in the early generation was too low to characterize such small molecules. In addition, severe interference of ion peaks derived from conventional matrices made the MALDI-MS analysis in the low mass range less attractive. Considering these factors, the analysis of low-molecular-weight compounds was predominantly performed using ESI.

However, a higher mass resolution has recent been achieved by improvement of TOF MS system and utilization of delayed extraction (Brown and Lennon 1995, Vestal *et al.* 1995). FTICR-MS is also suitable to MALDI, because the mass spectrum is generated with an individual shot of the laser, unlike ESI with continuous outlet and ionization (Hager 2004). Quantitative analysis of low molecular weight metabolites is one of the most major topic in metabolomics (Brown *et al.* 2005). Nowadays, MALDI is coupled a various MS instruments including TOF/TOF-MS (tandem TOF/TOF mass spectrometer) (Yergey *et al.* 2002, Fagerquist *et al.* 2010, Trimpin *et al.* 2007), QTOF-MS (quadrupole-time of flight mass spectrometer) (Hunnam *et al.* 2001), and QIT-TOF-MS (quadrupole ion trap time-of-flight mass spectrometer) (Suzuki *et al.* 2006), allowing extensive CID analysis, which is indispensable to peak annotation in non-chromatographic techniques like MALDI (Vestal and Campbell 2005). Hence, MALDI-MS is considered as a complement to other analytical techniques for low-molecular-weight compounds (Cohen and Gusev 2002, van Kampen *et al.* 2011).

1.5.2 Current uses in metabolomics

Application of MALDI-MS in drug discovery and biotechnology typically involves a target approach (Cohen and Gusev 2002, van Kampen *et al.* 2011, Wang *et al.* 2006). As well as high-molecular-weight molecules such as proteins or peptides, secondary metabolites was analyzed coincided with bacterial identification, *e.g.* microbial toxins (Erhard *et al.* 1997) or pigments in organelles (Persson *et al.* 2000). Nevertheless, metabolomic application of MALDI-MS is even recent, mainly because of the matrix ion interference on the low-mass range of the mass spectra. The most significant progress in metabolomic application was thus

matrix development including ionic liquid matrices (Vaidyanathan *et al.* 2006) or novel matrices such as 9-aminoacridine (9-AA) (Edwards and Kennedy 2005, Miura *et al.* 2010b).

LDI-based analysis such as MALDI also possesses its advantage in the two-dimensional visualization of molecular distribution in a section of biological tissue samples for remarkably novel insights into metabolism in higher organisms (Miura *et al.* 2010a, Miura *et al.* 2012, Svatoš 2010). MALDI-MS enabled the imaging MS (IMS) analysis of the endogenous and exogenous metabolic compounds including a drug and first-pass metabolites in whole body sections of animals (Miura *et al.* 2010a) and plant (Zaima *et al.* 2010). Since *in situ* metabolite identification in IMS usually cannot benefit from separation techniques, sophisticated identification methods based sorely on MS analysis are indispensable in MALDI-MS-based metabolomics.

In the present study, we utilized MALDI-MS for dynamic metabolomics. In Chapter 2, basic characteristics of the MALDI-MS-based high-throughput method for metabolite analysis were examined to show how the technique was applicable to tracing the dynamics of intracellular metabolites in bacterial cells. Since this method enabled facile acquisition of a series of time-course MS data, which were derived from hundreds of biological sample, we conducted a study for the dynamics of bacterial intracellular metabolism with a concept of correlation network in Chapter 3. We characterized metabolic responses to a nutritional fluctuation prior to transcriptional alteration. In Chapter 4, this approach was further applied to investigate the structural consensus of metabolite correlation networks under variety of nutritional fluctuations. Additionally, we attempted to expand the fundamental usability of MALDI-MS-based metabolomics through developing numerical models for MALDI events in Chapter 5. A quantitative structure-property relationship (QSPR) approach was employed

to elucidate the structural compatibility between analyte compounds and a matrix compound.

Chapter 2.

MALDI-MS-based High-throughput Metabolite Analysis for Intracellular Metabolic Dynamics

2.1 Introduction

Systems biology aims to represent and understand biology at a global scale where biological functions are recognized as a result of complex mechanisms (Kitano 2002). The whole cellular system is expected to be represented by an *in silico* model reconstituted by combining information about every molecular step in the system (Görke *et al.* 2010, Henry *et al.* 2010). Once all the system components are thoroughly understood, the dynamic behavior of the cellular system can be predicted through the reconstructed model. To accomplish this goal, it employs concepts from a wide range of fields such as mathematics, physics, engineering, and computer science besides biological science. The "building blocks" of systems biology models are knowledge and data acquired in biological experiment, and mathematical modeling provides the "cement" that links these "building blocks" (Kherlopian *et al.* 2008). In addition, analyses on dynamic behaviors of the molecular network are required because the cellular system cannot be understood through a static network structure alone (Kitano 2002). In this regard, reverse engineering using high-throughput experimental data and mathematical theories to infer underlying biological networks is the most challenging issues in systems biology (Katagiri 2003). However, large-scaled metabolomic analysis has been hindered because of in part the poor throughput of conventional analytical methods. The major hindrances of analytical throughput are not only analytical time consumption itself but also complicated pre-treatment processes including sample extraction, derivatization, filtration and concentration. These requirements are derived from low sensitivity of MS. To attack the problem, we introduce a high-throughput method for

metabolite analysis using MALDI-MS, which has been successfully applied to metabolic profiling analysis (Miura *et al.* 2010a). Using the analytical method, we then demonstrate its application for analyzing dynamics of metabolic system.

Metabolites represent the final downstream products of gene expression and cellular regulatory processes, and changes in metabolic levels may be regarded as the ultimate response of biological systems to environmental variations (Fiehn *et al.* 2000). There is significant research interest in understanding the organization and regulation of microbial metabolism[2]. In particular, the central metabolic pathway is often a major research target. This is because the central metabolic pathway represents a critical component of cellular metabolism that is responsible for both anabolic and catabolic functions that provide cofactors and building blocks for the synthesis of other macromolecules as well as energy production. Consequently, it provides indispensable data for metabolic engineering and metabolomics to characterize this pathway by quantifying time-dependent changes in the concentrations of metabolic intermediates and their corresponding cofactors present in the central metabolism pathway. Such quantification provides an approach to estimate the major metabolic processes of microbes under various environmental conditions.

For metabolite analysis, a wide variety of analytical methods including enzymatic assays (Theobald *et al.* 1997), liquid chromatography-MS (LC-MS) (Luo *et al.* 2007), GC-MS (Pasikanti *et al.* 2008), and NMR (Slupsky *et al.* 2007) have been employed. As phosphorylated intermediates in the central metabolic pathways share similarities in structure, polarity and non-characteristic UV absorption, MS is decisively employed for intracellular metabolite studies. However, it is difficult to quantify phosphorylated metabolic intermediates because they exist at such low concentrations in cells and are readily degraded

during sample preparation due to structural instability. Consequently, high-throughput sample preparation combined with a highly sensitive analytical method is essential. Although LC-MS and GC-MS, which are conventional methods for intracellular metabolite analysis, have proven to be powerful approaches for the simultaneous determination of different metabolites in a single run, both methods have relatively low throughput. The dominant factors that hinder the throughput of a whole analytical process are not only analytical time consumption but also the complications associated with pre-treatment processes including sample extraction, derivatization, filtration and concentration (when required).

Recently, MSLDI-MS has come to be applied for metabolite analysis (Becher *et al.* 2008, Shroff *et al.* 2007a, Sun *et al.* 2007, Vaidyanathan and Goodacre 2007a). MALDI is a direct ionization method that is characteristically a high-throughput technique. Compared with other MS-based analytical methods, MALDI-MS provides one of the most sensitive analysis tools and it requires absolutely minimal amounts of sample volume, *i.e.* sub-micro liter amounts or less (Amantonico *et al.* 2008a). We have recently developed a highly sensitive and high throughput metabolic profiling technique for cultured mammalian cells with MALDI-MS using 9-aminoacridine (9-AA) as a matrix (Miura *et al.* 2010a). We have also developed a high throughput sample preparation method for MALDI-MS that can quench cellular metabolism, extract intracellular metabolites and co-crystallize the matrix with metabolites in parallel.

In the present study, the quantitative performance of the MALDI-MS-based metabolite analysis was initially examined. The utility of this method for tracing intracellular metabolic dynamics of bacteria was subsequently investigated. As a model system, the time-dependent metabolite change during environmental carbon source perturbation

following the rapid relief from glucose limitation in *E. coli* was observed.

2.2 Results and Discussion

Analytical throughput is a critical aspect in particular research scenarios. Studies that involve samples with innate variances (e.g., biofluids of humans) require the analysis of a sufficiently large number of samples as possibly the only way to accurately assess the biological variation. For another example, monitoring the intracellular metabolism during bioproduction, such as bacterial fermentation, requires an on-time illustration of how the bacterial metabolism proceeds. The use of a developed high-throughput method, thus, possesses the potential to realize large-scale analysis that deals with tens of thousands of samples and real-time monitoring of intracellular metabolism. Such studies are not practically feasible using other analytical methods such as LC−MS. Although LC separation would reduce ion suppression effect, the developed MALDI−MS analysis exhibited high sensitivity and fair quantitative performance even when biological samples were analyzed. Additionally, the developed method involves only an *m/z* alignment while LC−MS usually includes a retention time alignment that requires a carefully arranged quality control during the analyses. While LC-separated information is of course not available in this system, this workflow can simplify data processing, which is particularly crucial when characterizing tens of thousands of samples. To further optimize the high-throughput nature of MALDI, biological samples were collected using minimal operations. These operations included a simplified extraction process and no concentration of the sample. For sampling, 5 μL of the cell suspension was released into the matrix solution (100% methanol), serving both as an immediate quenching

and extraction agent. Here, the bacterial cell would be disrupted when the cell suspension was mixed into the matrix/methanol solution resulting in the release of the intracellular metabolites into the extracellular environment. The released metabolites would readily mix with the matrix. The intact solution was then directly applied for MALDI−MS analysis. With only minimal sample preparation requiring a few seconds and within approximately 2 min of MALDI-MS analysis per sample, over a hundred of intracellular metabolites including target phosphorylated compounds could successfully be detected in a single analysis (Figure 2.1). The culture medium should not contain the compounds that possibly exist in the cell, because living cells are collected with the medium in this method. In this study, an extracellular environment was comprised only of water to eliminate any of factors that would interfere with the produced mass spectra. While the synthetic mineral medium or PBS buffer could be suitable to the system, pure water was employed because they affected the produced mass spectra with a rather higher background (Figure 2.2). The osmotic stress on the cell was not so significant that the metabolites spread to an extracellular environment (Figure 2.3). The high-throughput sampling method developed here, thus, provided a “crude” cell extract. To evaluate the sample quality for MALDI−MS analysis, mass spectra acquired from these crude samples were compared with samples that had undergone extensive extraction and were considered as “clean” samples, (Maharjan and Ferenci 2003) which would otherwise be suitable for LC−MS analysis (Luo *et al.* 2007). An identical cell suspension was collected at the same time and subjected to either the high-throughput sampling method or the conventional method using cold methanol. The samples were then analyzed by MALDI−MS under the same instrument conditions. Despite that direct cell analysis might negatively influence ionization efficiency due to the heterogeneous nature of the sample, nearly

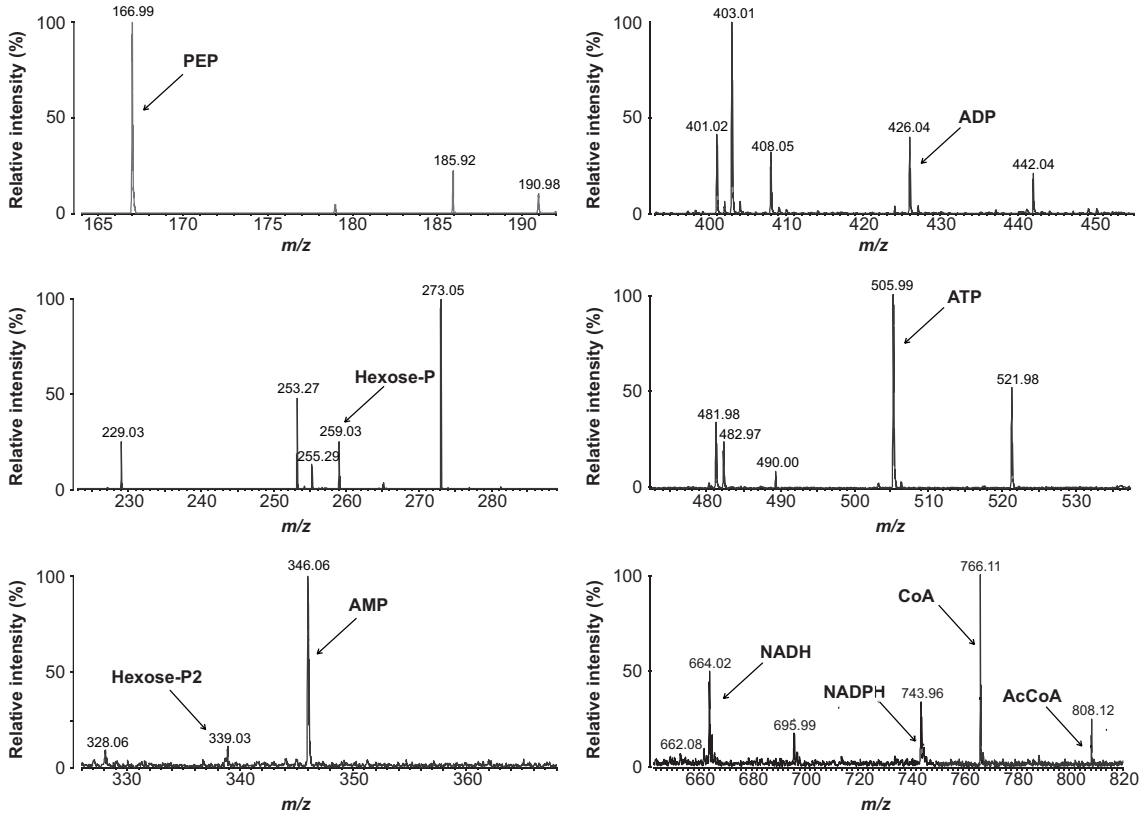


Figure 2.1 Mass spectra acquired by direct detection of metabolic intermediates and corresponding cofactors in central metabolic pathway from whole *E. coli* cells.

Mass spectra were obtained by analyzing 1 μL of the mixture of *E. coli* cell suspension and the matrix/metahnol solution on AXIMA Confidence in negative ion mode. Phosphorylated metabolic intermediates and corresponding cofactors representative of central metabolism were sensitively detected. PEP: Phosphoenolpyruvate, Hexose-P: Hexose phosphate, Hexose-P2: Hexose bisphosphate, AcCoA: Acetyl-CoA.

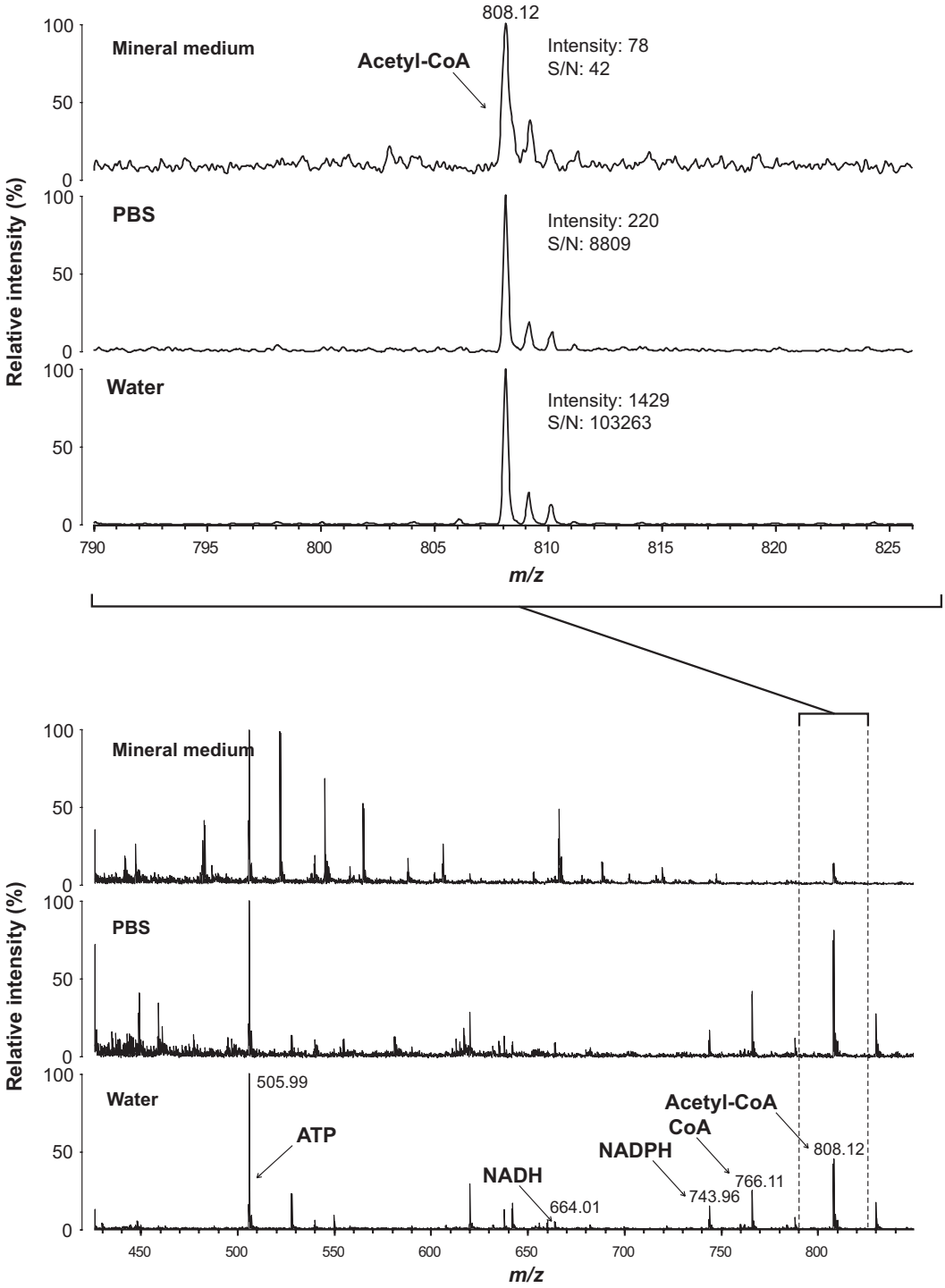


Figure 2.2 Mass spectra acquired by direct analysis of *E. coli* suspended either in water, in PBS buffer or in mineral medium.

Mass spectra were acquired by direct MALDI-MS analysis of *E. coli* cells suspended either in water, PBS buffer or synthetic mineral medium. The mass spectra acquired on each condition were aligned and magnified around the peak derived from acetyl-CoA. Single to noise ratio decreased when the mineral medium (upper mass spectrum) or PBS buffer (middle) was employed compared to one with pure water (lower).

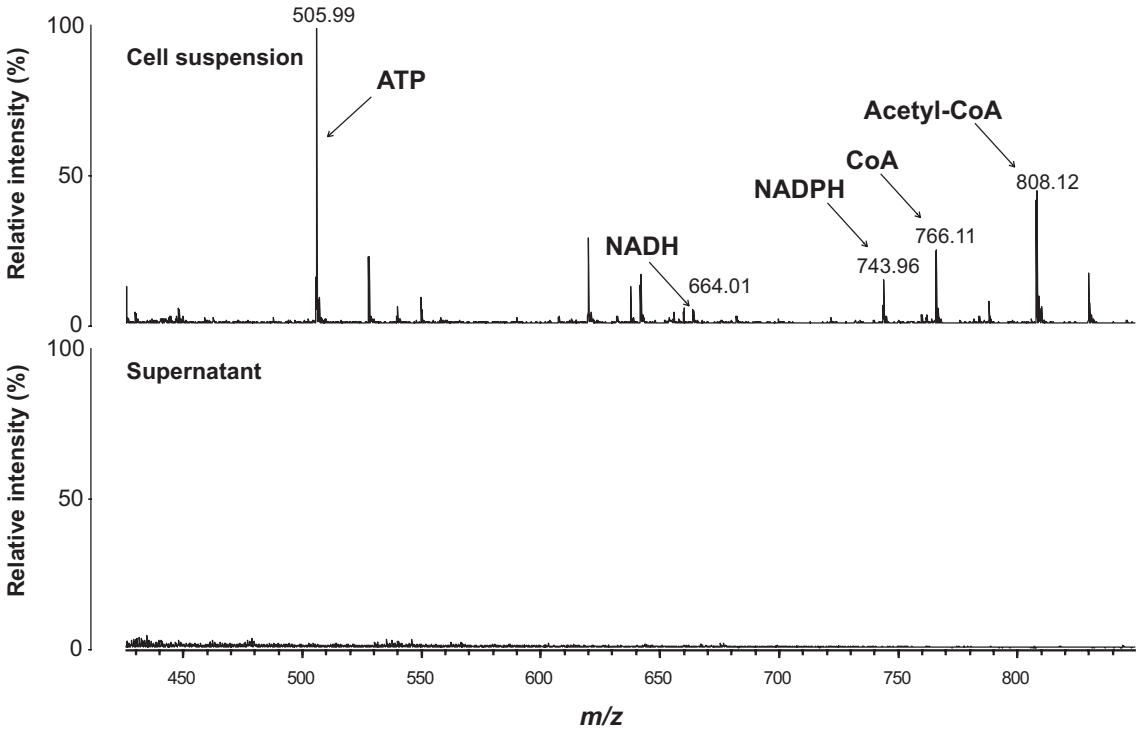


Figure 2.3 Mass spectra acquired by direct analysis of *E. coli* or the supernatant of the cell suspension after inducing glucose depletion.

The cells suspended in water for an hour were subjected to centrifugation and the supernatant was collected. The collected supernatant was analyzed by MALDI-MS in the same method as direct cell analysis. The mass spectrum acquired by direct cell analysis of identical *E. coli* cells (upper) was aligned with one from the supernatant (lower).

equivalent or more significant mass responses of the targeted phosphorylated metabolites were observed with the "crude" sample when compared with the data collected on the clean sample (Figure 2.4). These phosphorylated metabolites were considered to have been degraded or were not fully recovered during the more extensive extraction process required for the preparation of the "clean" sample. Consequently, in addition to the high throughput nature of this method, this observation exemplified another advantage that this fast and simple sampling method should minimize the analytical variations that originate in a series of experimental steps that are required for other MS-based methods.

Chromatographic separation and *m/z*-based separation function complementarily for the identification of detected product ions. Whilst MALDI-MS analysis omits chromatographic separation, which on the other hand contributes to the high-throughput of the analysis, MS/MS spectra and highly accurate *m/z* measurements should provide sufficient information to identify the elemental compositions of the product ions particularly in a low-mass range ($m/z < 1{,}000$). In this study, the product ions were identified according to their MS/MS pattern acquired in the MALDI-TOF-MS analysis. The reliability of the metabolite identification is further strengthened using FT-ICR-MS analysis, which realizes both a high mass-resolving power and a high mass-accuracy. As a result of the MALDI-FT-ICR-MS analysis, all the target metabolites that were observed in the MALDI-TOF-MS analysis were confirmed. In this study, over a hundred of peaks other than the targeted metabolites were detected, suggesting that this method should be applicable for non-target metabolite analysis.

As MALDI-MS has been primarily used for qualitative analysis of macromolecules, its quantitative performance has not been extensively examined. To investigate

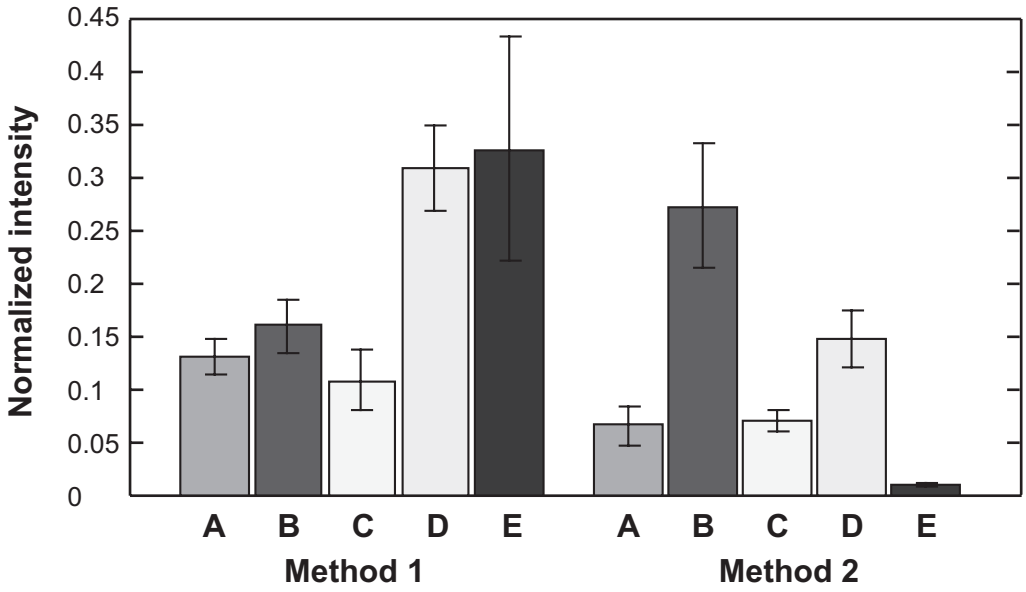


Figure 2.4 Mass responses of representative target phosphorylated metabolites under two different extraction methods.

A: Hexose phosphate, B: AMP, C: ADP, D: ATP, E: Acetyl-CoA. Method 1 indicates the rapid extraction developed in this study while Method 2 is the methanol extraction previously reported. Error bar indicates standard deviation. In the case of target phosphorylated metabolites, nearly equivalent or more significant mass responses were observed with Method 1.

time-dependent changes of intracellular metabolite concentrations, the quantitative performance of the MALDI-MS analysis approach and inter-assay precision was confirmed by spiking the metabolites into diluted (5 times with deionized water) cell extracts for the standard addition method. Raw MALDI-MS spectra could not be quantitatively compared because of analytical variance. Inter-sample reproducibility of MALDI-MS is known to be quite low (around 50% RSD) (Edwards and Kennedy 2005) compared with GC-MS (around 10% RSD) (Fiehn *et al.* 2000). Among several normalization strategies available, we decided to perform TIC normalization (Norris *et al.* 2007). The normalization resulted in good linear relationships between the deposited concentrations of metabolite standards and mass responses (Figure 2.5). While some deviations in peak intensities were observed, the quantitative performance of this analytical method was considered sufficient for a rough illustration of intracellular metabolite dynamics.

The rapid glucose relief to *E. coli* resulted in a dramatic time-dependent change of the mass spectral response of intracellular metabolites and corresponding cofactors (Figure 2.6). The observed time-dependent changes of target metabolite concentrations were mapped onto summary central metabolism pathways; glycolysis and the pentose phosphate pathway of *E. coli* (Figure 2.7). The most significant changes observed were the increasing levels of hexose phosphate, hexose bisphosphate and acetyl CoA. In contrast, the level of phosphoenolpyruvate (PEP) dropped after the relief from glucose depletion followed by a rather fast recovery. This metabolic behavior by the bacteria is most likely due to PEP being used initially for the phosphorylation of glucose resulting in the production of pyruvate (phosphoenolpyruvate:glucose phosphotransferase system) (Kaback 1968). Within the first hour of limited glucose availability, the bacterial cells respond to the glucose limitation by

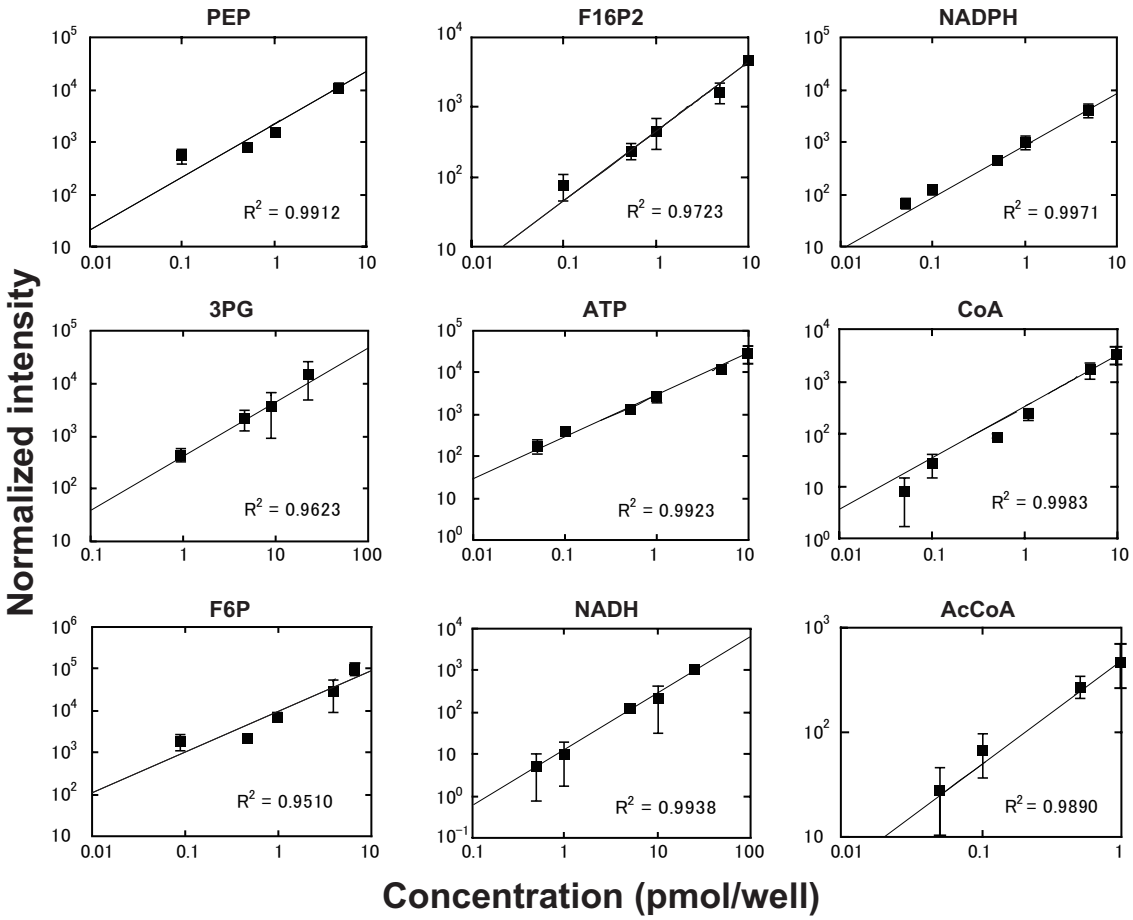


Figure 2.5 Calibration curves of targeted phosphorylated metabolic intermediates and corresponding cofactors.

The mass spectra were obtained by analyzing mixture of metabolite standards dissolved in diluted (5:1) cell extract on MALDI-TOF-MS in negative ion mode. Ion intensity was normalized to the total ion intensity of each analysis. Individual mass spectrum of metabolites was obtained by averaging 121 subspectra (5 shots per subspectra) and six mass spectra were averaged per sample. Error bars indicate standard deviation of analyses on the replicated sample spots. R^2 indicates the coefficient of determination. Fine linearity could be observed from 0.1 to 10 pmol/well in most cases. PEP: Phosphoenolpyruvate, 3PG: 3-phosphoglycerate, F6P: Fructose phosphate, F16P2: Fructose 1,6-bisphosphate, AcCoA: Acetyl-CoA.

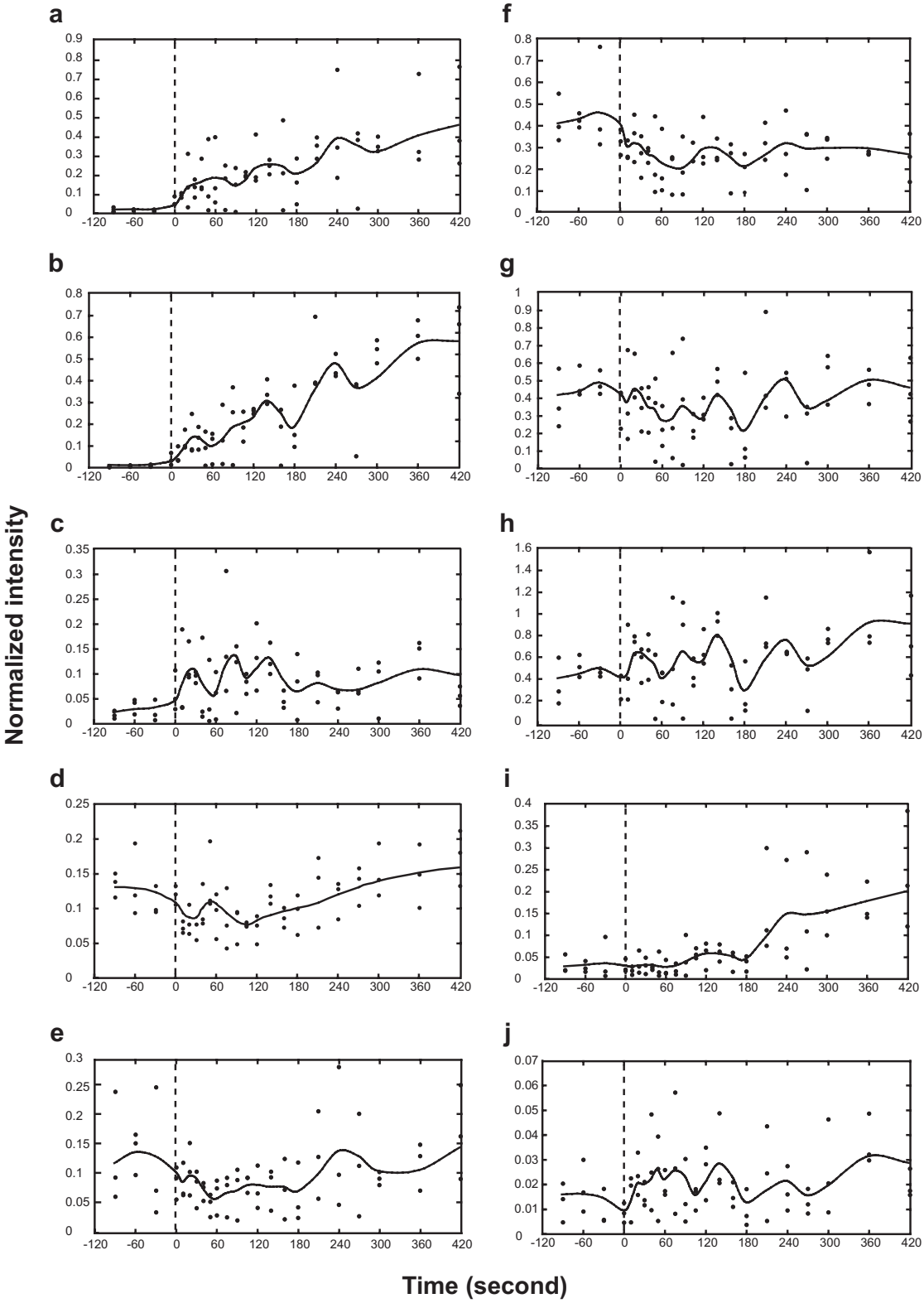


Figure 2.6 Time-dependent change of concentration of intracellular metabolites in *E. coli* before and after a carbon source perturbation.

Relief from glucose limitation was caused at time 0 indicated by a broken line. The plot originated from three experiments operated independently. Solid curves indicate moving average of mean value of the experiments operated at each time point. (a) Hexose

phosphate, (b) Hexose bisphosphate, (c) Acetyl-CoA, (d) PEP, (e) cAMP, (f) AMP, (g) ADP, (h) ATP, (i) 6-Phosphogluconate, (j) NADPH.

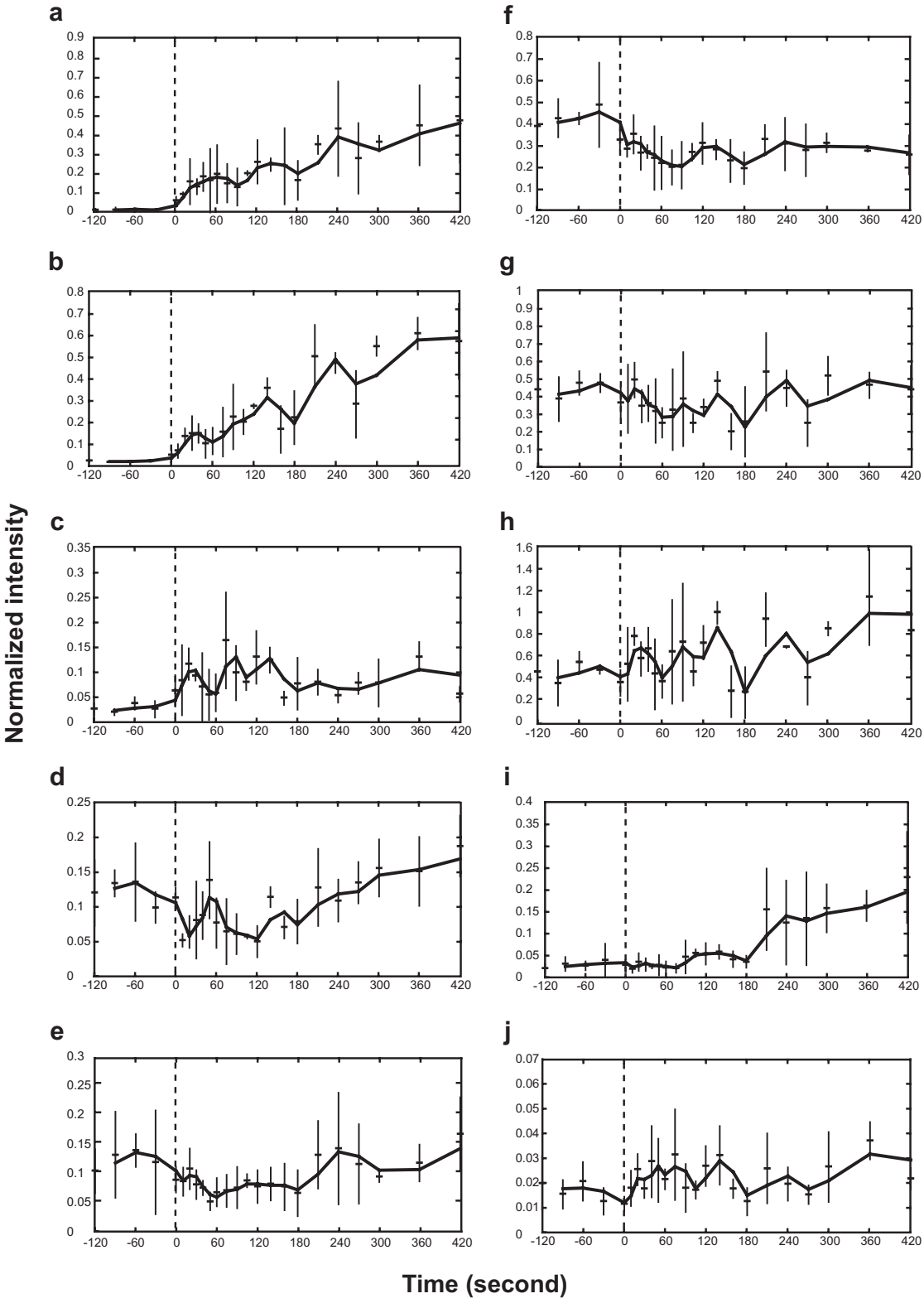


Figure 2.6 Time-dependent change of concentration of intracellular metabolites in *E. coli* before and after a carbon source perturbation. (Another version)

Error bars indicate SD.

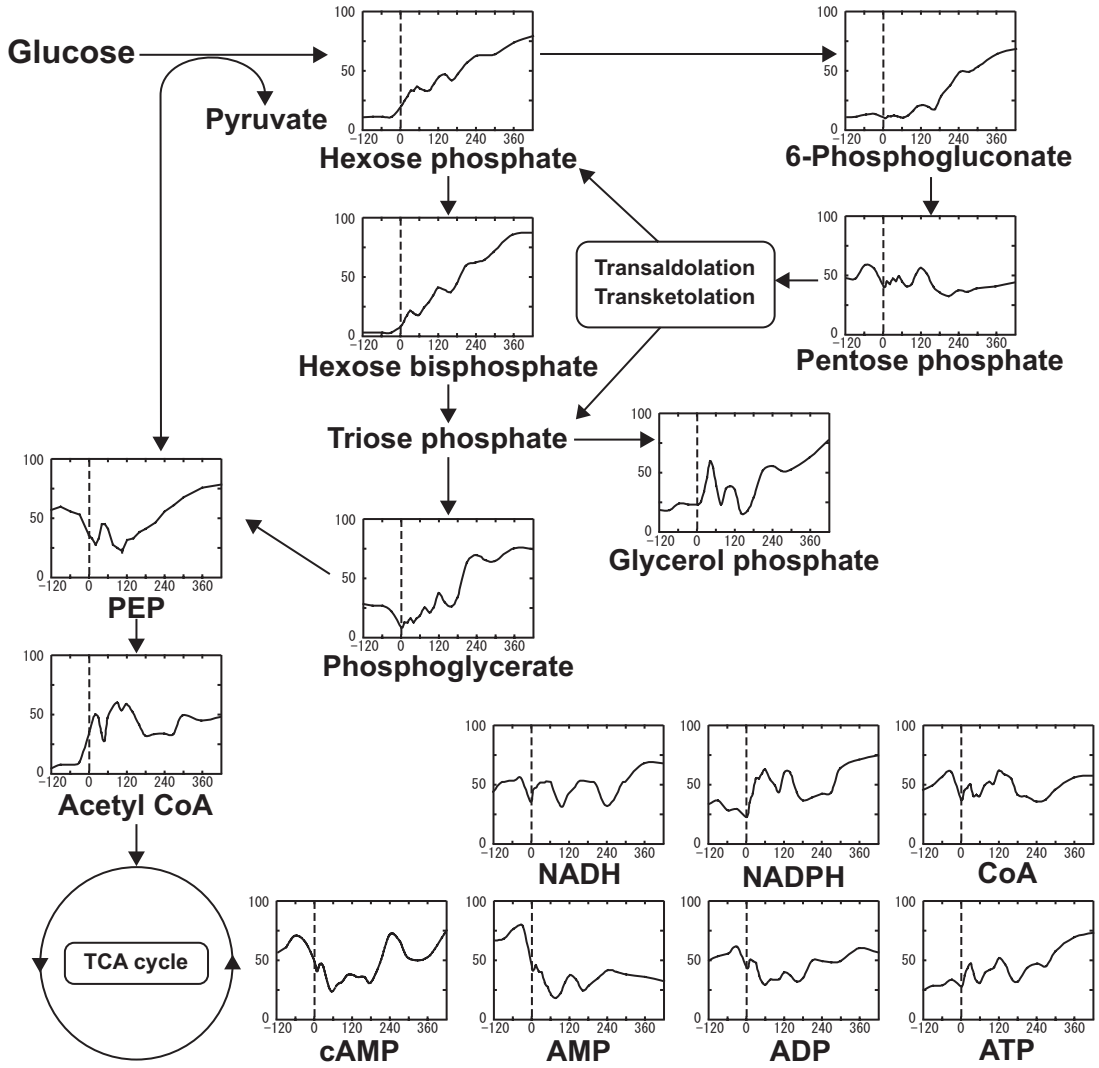


Figure 2.7 Time-dependent change of concentration of intracellular metabolites in bacteria mapped on summary central metabolism pathway, glycolysis and pentose phosphate pathway.

Relative ion intensity of each metabolite was plotted as a function of time (s). Instant relief from glucose limitation was caused at time 0.

making a general effort to increase the ability to scavenge and utilize different carbon/energy substrates (Wick *et al.* 2001). Similarly, lower concentrations of environmental glucose accelerated the levels of cyclic adenosine 5'-monophosphate (cAMP) which induces the sugar transport systems that improve the scavenging potential for glucose or other carbon sources (Ferenci 1996). This observation is supported by cAMP levels immediately decreasing after the glucose pulse. Central metabolism, especially glycolysis in bacteria could be regulated by ATP demand rather than the relationship among its intermediates (Koebmann *et al.* 2002). The present study has shown that this method is quite sensitive for the detection of cofactors associated with central metabolism, including ATP. The AMP levels were observed to suddenly drop in response to the glucose pulse while ATP levels increased and ADP levels decreased moderately. These behaviors were opposite to an observation on *Saccharomyces cerevisiae* where intracellular ATP levels and ADP levels immediately decreased and AMP levels temporally increased after a glucose pulse (Theobald *et al.* 1997, Kresnowati *et al.* 2006). Interestingly, the level of 6-phosphogluconate (6PG) increased 180 seconds after glucose relief and, on the other hand, the level of NADPH immediately increased after the glucose pulse and returned to the original level after 180 seconds had lapsed. NADPH is synthesized when glucose 6-phosphate or 6-phosphogluconate is oxidized in the initial part of the pentose phosphate pathway and this pathway is activated when nucleotide synthesis demands surge. As such, the up-regulated flux through the oxidative part of the pentose phosphate pathway during the time from 0 to 180 seconds probably resulted in NADPH production along with a dramatic change in the level of 6PG. Moreover, in the initial 180 seconds following glucose release, the level of nucleotides (ATP, GTP, CTP and UTP) increased. These observations indicate that the glucose pulse induced cell growth that caused

immediate demands for nucleotide synthesis and thus activation of the pentose phosphate pathway.

Based upon observations under a single experimental condition, it is challenging to discuss detailed mechanisms that explain the observed fluctuations in intracellular metabolite concentrations derived from a rapid change of environmental glucose concentrations. When isotope labeled substrate such as 13C-glucose is employed, more precise flux balance in the cell can be elucidated. As for a practical purpose; however, this method could further be used for real-time monitoring of intracellular metabolism in combination with on-line analytical applications. This system, in combination with metabolic pathway analysis, should represent a valuable approach to investigate metabolic regulatory systems that are responsible for the rapid response toward environmental perturbation.

2.3 Materials and Methods

All solvents, metabolite standards and other chemicals were purchased from Sigma Aldrich (St. Louis, MO, USA). 9-Aminoacridine hydrochloride was purified and recrystallized prior to use (Shroff *et al.* 2007a). Deionized water was obtained from a Milli-Q system (Millipore, Schwalbach, Germany).

E. coli strain JM109 was used for the metabolite analysis experiments. Incubation was carried out in a 50 mL test tube (12 h, 150 rpm, 37 °C) containing 20 mL LB medium. Bacterial cells were collected by centrifugation (5,500 *g*, 2 min) and the collected cells were washed once with water. The cells were centrifuged again and resuspended in 1 mL water. The cell suspension was then incubated in a 2 mL tube for an hour at 37 °C.

Matrix solution (6 mg/mL 9-AA in methanol containing 1 μM 8-anilino-1-naphtalenesulfonic acid as an internal standard) was used to quench intracellular metabolism and extract metabolites. A highly concentrated glucose solution was added to the shaking cell suspension in an aerobic environment with a final concentration of 1 g/L. Sampling was performed by taking 5 μL of suspension and mixing into 15 μL of the pre-cooled matrix solution (40 °C) in swift succession before and after the glucose pulse. The sampling was performed at the indicated intervals. Sample preparation and processing was performed in triplicate. To evaluate the sample quality, the metabolite extract was also prepared using cold methanol as reported (Maharjan and Ferenci 2003) with a slight

modification as a control sample. The final content of cells and matrix was the same as the sample preparation described above. The collected supernatant was subjected to MALDI-MS analysis.

A mixture of metabolite standards was prepared in 60% methanol/water and was used for external mass spectrum calibration. Intra-assay precision was confirmed by spiking the metabolites into diluted (5 times with deionized water) cell extracts. The calibration curves were obtained by analyzing standard solutions ranging from 100 nM to 25 μM.

In this study, three types of MS instruments were used. For quantitative metabolite analysis, single reflectron-type MALDI-TOF-MS (AXIMA Confidence, Shimadzu, Japan) was used. Analysis time was less than 20 second/spot. For identification of metabolites by MS/MS analysis, quadrupole ion trap (QIT)-type (AXIMA QIT, Shimadzu, Japan) instruments were used. CID power parameter was set to almost eliminate the mass response of the precursor. These instruments were equipped with a 337 nm N_2 laser. For the determination of the elemental composition of the metabolites by ultra accurate mass, a Fourier transform ion cyclotron resonance (FT-ICR)-type instrument (Apex-Q94e, Bruker Daltonics, USA) with an Apollo II ionization source equipped with a 355 nm Nd:YAG laser was used. In all analyses, 1 μL of the analyte was applied onto a ground-steel MALDI sample plate and air-dried. The samples were irradiated at a laser power that gives satisfactory ion intensity and all analyses were performed using the same laser power in the negative ionization mode. All the metabolites were identified according to their accurate *m/z* and MS/MS spectra acquired on the MALDI-TOF-MS and/or MALDI-FT-ICR-MS instruments.

Mass spectra were obtained by MALDI-MS analysis where two laser shots were accumulated and 121 spectra were averaged per spot. Six spots were deposited from a sample

and averaged to apply for further data analysis. Mass spectra were internally calibrated using the mass peaks of the matrix (9-AA; *m/z* = 193.0776), the internal standard (8-anilino-1-naphtalenesufonic acid; *m/z* = 298.0538) and the constantly detected metabolites such as ATP (*m/z* = 505.9885) and/or acetyl CoA (*m/z* = 808.1179). The ion intensities of individual peaks were normalized to the total ion count (TIC) of each analysis.

Mass responses of each metabolite at each time point were plotted as a function of time (s). To outline the time-dependent transition of a metabolite level, the mean value of three experiments at each time point was interpolated by a simple moving average (Chou 1975).

Chapter 3.

Temporal Metabolite Network Analysis of Bacterial Metabolic Fluctuation in an Initial Action

3.1 Introduction

Adapting to fluctuations in carbon source availability is crucial for microorganisms to survive. The adaptation system comprised of molecular regulatory network is firstly invoked by recognizing the fluctuation. We previously demonstrated that the variation of the intracellular metabolite levels in *E. coli* progressed on a sub-minute scale following a glucose pulse (Yukihira *et al.* 2010). This preliminary analysis clearly indicated that the variation of the intracellular metabolite levels progressed on a second scale. Such a time-scale was relatively short compared with that of gene expression alterations, where Dikicioglu *et al.* reported a strong response in mRNA levels in *E. coli* a few minutes after changes in nutrient availability (Dikicioglu *et al.* 2011). It was also true of variations in protein levels, where a temporal surge of transcription immediately promoted an increase in protein levels in yeast, resulting in a minute- to hour-scale variation (Lee *et al.* 2011). It has been suggested that the sensing system is associated with not only the abundance of carbon sources themselves but also the consequent perturbations in the intracellular metabolism (Kotte *et al.* 2010, Kochanowski *et al.* 2012). We thus assumed that, although such an immediate metabolic fluctuation progressed in a passive manner, the cellular system should be organized to buffer and recognize the fluctuation. However, the temporal and structural characteristics of the fluctuation propagation in its intrinsic time-scale is still unclear because most studies of biological responses have not concentrated on the time-scales of biochemical processes of metabolites (Steuer 2006). Dynamic behaviors of the molecular network are of particular importance because the cellular system cannot be understood through a static network

structure alone (Kitano 2002). Therefore, these circumstances encouraged further investigation of the time-dependent structural variation of the metabolic network in an early phase of metabolic fluctuations.

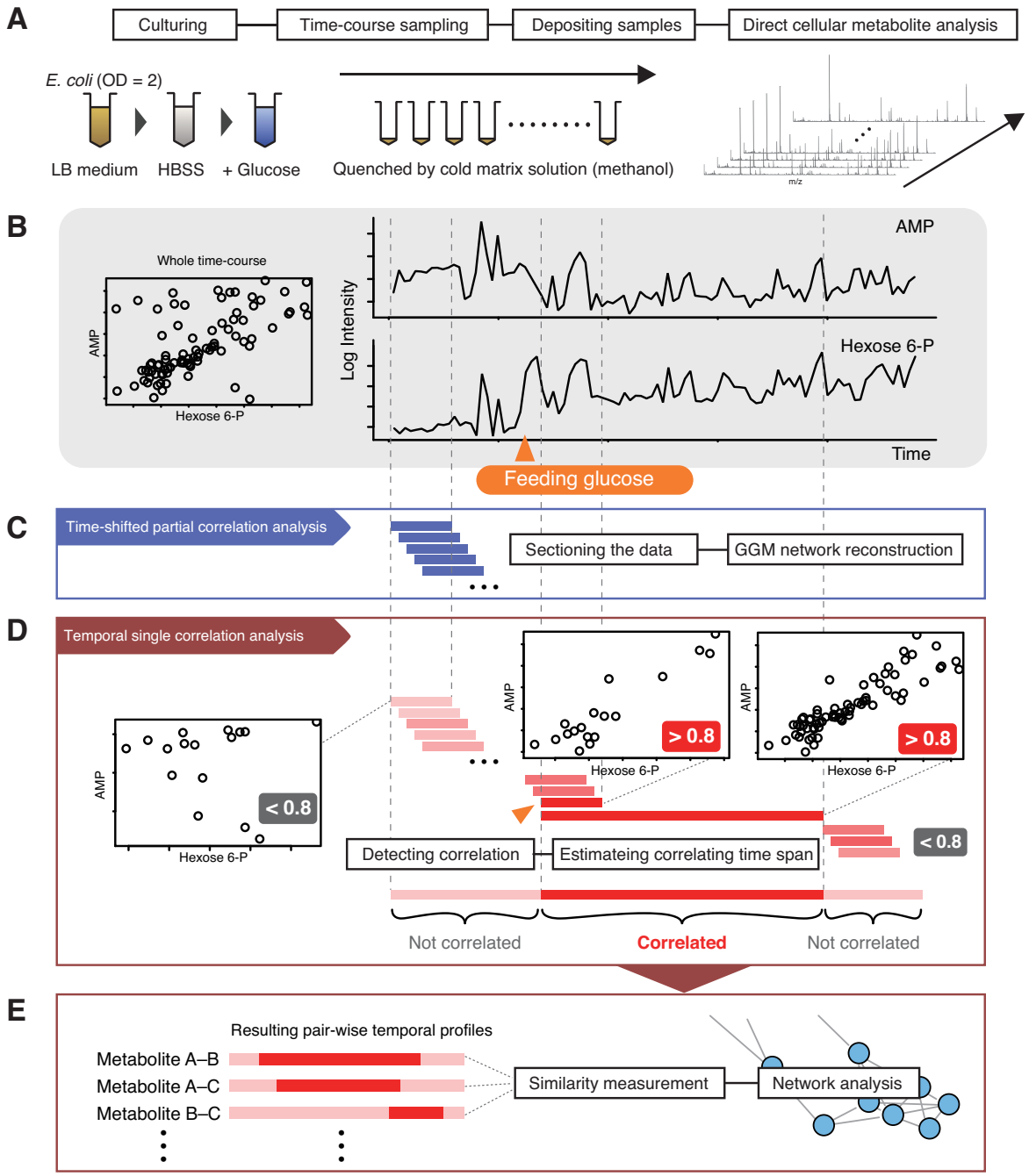
The primary focus of this study was to understand characteristics of the metabolic system of *E. coli* in the transient state based on the level of intracellular metabolites. It is important to investigate interdependencies of metabolite levels, referred as metabolite network, because rather than metabolite level snapshots they should reflect the network structure of the underlying complex system. However, correlation itself does not provide temporal or contextual information about observed relationships, because it is a static representation under a given condition. To address the issue, we formulated the temporal behavior of metabolic fluctuation as an evolving network of metabolite correlation, where a sliding window was used for evaluating the transient correlation structure of metabolites in response to a perturbation. Starving *E. coli* cells were exposed to an instant glucose relief to induce a strong metabolic fluctuation. Firstly, a time-dependent alteration of the metabolite correlation structure was overviewed as a phenomenon that reflects the transient metabolic fluctuation. The sequential propagation of the metabolic fluctuation was then investigated based on the temporal similarity of the variations in metabolite correlations. We discuss the usefulness of the temporal information for interpreting the origin of the observed metabolite correlations.

3.2 Results and Discussion

We previously developed a high-throughput analytical method that employed MALDI-MS utilizing its semi-quantitative performance as well as high-throughput characteristic for acquiring detailed time course data on intracellular metabolites (Scheme 3.1A, *MALDI-MS-based high-throughput metabolite analysis*). With minimal experimental work, this method can trace the levels of phosphorylated metabolites such as sugar phosphates, nucleotides, nucleotide sugars, and cofactors, which play important roles in cellular metabolism (Miura *et al.* 2010b). In this study, *E. coli* was exposed to a nutritional perturbation and its response was characterized by the temporal variation in metabolite levels. In the direct cell analysis using MALDI-MS, about 100 mass peaks were frequently detected. Of these, we identified 28 metabolites that were detected reproducibly throughout the time course (Table 3.1). These metabolites included a variety of nucleotides, nucleotide sugars, and CoA compounds that had historically proven difficult to quantify in LC-MS analyses, in spite of their biological importance (Jansen *et al.* 2009).

3.2.1 Metabolic Pathway Served as The Source for Initial Metabolite Correlation

The time-scale of cellular metabolism alteration was firstly checked. The energy charge of the cells could be represented by ATP-ADP ratio. The time course of energy charge indicated that the metabolic state of the cells changed on a second scale in response to a



Scheme 3.1. Workflow summary of the present study.

A. MALDI-MS-based high-throughput metabolite analysis. Cell suspension was continuously harvested. Intracellular metabolites were detected by directly analyzing the cells. B. Time-shifted evaluation. Example of a time course pair (F6P and AMP). The scatter plot on the left was constructed using the overall time point of the data set. As the correlation appeared to be non-stationary, the short span correlation was evaluated using a sliding window technique. C. Time-shifted partial correlation analysis. Temporal correlation network analysis based on the GGM technique was performed to illustrate the shift in the correlation structure of metabolite levels. Because GGM uses a partial correlation, the direct correlation network is provided. D. Temporal single correlation analysis. Maximum length for the span of the single correlation was evaluated to compare the temporal similarity of the correlation

profile. E. Temporal similarity analysis. Based on the temporal similarity of the correlations, a meta-correlation network was constructed. Network analysis methods were applied to extract the temporal traits of correlation profiles. See *Materials and Methods* in Chapter 3 for details of the following analytical workflow.

Table 3.1 List of detected peaks and identified or estimated metabolites.

Observed *m/z*	Metabolite name	Abbreviation
259.0	Fructose 6-phosphate (hexose phosphate)	F6P
275.0	6-Phosphogluconate	6PG
306.1	Glutathione (reduced form)	GSH
321.0	Thymidine monophosphate	dTMP
322.1	Cytidine monophosphate	CMP
323.1	Uridine monophosphate	UMP
339.0	Fructose 1,6-bisphosphate	F16P
346.1	Adenosine monophosphate	AMP
362.1	Guanosine monophsphate	GMP
401.1	Thymidine diphosphate	dTDP
402.1	Cytidine diphosphate	CDP
403.0	Uridine diphosphate	UDP
426.1	Adenosine diphosphate	ADP
442.0	Guanosine diphsphate	GDP
481.0	Thymidine triphosphate	dTTP
482.0	Cytidine triphosphate	CTP
483.0	Uridine triphosphate	UTP
506.0	Adenosine triphosphate	ATP
522.0	Guanosine triphsphate	GTP
540.1[a]	Nicotinamide adenine dinucleotide	NADH
545.1	Thymidine diphosphate 4-oxo-6-deoxy-glucose	dTDPg
565.1	Uridine diphosphate glucose	UDPG
588.1	Thymidine diphosphate 3-Acetamido-3,6-dideoxy-galactose	dTDPa
606.1	Uridine diphosphate *N*-acetylglucosamine	UDPGN
611.1	Glutathione (oxydated form)	GSSH
620.1[a]	Nicotinamide adenine dinucleotide phosphate	NADPH
766.1	Coenzyme A	CoA
808.2	Acetyl coenzyme A	AcCoA

[a]Fragmented ion

nutritional fluctuation (Figure 3.1). Therefore, supervised models of the behavior of metabolite levels were unavailable because the state of the underlying system should be non-stationary. As an unsupervised method, metabolite correlation has been used to characterize the effects of environmental or gene variation as a fingerprint (Görke *et al.* 2010). However, the non-stationary metabolic system could lead a transient correlation structure, where significant correlations observed at a given time might disappear at a later stage, and *vice versa*. As a straightforward way to address this situation, a short sequence of time course data was subjected to the correlation analysis with one time point shifts forward (Scheme 3.1B, *Time-shifted evaluation*). This representation is often termed evolving network, a natural extension of network analysis onto a temporal context. Firstly, we performed a network analysis based on partial correlation coefficient to confirm the time scale of the metabolic fluctuation and to extract relevant metabolite correlations. Partial correlation is just one of several possibilities for estimating global regulatory interaction structures (Andorf *et al.* 2010). When partial correlations are measured, indirect correlations are explicitly excluded. As this approach is recommended to reveal the molecular interaction of cellular regulatory networks (Werhli *et al.* 2006), we first estimated the partial correlation using time-shifted sequential data (Scheme 3.1C, *Time-shifted partial correlation analysis)*. In the significance test, the threshold of local false discovery rate (fdr) was set to be flexible (up to 0.4) to keep the temporal context as consistent as possible (see *Materials and Methods*). As a result, 28 out of 378 pairs of metabolites were significantly correlated with a specific time-range window (Figure 3.2A). The timings of transient correlations were also informative: numerous correlations appeared in response to the glucose pulse, indicating that apparent shifts in metabolite correlations were induced. This result indicated that an environmental perturbation immediately altered the state of the metabolic system.

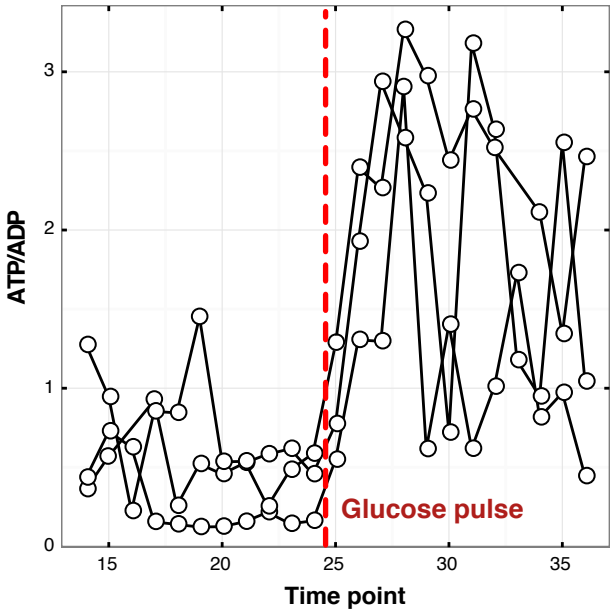


Figure 3.1 Time course of ATP-ADP ratio before and after the glucose pulse.

For each time point, the peak intensity of ATP was divided by that of ADP. Triplicate data are shown. Not available (NA) points were omitted. Following the glucose pulse, the ATP-ADP ratio reached a maximum in three to four time points (corresponding to 30–40 s).

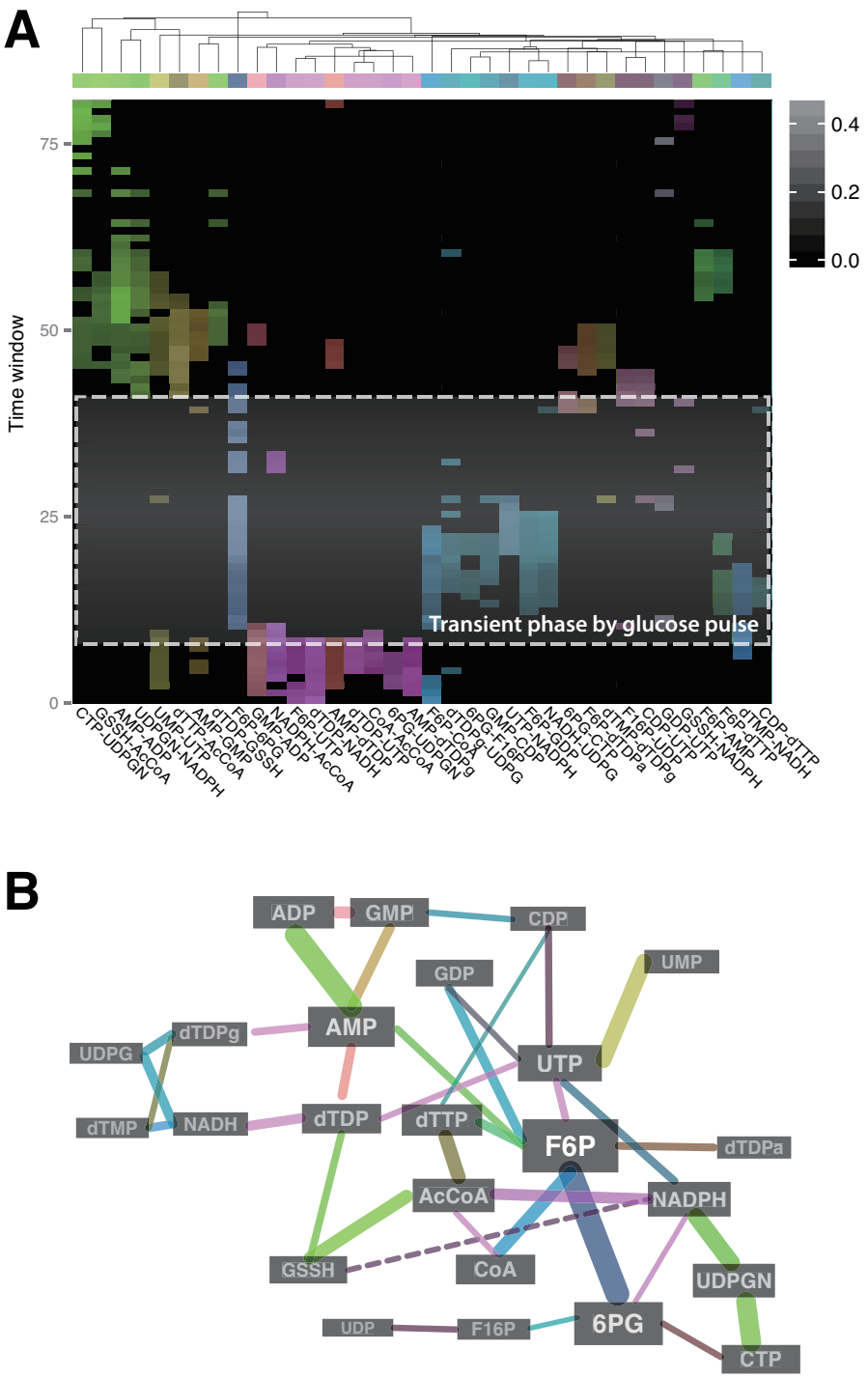


Figure 3.2 Temporal profile of the partial correlation structure and network representation.

A. A partial correlation coefficient for each relationship was calculated using the sectioned time-shifted longitudinal data of metabolite levels. The profiles were visualized as a heat map and clustered by hierarchical clustering using a complete linkage method. A glucose pulse was applied just prior to the 25th time point, and time windows that included the time point (time windows 9–40) are highlighted (dashed line). The color of each profile was determined by mapping the three-dimensional coordination of each relationship in the similarity space of

the temporal profile onto a red-green-blue color space. The brightness in the map indicates partial correlation coefficients. **B.** A metabolite correlation network based on the temporal correlation profile. The width of the edges indicates the time span of correlation.

Considering the time scale of biological events associated with gene and protein expression (Dikicioglu *et al.* 2011, Lee *et al.* 2011), the initial variation of the correlation network indicates a passive fluctuation in the metabolic system, which was dominantly associated with metabolites. In this phase, a collapse of metabolic equilibrium could be buffered to prepare a new state of metabolic balance. While transcription alternation is most likely to cooperate with metabolites, the variations in the protein levels following transcription alteration should have little effect on the metabolite-level correlation in this time phase. Secondly, metabolic shifts were brought about by binding of allosteric effectors and temporal change in protein levels for adaptation to the environment, which might be represented as a more gradual change in the correlation profile in a minute-scale. The transient correlation profile was then reconstructed as a metabolite network to review the evolution of the correlation structure (Figure 3.2B). As several metabolites have edges with distinct colors indicating the temporal pattern of the correlation, it was confirmed that a single metabolite could participate in more than one correlation at distinct timings. Such correlations with different appearance times, which could be derived from different factors, at least in a temporal context, would have been overlooked in the ordinary correlation analysis. In this study, expressive correlation was observed between fructose 6-phosphate (F6P) and 6-phosphogluconate during the transition phase in response to the glucose pulse (Figure 3.2A and B). Because these metabolites are intermediates of glycolysis and the pentose phosphate pathway (PPP) respectively, their correlation could be interpreted as the coordinated supply of a carbon source to both pathways. This result agreed with a report that the back-flux from PPP scales with the glycolytic flux (Haverkorn van Rijsewijk *et al.* 2011). Furthermore, the correlation pairs that responded faster tended to have less minimum path lengths in the reference metabolic pathway, implying that metabolite correlations could be derived from the

activation of corresponding metabolic pathway. F6P showed correlations with various metabolites (Figure 3.2B). Assuming that F6P was the initial indicator of glucose utilization, its correlation partners specific to the phase of metabolic shift might provide a simple indicator for the distance in the metabolic pathway.

3.2.2 Temporal Analysis using a Single Correlation Profile Provides a Straightforward View for Metabolite Networks

When single correlations like Pearson product-moment correlation or Spearman's rank correlation are measured to describe a correlation network of biomolecules, its global structure is usually too complicated for clear interpretation. In the context of metabolite levels, however, indirect correlations are still useful to extract relationships regulated by missing factors. Szymanski *et al.* conducted metabolite pair-wise correlation analysis under various stress conditions to allow advanced observation beyond the change in metabolite concentration (Szymanski *et al.* 2009). They found that the stable network, a commonly observed network under various stresses, had some components enriched for functionally related biochemical pathways. On the other hand, Müller-Linow *et al.* reported that closeness in metabolomic correlation was not an indicator of closeness in biochemical networks (Müller-Linow *et al.* 2007). These reports imply that, whereas it is difficult to understand the correlation profile based on a known metabolic pathway, the metabolite correlation itself is important information to estimate functionality in the metabolic system. We thus attempted to extract module structure of the interdependency concealed within a complex correlation network through comparing its temporal traits, namely the simultaneity of relationships, which was one of the properties of the temporal profiles that could not be investigated by static methods. Compared to the partial correlation analysis, this approach rather concentrated

on elucidating the temporally clustered alteration of the metabolic network, which was expected to associate with a similar phase of the regulatory system. To perform a temporal analysis of correlation profiles based on the single correlation coefficient, we examined the time course with a minimum length of time points to detect transient correlation, followed by evaluating a maximum length of correlation (Scheme 3.1D, *Temporal single correlation analysis*). In the construction of a transient correlation network, the appropriate adjustment of parameters is important. Although the length of the detection probe should be as short as possible to evaluate a short-term correlation, the sample size itself influences the quality of the detected correlation. The threshold level of correlation coefficients is also critical for the resulting correlation network. These two parameters were optimized to give an ideal balance of graph theory properties of the resulting network, *i.e.* graph density and modularity (Figure 3.3). Significant variations in metabolite correlation were detected for each metabolite pair and expressed as a time course profile (Figure 3.4A). Numerous correlations emerged immediately in response to glucose pulse at various temporal durations.

3.2.3 Variation of Degrees in the Metabolite Correlation Network Summarizes the Transience of the Correlation Profile

The temporal correlation profile based on the single correlation was highly complicated and required further analysis from different perspectives for better interpretation. To characterize the temporal trend of the metabolic shift with a viewpoint of the metabolites themselves, we examined the time-dependent variation in the connection degree of each metabolite node (the number of significant correlations that the metabolite had with other metabolites). Centrality is one of measures for importance of given nodes in the network. Centrality of each node in a correlation network was evaluated for each time point using the

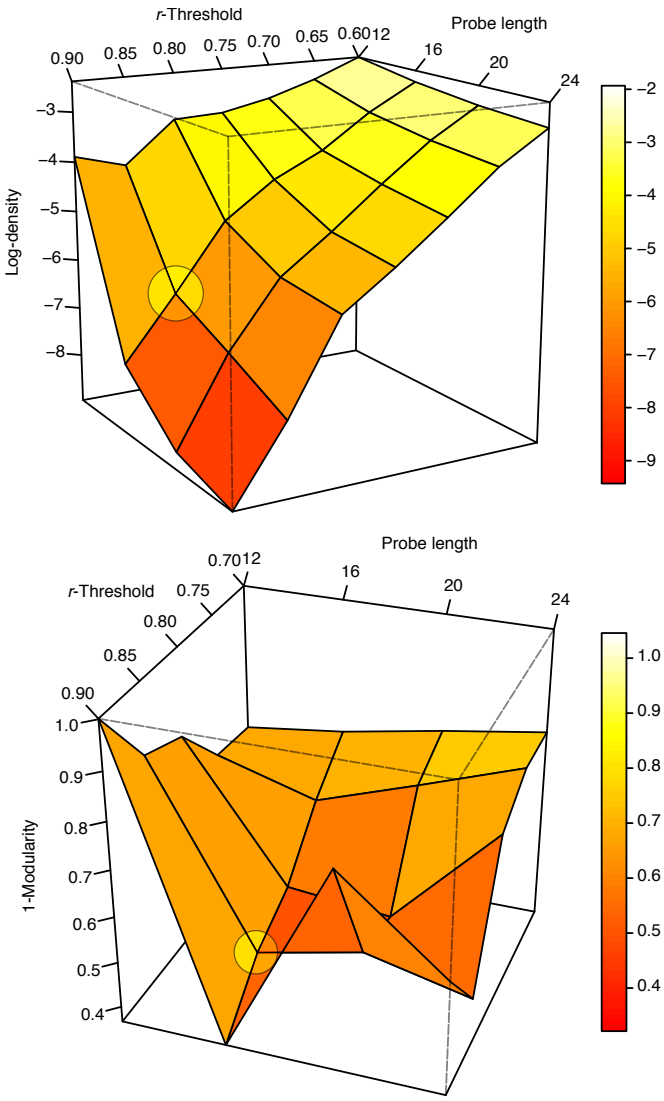


Figure 3.3 Relationship of parameters for correlation analysis to the properties of the resulting similarity network.

A. Density of the similarity network corresponding to a given correlation coefficient threshold and significance level. As network density monotonically decreased along with the increase of parameters, we estimated the optimum parameters to initiate a drop in density (r = 0.85, k = 16). **B.** Modularity achieved on the similarity networks under the same set of conditions as **A**. Higher modularity was observed when the estimated optimum parameters were applied, supporting the validity of parameter estimation. The lattice package (Sarkar, 2008) was used to illustrate 3D perspective views.

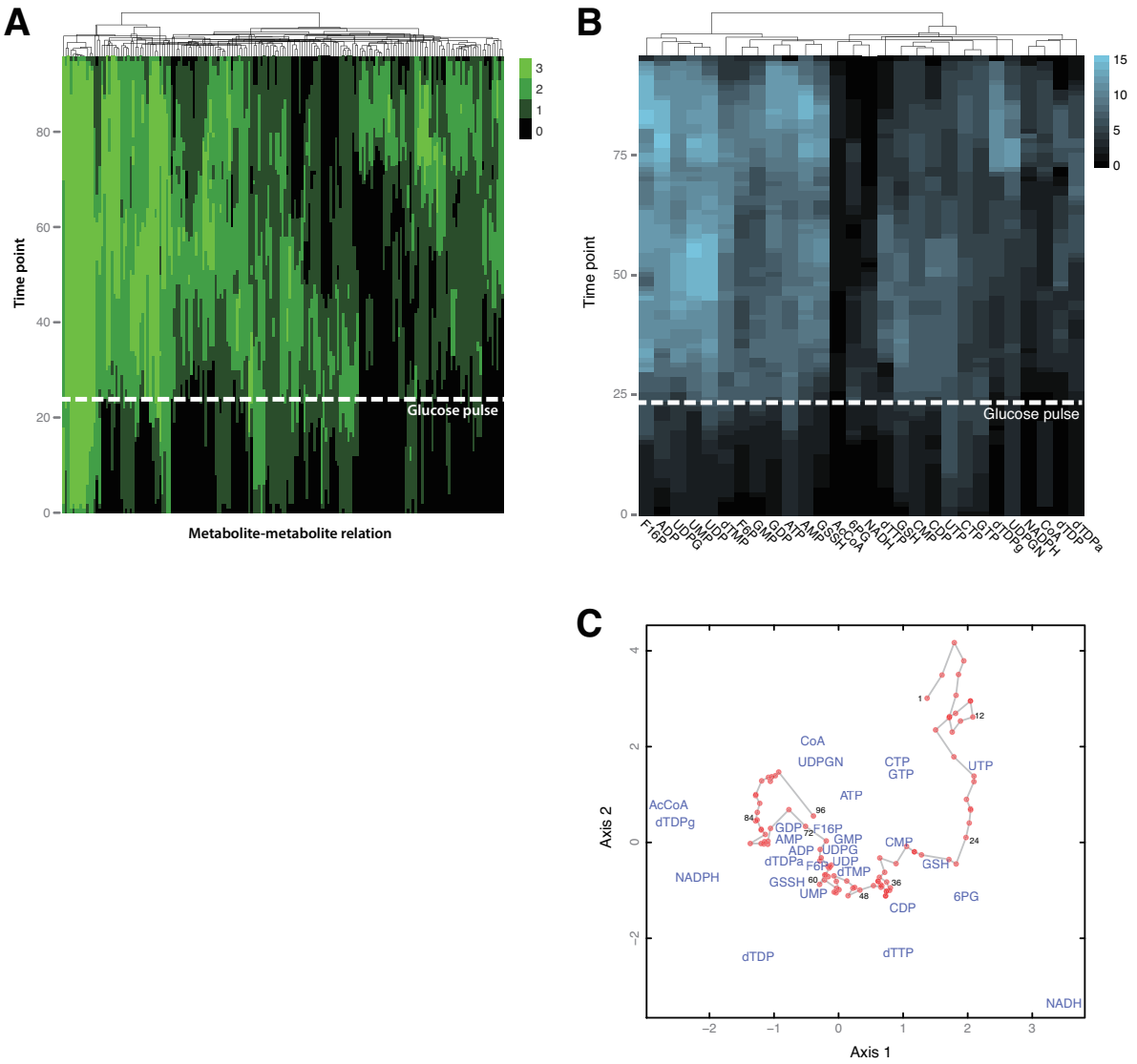


Figure 3.4 Temporal profile of single correlation structure represented by a time course of correlation indicator and centrality analysis.

A. Each slot indicates the maximum time span when significant correlation could be detected for a pair of metabolites. The white dashed line indicates the time point when glucose was added. Triplicate results were overlaid. The profiles were clustered by hierarchical clustering using a complete linkage method. **B.** A comprehensive view of the time course variation of centrality. As centrality can be evaluated when a network is given, time-dependent correlation networks were constructed for each time point of the correlation indicator matrix. **C.** CRA plot of the degree centrality. The lined plot represents the time point. Metabolites were located to indicate relevance to the time points.

time course data of the correlation indicator, as shown in Figure 3.4A, and time-dependent variances in centrality were obtained for all nodes. In the present study, several metabolites exhibited significant variation of centrality in response to the glucose pulse across a variety of durations (Figure 3.4B). Considered inversely, the variation of the correlation network was compressed into the variation of node centrality. To better understand this, we further employed centering resonance analysis (CRA). CRA is a method of network analysis that was originally designed to study complex discourse systems derived from a wide range of sociological or psychological phenomena (Corman *et al.* 2002), and is applicable to evolving networks (Brandes and Corman 2003). Briefly, CRA can be conducted as a correspondence analysis (CA) of network centrality of a set of nodes evaluated under different network structures. However, a potential problem with CA for 2-mode networks has been reported: the distances in CA can be misleading because they are not Euclidean (Borgatti and Everett 1997). In addition, on the two-dimensional (2D) map, it is difficult to determine which relationship appears at each time point. Nevertheless, these limitations are not critical for understanding the trend of the correlation profiles, as long as the major interest is to assess structural equivalence of the network, rather than component association itself (Roberts Jr. 2000). In the present study, CRA was useful for visualizing the time when each metabolite became the center of the correlations, along with the metabolic shift (Figure 3.4C). While the first 24 times prior to the glucose pulse remained within a narrow region with regard to Axis 1, the time points post-induction shifted away from the initial region. The metabolic shift was initiated in accordance with a surge in the degree centrality of the number of metabolites. The CRA plot indicated that early shifts were related to the variation in the centrality of glutathione, CMP, and CDP. The wide-spreading correlating behavior of glutathione is reasonable in terms of redox balance maintenance during oxidative respiration coincided with

the active utilization of nutrition. It is also known that pyrimidine nucleotides are more responsive to the growth phase than purine nucleotides in *E. coli* (Huzyk and Clark 1971, Buckstein *et al.* 2008). Indeed, UDP, UMP, UDP-glucose UDP-ribose showed the highest centrality while purine nucleotides including ADP, AMP, GDP, GMP showed rather delayed variations in their centrality. Whilst nucleoside triphosphates themselves do not exhibit any correlative behavior in response to the fluctuation, synthesis of pyrimidine nucleotides and their phosphorylation level could be sensitively coupled with the developmental conditions, implying the structural characteristics of the metabolic pathway. The subsequent metabolic alteration was characterized by the centrality of various sugar phosphates and nucleotide. Considering the sugar phosphates being the representative intermediates in the central carbon metabolism, their correlation partners should indicate the distribution of carbon flux.

3.2.4 Simultaneity of Correlations in the Profile Indicates Potential Relevance in Biological Events

Generally, the significant metabolite correlations observed at a given time range do not immediately imply any relationship of the metabolites in a biological context. Nevertheless, concurrent emergence of correlations could be expected to be under the influence of a similar regulation phase. The similarity among correlation profiles was evaluated to examine the simultaneity of correlation (Scheme 3.1E, *Temporal similarity analysis*). The resulting similarity matrix was then reconstructed as a temporal similarity network (Figure 3.5A). Unlike 2D projection of a multidimensional similarity space, where the variation is often poorly explained, this simplified network representation was useful because it was obvious which node (a correlating pair of metabolites) was connected to others. We then extracted communities, which represented sub-networks having higher

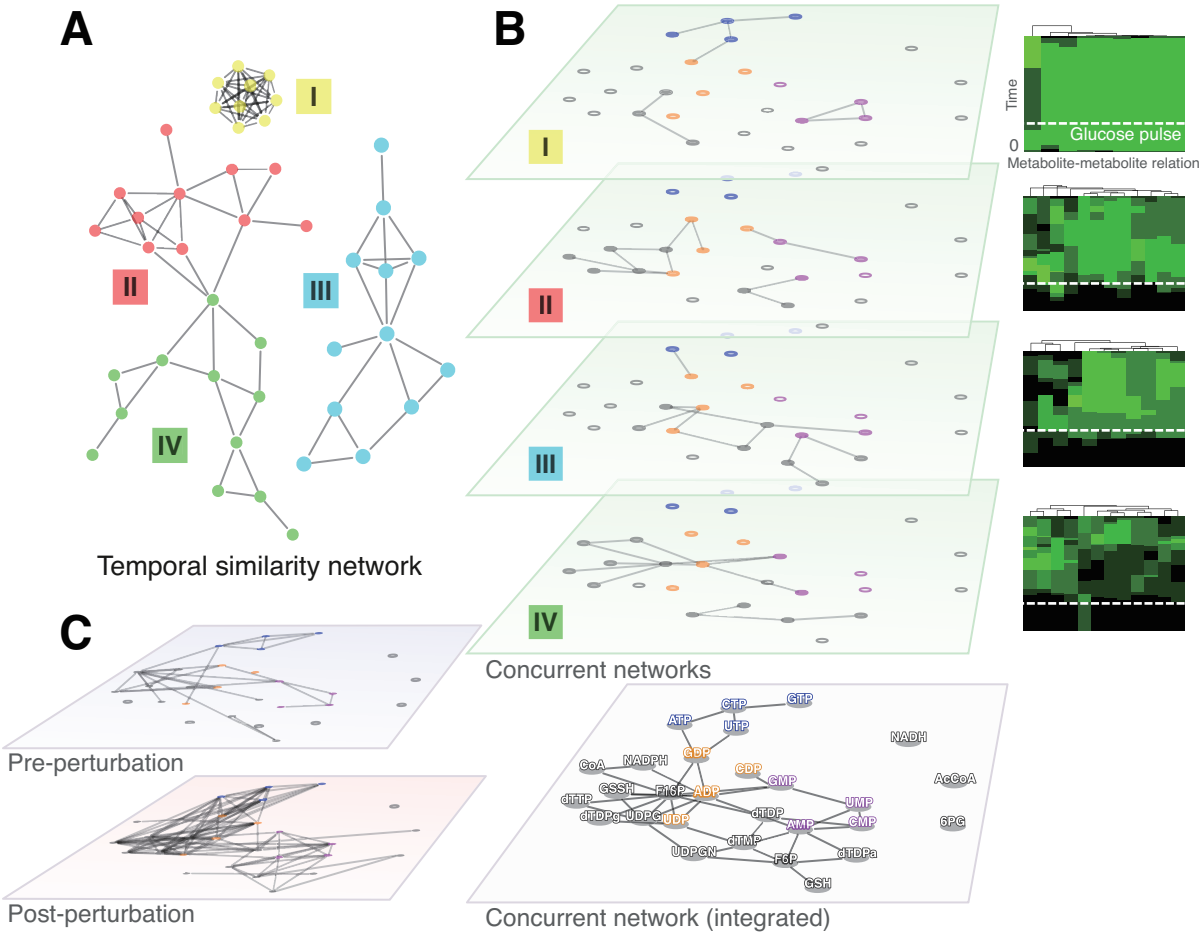


Figure 3.5 Concurrent modules in the metabolite-metabolite network determined by community in the similarity network.

A. Temporal similarity network of metabolite correlations. The similarity index among temporal correlation profiles was reconstructed as a similarity network. The nodes in this network represented unique pairs of metabolites. The modules composed of more than four metabolite pairs were individually colored or otherwise left blank. B. Subsets of the correlation network (concurrent modules) that were reconstructed from the communities in the similarity network. The module numbers I–IV, indicated at the bottom left of the concurrent networks, correspond to those of the temporal similarity network. Corresponding slots of the temporal profile (Figure 2) are displayed on the right. Overlapped concurrent modules are displayed on the bottom, with metabolite names on the nodes. Ribonucleotides are colored purple, orange, and blue, respectively. C. Metabolite-metabolite network based on the single correlation was evaluated using time point 1-24 (upper) and 25-96 (lower), corresponding to pre- and post-perturbation, respectively.

modularity in the similarity network. This analysis was almost equivalent to clustering analysis for finding correlations with temporal similarity. Although there were modules moderately connecting to each other, a clear module structure was observed. These results suggested that the time-dependent metabolite correlation had highly concurrent characteristics. As the nodes in the similarity network corresponded to the correlation edges in the metabolite correlation network, the modules in the similarity network were reconstructed as metabolite correlation networks, which appeared at a specific timing. We then examined one of each of the modules in terms of interrelation with the system of metabolic dynamics (Figure 3.5B). The metabolite correlation networks evaluated using the whole time course data in pre- and post-perturbation were shown in Figure 3.5C, representing a non-temporal correlation network analysis. In Figure 3.5B, one of the communities consisted of constant correlations regardless of the glucose pulse (Figure 3.5B, module I). These correlations were composed of nucleotides, and were strongly associated with known biological events, namely glycolysis and nucleotide equilibrium. Hexose bisphosphates, UDP-glucose, and UDP-GlcNAc are closely related to glycolysis and sugar nucleotide metabolism. These three metabolites are intermediates at the metabolic pathway branching from the hexose phosphate pool. It has also been previously determined that levels of nucleoside triphosphates are related to growth phase, and that the level of each nucleotide is more or less correlated (Buckstein *et al.* 2008). Such relationships were intrinsically maintained even under the nutritional perturbation, while the degree of correlation could change. In contrast, the other modules represented emerging correlations responding to the glucose pulse, at various times and durations (Figure 3.5B, modules II–IV). In these modules, nucleoside triphosphates rarely participated in the correlation network, while a large number of nucleoside mono- or di- phosphates participated in various temporal modules. This pattern

is also shown in Figure 3.4C, where the metabolites clustered around the time course plot following glucose pulse, and nucleoside triphosphates were located at quite separate points from the cluster. Module II especially was composed of correlations that instantly appeared following the glucose pulse. Furthermore, this module included most of nucleoside mono- and di- phosphates, as well as the sugar phosphates. In contrast, none of nucleoside diphosphates were included in module IV. Instead, sugar phosphates (F6P and F16P) were central nodes. Assuming that the immediate fluctuation in the metabolic network is independent of the consequent transcriptional regulations, the initial and transient variation of the metabolite correlation structure (module II) could be buffering components in the metabolic system. These variations were inevitable for the metabolic system, but to be resolved by the following regulation, resulting in their disappearance. Although there might be missing partners that concurrently correlate with nucleotides, it was implied that nucleoside mono- and di- phosphates first buffered the metabolic fluctuation, while the nucleoside triphosphate balance was basically maintained. This behavior could further lead a speculation that preparing various nucleoside mono- and di- phosphates was versatile in the initial action as well as necessary for a stable supply of nucleotide triphosphates required for growth. As module IV was composed of correlations that appeared at relatively late time points compared to module II, the variations in this phase could be influenced by the transcriptional regulation. As Kochanowski *et al.* reported that utilizing the intermediate metabolite in the central metabolism could be an effective way to uniformly detect the availability of distinct carbon sources (Kochanowski *et al.* 2013), the variations in metabolite correlations associated with sugar phosphates could possibly provides further implication for the metabolic sensing system.

3.3 Conclusion

This study revealed the temporal and structural behavior of metabolite correlation network in a fast and short span following to the nutritional perturbation. The evolving network structure provided novel information for interpreting the metabolite correlation that would have been obscured by evaluating a static correlation structure. We observed that several metabolites transiently correlated with distinct partners time-dependently in response to the glucose pulse, suggesting that some of these relations could be used as the marker in immediate sensing of the metabolic state. The temporal profiling of metabolite levels introduced in the present study should be considered as a methodological proposal that complements metabolomic studies, aimed at developing a broader view of quantitative metabolite levels for the price of a limited sample size. As long as a cellular system is dynamic, temporal information also benefits advanced interpretation in other omics studies that deal with multivariate data individually influenced by a variety of factors. Temporally specific patterns of variation in the biological network remain to be investigated, but should be a rich source of novel perspective for understanding the network dynamics.

3.4 Materials and Methods

Workflow of the experiments and data analysis is summarized in Scheme 3.1.

3.4.1 Chemicals

All solvents, metabolite standards, and other chemicals were purchased from Sigma Aldrich (St. Louis, MO, USA). Deionized water was obtained from a Milli-Q system (Millipore, Schwalbach, Germany).

3.4.2 Culture and Induction of Nutritional Perturbation

E. coli strain JM109 was used for the direct metabolite analysis. Cultures were incubated Luria-Bertani medium (4 h, 150 rpm, 37°C). Bacterial cells were collected by centrifugation (6,000 *g*, 5 min, 37°C) and resuspended in Hank's balanced salt solution (HBSS) containing 5 μM phenol red (OD_{600} = 2). The cell suspension was further incubated in a water bath (37°C) with constant stirring. A pulse of glucose was added to give a final concentration of 5% (w/v), and cell samples were harvested from the suspension both before and after glucose addition.

3.4.3 Sampling

Matrix solution (6 mg/mL 9-AA in 80% methanol) was used to quench intracellular metabolism. Each sampling was performed by mixing 10 μL of suspension with 60 μL of the

pre-cooled matrix solution (−40°C). The sampling interval was fixed at 10 s. For each time-course sample acquisition, 24 samples were taken prior to the nutritional perturbation induction and 72 post-induction, resulting in a sample set of 96 time points over 16 min.

3.4.4 Mass Spectrometry

For time-course metabolite analysis, a time-of-flight type MALDI-MS instrument (AXIMA Performance, Shimadzu, Japan) was used. The technique was previously introduced as a high-throughput and highly sensitive metabolite analysis. In brief, 1 μL of the analyte was applied onto a ground-steel MALDI sample plate and air-dried to give a sample spot. The spots were irradiated at a laser power that gave satisfactory ion intensity, and all analyses were performed using the same laser power in the negative ionization mode. Mass spectra were obtained by MALDI-MS analysis where five laser shots were accumulated and 256 spectra were averaged per spot. Analysis time was less than 20 s/spot. Four spots were deposited from an individual sample and averaged to apply to further data analyses. Mass spectra were internally calibrated using the internal standard and peaks that appear constantly.

3.4.5 Raw Data Processing

Peak pick, normalization, peak alignment, and scaling were conducted using an in-house Perl script. The cut-off threshold was 30-fold of noise intensity and mass error tolerance was 200 ppm. Ultimately, 100–200 peaks were detected per spectrum. Peaks that appeared in blank sample (HBSS + 9-AA) or that were detected fewer times than half of the number of acquired spectra were excluded from the following statistical analysis. Peak intensity in a spectrum was normalized to give a zero mean and unit variance throughout the

time course. Missing values were designated not available.

3.4.6 Partial Correlation Analysis using Sliding Window

The following statistical analysis was conducted using R language (R Core Team 2012). A set of correlation coefficients among observed metabolites was denoted as *correlation profile*. A matrix **X** is the time course data of *M* metabolites with *T* discrete time points for observation. A temporal subset of **X** starting from time point *t* with length *k* is denoted as:

$$\mathbf{X}_t^k = [\mathbf{x}_t, \mathbf{x}_{t+1}, ..., \mathbf{x}_{t+k-1}] \quad \text{where} \quad 0 \le t \le T - k + 1$$

$$\mathbf{x} = (x_1, x_2, ..., x_M)^T$$

Vector $\mathbf{x}_t$ represents the peak intensities of M metabolites in a mass spectrum observed at time point *t*. While parameter *k* can be arbitrarily defined, it controls the trade-off between the correlation detection power and the shortest detectable correlation span. The parameter was set in accordance with the following analytical section (Scheme 3.1E, *Temporal Similarity Analysis*). The graphical Gaussian modeling (GGM) framework (Edwards 2000) was employed to eliminate indirect interrelations. GGMs, also known as covariance selection models, are undirected graphical models where each relationship is conditioned on all remaining metabolites simultaneously. GGM modeling is based on partial Pearson correlation scores, simply calculated by inversion and normalization of the Pearson correlation matrix (Schäfer and Strimmer 2005). We estimated the partial correlations using GeneNet (Schäfer *et al.* 2012), an R package that employs a shrinkage approach, which is suitable for data with a small sample size and a large number of variables. For the first time window, we liberally set the threshold of local false discovery rate (fdr) to give five to ten edges in the correlation network. To clearly illustrate the time-dependent variance of partial

correlation, cutoff values for the following time windows were determined depending on the previous correlation profile. The cutoff value at the *t*-th time window F_t is given as

$$F_t = \max\{0 \le x \le 0.4 \mid S_t(x) \ge 0.8, n_{11} \ge 5\}$$

$$S_t = \frac{n_{11}}{n_{11} + n_{10} + n_{01}}$$

The similarity index S_t is the degree of identification with regard to the *t*-th and the previous time windows. The term n_{11} is the number of edges that are significant under a given fdr threshold, both in the *t*-th and the previous time window. The terms n_{10} and n_{00} indicate the number of edges of either or neither of the adjacent time windows, respectively.

3.4.7 Temporal Similarity Analysis of Correlation Profile

In addition to the GGM approach, we performed single correlation analysis to examine all correlations, including indirect ones, followed by temporal similarity network analysis to extract correlation modules. For every possible pair (*i*, *j*) from *M* metabolites, Spearman's rank correlation coefficient was calculated to give a temporal correlation profile matrix **Y**.

$$(\mathbf{Y})_{i,j,t} = r_{i,j}(X_t^k) = rankcor(X_t^k(i), X_t^k(j)) \quad \text{where} \quad X_t^k(l) = (x_{l,t}, x_{l,t+1}, ..., x_{l,t+k-1})$$

The statistical significance of each pair was then tested. Here, H_0 denotes the null hypothesis that $|r| = 0.85$. Using the alternative hypothesis, H_1, that $|r| > 0.85$, we performed a one-way *t*-test with the alpha level at 0.05. A *t*-statistic Z_0 was calculated using the following formula:

$$Z_0 = \frac{\left|\xi_r - \xi_p\right|}{1/\sqrt{n-3}}$$

Here, *n* indicates the sample size ($n \ge k_0$, otherwise N/A). ξ_r and ξ_p denote a

Z-transformed score of population correlation coefficient 0.85 and absolute sample correlation coefficient $|r|$ using the formula:

$$Z = \frac{1}{2}\ln\frac{1+r}{1-r}$$

$$t_c = \min\left\{t_0 \le t \le T - k_0 \mid P(Z_0(X_t^{k_0})) \le 0.005\right\}$$

$$t_i = \min\left\{t_0 \le t \le T - k_0 \mid P(Z_0(X_t^{k_0})) \le 0.005 \cap P(Z_0(X_{t_c}^{t-t_c})) \le 0.005\right\}$$

Initial t_0 is 0. The value of k_0 can be considered to be the length of a probe for evaluating transient correlation. As an appropriate probe length may depend on the temporal resolution of the time course data, k_0 was set as 16 for the current study, which is equivalent to 160 s. t_0 is then updated to t_c and used in the successive iteration to find the next slot.

Time points from t_c to $t_i - 1$ represent a correlating slot. Here we denote the series of resulting slots as a binary correlation profile matrix **B**, representing whether a metabolite pair p correlates at time point t.

$$(\mathbf{B})_{t,p} = \begin{cases} 1 & (t_c \le t \le t_i) \\ 0 & \text{otherwise} \end{cases} \quad \text{where} \quad 0 \le t \le T - k_0 + 1$$

For each temporal network produced by $(\mathrm{B})_{t,\cdot}$, the degree centrality of nodes, which represent metabolites, was evaluated. The resulting matrix of the centrality was applied to a correspondence analysis.

The slots $(\mathrm{B})_{\cdot,p}$ were compared to each other to measure the similarity, represented as follows:

$$\mathbf{S}_{p_1,p_2} = \min\left\{D(p_1,p_2)/C(p_1,p_2)\right\}$$

The function D represents the difference of time points when the correlation indicators of two metabolite pairs changed from negative to positive. The function C represents the number of time points when both of the indicators of the two metabolites were

positive. Because **S** is produced for every experimental replication, replicated similarity values were represented by the minimum one, which represents the worst case of the simultaneity, and was regarded as an 'honest' estimation to prevent a chance coincidence. A similarity network of the temporal correlations was constructed using **S** with a specified threshold for edge selection. In the similarity network, communities were extracted by deleting nodes with the highest degree of betweenness at the corresponding iteration step to achieve the highest modularity. Here, we determined the parameters, the probe length *k* and the correlation coefficient threshold *r*, with regard to the influences on the resulting network, based on network characteristics such as number of edges, graph density, and modularity. The nodes in the module network, which represent the correlation between the corresponding two metabolites, were reconstructed as metabolite networks. Correspondence analysis was conducted using the MASS package (Venables and Ripley 2002). Graphs were visualized and evaluated using the igraph package (Csárdi and Nepusz 2006).

3.4.8 Genome-scale model of *E. coli* metabolic network

A previously reported genome-scale network (GEM) of *E. coli*, *i*JO1366 (Orth *et al.* 2011) was used as the reference network of metabolism. This model comprised of 1136 compounds and 2551 reactions, reconstructed based on 1366 genes. SBML-type GEM data was converted into a stoichiometric matrix using COBRA Toolbox (Schellenberger *et al.* 2011) on MATLAB version R2012b. We excluded ubiquitous compounds that incorporate various biochemical reactions, which servie as hub and reducing the path lengths in the network, based on their connective degree in the network. The matrix was imported to R environment and treated as a graph. Shortest path length (SPL) was calculated using igraph package (Csárdi and Nepusz 2006).

Chapter 4.

Consensus Patterns in Metabolite Correlation Network of Escherichia coli during Metabolic Reorganization in Response to Nutritional Perturbations

4.1 Introduction

The cellular regulatory system is capable of adapting to immediate environmental changes. Availability of nutrition is crucial for cells for survival, but it is unlikely that specific regulation pathways are prepared for every possible substrates. Kochanowski et al. reported that flux sensors were responsible for appropriate metabolic regulation responding to nutritional environment (Kochanowski *et al.* 2013). Fundamental concept of the sensing system is that multiple nutrient 'inputs' are directly integrated to inform for environmental changes (Perkins and Swain 2009). The fact that a limited number of metabolites can serve as indirect information for nutritional environment can be a natural consequence of metabolic network structure. Inversely, metabolic network has possibly evolved to converge the nutritional input into limited but specific signals. Considering a few-minute delay of mRNA variation, such a sensing system work as the first responder through passive recognition of changes in extracellular nutrient abundances. For nutrition-dependent cellular responses, a number of studies focused on the variation at metabolite level in various time scales. However, nutritional pulse was mostly introduced by adding glucose as the representative carbon source, providing no differential information accounting for the sort of substrates. Furthermore, few studies discussed the metabolomic dynamics in an immediate response to fluctuation despite the fact that the metabolic biochemistry of *E. coli* alters during its growth (Buckstein *et al.* 2008, Allen *et al.* 2003, Baev *et al.* 2006).

A correlation analysis is one of unbiased method to examine system-level associations among biomolecules, and is a basic approach in the systems biology (Steuer *et al.* 2003, Camacho *et al.* 2005). Metabolite-metabolite correlation is recently considered as a rich

source of information for cellular physiological state (Steuer 2006). In Chapter 3, we discussed the transitional behavior of metabolite-metabolite correlation network in response to a glucose pulse. It was indicated that the correlation profile of *E. coli* evolved immediately after glucose pulse and seemingly reached the next stationary phase within a few minutes. This result implied that metabolic network could be inherently capable of curtailing sudden metabolic perturbations before the alteration of enzyme abundances.

This study clearly showed that the structure of correlation network dramatically evolved with an immediate manner in response to the nutritional fluctuation, and that the source of the correlation could be the topological structure of the metabolic pathway. The closeness between metabolite-metabolite correlation and the distance in a network is a straightforward relation. However, observed uncorrelated relations between the metabolite-metabolite correlation and the network distance still remain elusive.

The primary aim of this study is to understand the rule of variance in the metabolic balance with which the cellular system can immediately recognize the environmental nutrition abundances. The key points of the present study are that; 1) a wide range of compounds were used to give nutritional pulses to *E. coli* cells; 2) metabolic fluctuation at the metabolite level was traced for a few minutes, when a transcriptional regulation could have a minor influence on the metabolic state, in order to investigate the innate structural characteristics of metabolic pathway; 3) correlation network analysis was employed to systemically extract unexpected propagation of metabolic fluctuation.

MALDI-MS-based high-throughput metabolite analysis is capable of direct detection of intracellular metabolite from nearly intact cells with a high sensitivity. Using MALDI-MS, Ibáñez *et al.* characterized the metabolomic variability in single cells (Ibáñez *et al.* 2013). Such a high-throughput system would allow a large-scale analysis as a foundation for the

phenotype profiling without non-natural perturbation such as gene modification that is directed to one favored phenotype. Using MALDI-MS as the high-throughput analytical platform, temporal variation of metabolite-metabolite balance was evaluated during immediate response to nutritional fluctuations.

4.2 Results and Discussion

4.2.1 Experimental design

High-throughput technique becomes more important in proportion to the exponentially increasing number of experimental conditions to examine. In the present experiment, nearly 2,000 of time-course samples were subjected to MS analysis to perform temporal correlation network analysis to trace a short-time variation of the metabolic state. Such a scale of experiments should be unpractical as long as conventional methods were used, in either terms of time, quality control or continuous labor. The MALDI-MS-based high-throughput method enabled the present study, finishing all the analysis within a few days. The analytical stability and specificity for temporal and conditional aspects was confirmed by PCA analysis of replicated experiments.

4.2.2 Temporal variances of individual metabolite levels

As we observed in Chapter 2, PEP level (m/z = 167.0) decreased in most of cases other than those of acetic acid, alanine and proline (Figure 4.1). For recognition of environmental carbon sources, two major systems are known in bacteria. Some sugars (*e.g.* fructose, galactitol, mannitol, mannose, sorbitol) are recognized through the phosphotransferase sugar uptake system (PTS) (Roseman 1969, Janausch *et al.* 2002, Plumbridge 2002, Schlegel *et al.* 2002, Bijlsma and Groisman 2003, Bettenbrock *et al.* 2007). On the other hand, other sugars (*e.g.* arabinose, glycerol, galactose, lactose, maltose, melibiose, fucose) are recognized transcription factors (TFs), intracellular regulatory proteins

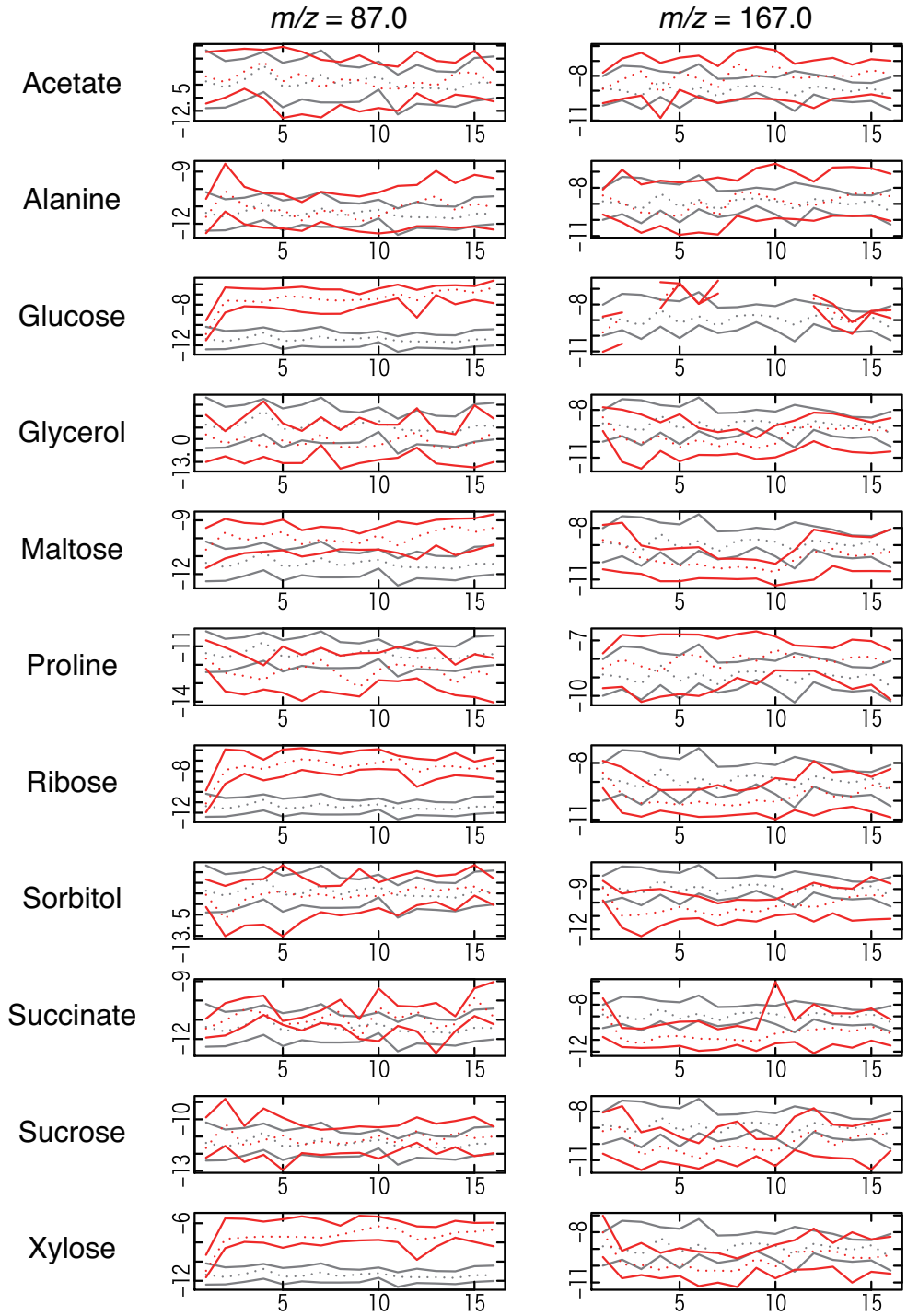


Figure 4.1 Time courses of metabolite levels in response to nutritional fluctuations.

Red lines indicate the fluctuated samples and gray lines indicate the control samples. Solid lines indicate first and third quantile of experimentally or analytically replicated samples and dotted lines indicate the second quantile, or median. X-axis corresponds to time point that indicates sampling point with a 10-sec interval. Y-axis indicates $\log_2$ scaled intensity normalized to TIC.

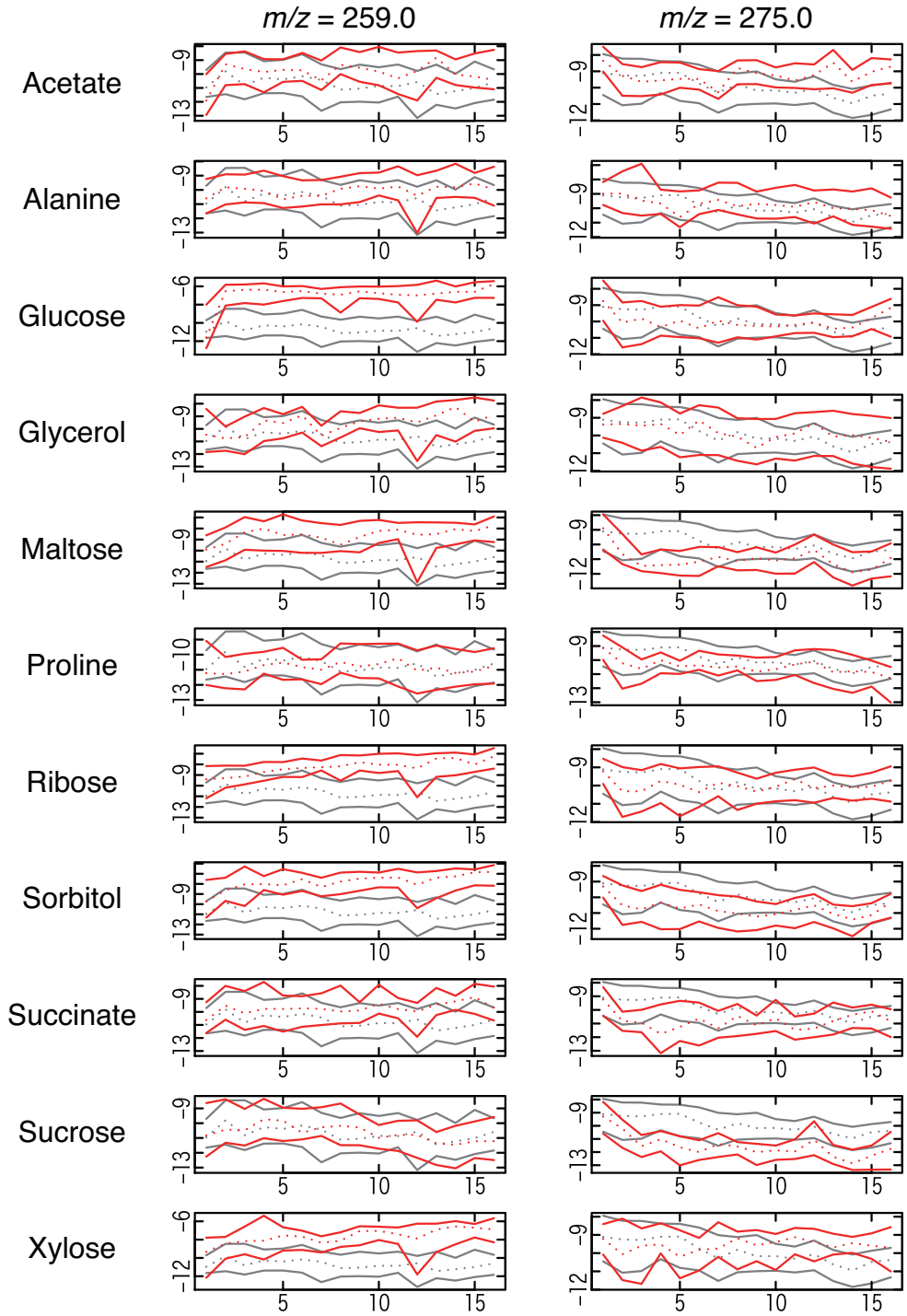


Figure 4.1 Time courses of metabolite levels in response to nutritional fluctuations (continued).

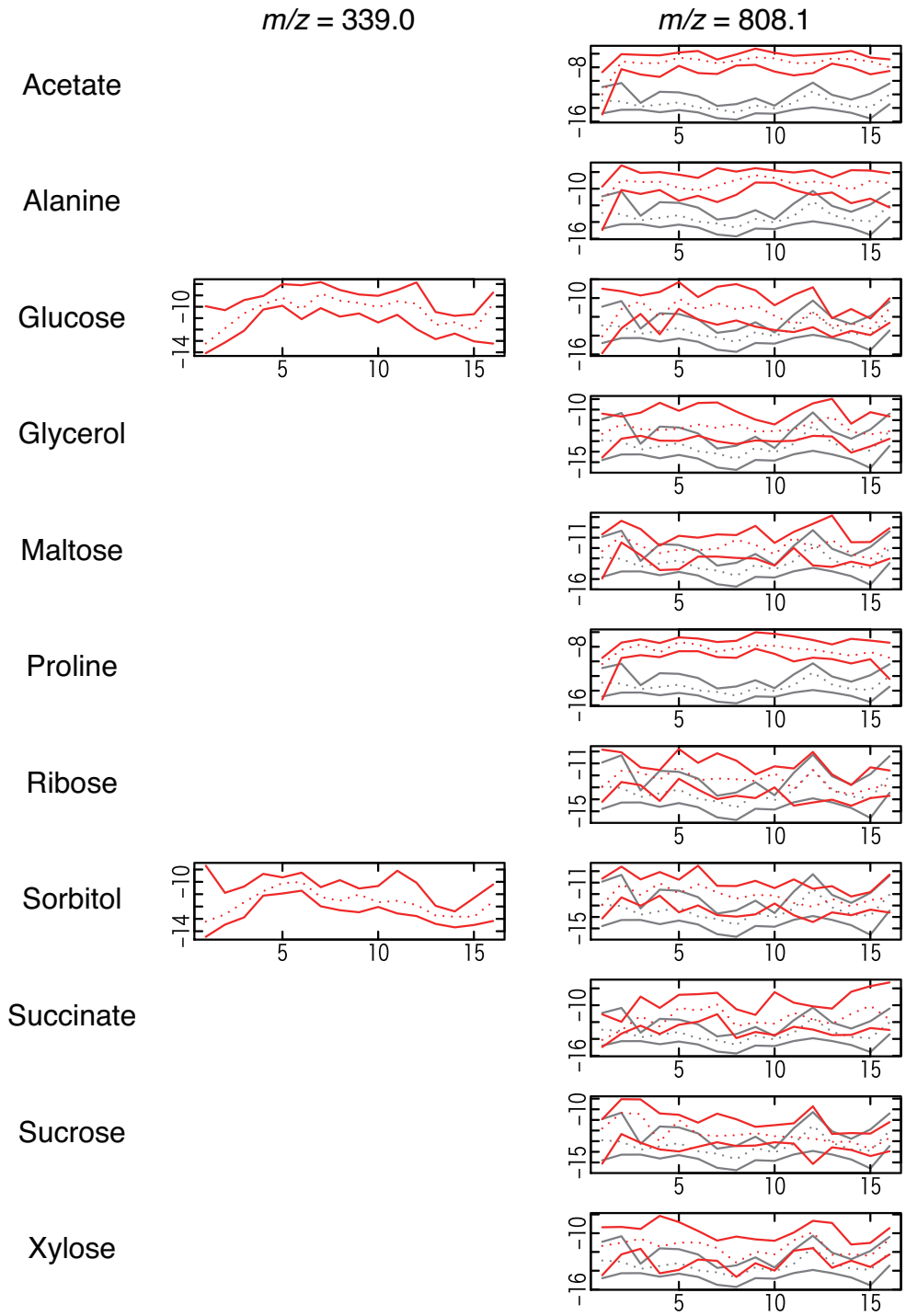


Figure 4.1 Time courses of metabolite levels in response to nutritional fluctuations (continued).

Blank spaces indicate that the corresponding peaks were not detected.

(Cohen and Monod 1957, Seshasayee *et al.* 2006, Ozbudak *et al.* 2004, Martínez-Antonio *et al.* 2012). For other carbon source including organic acids (*e.g.* acetate or succinate), no such transmembrane sensors or regulatory protein has been reported. However, even though carbon source-specific transcription factors are responsible for achieving homeostasis in response to new nutritional environments, immediate imbalance of intracellular metabolite profile should be controlled firstly without transcriptional regulations.

The observed decreases in PEP levels could be attributed to the PTS, typically in the case of glucose or sorbitol. However, glycerol and maltose have been reported as inactive substrates for PTS (Saier 1989). PTS consumes PEP and produce equimolar pyruvate, and in the present study, the pyruvate levels (m/z = 87.0) surged in the case of glucose, ribose and xylose. Although catabolic in-flux could also increase pyruvate levels, and the case of sorbitol could not be well explained, these results implied that these three substrates were involved with the PTS. Therefore, the decrease of PEP level observed in the other cases was assumingly due to PEP-dependent kinase activity, which would phosphorylate the intermediates in other metabolic pathways with a consumption of PEP.

We examined the temporal variances of other metabolite concentrations to confirm the utilization of given substrates by *E. coli* cells. Apparent changes were found in F6P (m/z = 259.0) and F16P (m/z = 339.0) when glycolytic substrates (glucose and sorbitol) were added. It has been reported that F16P is relevant to a flux sensor that regulates the activity of glycolysis. F16P level increased for about 50 sec after the nutrition pulse, while F6P level surged within 20 sec, in both case of these substrates. However, although the F6P level also increased in case of some of the other substrates, F16P was not even observed in any other cases, including the control. This result implied the glycolytic flux was remarkably greater in the case of hexose substrates than the others. Similarly, adding sucrose/maltose led to

increases in sucrose/maltose phosphate level (*m/z* = 421.0), while F6P level was stable. Sucrose/maltose phosphate is consequently hydrolyzed to glucose 6-phsphate and fructose or glucose, respectively. Supply of hexose phosphate was considered balanced with its consumption.

In contrast, a surge of 3PG level was observed exclusively when pentose substrates (xylose and ribose) were added. Flux from pentose substrates directly flows into GAP3, bypassing the FBP pathway that controls the glycolytic flux, and then being oxidized into 3PG. Although the flux from glycerol could flow into GAP3, glycerol might be consumed for glycerolipid synthesis with higher ratio compared to the glycolysis.

Alternatively, a surge of AcCoA was observed in the case of amino acids (alanine and proline) and organic acids (acetic acid and succinic acid), while only a moderate increase of AcCoA was observed for other cases, reflecting relatively less flux from these substrates to the endpoint of the glycolysis.

4.2.3 Temporal Variances of Metabolite Levels in Accordance with ATP/ADP

In the present study, we employed various kinds of substrates including monosaccharide, sugar alcohol, disaccharides, amino acids and organic acids as the source of nutritional pulse to investigate specific or common metabolic perturbations that could inform the given environment. Firstly, the temporal progress of nutrition utilization was confirmed by ATP-ADP ratio (ATP/ADP). ATP/ADP represents adenylate energy charge that should indicate actual free energy of ATP hydrolysis available for cellular reactions (Atkinson 1968). The concept of the index is based on the assumption that, although the absolute individual amounts of ATP, ADP and AMP might vary widely, the ratios of ATP and ADP, or ATP and AMP are more reliable indicator of metabolism. In the present study, the ATP/ADP increased

significantly when sugars were added, typically within 30-40 sec (Figure 4.2). Amino acids led to modest increase of ATP/ADP with a longer time span. This time scale of energy metabolism was equivalent to the previously reported behavior in response to nutritional pulses. We further examined the correlative variation of other metabolites with ATP/ADP, which can be the consequential relationships of cellular metabolism (Ibáñez *et al.* 2013). In control case, basically no metabolite level correlated with ATP/ADP. Amino acids did not lead to any significant increase of ATP/ADP, but the level of AMP negatively correlated with ATP/ADP. It was thus assumed that AMP was converted into ADP and then ATP with maintaining ATP/ADP. On the other hand, CTP, UTP, GTP and succinyl CoA were found to positively correlate with ATP/ADP. It should be noted that the levels of these nucleotides were seemingly stable in both cases. Such balancing would allow an efficient utilization of possible sugar substrates that involves with glycolytic pathways influenced by ATP/ADP. Throughout the cases, many other metabolites also negatively correlated with ATP/ADP, while most of them remained unknown. There was no positive correlation with ATP/ADP. Such trends could indicate that a surplus of intracellular metabolites was directed to synthesis of ATP. The structure of the metabolic pathway should determine which metabolites to be consumed or saved.

4.2.4 Centrality analysis of evolving networks

We so far discussed the temporal behavior of metabolites in terms of their abundances and individual correlative relationships. In the following, we focused on the structures of metabolite-metabolite correlation networks observed either commonly or specifically in certain cases. As discussed in Chapter 3, alterations in the temporal correlation profile could be associated with the variations of adenylate energy charge. We thus constructed a GGM for

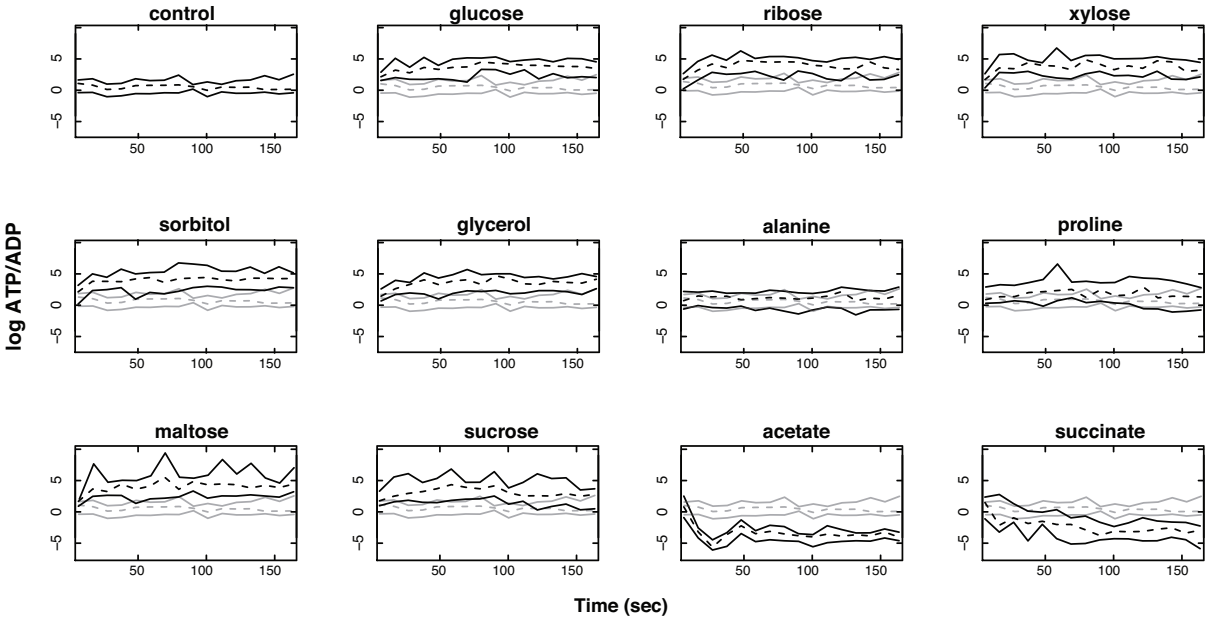


Figure 4.2 Time course of ATP/ADP ratio

Time-dependent variations of ATP-ADP ratio (ATP/ADP) in *E. coli* cells were displayed. Each chart represents the time-course of ATP/ADP variationThe variation chart for control sample was displayed in all the panes by gray lines.

each substrate using the time-course data acquired at the time points when the energy charge reached a maximum (Figure 4.2). Graphical Gaussian networks were constructed for every cases of substrate and time point with uniform criteria, and then the eigenvector centrality of every node was calculated to summarize key the structural components in the network (Figure 4.3). In reconstruction of network, the parameter setting is crucial: the threshold of partial correlation coefficient for edge selection would dramatically affect the property of resulting network. We checked such influence on the correlation network with increasing threshold of correlation coefficients based on the centrality profile of the resulting networks. With $\rho > 0.05$, substrate-specific centrality patterns were observed. As concerned, such patterns got dim with $\rho > 0.03$ (data not shown). On the other hand, they almost disappeared with $\rho > 0.08$. It was assumed that irrelevant relations contaminated to the centrality profile due to the loose threshold in the former case, and the latter lacked relevant relations to illustrate network characteristics due to the high threshold. Nevertheless, the trend was considered at a holistic view, and different thresholds led to different but seemingly significant centrality patterns. When the threshold $\rho > 0.05$, a clear centrality of UMP (*m/z* = 323.1) was observed with a moderate progression in the case of pentose sugars (Figure 4.4A). A similar pattern was observed for a signal of *m/z* = 59.1, which remained unknown.

As discussed in Chapter 3, it is known that pyrimidine nucleotides are more responsive to the growth phase than purine nucleotides in *E. coli* (Huzyk and Clark 1971). Pentose sugars directly supply the base structure of nucleotides, and thus it was assumed that this process was tightly linked with other metabolite levels. The connectivity (or partial correlation coefficients) between UMP and other metabolites was also similar (Figure 4.4A). When the threshold $\rho > 0.08$, a similar pattern was also observed for sucrose/maltose phosphate in the case of sucrose and maltose, along with an unknown signal of *m/z* 404.0

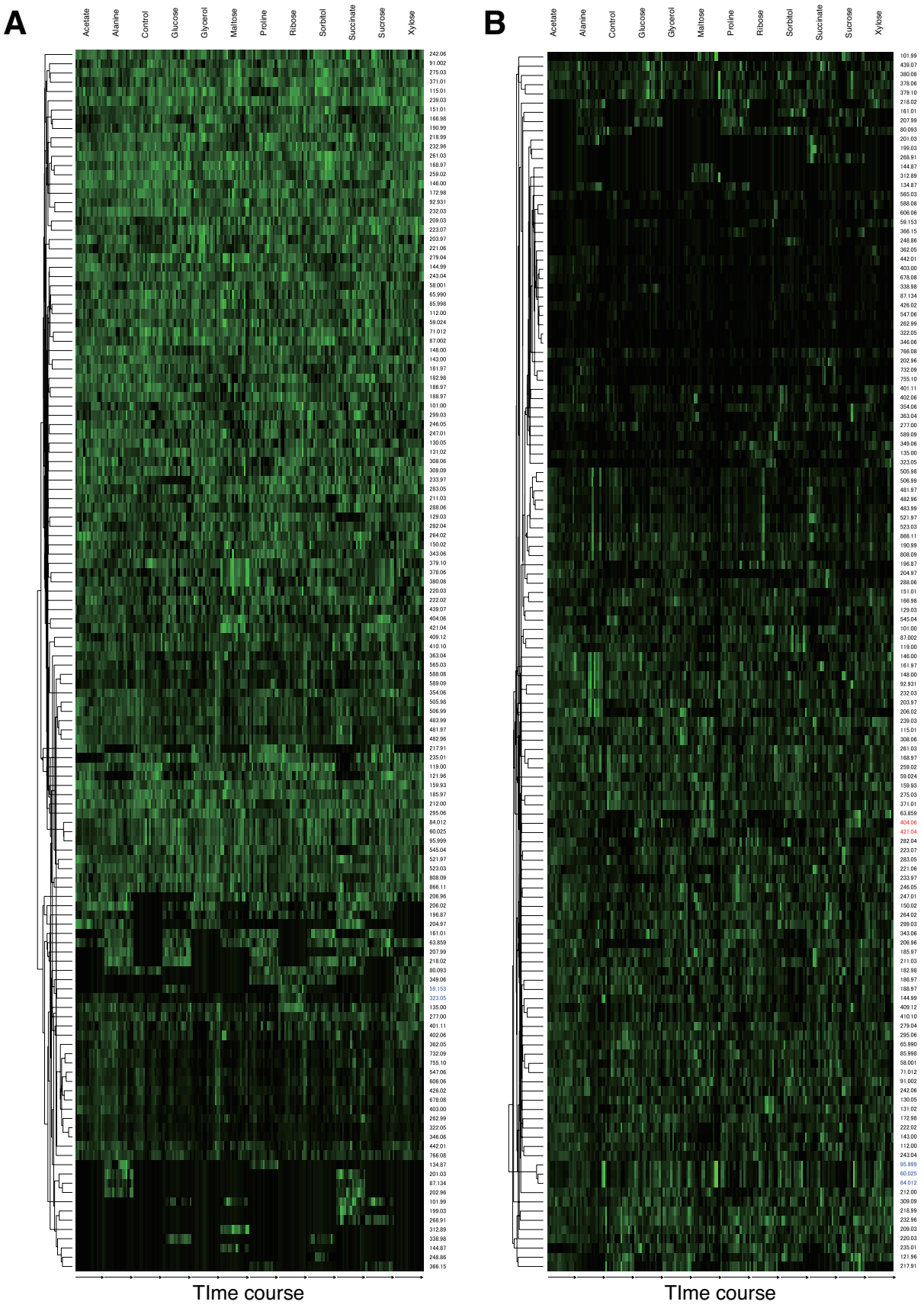


Figure 4.3 Centrality profiles of metabolite-metabolite correlation networks.

Eigenvector centrality of metabolites was indicated by the brightness in the heat map. The

substrate condition was displayed on the top, and time course was ordered from the left to the right as indicated by arrows on the bottom. The row was rearranged through hierarchical clustering using Euclid distances. Several *m/z* mentioned in the text were indicated by color (see also Figure 4.4). **A.** Threshold was $\rho > 0.05$. **B.** $\rho > 0.08$.

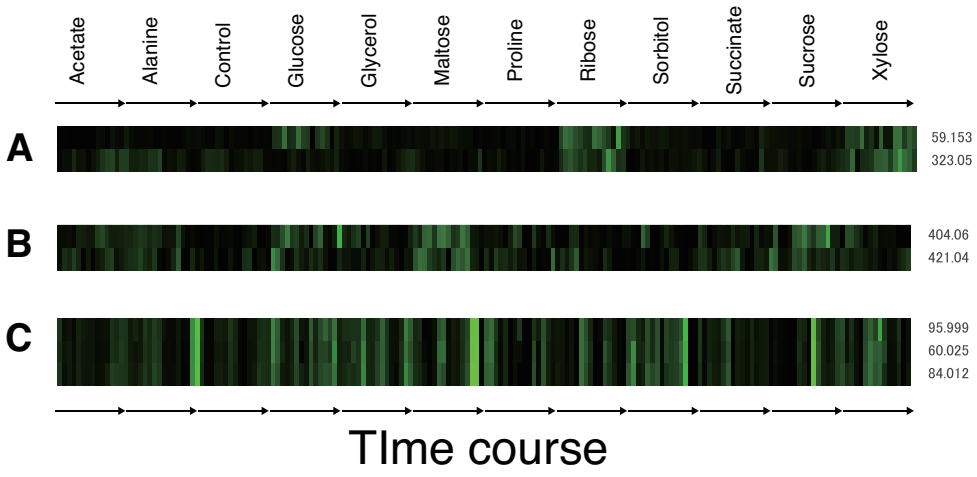


Figure 4.4 Excerpts of Centrality profiles of metabolite-metabolite correlation networks.

(Figure 4.4B). Since sucrose/maltose phosphate is a direct product from sucrose or maltose, its level could naturally influence the levels of other metabolites. One of centrality patterns with significant difference than the control distributed among several substrate cases, but these low molecular weight signals were unknown (Figure 4.4C). These centrality variations were not observed for the case of acetic acid and succinic acid, implying that this pattern was characteristic to glycolytic substrates. Such a pattern might serve as indirect information for the existence of certain class of substrates.

4.2.5 Consensus network involved with various nutritional fluctuations

The edges that showed significant differences in their correlations were selected by edge-wise t tests, using GGMs derived from the used substrates and ones from control data of every time points. The each selected edge was further filtered with a criterion that the edge had significant correlation coefficients with at least five substrates. Based on p-value of t test, top 2% of edges were selected. These parameters naturally affect the structure of the resulting differential network (data not shown), but following discussion was basically consistent even with 2-fold changes of the parameters. Although this approach might not be capable of capturing significantly altered correlations for a specific substrates, the main objective of the present study is to extract a consensus motif of the metabolite correlation network that altered from the control condition. It should also be noted that true specificity of one substrate could be elusive because the comprehensiveness of substrates was limited. The constructed differential networks derived from individual substrates (with an identical edge composition) were then expressed as a heatmap of correlation coefficients for the graph edges (Figure 4.5). The dendrogram of the substrates in the heatmap implied that the similarity of network structures could be associated with the classification of substrate compounds, *e.g.* ribose and

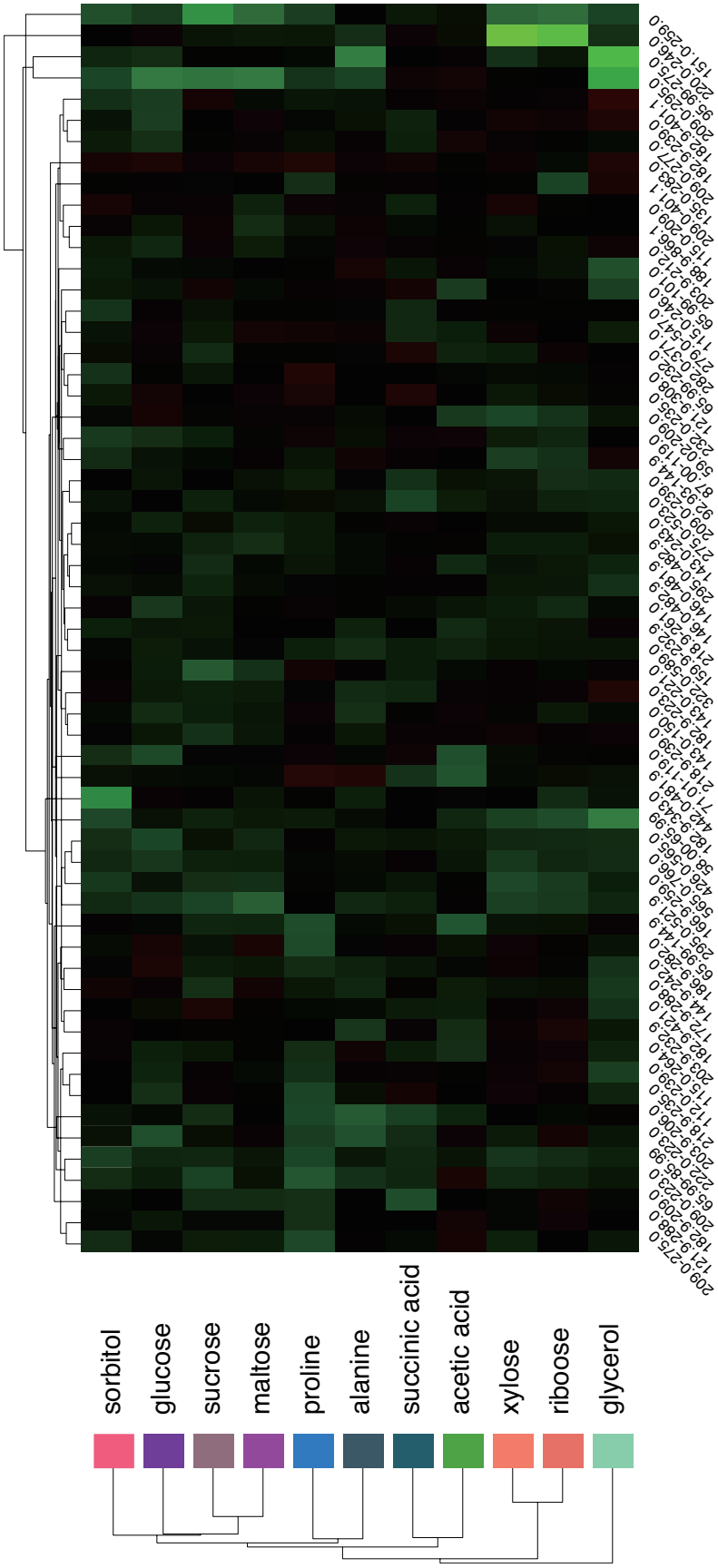


Figure 4.5 Differential correlation profiles

A partial correlation coefficient for each relationship was calculated using the time-course data of metabolite levels sectioned around a maxima of the adenylate energy charge. The profiles were visualized as a heat map and clustered by hierarchical clustering using a complete linkage method. The index color of each substrate was determined by mapping the three-dimensional coordination of each relationship in the similarity space of the differential profile onto a red-green-blue color space.

xylose led to similar differential networks. However, some typical carbon sources including glucose, glycerol and sorbitol exhibited rather distinct differential network structures. Such a tendency might indicate that specific responses were prepared for these substrates for realizing optimized adaptation. It was rather astonishing that the amino acids (alanine and proline) led to relatively similar differential networks, even though they possess fundamentally different chemical structures and involve with distinct metabolic pathways. A decisive difference between the amino acids and the other carbon sources was nitrogen element, implying that the correlation network was involved with a nitrogen-specific metabolic response. The differential correlation profile was then reconstructed into a network, where edges were colored to indicate the conformity with the substrates (Figure 4.6). Closeness of the hue represented the distances in the similarity space (up to the third principal component). There was an obvious hub metabolite (*m/z* = 209.0, referred as CpdA) possessing edges with various colors, implying that the metabolite exhibited condition-dependent correlations with various metabolites. The green edge between CpdA and the vertex of *m/z* 59.02 corresponded to the condition of acetic acid (*m/z* = 59.0). The edge between CpdA and the vertex of *m/z* 277.0 was colored purple, corresponding to glucose. The peak of *m/z* 277.0 might be derived from a water-adduct ion ($[M + H_2O]^-$) of F6P (*m/z* = 259.0). The vertex of *m/z* 115.0 could be derived from fumaric acid, a direct product of succinic acid in the TCA cycle. The edge between the putative fumaric acid and CpdA was colored with deep blue, roughly corresponding to amino acids and succinic acid. These observations thus imply that CpdA could be a mediator that informs the presence of various kinds of substrates. Although the identity of CpdA was unfortunately yet unknown, the consensus network suggested that a certain metabolite could serve as a key point where the information fir the availability of nutritional sources was converged.

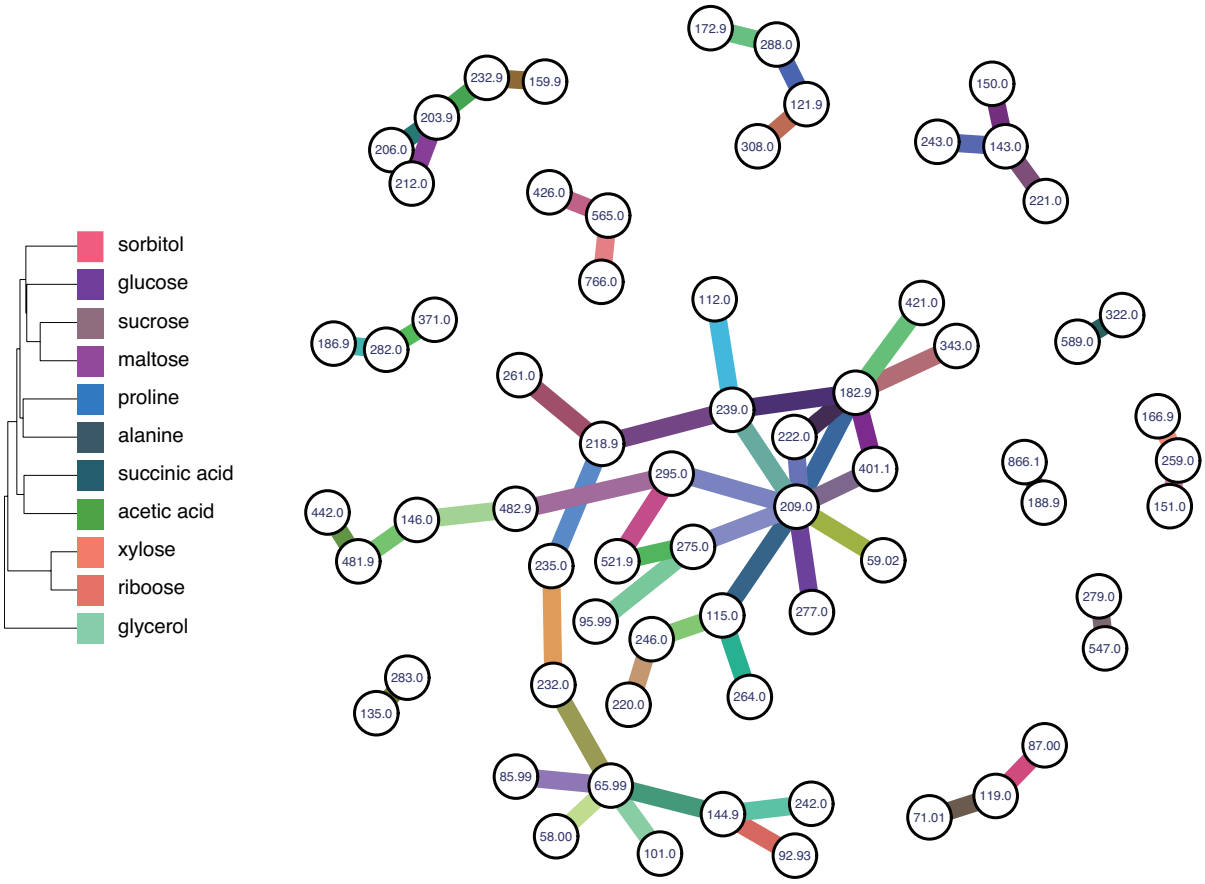


Figure 4.6 Consensus network in response to nutritional perturbations

The numbers in vertexes indicate *m/z* values of detected peaks. The colors of edges indicate the conformity with the substrates.

4.3 Conclusion

This study revealed the condition-dependent behavior of metabolite correlation network structure. Starting from the metabolites known to participate to the metabolic sensing of nutritional environment and its behavior in the correlation network, we extracted components that exhibited analogous network characteristics. Although preliminary, it was demonstrated that dynamic metabolomic analysis could delineate general network motif responsible for immediate adaptation to drastically changing environment. Still, more sophisticated methods for efficient accumulation of dynamic metabolome data was also indispensable. With on-going progression of analytical platforms for high-throughput metabolomics the scale of dynamic metabolome data would further grow, but our proposed calculative method based on network theory should yet applicable.

4.4 Materials and Methods

4.4.1 Chemicals

Tested substrates included glucose, ribose, xylose, sorbitol, glycerol, sucrose, maltose, alanine, proline, succinic acid and acetic acid, which were expected to serve as carbon source (Orth *et al.* 2011). All solvents including ultra pure water, metabolite standards and other chemicals were purchased from Sigma Aldrich (St. Louis, MO, USA).

4.4.2 Culture and induction of nutritional perturbation

E. coli strain JM109 was used for the direct metabolite analysis. Cultures were incubated Luria-Bertani medium (4 h, 150 rpm, 37°C). Bacterial cells were collected by centrifugation (6,000 g, 5 min, 37°C) and resuspended in water on a 96-well PCR plate (OD600 = 2). The cell suspension was further incubated in a block heater (37°C). A nutrition pulse was induced by adding either of substrates to give a final concentration of 1 g/L, and cell samples were harvested from the suspension both before and after the nutritional fluctuation.

4.4.3 Sampling

Matrix solution (6 mg/mL 9-AA in 80% methanol) was used to quench intracellular metabolism. Each sampling was performed by mixing 10 μL of suspension with 60 μL of the pre-cooled matrix solution (−40°C). The sampling interval was fixed at 10 s. For each time-course sample acquisition, 1 samples were taken prior to the nutritional perturbation

induction and 15 post-induction, resulting in a sample set of 15 time points over 150 sec. The cell culturing and sampling was replicated by ten times.

4.4.4 Mass spectrometry

For time-course metabolite analysis, a time-of-flight type MALDI-MS instrument (AXIMA Performance, Shimadzu, Japan) was used. The technique was previously introduced as a high-throughput and highly sensitive metabolite analysis. In brief, 1 μL of the analyte was applied onto a ground-steel MALDI sample plate and air-dried to give a sample spot. The spots were irradiated at a laser power that gave satisfactory ion intensity, and all analyses were performed using the same laser power in the negative ionization mode. Mass spectra were obtained by MALDI-MS analysis where five laser shots were accumulated and 256 spectra were averaged per spot. Analysis time was less than 20 sec/spot. Four spots were deposited from an individual sample and averaged to apply to further data analyses. Mass spectra were internally calibrated using the internal standard and peaks that appear constantly.

4.4.5 Raw data processing

The following data processing was conducted using R language (R Core Team 2012). Peak pick, normalization and peak alignment were conducted using an R package MALDIquant (Gibb and Strimmer 2012). The cut-off threshold was signal-to-noise ratio of 5 and mass error tolerance was 0.05 Da. Peak intensity in a spectrum was normalized to give a zero mean and unit variance throughout the time course. Missing values were replaced with zero. The background and unaffected metabolite peak was excluded by student's *t*-test using the control case, where *E. coli* cells were prepared in the same manner as the other cases but

no nutritional pulse was applied. Resulting data matrix has the time points in row, and the detected peaks (132 in total) in column (a time-course metabolite profile; TC). TC is reconstructed for each case of substrate (12 in total including the control) and experimental replication (n = 10).

4.4.6 Evolving partial correlation network

The graphical Gaussian modeling (GGM) framework was employed as described in Chapter 3. A temporal correlation network was constructed using TCs of two successive time points for each substrate, resulting in [total number of time points – 1] networks. These networks are denoted as *temporal correlation networks*. Centrality profile of temporal correlation networks was constructed as described in Chapter 3.

Chapter 5.

Ion Yield in MALDI-MS Analysis of Metabolites Quantitatively Associated with the Structural Properties of Participating Compounds: A quantitative structure-property relationship (QSPR) Approach

5.1 Introduction

MALDI-MS has come to play a unique role in the analysis of low-molecular-weight biological compounds, principally metabolites (Edwards and Kennedy 2005, Dally *et al.* 2003). It is well known that the scope of detectable compounds in the MALDI-MS analysis is strongly associated with the molecular species of the matrix.

Many compounds have been reported as the MALDI matrices. They are typically aromatic acids in the case of the positive ion mode, and aromatic bases in negative. They commonly have optical absorption near the wavelength of the laser and contain a chromophor, though it is not always necessary. In the positive ion mode, a storng acid such as trifluoroacetic acid is added to matrix solution as a proton donor. The most representative matrix compounds are 2,5-dihydroxy benzoic acid (DHB) and α-cyano-4-hydroxy cinnamic acid (CHCA). Sinapinic acid or ferulic acid are also occasionally used. These matrix compounds are utilized mostly for the positive ionization of microorganisms and other class of compounds mentioned above. For metabolite analysis, 9-AA is the most popular matrix compound because it can support negative ionization and most of intracellular metabolites tend to have negative charges.

Co-use of different compounds other than matrix compounds is frequently attempted, *e.g.* a surfactant to control the interference of the matrix on the mass spectrum (Guo *et al.* 2002). A high molecular weight is also used as alternative approach (Ayorinde *et al.* 1999, Ayorinde *et al.* 2000). In addition to low-molecular-weight organic compounds, a variety of inorganic materials are utilized as MALDI matrix (Sunner *et al.* 1995, Chen *et al.* 1998). As for different kinds of carbon-based materials, other studies utilized graphite plates (Sunner *et*

al. 1995), graphite suspensions in different solvents (Sunner *et al.* 1995, Dale *et al.* 1996), graphite trapped in a silicone polymer, (Li *et al.* 2005), pencil lead (Langley *et al.* 2007, Black *et al.* 2006), and grapheme (Dong *et al.* 2010). Furthermore, Matrix-free LDI is yet another MALDI technique that use a photoactive but non-desorbable materials on which samples are embedded and analyzed without any matrix compound co-crystallized (Cohen, Go, & Siuzdak, 2007; Najam-ul-Haq et al., 2007; Petterson, 2007). These strategies are in continual development to find more amenable analysis of small molecules without matrix interferences. Although no matrix is used and some differences exist in the mechanism of ion formation, this approach may be regarded as a derivative of MALDI. The main advantage of this approach is that little or no background signals are observed in the mass spectrum. In addition, sample preparation is simplified because no matrix is involved. Recently, many efforts have been made to eliminate matrix ion interference by using different matrix substances, such as desorption/ionization on porous silicon (Zou *et al.* 2002, Zhang *et al.* 2001, Lewis *et al.* 2003, Wei *et al.* 1999). Nevertheless, conventional MALDI that utilizing low-molecular-weight compounds as matrix compounds still has potential advantage for controlling the sensitivity and specificity, because it is empirically well known that different matrix compounds supports distinct class of compounds with various sensitivity. However, although the modification of matrix compounds would alter their ionization capability, there is no general strategy to design the molecular structure for desired ionization property due to insufficient understanding of MALDI events.

In the present study, we aimed to model the relationship between the structural properties of the metabolites and matrix compounds that contribute to MALDI efficiency. QSPR models have been investigated in a variety of fields to summarize supposed relationships between chemical structures and corresponding bioactivities or physical

properties (*e.g.* metabolic activation (Liew *et al.* 2012) or azeotropic boiling points (Katritzky *et al.* 2011)). The chemical structure could be represented by molecular descriptors that represent certain perspectives of the structure through specialized mathematical procedures.

The present study dedicated to constructing a metabolite QSPR for MALDI efficiency, employing 9-AA as a representative matrix compound for metabolite analysis (Vermillion-Salsbury and Hercules 2002). The MALDI-MS analyses with 9-AA (9-AA-MALDI-MS) have been utilized for various studies, including high-throughput and highly sensitive metabolite analyses (Miura *et al.* 2010b, Yukihira *et al.* 2010, Amantonico *et al.* 2008b) as well as metabolite MS imaging (Miura *et al.* 2010a, Miura *et al.* 2012). In the targeted analyses, the merit of metabolite property modeling lies in the prediction of the probability of the ionization of metabolites yet to be analyzed in MALDI-MS. In the non-targeted analyses, on the other hand, the model would work to screen chemical structures plausibly assigned to a detected peak, even if compounds with similar m/z values are not distinguishable. A metabolite QSPR model was constructed using a 9-AA-MALDI-MS ionization profile of 200 metabolite standard compounds, which were selected to cover a wide range of structural diversity and biological importance. The importance of the descriptors was estimated and discussed with regard to the relevance to the ionizability and ionization efficiency of the compounds.

5.2 Results and Discussion

5.2.1 Ionization profile of metabolites in 9-AA-MALDI-MS analysis

First, we investigated the ionizability and ionization efficiency of 200 compounds to clarify the coverage of 9-AA-MALDI-MS for the metabolite analysis. As a result of the test analysis, 100 out of 200 compounds were detected as deprotonated peaks. The LOD value ranged from 0.00125 ppm to 100 ppm. As the chemical diversity defines the applicability of models constructed using the dataset, the taxonomy superclass of the metabolites in the sample set was summarized in Table 5.1. In general, phosphorylated compounds, such as nucleotides and sugar phosphates, exhibited excellent ionization efficiency (Table 5.1, Minimum LOD of the *"Nucleosides, Nucleotides, and Analogues"* class and "*Carbohydrates and Carbohydrate Conjugates"* class. Refer to Online Resource 2 for full information). This observation was understandable, because negatively charged functional groups easily produce negative ions through deprotonation. Indeed, these compounds have been detected in previous studies using 9-AA-MALDI-MS (Vaidyanathan and Goodacre 2007b, Miura *et al.* 2010b, Yukihira *et al.* 2010, Amantonico *et al.* 2008b, Shroff *et al.* 2007b). On the other hand, the compounds in the "*Carbohydrates and Carbohydrate Conjugates"* class other than sugar phosphates, *i.e.* oligosaccharides and sugar acids, exhibited poor ionization efficiencies. Lipids were scarcely detected, except for cholates (6%, 2/36). Amino acids showed the best ionizability (76%, 45/59) at a broad range of LODs (from 0.00125 ppm to 100 ppm). Although many compounds in the sample set contained one or more carboxylic group to produce negative ions, the ionization efficiencies of these compounds were not directly

Table 5.1 The ionization profiles of metabolites in the 9-AA-MALDI-MS analysis and the predictive accuracy of the Random forest ionizability models.

			Correct rate of prediction model					LOD (ppm)	
Superclass	#	Ionized	*Global*	*2D*	*3D*	*KRFP*[c]	*MACCSFP*[c]	Minimum	Max
Aliphatic acyclic compounds	10	1 (10%)	1.000	1.000	0.900	1.000	1.000	0.00125	0.00125
Amino acids/peptides	59	45 (76%)	0.763[a] 0.932[b]	0.797[a] 0.949[b]	0.932[a] 1.000[b]	0.763[a] 0.847[b]	0.746[a] 0.780[b]	0.0125	100
Aromatic heteromonocyclic compounds	12	6 (50%)	0.917	0.833	1.000	0.833	0.667	0.157	2.50
Aromatic heteropolycyclic compounds	8	3 (38%)	0.875	0.875	1.000	0.750	0.750	0.0313	50.0
Aromatic homomonocyclic compounds	12	8 (67%)	0.750	0.750	0.750	0.750	0.667	0.050	100
Carbohydrates	13	6 (46%)	0.846	0.846	0.846	0.769	0.769	0.0250	25.0
Lipids	36	2 (6%)	0.972	0.944	0.917	0.917	0.944	0.157	0.312
Nucleosides/nucleotides	27	19 (70%)	0.963	0.963	0.889	0.889	0.704	0.00625	6.25
Organic acids	14	8 (57%)	0.571	0.643	0.857	0.714	0.500	0.0500	12.5
Others	9	2 (22%)	0.889	0.889	1.000	0.889	0.778	0.625	100
Whole compounds	200	100 (50%)	0.850	0.855	0.910	0.825	0.765	0.00125	100

[a] Predicted by whole-data model.

[b] Predicted by amino acid model.

[c] The other available types of fingerprint were omitted because of their moderate performance.

associated with the existence of a carboxylic group (Table 5.1, 5.2). Nevertheless, the compounds in the "*Organic Acids and Derivatives"* class showed moderate ionization efficiencies (from 0.0500 ppm to 12.5 ppm).

Interestingly, distinct ionization profile was observed even in compounds with a similar structure (*e.g.* β-alanine and sarcosine, or proline and pipecolic acid, Figure 5.1a). In these cases, β-alanine and isoleucine exhibited concentration-dependent peak intensity in MALDI-MS analysis, while sarcosine and pipecolic acid were not detected (data not shown). Generally, structural similarity of low-molecular-weight compounds should give similar physicochemical properties. In contrast, these observations strongly indicated that apparent properties of the molecule, such as the presence of functional groups, are insufficient to explain the diverse ionization profiles of the compounds.

5.2.2 QSPR model for ionizability and relevance of descriptor class

The physicochemical factors of the metabolites that influenced the ionization profiles were of interest. To address these factors, we performed non-hypothesis-based statistical modeling, where the source of efficient MALDI was sought by molecular descriptors of target compounds. First, we constructed a Random forest QSPR model for the ionizability prediction (*ionized* or *not ionized*) using the whole descriptor provided by the PaDEL-Descriptor (*Global* model). The overall accuracy of the prediction was 85.0%, and there were no significant biases with regard to the estimation error and the metabolite class (Table 5.1, *Global* model for whole compounds).

The prediction model was then investigated to estimate the prerequisite properties for the ionization of a compound in a 9-AA-MALDI-MS analysis. In the *Global* model, the descriptors with higher importance indicated the electrotopological state of strength for

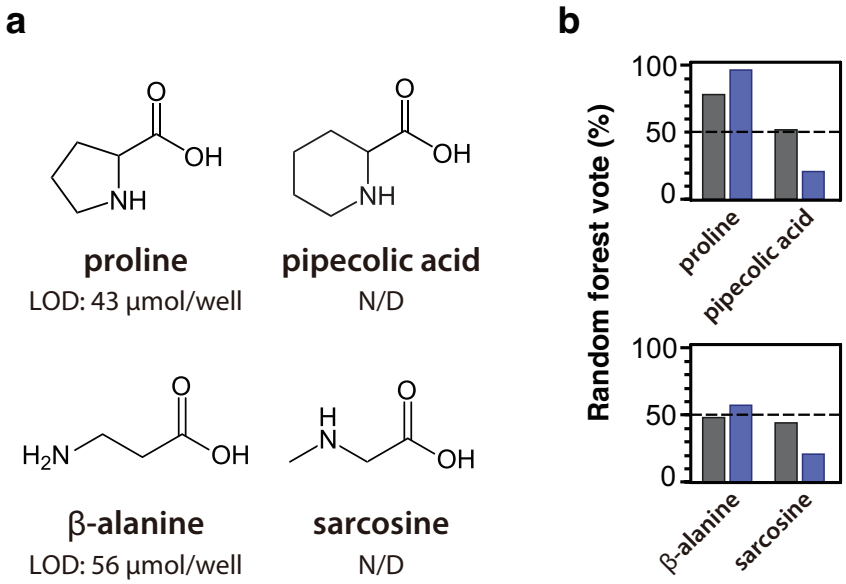


Figure 5.1 Distinct ionization profiles of structurally similar compounds in MALDI-MS analysis and their Random forest prediction.

a. Structural formulas and LODs of four representative compounds with similar structures but distinct ionization profiles. **b.** The prediction of ionizability by the 3D descriptor model for whole compounds (gray bar) and by the 3D descriptor amino-acid-specific model (blue bar) represented as the votes of the ensemble trees. When the ratio of positive vote (ionizable) exceeds 50%, the corresponding compound is predicted to be ionizable.

potential hydrogen bonds and the area of the negatively-charged surface (Figure 5.2a and Table 5.2). These descriptors belong to the 2D and 3D descriptors, respectively. The electrotopological state value (E-state value) is a kind of 2D descriptor that combines both the electronic characteristics and the topological environment of each skeletal atom in the molecule (Hall and Kier 1995). The importance of the E-state value indicated that the strength of possible hydrogen bonds positively correlated with the ionizability in MALDI. It was clear that the ionization profiles were strongly influenced by the interaction between molecules. In addition to the global model, which incorporated all the type of descriptors available, the respective types of descriptors were applied to construct Random forest prediction models to investigate the relevance of each descriptor types to the prediction performance (Table 5.1). As the result, *3D* model exhibited the highest performance followed by *2D* model (91.0% and 85.5% accuracy rate for whole compounds, respectively). Considering the variable importance of these models (Figure 5.2b, c), although the strength of hydrogen bonds well represented the ionization profile, the information of charged surface area led to a better ionizability model. This result was reasonable because the charged surface area indicated the electron distribution within the molecules that should cover the effect of hydrogen bond acceptors. The further functioning of the negatively charged surface area could be the effectiveness of proton abstraction in the interaction with 9-AA.

The constructed prediction models for amino acids (*"Amino Acids, Peptides, and Analogues"* class) exhibited relatively poor accuracy, even though they were a major class in our data set. Our models were effective for a broad spectrum of metabolites, but they still lacked the ability to model rather faint structural differences of amino acids. The reason of this defect could be strongly attributed to the relevance of hydrogen bonds. As both amines and carboxyl groups in amino acids can form hydrogen bonds, the ionizatilities of amino

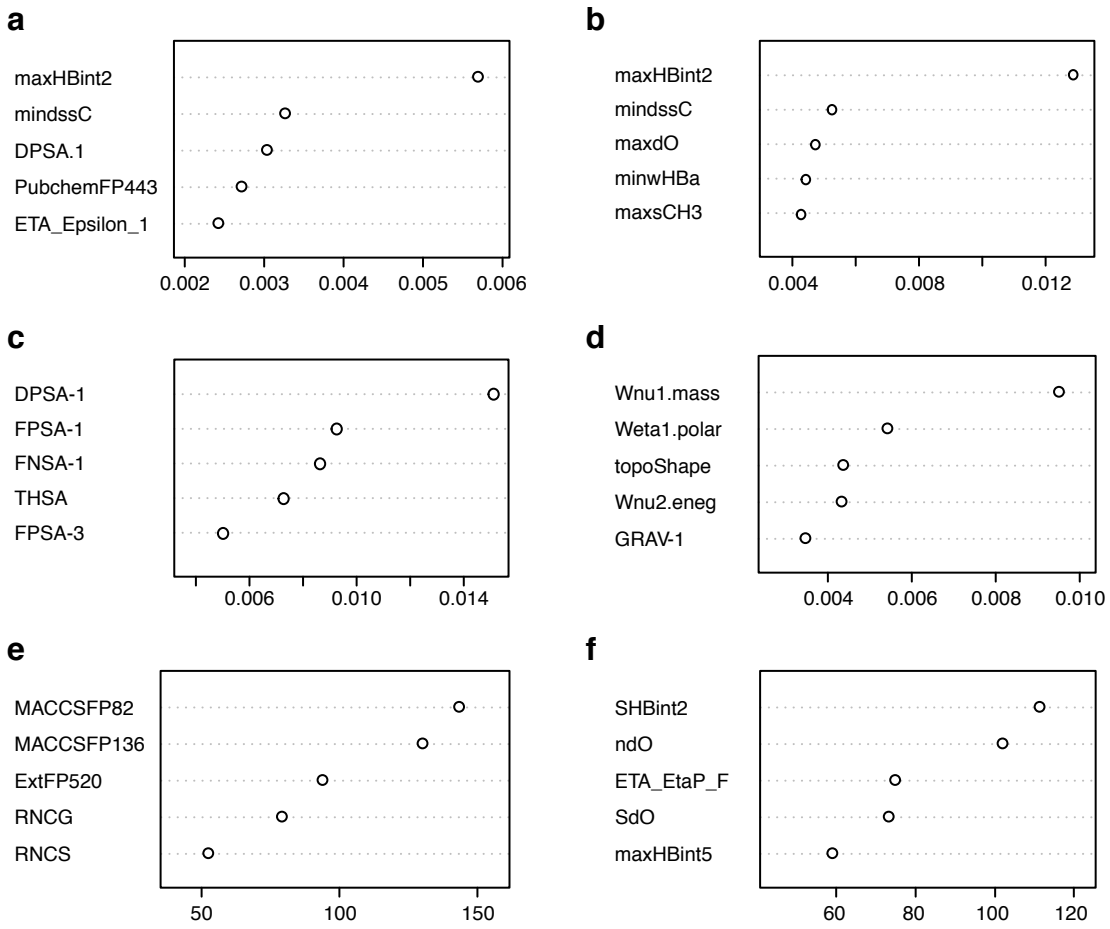


Figure 5.2 The variable importance of the Random forest models.

Each panel indicates the variable importance for the following models. The descriptions for individual descriptors can be found in Table 5.2. **a.** The *Global* ionizability model for the whole compounds. **b.** The *2D* ionizability model for the whole compounds. **d.** The *3D* ionizability model for the whole compounds. **d.** The *3D* ionizability model for amino acids. **e.** The *Global* ionization efficiency model for the whole compounds. **f.** The *2D* ionization efficiency model for the whole compounds.

Table 5.2 The list of the descriptors with the higher importance for each model.

Descriptor	Description
Global Ionizability model	
maxHBint2	Maximum E-State descriptors of strength for potential Hydrogen Bonds of path length 2
mindssC	Minimum atom-type E-State: =C<
DPSA-1	Difference of PPSA-1 (Partial positive surface area -- sum of surface area on positive parts of molecule) and PNSA-1 (Partial negative surface area -- sum of surface area on negative parts of molecule)
PubchemFP443	C(-C)(=O)
ETA_Epsilon_1	A measure of electronegative atom count
2D ionizability model	
maxHBint2	Maximum E-State descriptors of strength for potential Hydrogen Bonds of path length 2
mindssC	Minimum atom-type E-State: =C<
maxdO	Maximum atom-type E-State: =O
minwHBa	Minimum E-States for weak Hydrogen Bond acceptors
maxsCH3	Maximum atom-type E-State: -CH3
3D ionizability model	
DPSA-1	Difference of PPSA-1 (partial positive surface area -- sum of surface area on positive parts of molecule) and PNSA-1 (partial negative surface area -- sum of surface area on negative parts of molecule)
FPSA-1	PPSA-1 / total molecular surface area
FNSA-1	PNSA-1 / total molecular surface area
THSA	Sum of solvent accessible surface areas of atoms with absolute value of partial charges less than 0.2
FPSA-3	PPSA-3 (charge weighted partial positive surface area) / total molecular surface area
3D amino acid ionizability model	
Wnu1.mass	Directional WHIM, weighted by atomic masses
Weta1.polar	Directional WHIM, weighted by atomic polarizabilities
topoShape	Petitjean topological shape index
Wnu2.eneg	Directional WHIM, weighted by Mulliken atomic electronegativites
GRAV-1	Gravitational index of heavy atoms
Global ionization efficiency model	
MACCSFP82	ACH2QH

MACCSFP136	O=A > 1
RNCS	Relative negative charge surface area -- most negative surface area * RNCG
RNCG	Relative negative charge -- most negative charge / total negative charge
2D ionization efficiency model	
ndO	Count of atom-type E-State: =O
ETA_EtaP_F	Functionality index EtaF relative to molecular size
SHBint2	Sum of E-State descriptors of strength for potential Hydrogen Bonds of path length 2
SdO	Sum of atom-type E-State: =O
maxHBint5	Maximum E-State descriptors of strength for potential Hydrogen Bonds of path length 5

acids could be overestimated. To address these issues, we attempted to improve the prediction performance for amino acids because they are one of the most important classes in the metabolite analysis because of their significant metabolic and regulatory versatility (Wu 2009). We thus developed new models specific for amino acids to improve the predictive accuracy and investigate the relevant structural properties. Again, the models were constructed using the whole or the individual types of descriptors. As a result, the accuracy of model prediction improved for all types of descriptors (Table 5.1). Especially, the *3D* model achieved a perfect prediction of the ionizability, even for the above-mentioned pairs of structurally similar amino acids (Figure 5.1b). Fingerprinting descriptors provided still a moderate accuracy (84.7% correct rate for the highest value by the *KRFP* model), indicating that the presence of substructures was insufficient to fully represent the ionizability of amino acids. Unlike the class-independent model (whole-data model), the relevant 3D descriptors were not involved with the charged surface areas, but Weighted Holistic Invariant Molecular (WHIM) descriptors (Todeschini *et al.* 1994) (Figure 5.2d). WHIM descriptors provide information about the whole 3D-molecular structure in terms of the size, shape, symmetry and atom distribution. This result was intriguing because the shape of the molecules itself was relevant rather than electronic properties. It has been reported that cation affinities of amino acids were associated with degree of linearity (Siu and Che 2006), which is a direct index of the flexibility of molecule (Devillers and Balaban 1999). Hence, it was suggested that the shape properties of target compounds affect their interaction with other molecules to promote or inhibit their ionization.

5.2.3 QSPR model for ionization efficiency

The Random forest method is applicable to a regression, averaging the output of

decision trees (Breiman 2001). The experimentally evaluated ionization efficiency, indicated by LOD values, was also modeled by the Random forest method using individual types of descriptors. While the *Global* ionization efficiency model reached $\rho = 0.69$ (Figure 5.3a, and the variable importance was shown in Figure 5.2e), the best predictive performance was achieved with 2D descriptors, evaluated as $\rho = 0.73$ (*2D* model, Figure 5.4, and the variable importance was shown in Figure 5.2f). It was supposed that the fundamental trend of the ionization efficiency was reasonably modeled. The MACCSFP also provided a highly accurate model compared to the *2D* and *Global* models ($\rho = 0.66$, Figure 5.3b). The *3D* model showed an inferior performance ($\rho = 0.60$) to the above-mentioned models, in spite of the relevance of 3D descriptors in the *Global* model (Figure 5.3c). The *2D* model indicated that the quantitative extent of ionization was mainly associated with E-state index of double-bonded oxygen and the strength of the potential hydrogen bonds (Figure 5.2f). Hence, overall results indicated that the partial negative charge in the molecule could be a prerequisite for ionization, and that the richness of carbonyl oxygen should be preferable for efficient negative MALDI because of the basic condition brought by 9-AA. Noteworthy, structural flexibility of the target compounds might play a special role to specific interaction with other molecules, presumably the matrix molecules to ruduce ionization energies (Kinsel *et al.* 2002), which determine the fate of their ionization profiles.

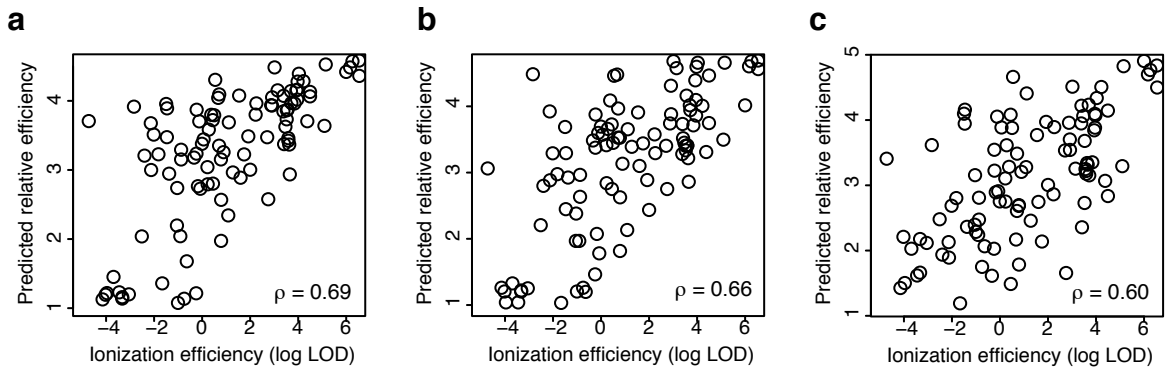


Figure 5.3 The prediction of Random forest ionization efficiency models.

The predicted ionization efficiencies provided by the following models were plotted against the measured ionization efficiencies. See main text for the *2D* model, which achieved the best performance (ρ = 0.73). **a.** The *Global* model. **b.** The *MACCSFP* model. **c.** The *3D* model.

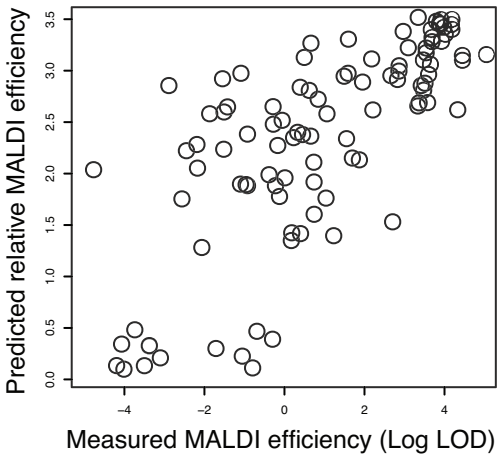


Figure 5.4 The Random forest regression model for the ionization efficiency in 9-AA-MALDI.

The *2D* model showed the best performance in terms of the regression for ionization efficiency. The rank correlation coefficient for the plot was indicated as ρ. The models of other types (Global, MACCSFP and 3D) can be found in Figure 5.3.

5.3 Conclusion

T his study was primarily intended to lead to more rational and predictive MALDI-MS analyses. In contrast to empirical approaches, this study employed a systematic analysis of the ionization profile in 9-AA-MALDI-MS for the first time. In the MALDI-MS analysis, the ionizability prediction model evaluates the likelihood of peak identification. On the other hand, the ionization efficiency model would help to estimate the abundance of the metabolite based on the observed signal intensity. The relevant descriptors found in this study can be interpreted as the structural preference specific to 9-AA and/or negative mode MALDI-MS analysis. The QSPR approach should also be applicable for other MALDI matrices to characterize the structural properties of target compounds for preferred ionization. Such information will play an indispensable role in the strategic development of MALDI-MS-based studies.

5.4 Materials and Methods

5.4.1 MALDI-MS analysis of metabolite standards

The ionizability and ionization efficiency in MALDI-TOF-MS (AXIMA Confidence, Shimadzu, Japan) analysis for each standard compound was assessed using 9-AA as the matrix. Ionization efficiency was represented as limit of detection (LOD) value in ppm.

All the metabolites used in the present study were purchased from Sigma Aldrich (MO, USA) or Wako Pure Chemicals (Osaka, Japan). Individual standard compounds were dissolved in water or DMSO, and diluted to give graded concentrations (from 0.00125 ppm to 100 ppm) and mixed with a 9-AA/methanol solution (10 mg/mL) at a ratio of 1:1 (*v/v*). Deionized water was obtained from a Milli-Q system (Millipore, Schwalbach, Germany). One milliliter of the sample was spotted onto the ground-steel MALDI plate and air-dried. Four spots were deposited from an individual sample and averaged to apply for the further data analyses. A MALDI-TOF-MS (AXIMA Confidence, Shimadzu, Japan) was used for all the analyses. A mass spectrum was acquired with five laser shots. For each sample spot, 256 spectra were averaged. For each metabolite sample, deprotonated peaks were sought with a threshold signal-to-noise ratio (more than 5) to confirm its ionizability and the limit of detection (LOD) of the metabolite. The ionization of each metabolite was confirmed by a deprotonated peak ($[M-H]^-$), because deprotonated ions are exclusively generated in the negative mode MALDI, while alkali-metal adduct ions and protonated ions are generated in the positive mode. The full list of examined compounds can be found in Table 5.3.

Table 5.3 Limit of detection (LOD) of metabolites measured in 9-AA-MALDI-MS analysis.

Compound	Class	LOD
PEP	Aliphatic Acyclic Compounds	0.00125
agmatine	Aliphatic Acyclic Compounds	N/D
betaine aldehyde	Aliphatic Acyclic Compounds	N/D
choline	Aliphatic Acyclic Compounds	N/D
Diethanolamine	Aliphatic Acyclic Compounds	N/D
ethanolamine	Aliphatic Acyclic Compounds	N/D
putrescine	Aliphatic Acyclic Compounds	N/D
spermidine	Aliphatic Acyclic Compounds	N/D
Spermine	Aliphatic Acyclic Compounds	N/D
urea	Aliphatic Acyclic Compounds	N/D
5-oxoproline	Aliphatic Heteromonocyclic Compounds	0.625
ascorbate (Vitamin C)	Aliphatic Heteromonocyclic Compounds	100
allantoin	Aliphatic Heteromonocyclic Compounds	N/D
erythrose	Aliphatic Heteromonocyclic Compounds	N/D
gammma-Butyrolactone	Aliphatic Heteromonocyclic Compounds	N/D
1-6-Anhydro-β-D-Glucose	Aliphatic Heteropolycyclic Compounds	N/D
myo-inositol	Aliphatic Homomonocyclic Compounds	N/D
trans-4-hydroxyproline	Amino Acids, Peptides, and Analogues	0.0125
L-Cysteate	Amino Acids, Peptides, and Analogues	0.025
N-acetylglutamate	Amino Acids, Peptides, and Analogues	0.025
xanthurenate	Amino Acids, Peptides, and Analogues	0.04
kynurenate	Amino Acids, Peptides, and Analogues	0.05
L-aspartate	Amino Acids, Peptides, and Analogues	0.05
L-glutamate	Amino Acids, Peptides, and Analogues	0.1
N-acetylaspartate	Amino Acids, Peptides, and Analogues	0.1
N-acetylleucine	Amino Acids, Peptides, and Analogues	0.1
O-acetylserine	Amino Acids, Peptides, and Analogues	0.1
Nicotinurate	Amino Acids, Peptides, and Analogues	0.15625
L-asparagine	Amino Acids, Peptides, and Analogues	0.2
L-glutamine	Amino Acids, Peptides, and Analogues	0.2
L-histidine	Amino Acids, Peptides, and Analogues	0.2

N-acetyl-aspartyl-glutamic acid	Amino Acids, Peptides, and Analogues	0.2
5-Aminolevulinate	Amino Acids, Peptides, and Analogues	0.3125
phenylacetylglycine	Amino Acids, Peptides, and Analogues	0.3125
L-Homoserine	Amino Acids, Peptides, and Analogues	0.5
citrulline	Amino Acids, Peptides, and Analogues	0.625
GSH	Amino Acids, Peptides, and Analogues	0.625
homocitrulline	Amino Acids, Peptides, and Analogues	0.625
N-acetylasparagine	Amino Acids, Peptides, and Analogues	0.625
N-acetylproline	Amino Acids, Peptides, and Analogues	0.625
cysteine-glutathione disulfide	Amino Acids, Peptides, and Analogues	2.5
glycine	Amino Acids, Peptides, and Analogues	2.5
L-phenylalanine	Amino Acids, Peptides, and Analogues	2.5
N-acetylphenylalanine	Amino Acids, Peptides, and Analogues	2.5
Quinaldic acid	Amino Acids, Peptides, and Analogues	2.5
2-aminobutyrate	Amino Acids, Peptides, and Analogues	5
L-alanine	Amino Acids, Peptides, and Analogues	5
L-isoleucine	Amino Acids, Peptides, and Analogues	5
L-leucine	Amino Acids, Peptides, and Analogues	5
L-serine	Amino Acids, Peptides, and Analogues	5
L-tryptophan	Amino Acids, Peptides, and Analogues	5
ornithine	Amino Acids, Peptides, and Analogues	5
1-methylhistidine	Amino Acids, Peptides, and Analogues	10
3-methylhistidine	Amino Acids, Peptides, and Analogues	10
L-arginine	Amino Acids, Peptides, and Analogues	10
L-cysteine	Amino Acids, Peptides, and Analogues	10
L-methionine	Amino Acids, Peptides, and Analogues	10
L-proline	Amino Acids, Peptides, and Analogues	10
L-tyrosine	Amino Acids, Peptides, and Analogues	10
L-valine	Amino Acids, Peptides, and Analogues	10
N-acetylcysteine	Amino Acids, Peptides, and Analogues	10
N-acetylvaline	Amino Acids, Peptides, and Analogues	10
β-alanine	Amino Acids, Peptides, and Analogues	5
5-aminovalerate	Amino Acids, Peptides, and Analogues	N/D
L-pipecolate	Amino Acids, Peptides, and Analogues	N/D

sarcosine (N-Methylglycine)	Amino Acids, Peptides, and Analogues	N/D
(R)-S-Lactoylglutathione	Amino Acids, Peptides, and Analogues	N/D
1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide	Amino Acids, Peptides, and Analogues	N/D
D-Alanyl-D-alanine	Amino Acids, Peptides, and Analogues	N/D
L-Homocystein	Amino Acids, Peptides, and Analogues	5
L-lysine	Amino Acids, Peptides, and Analogues	5
L-threonine	Amino Acids, Peptides, and Analogues	5
N-acetylglutamine	Amino Acids, Peptides, and Analogues	N/D
N-acetylmethionine	Amino Acids, Peptides, and Analogues	N/D
N-acetyltyrosine	Amino Acids, Peptides, and Analogues	N/D
quinolinate	Amino Acids, Peptides, and Analogues	N/D
nicotinate	Aromatic Heteromonocyclic Compounds	0.15625
uracil	Aromatic Heteromonocyclic Compounds	0.15625
urocanate	Aromatic Heteromonocyclic Compounds	0.2
histamine	Aromatic Heteromonocyclic Compounds	0.3125
4-imidazoleacetate	Aromatic Heteromonocyclic Compounds	0.625
pyridoxal	Aromatic Heteromonocyclic Compounds	2.5
1-methylnicotinamide	Aromatic Heteromonocyclic Compounds	N/D
5-methylcytosine	Aromatic Heteromonocyclic Compounds	N/D
cytosine	Aromatic Heteromonocyclic Compounds	N/D
nicotinamide	Aromatic Heteromonocyclic Compounds	N/D
Pyridoxamine	Aromatic Heteromonocyclic Compounds	N/D
thymine	Aromatic Heteromonocyclic Compounds	N/D
xanthine	Aromatic Heteropolycyclic Compounds	0.03125
5-hydroxyindoleacetate	Aromatic Heteropolycyclic Compounds	10
riboflavin (Vitamin B2)	Aromatic Heteropolycyclic Compounds	50
5MeTHF	Aromatic Heteropolycyclic Compounds	N/D
biliverdin	Aromatic Heteropolycyclic Compounds	N/D
Folate	Aromatic Heteropolycyclic Compounds	N/D
thiamin (Vitamin B1)	Aromatic Heteropolycyclic Compounds	N/D
Thiamine diphosphate	Aromatic Heteropolycyclic Compounds	N/D
anthranilate	Aromatic Homomonocyclic Compounds	0.05
gentisate	Aromatic Homomonocyclic Compounds	0.2

4-Coumarate	Aromatic Homomonocyclic Compounds	1.25
3-(3-hydroxyphenyl)propionate	Aromatic Homomonocyclic Compounds	5
3-hydroxyphenylacetate	Aromatic Homomonocyclic Compounds	10
4-Aminobenzoate	Aromatic Homomonocyclic Compounds	10
4-hydroxyphenylpyruvate	Aromatic Homomonocyclic Compounds	10
4-hydroxyphenylacetate	Aromatic Homomonocyclic Compounds	100
3,4-dihydroxyphenylacetate	Aromatic Homomonocyclic Compounds	N/D
Benzoate	Aromatic Homomonocyclic Compounds	N/D
Homogentisate	Aromatic Homomonocyclic Compounds	N/D
tyramine	Aromatic Homomonocyclic Compounds	N/D
α-D-Glucose 6-phosphate	Carbohydrates and Carbohydrate Conjugates	0.025
ribose 5P	Carbohydrates and Carbohydrate Conjugates	0.025
gluconate	Carbohydrates and Carbohydrate Conjugates	0.3125
pseudouridine	Carbohydrates and Carbohydrate Conjugates	0.3125
N-acetylneuraminate	Carbohydrates and Carbohydrate Conjugates	1.25
sucrose	Carbohydrates and Carbohydrate Conjugates	25
15-AG	Carbohydrates and Carbohydrate Conjugates	N/D
glycerol	Carbohydrates and Carbohydrate Conjugates	N/D
gulono-1,4-lactone	Carbohydrates and Carbohydrate Conjugates	N/D
raffinose	Carbohydrates and Carbohydrate Conjugates	N/D
ribitol	Carbohydrates and Carbohydrate Conjugates	N/D
stachyose	Carbohydrates and Carbohydrate Conjugates	N/D
xylonate	Carbohydrates and Carbohydrate Conjugates	N/D
phosphate	Homogeneous Non-metal Compounds	N/D
fumarate	Lipids	0.15625
4-acetamidobutanoate	Lipids	0.3125
3-Methyladipic acid	Lipids	N/D
6-Aminohexanoate	Lipids	N/D
7-dehydrocholesterol	Lipids	N/D
acetylcarnitine	Lipids	N/D
alpha-tocopherol	Lipids	N/D
beta-sitosterol	Lipids	N/D
caprate (10:0)	Lipids	N/D
caproate (6:0)	Lipids	N/D

caprylate (8:0)	Lipids	N/D
cholate	Lipids	N/D
cholesterol	Lipids	N/D
corticosterone	Lipids	N/D
D-Mannitol	Lipids	N/D
ethylmalonic acid	Lipids	N/D
glycerophosphorylcholine	Lipids	N/D
heptanoate (7:0)	Lipids	N/D
hyodeoxycholate	Lipids	N/D
isovalerate	Lipids	N/D
laurate (12:0)	Lipids	N/D
linoleate (18:2n6)	Lipids	N/D
linolenate [18:3n3 or 6]	Lipids	N/D
maleic acid	Lipids	N/D
margarate (17:0)	Lipids	N/D
Monomethyl glutarate	Lipids	N/D
myristate (14:0)	Lipids	N/D
myristoleate (14:1n5)	Lipids	N/D
palmitate (16:0)	Lipids	N/D
palmitoleate (16:1n7)	Lipids	N/D
pentadecanoate (15:0)	Lipids	N/D
Pregnenolone	Lipids	N/D
sphingosine	Lipids	N/D
Stigmasterol	Lipids	N/D
Testosterone	Lipids	N/D
valerate	Lipids	N/D
ADP	Nucleosides, Nucleotides, and Analogues	0.00625
CDP	Nucleosides, Nucleotides, and Analogues	0.00625
dCMP	Nucleosides, Nucleotides, and Analogues	0.00625
UTP	Nucleosides, Nucleotides, and Analogues	0.00625
ATP	Nucleosides, Nucleotides, and Analogues	0.0125
CTP	Nucleosides, Nucleotides, and Analogues	0.0125
dGDP	Nucleosides, Nucleotides, and Analogues	0.0125
UMP	Nucleosides, Nucleotides, and Analogues	0.0125

cytidine 5'-monophosphate	Nucleosides, Nucleotides, and Analogues	0.05
thymidine	Nucleosides, Nucleotides, and Analogues	0.05
GMP	Nucleosides, Nucleotides, and Analogues	0.15625
GTP	Nucleosides, Nucleotides, and Analogues	0.15625
inosine	Nucleosides, Nucleotides, and Analogues	0.15625
UDP	Nucleosides, Nucleotides, and Analogues	0.15625
dATP	Nucleosides, Nucleotides, and Analogues	0.3125
AMP	Nucleosides, Nucleotides, and Analogues	0.625
Deoxyguanosine	Nucleosides, Nucleotides, and Analogues	1.25
uridine	Nucleosides, Nucleotides, and Analogues	3.125
2'-deoxyinosine	Nucleosides, Nucleotides, and Analogues	6.25
2'-deoxycytidine	Nucleosides, Nucleotides, and Analogues	N/D
5-Methyl-2'-deoxycytidine	Nucleosides, Nucleotides, and Analogues	N/D
5-methylcytidine	Nucleosides, Nucleotides, and Analogues	N/D
adenosine	Nucleosides, Nucleotides, and Analogues	N/D
cytidine	Nucleosides, Nucleotides, and Analogues	N/D
cytidine 5'-diphosphocholine	Nucleosides, Nucleotides, and Analogues	N/D
Deoxyadenosine	Nucleosides, Nucleotides, and Analogues	N/D
Guanosine	Nucleosides, Nucleotides, and Analogues	N/D
glutarate (pentanedioate)	Organic Acids and Derivatives	0.05
3-hydroxydecanoic acid	Organic Acids and Derivatives	0.3125
pimelate (heptanedioate)	Organic Acids and Derivatives	1.25
itaconate	Organic Acids and Derivatives	3.125
lactate	Organic Acids and Derivatives	5
beta-hydroxyisovalerate	Organic Acids and Derivatives	10
hypotaurine	Organic Acids and Derivatives	10
citrate	Organic Acids and Derivatives	12.5
3-hydroxybutyrate (BHBA)	Organic Acids and Derivatives	N/D
3-hydroxyoctanoic acid	Organic Acids and Derivatives	N/D
adipate	Organic Acids and Derivatives	N/D
malate	Organic Acids and Derivatives	N/D
sebacate (decanedioate)	Organic Acids and Derivatives	N/D
Taurocyamine	Organic Acids and Derivatives	N/D
phosphoethanolamine	Organophosphorus Compounds	N/D

5.4.2 Summary of the QSPR analysis

MDL Molfiles of individual metabolites were acquired from the PubChem website (http://pubchem.ncbi.nlm.nih.gov), using a list of PubChem Compound IDs (CIDs) as the query. The acquired MDL Molfiles were applied for the calculation of the molecular descriptors by the PaDEL-Descriptor software program (Yap 2011). The types of molecular descriptors included 1-2D and 3D type descriptors and fingerprints. Descriptors with zero variance or 95% identical values (including NAs) were excluded from the subsequent analysis.

The LOD was used as the response variable, which could be considered as an inverse measure of the ionization efficiency. In the classification model, the responsive variable was converted to a categorical value denoted as *ionized* or *not ionized*, corresponding to whether the LOD value could be evaluated or not. In the regression model, where *not ionized* observations were eliminated, the LOD values were used in the molar concentrations. Modeling of the interrelationships between the descriptors and the ionization profiles of metabolites was conducted using the Random forest method (Breiman 2001) . The importance of variables for constructing a model was evaluated as the mean decrease in accuracy. Decision tree models were constructed using the descriptors with the highest importance. All of the analyses were performed using the R language (R Core Team 2012). Random forest and decision tree models were constructed by the party package (Hothorn *et al.* 2006). The accuracy of the prediction model was evaluated based on the correct rate given as a fraction of the number of correct predictions to the number of the examined metabolites. The performance of a regression model was evaluated by Spearman's rank correlation coefficients between the measured LODs and the fitted values.

In two-way QSPR modeling, the descriptor for a mixture of metabolite and matrix

compound was represented through numerical functions as follows:

5.4.3 *Note:* Cheminformatics and QSPR

A chemical structure can be expressed as a graph in 2D, or an atomic coordination in 3D. Once a molecule is encoded into a symbolic representation, it can be transformed into a simple yet useful number through a mathematical procedure (Todeschini and Consonni 2000). An MDL Molfile is one of the chemical table formats that contains information about the constituent atoms and their connectivity and coordinates of a molecule (Dalby *et al.* 1992). This format of the chemical structure is applicable to numbers of cheminformatics software applications, including PaDEL-Descriptor (Yap 2011). The QSPR analysis is constituted of five steps: Experimental data collection, descriptor calculation, variable selection, predictive model construction through cross validation, and interpretation of the model. An experimental data set is used as the response values in the QSPR model. It is usually comprised of biological activities or physiochemical properties expressed as both potency and categorical values, which should be modeled by the regression or the classification, respectively. In the present study, a PaDEL-descriptor was employed for the QSPR analysis, because every descriptor was calculated by open-source programs, which allowed for an understanding of the actual procedure for calculating the descriptors. A PaDEL-descriptor provides up to 733 1D-2D and 3D type descriptors and 10 types of fingerprinting descriptors for each compound, thus amounting to several thousand variables. In QSPR modeling, variable selection is a very important process for producing a reasonable model, particularly when a large number of variables are available. Although some simple variable-filtering processes were performed on the descriptor set, hundreds of variables still remained applicable to the model. Considering such circumstances, Random forest was applied to

develop QSPR models.

5.4.4 *Note:* How Random forest works in brief

Random forest is a kind of machine learning method first introduced by Breiman (Breiman 2001). The method is robust for sparse and high-dimensional data, and has been utilized in many QSPR studies. A Random forest model is an aggregation of large number (*ntree*) of decision trees constructed from training data sets. For each tree construction, *m* bootstrap samples were drawn from the original data set. Then, leaving about one third of the subset as the test data set (out-of-bag observation, OOB), a decision tree is then grown using *mtry* randomly selected variables from *p* original variables. As the result, the model is internally validated like cross-validation to yield a consensus prediction of the response. The prediction is given as a majority vote for classification and an average for regression. It is recommended that *mtry* is the square root of *p* for the classification, and *p*/3 for the regression. In the present study, we performed the Random forest model construction with the *ntree* set as 3,000 and the default *mtry* values. As a Random forest model is comprised of numerous small decision tree models of randomly selected variables and observations, neither decisive predictors nor their threshold values to predict the response are determined. The variable importance was thus estimated according to the 'mean decrease in accuracy'. This measure indicates the decrease in the model accuracy when a specific descriptor is removed from the tree construction. Hence, a higher mean decrease in accuracy indicates a higher importance for the model.

Chapter 6.

Conclusive Remarks

Although there has been indeed increasing interest in metabolomic approaches, with only a limited extent has the scene of the 'omics' research field experienced a fundamental progress, unlike the emergence of the notion of genome. It is advantageous to directly monitor the substrates and products during the cellular metabolism. However, true biochemistry is far more complex than a metabolic pathway represents, forcing sometimes an unreasonable simplification or abstraction of interpretation for observed phenomena. Diagnostic biomarker development could be a straightforward application to exclude such barriers, while biological evidences are still required. Assumingly, metabolomics was expected to work as fundamental information for bridging the phenotype and genotype, which is the ultimate goal of systems biology. More deductively, phenotypic modeling would lead to a deeper insight into the principle of dynamics or economics of cellular biochemistry, which could be partly parallel to the known metabolic pathway. Such ambition has been however hindered by numerous problems as mentioned, *e.g.* the identity of detected signals, coverage of molecular species, or absence of experimental and computational method for exploring the additional dimension of metabolome such as time or space.

Quantitative observation of compound-level phenotype also poses a serious question, *i.e.*, what is it like to understand the dynamics? When avoiding reductionism, there is no guarantee that underlying mathematics is rational to us. The mechanism of the biological system could be unforgiving to predict or reproduce its behavior as a whole. For solving such a challenging problem, whilst the primary interest seems to be focusing on the development of 'elegant' algorithms, we guess the elaborated, multimodal and precise quantitation of biomolecules and integrative approaches should have the most significant relevance, just as biological validation being prior to statistical validation. Furthermore, the system should not only be analyzed, but also synthesized to lead a full-length understand of the underlying

principle. Metabolomics will serve as a model plantation of multidisciplinary science, when participants desire it to be.

References

Aharoni, A., Ric de Vos, ,C.H., Verhoeven, H.A., Maliepaard, C.A., Kruppa, G., Bino, R., Goodenowe, D.B. (2002) "Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry", *Omics : a journal of integrative biology,* **6** (3), 217-34.

Alexander, M.L., Hemberger, P.H., Cisper, M.E., Nogar, N.S. (1993) "Laser desorption in a quadrupole ion trap: mixture analysis using positive and negative ions", *Anal. Chem.,* **65** (11), 1609-1614.

Allen, J., Davey, H.M., Broadhurst, D., Heald, J.K., Rowland, J.J., Oliver, S.G., Kell, D.B. (2003) "High-throughput classification of yeast mutants for functional genomics using metabolic footprinting", *Nat. Biotechnol.,* **21** (6), 692-6.

Alm, E. and Arkin, A.P. (2003) "Biological networks", *Curr. Opin. Struct. Biol.,* **13** (2), 193-202.

Altshuler, D., Daly, M., Kruglyak, L. (2000) "Guilt by association", *Nat. Genet.,* **26** (2), 135-137.

Amantonico, A., Oh, J.Y., Sobek, J., Heinemann, M., Zenobi, R. (2008a)*Angew. Chem. , Int. Ed. Engl.,* **47** , 5382.

Amantonico, A., Oh, J., Sobek, J., Heinemann, M., Zenobi, R. (2008b) "Mass Spectrometric Method for Analyzing Metabolites in Yeast with Single Cell Sensitivity", *Angew. Chem. Int. Ed.,* **47** (29), 5382-5385.

Andorf, S., Selbig, J., Altmann, T., Poos, K., Witucka-Wall, H., Repsilber, D. (2010)

"Enriched partial correlations in genome-wide gene expression profiles of hybrids (A. thaliana): a systems biological approach towards the molecular basis of heterosis", *Theor. Appl. Genet.,* **120** (2), 249-259.

Atkinson, D.E., (1968) "Energy charge of the adenylate pool as a regulatory parameter. Interaction with feedback modifiers", *Biochemistry,* **7** (11), 4030-4.

Ayorinde, F.O., Hambright, P., Porter, T.N., Keith, Q.L. (1999) "Use of meso-tetrakis(pentafluorophenyl)porphyrin as a matrix for low molecular weight alkylphenol ethoxylates in laser desorption/ ionization time-of-flight mass spectrometry", *Rapid Commun. Mass Spectrom.,* **13** (24), 2474-2479.

Ayorinde, F.O., Garvin, K., Saeed, K. (2000) "Determination of the fatty acid composition of saponified vegetable oils using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry", *Rapid Commun. Mass Spectrom.,* **14** (7), 608-615.

Baev, M.V., Baev, D., Radek, A.J., Campbell, J.W. (2006) "Growth of Escherichia coli MG1655 on LB medium: determining metabolic strategy with transcriptional microarrays", *Appl. Microbiol. Biotechnol.,* **71** (3), 323-8.

Bahr, U., Deppe, A., Karas, M., Hillenkamp, F., Giessmann, U. (1992) "Mass spectrometry of synthetic polymers by UV-matrix-assisted laser desorption/ionization", *Anal. Chem.,* **64** (22), 2866-2869.

Barabasi, A. and Oltvai, Z.N. (2004) "Network biology: understanding the cell's functional organization", *Nat. Rev. Genet.,* **5** (2), 101-113.

Becher, J., Muck, A., Mithofer, A., Svatos, A., Boland, W. (2008) *Rapid Commun. Mass Spectrom.,* **22** , 1153.

Beckmann, M., Parker, D., Enot, D.P., Duval, E., Draper, J. (2008) "High-throughput, nontargeted metabolite fingerprinting using nominal mass flow injection electrospray

mass spectrometry", *Nat. Protoc.,* **3** (3), 486-504.

Bergmeyer, H.U., Bergmeyer, J., Grassl, M. (1985) "Methods of enzymatic analysis. Metabolites 3: lipids, amino acids and related compounds", .

Bernini, P., Bertini, I., Luchinat, C., Nepi, S., Saccenti, E., Schäfer, H., Schütz, B., Spraul, M., Tenori, L. (2009) "Individual human phenotypes in metabolic space and time", *J. Proteome Res.,* **8** (9), 4264-71.

Bettenbrock, K., Sauter, T., Jahreis, K., Kremling, A., Lengeler, J.W., Gilles, E. (2007) "Correlation between growth rates, EIIACrr phosphorylation, and intracellular cyclic AMP levels in *Escherichia coli* K-12", *J. Bacteriol.,* **189** , 6891-6900.

Bijlsma, J.J. and Groisman, E.A. (2003) "Making informed decisions: regulatory interactions between two-component systems", *Trends Microbiol.,* **11** , 359-366.

Bino, R.J., Hall, R.D., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B.J., Mendes, P., Roessner-Tunali, U., Beale, M.H., Trethewey, R.N., Lange, B.M., Wurtele, E.S., Sumner, L.W. (2004) "Potential of metabolomics as a functional genomics tool", *Trends Plant Sci.,* **9** (9), 418-425.

Black, C., Poile, C., Langley, J., Herniman, J. (2006) "The use of pencil lead as a matrix and calibrant for matrix-assisted laser desorption/ionisation", *Rapid Commun. Mass Spectrom.,* **20** (7), 1053-1060.

Borgatti, S.P. and Everett, M.G. (1997) "Network analysis of 2-mode data", *Soc. Networks,* **19** , 243-269.

Box, G.E.P., Hunter, W.G., Hunter, J.S. (1978) *Statistics for experimenters: an introduction to design, data analysis, and model building.*

Brandes, U. and Corman, S.R. (2003) "Visual Unrolling of Network Evolution and the Analysis of Dynamic Discourse", *Inform. Vis.,* **2** (1), 40-50.

Breiman, L., (2001) "Random Forests", *Mach. Learning,* **45** (1), 5-32.

Broadhurst, D.I. and Kell, D.B. (2006) "Statistical strategies for avoiding false discoveries in metabolomics and related experiments", *Metabolomics,* **2** (4), 171-196.

Brown, M., Wedge, D.C., Goodacre, R., Kell, D.B., Baker, P.N., Kenny, L.C., Mamas, M.A., Neyses, L., Dunn, W.B. (2011) "Automated workflows for accurate mass-based putative metabolite identification in LC/MS-derived metabolomic datasets", *Bioinformatics,* **27** (8), 1108-12.

Brown, R.S. and Lennon, J.J. (1995) "Mass Resolution Improvement by Incorporation of Pulsed Ion Extraction in a Matrix-Assisted Laser Desorption/Ionization Linear Time-of-Flight Mass Spectrometer", *Anal. Chem.,* **67** (13), 1998-2003.

Brown, S.C., Kruppa, G., Dasseux, J. (2005) "Metabolomics applications of FT-ICR mass spectrometry", *Mass Spectrom. Rev.,* **24** (2), 223-231.

Bruce, S.J., Rochat, B., Béguin, A., Pesse, B., Guessous, I., Boulat, O., Henry, H. (2013) "Analysis and quantification of vitamin D metabolites in serum by ultra-performance liquid chromatography coupled to tandem mass spectrometry and high-resolution mass spectrometry--a method comparison and validation", *Rapid Commun. Mass Spectrom.,* **27** (1), 200-6.

Buchholz, A., Hurlebaus, J., Wandrey, C., Takors, R. (2002) "Metabolomics: quantification of intracellular metabolite dynamics", *Biomol. Eng.,* **19** (1), 5-15.

Buckstein, M.H., He, J., Rubin, H. (2008) "Characterization of nucleotide pools as a function of physiological state in *Escherichia coli*", *J. Bacteriol.,* **190** (2), 718-726.

Buziol, S., Bashir, I., Baumeister, A., Claassen, W., Noisommit-Rizzi, N., Mailinger, W., Reuss, M. (2002) "New bioreactor-coupled rapid stopped-flow sampling technique for measurements of metabolite dynamics on a subsecond time scale", *Biotechnol.*

Bioeng., **80** (6), 632-6.

Cai, J. and Henion, J. (1995) "Capillary electrophoresis-mass spectrometry", *Journal of Chromatography A,* **703** (1), 667-692.

Cakır, T., Hendriks, M.M.W.B., Westerhuis, J.a., Smilde, A.K. (2009) "Metabolic network discovery through reverse engineering of metabolome data", *Metabolomics : Official journal of the Metabolomic Society,* **5** (3), 318-329.

Cakir, T., Patil, K.R., Onsan, Z.i., Ulgen, K.O., Kirdar, B., Nielsen, J. (2006) "Integration of metabolome data with metabolic networks reveals reporter reactions", *Molecular systems biology,* **2** , 50-50.

Camacho, D., de la Fuente, A., Mendes, P. (2005) "The origin of correlations in metabolomics data", *Metabolomics,* **1** (1), 53-63.

Canelas, A.B., Ras, C., ten, P., Angela, van, D., Jan C., Heijnen, J.J., van, G., Walter M. (2008) "Leakage-free rapid quenching technique for yeast metabolomics", *Metabolomics,* **4** (3), 226-239.

Castrillo, J.I., Hayes, A., Mohammed, S., Gaskell, S.J., Oliver, S.G. (2003) "An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry", *Phytochemistry (Elsevier),* **62** (6), 929-937.

Castro, J.A., Köster, C., Wilkins, C. (1992) "Matrix-assisted laser desorption/ionization of high-mass molecules by Fourier-transform mass spectrometry", *Rapid communications in mass spectrometry : RCM,* **6** (4), 239-41.

Chang, R.L., Ghamsari, L., Manichaikul, A., Hom, E.F.Y., Balaji, S., Fu, W., Shen, Y., Hao, T., Palsson, B.Ø., Salehi-Ashtiani, K., Papin, J.A. (2011) "Metabolic network reconstruction of Chlamydomonas offers insight into light-driven algal metabolism", *Molecular systems biology,* **7** (1), 518-518.

Chassagnole, C., Noisommit-Rizzi, N., Schmid, J.W., Mauch, K., Reuss, M. (2002) "Dynamic modeling of the central carbon metabolism ofEscherichia coli", *Biotechnol. Bioeng.,* **79** (1), 53-73.

Chen, H., Venter, A., Cooks, R.G. (2006) "Extractive electrospray ionization for direct analysis of undiluted urine, milk and other complex mixtures without sample preparation", *Chemical communications (Cambridge, England),* (19), 2042-4.

Chen, Y., Shiea, J., Sunner, J. (1998) "Thin-layer chromatography-mass spectrometry using activated carbon, surface-assisted laser desorption/ionization", *J. Chromatogr. A,* **826** (1), 77-86.

Chou, Y. (1975) *Statistical Analysis.*

Cohen, G.N. and Monod, J. (1957) "Bacterial permeases", *Bacteriol. Rev.,* **21** , 169-194.

Cohen, L. and Gusev, A. (2002) "Small molecule analysis by MALDI mass spectrometry", *Analytical and Bioanalytical Chemistry,* **373** (7), 571-586.

Cole, H.A., Wimpenny, J.W.T., Hughes, D.E. (1967) "The ATP pool in Escherichia coli. I. Measurement of the pool using a modified luciferase assay", *Biochimica et Biophysica Acta (BBA) - Bioenergetics,* **143** (3), 445-453.

Cook, G.A., O'Brien, W.E., Wood, H.G., Todd King, M., Veech, R.L. (1978) "A rapid, enzymatic assay for the measurement of inorganic pyrophosphate in animal tissues", *Anal. Biochem.,* **91** (2), 557-565.

Corman, S.R., Kuhn, T., Mcphee, R.D., Dooley, K.J. (2002) "Studying Complex Discursive Systems.", *Human Communication Research,* **28** (2), 157-206.

Csárdi, G. and Nepusz, T. (2006) "The igraph software package for complex network research", *Inter Journal Complex Systems,* **1695** , 1-9.

Dalby, A., Nourse, J.G., Hounshell, W.D., Gushurst, A.K.I., Grier, D.L., Leland, B.A.,

Laufer, J. (1992) "Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited", *J. Chem. Inf. Comput. Sci.,* **32** (3), 244-255.

Dale, M.J., Knochenmuss, R., Zenobi, R. (1996) "Graphite/Liquid Mixed Matrixes for Laser Desorption/Ionization Mass Spectrometry", *Anal. Chem.,* **68** (19), 3321-3329.

Dally, J.E., Gorniak, J., Bowie, R., Bentzley, C.M. (2003) "Quantitation of Underivatized Free Amino Acids in Mammalian Cell Culture Media Using Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry", *Anal. Chem.,* **75** (19), 5046-5053.

Danis, P.O., Karr, D.E., Mayer, F., Holle, A., Watson, C.H. (1992) "The analysis of water-soluble polymers by matrix-assisted laser desorption time-of-flight mass spectrometry", *Org. Mass Spectrom.,* **27** (7), 843-846.

Davey, M.P., Burrell, M.M., Woodward, F.I., Quick, W.P. (2008) "Population-specific metabolic phenotypes of Arabidopsis lyrata ssp. petraea", *The New phytologist,* **177** (2), 380-8.

De Vos, R.,C.H., Moco, S., Lommen, A., Keurentjes, J.J.B., Bino, R.J., Hall, R.D. (2007) "Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry", *Nat. Protoc.,* **2** (4), 778-91.

de, K.W. and van, D.K. (1992) "A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH", *Anal. Biochem.,* **204** (1), 118-23.

Devillers, J., Balaban, A.T. (1999) *Topological Indices and Related Descriptors in QSAR and QSPR.* Gordon and Breach.

Dey, M., Castoro, J.A., Wilkins, C.L. (1995) "Determination of Molecular Weight

Distributions of Polymers by MALDI-FTMS", *Anal. Chem.,* **67** (9), 1575-1579.

Dikicioglu, D., Karabekmez, E., Rash, B., Pir, P., Kirdar, B., Oliver, S.G. (2011) "How yeast re-programmes its transcriptional profile in response to different nutrient impulses", *BMC systems biology,* **5** (1), 148-148.

Dixon, S.J., Brereton, R.G., Soini, H.A., Novotny, M.V., Penn, D.J. (2006) "An automated method for peak detection and matching in large gas chromatography-mass spectrometry data sets", *J. Chemometrics,* **20** (8-10), 325-340.

Domon, B. and Aebersold, R. (2006) "Mass spectrometry and protein analysis", *Science (New York, N.Y.),* **312** (5771), 212-7.

Dong, X., Cheng, J., Li, J., Wang, Y. (2010) "Graphene as a Novel Matrix for the Analysis of Small Molecules by MALDI-TOF MS", *Anal. Chem.,* **82** (14), 6208-6214.

Dunn, W.B., Overy, S., Quick, W.P. (2005) "Evaluation of automated electrospray-TOF mass spectrometryfor metabolic fingerprinting of the plant metabolome", *Metabolomics,* **1** (2), 137-148.

Dunn, W.B., Broadhurst, D.I., Atherton, H.J., Goodacre, R., Griffin, J.L. (2011) "Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy", *Chem. Soc. Rev.,* **40** (1), 387-426.

Edwards, D. (2000) *Introduction to graphical modelling.* 2 edn. Berlin: Springer-Verlag.

Edwards, J.L. and Kennedy, R.T. (2005)*Anal. Chem.,* **77** , 2201.

Edwards, J.S., (1999) "Systems Properties of the Haemophilus influenzae Rd Metabolic Genotype", *J. Biol. Chem.,* **274** (25), 17410-17416.

Edwards, J.L. and Kennedy, R.T. (2005) "Metabolomic Analysis of Eukaryotic Tissue and Prokaryotes Using Negative Mode MALDI Time-of-Flight Mass Spectrometry", *Anal. Chem.,* **77** (7), 2201-2209.

Ellis, D.I., Dunn, W.B., Griffin, J.L., Allwood, J.W., Goodacre, R. (2007) "Metabolic fingerprinting as a diagnostic tool", *Pharmacogenomics,* **8** (9), 1243-66.

Enot, D.P., Lin, W., Beckmann, M., Parker, D., Overy, D.P., Draper, J. (2008) "Preprocessing, classification modeling and feature selection using flow injection electrospray mass spectrometry metabolite fingerprint data", *Nat. Protoc.,* **3** (3), 446-70.

Erhard, M., von Dohren, H., Jungblut, P. (1997) "Rapid typing and elucidation of new secondary metabolites of intact cyanobacteria using MALDI-TOF mass spectrometry", *Nat. Biotech.,* **15** (9), 906-909.

Fagerquist, C.K., Garbus, B.R., Williams, K.E., Bates, A.H., Harden, L.A. (2010) "Covalent attachment and dissociative loss of sinapinic acid to/from cysteine-containing proteins from bacterial cell lysates analyzed by MALDI-TOF-TOF mass spectrometry", *J. Am. Soc. Mass Spectrom.,* **21** (5), 819-32.

Faijes, M., Mars, A.E., Smid, E.J. (2007) "Comparison of quenching and extraction methodologies for metabolome analysis of Lactobacillus plantarum", *Microb. Cell Fact.,* **6** (1), 27-27.

Feist, A.M. and Palsson, B.Ø. (2008) "The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli", *Nat. Biotechnol.,* **26** (6), 659-67.

Ferenci, T., (1996) *FEMS Microbiol. Rev.,* **18** , 301.

Fiehn, O., (2001) "Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks", *Comp. Funct. Genomics,* **2** (3), 155-68.

Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N., Willmitzer, L. (2000) *Nat. Biotechnol.,* **18** , 1157.

Forsythe, I.J. and Wishart, D.S. (2009) "Exploring human metabolites using the human

metabolome database", *Curr. Protoc. Bioinformatics*

Franz, A.H., Molinski, T.F., Lebrilla, C.B. (2001) "MALDI-FTMS characterization of oligosaccharides labeled with 9-aminofluorene", *J. Am. Soc. Mass Spectrom.,* **12** (12), 1254-61.

Frazier, M.E., Johnson, G.M., Thomassen, D.G., Oliver, C.E., Patrinos, A. (2003) "Realizing the Potential of the Genome Revolution: The Genomes to Life Program", *Science,* **300** (5617), 290-293.

Geier, F.M., Want, E.J., Leroi, A.M., Bundy, J.G. (2011) "Cross-platform comparison of Caenorhabditis elegans tissue extraction strategies for comprehensive metabolome coverage", *Anal. Chem.,* **83** (10), 3730-6.

Gibb, S. and Strimmer, K. (2012) "MALDIquant: a versatile R package for the analysis of mass spectrometry data", *Bioinformatics,* **28** (17), 2270-2271.

Glish, G.L., Goeringer, D.E., Asano, K.G., McLuckey, S.A. (1989) "Laser desorption mass spectrometry and MS/MS with a three-dimensional quadrupole ion trap", *Int. J. Mass Spectrom. Ion. Proc.,* **94** (1), 15-24.

Goldsmith, P., Fenton, H., Morris-Stiff, G., Ahmad, N., Fisher, J. and Prasad, K.R., 2010. *Metabonomics: A Useful Tool for the Future Surgeon.* Academic Press.

Gonzalez, B., Francois, J., Renaud, M. (1997) "A rapid and reliable method for metabolite extraction in yeast using boiling buffered ethanol", *Yeast,* **13** (14), 1347-55.

Goodacre, R., Vaidyanathan, S., Dunn, W.B., Harrigan, G.G., Kell, D.B. (2004) "Metabolomics by numbers: acquiring and understanding global metabolite data", *Trends Biotechnol.,* **22** (5), 245-252.

Görke, R., Meyer-Bäse, A., Wagner, D., He, H., Emmett, M.R., Conrad, C.a. (2010) "Determining and interpreting correlations in lipidomic networks found in

glioblastoma cells", *BMC Syst. Biol.,* **4** , 126.

Griffin, J.L., (2006) "The Cinderella story of metabolic profiling: does metabolomics get to go to the functional genomics ball?", *Phol. Trans. R. Soc. B,* **361** (1465), 147-61.

Guo, Z., Zhang, Q., Zou, H., Guo, B., Ni, J. (2002) "A method for the analysis of low-mass molecules by MALDI-TOF mass spectrometry", *Anal. Chem.,* **74** (7), 1637-1641.

Hageman, J.A., Hendriks, M.M.W.B., Westerhuis, J.A., van der Werf, M.,J., Berger, R., Smilde, A.K. (2008) "Simplivariate models: ideas and first examples", *PloS one,* **3** (9), e3259-e3259.

Hager, J.W., (2004) "Recent trends in mass spectrometer development", *Analytical and bioanalytical chemistry,* **378** (4), 845-50.

Hajjaj, H., Blanc, P.J., Goma, G., FranÃ§ois, J. (1998) "Sampling techniques and comparative extraction procedures for quantitative determination of intra- and extracellular metabolites in filamentous fungi", *FEMS Microbiol. Lett.,* **164** (1), 195-200.

Hall, L.H. and Kier, L.B. (1995) "Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information", *J. Chem. Inf. Comput. Sci.,* **35** (6), 1039-1045.

Haverkorn van Rijsewijk, B.,R.B., Nanchen, A., Nallet, S., Kleijn, R.J., Sauer, U. (2011) "Large-scale 13C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in *Escherichia coli*", *Mol. Syst. Biol.,* **7** .

Henry, C.S., DeJongh, M., Best, A.a., Frybarger, P.M., Linsay, B., Stevens, R.L. (2010) "High-throughput generation, optimization and analysis of genome-scale metabolic models", *Nat. Biotechnol.,* **28** (9), 969-974.

Herbert, R.B. (1989) *The Biosynthesis of Secondary Metabolites.* Second edn. Chapman &

Hall.

Hood, E., (2003) "A big circuit model", *Environ. Health Perspect.,* **111** (11), A576-7.

Hothorn, T., Bühlmann, P., Dudoit, S., Molinaro, A., Van, D.L. (2006) "Survival ensembles", *Biostatistics,* **7** (3), 355-373.

Hunnam, V., Harvey, D.J., Priestman, D.A., Bateman, R.H., Bordoli, R.S., Tyldesley, R. (2001) "Ionization and fragmentation of neutral and acidic glycosphingolipids with a Q-TOF mass spectrometer fitted with a MALDI ion source", *J. Am. Soc. Mass Spectrom.,* **12** (11), 1220-1225.

Huzyk, L. and Clark, D.J. (1971) "Nucleoside triphosphate pools in synchronous cultures of *Escherichia coli*", *J. Bacteriol.,* **108** (1), 74-81.

Ibáñez, A.,J., Fagerer, S.R., Schmidt, A.M., Urban, P.L., Jefimovs, K., Geiger, P., Dechant, R., Heinemann, M., Zenobi, R. (2013) "Mass spectrometry-based metabolomics of single yeast cells", *Proc. Natl. Acad. Sci. U. S. A.,* **110** (22), 8790-4.

Iijima, Y., Nakamura, Y., Ogata, Y., Tanaka, K., Sakurai, N., Suda, K., Suzuki, T., Suzuki, H., Okazaki, K., Kitayama, M., Kanaya, S., Aoki, K., Shibata, D. (2008) "Metabolite annotations based on the integration of mass spectral information", *Plant J.,* **54** (5), 949-62.

Ikonomou, M.G., Blades, A.T., Kebarle, P. (1990) "Investigations of the electrospray interface for liquid chromatography/mass spectrometry", *Anal. Chem.,* **62** (9), 957-967.

Ito, E., Nakajima, K., Waki, H., Miseki, K., Shimada, T., Sato, T., Kakehi, K., Suzuki, M., Taniguchi, N., Suzuki, A. (2013) "Structural characterization of pyridylaminated oligosaccharides derived from neutral glycosphingolipids by high-sensitivity capillary electrophoresis-mass spectrometry", *Anal. Chem.,* **85** (16), 7859-65.

Jackson, J.E. (2005) *A User's Guide to Principal Components.* John Wiley & Sons.

Jain, M., Nilsson, R., Sharma, S., Madhusudhan, N., Kitami, T., Souza, A.L., Kafri, R., Kirschner, M.W., Clish, C.B., Mootha, V.K. (2012) "Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation", *Science,* **336** (6084), 1040-4.

Janausch, I.G., Zientz, E., Tran, Q.H., Krsger, A., Unden, G. (2002) "C4-dicarboxylate carriers and sensors in bacteria", *Biochim. Biophys. Acta,* **1553** , 39-56.

Jansen, R.S., Rosing, H., Schellens, J.H.M., Beijnen, J.H. (2009) "Simultaneous quantification of 2′ ,2′ -difluorodeoxycytidine and 2′ ,2′ -difluorodeoxyuridine nucleosides and nucleotides in white blood cells using porous graphitic carbon chromatography coupled with tandem mass spectrometry", *Rapid Commun. Mass Spectrom.,* **23** (19), 3040-3050.

Jellum, E., (1977) "Profiling of human body fluids in healthy and diseased states using gas chromatography and mass spectrometry, with special reference to organic acids", *J. Chromatogr. B,* **143** (5), 427-462.

Jiye, A., Trygg, J., Gullberg, J., Johansson, A.I., Jonsson, P., Antti, H., Marklund, S.L., Moritz, T. (2005) "Extraction and GC/MS analysis of the human blood plasma metabolome", *Anal. Chem.,* **77** (24), 8086-94.

Jonsson, P., Gullberg, J., Nordström, A., Kusano, M., Kowalczyk, M., Sjöström, M., Moritz, T. (2004) "A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS", *Anal. Chem.,* **76** (6), 1738-45.

Jonsson, P., Johansson, A.I., Gullberg, J., Trygg, J., A, J., Grung, B., Marklund, S., Sjöström, M., Antti, H., Moritz, T. (2005) "High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses", *Anal. Chem.,* **77** (17), 5635-42.

Junot, C., Madalinski, G., Tabet, J., Ezan, E. (2010) "Fourier transform mass spectrometry for metabolome analysis", *Analyst,* **135** (9), 2203-19.

Kaback, H.R., (1968) *J. Biol. Chem.,* **243** , 3711.

Karas, M., Bahr, U., Ingendoh, A., Nordhoff, E., Stahl, B., Strupat, K., Hillenkamp, F. (1990) "Principles and applications of matrix-assisted UV-laser desorption/ionization mass spectrometry", *Anal. Chim. Acta,* **241** (2), 175-185.

Karas, M. and Hillenkamp, F. (1988) "Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons", *Anal. Chem.,* **60** (20), 2299-2301.

Karas, M., Bahr, U., Ingendoh, A., Hillenkamp, F. (1989) "Laser Desorption/Ionization Mass Spectrometry of Proteins of Mass 100 000 to 250 000 Dalton", *Angew. Chem. Int. Ed.,* **28** (6), 760-761.

Katagiri, F., (2003) "Attacking Complex Problems with the Power of Systems Biology", *Plant Physiol.,* **132** (2), 417-417.

Katritzky, A.R., Stoyanova-Slavova, I., TaÌˆmm, K., Tamm, T., Karelson, M. (2011) "Application of the QSPR Approach to the Boiling Points of Azeotropes", *J Phys Chem A,* **115** (15), 3475-3479.

Kell, D.B. and King, R.D. (2000) "On the optimization of classes for the assignment of unidentified reading frames in functional genomics programmes: the need for machine learning", *Trends Biotechnol.,* **18** (3), 93-98.

Keller, B.O. and Li, L. (2001) "Detection of 25,000 molecules of substance P by MALDI-TOF mass spectrometry and investigations into the fundamental limits of detection in MALDI", *J. Am. Soc. Mass Spectrom.,* **12** (9), 1055-1063.

Kherlopian, A.R., Song, T., Duan, Q., Neimark, M.a., Po, M.J., Gohagan, J.K., Laine, A.F. (2008) "A review of imaging techniques for systems biology", *BMC Syst. Biol.,* **2** ,

74-74.

Kind, T. and Fiehn, O. (2010) "Advances in structure elucidation of small molecules using mass spectrometry", *Bioanal. Rev.,* **2** (1-4), 23-60.

Kind, T. and Fiehn, O. (2007) "Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry", *BMC Bioinformatics,* **8** (1), 105-105.

Kind, T. and Fiehn, O. (2006) "Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm", *BMC Bioinformatics,* **7** (1), 234-234.

Kinsel, G.R., Knochenmuss, R., Setz, P., Land, C.M., Goh, S., Archibong, E.F., Hardesty, J.H., Marynick, D.S. (2002) "Ionization energy reductions in small 2,5-dihydroxybenzoic acid?proline clusters", *J. Mass Spectrom.,* **37** (11), 1131-1140.

Kitano, H., (2002) "Computational systems biology", *Nature,* **420** (6912), 206-210.

Kochanowski, K., Volkmer, B., Gerosa, L., Haverkorn van Rijsewijk, B.,R., Schmidt, A., Heinemann, M. (2013) "Functioning of a metabolic flux sensor in Escherichia coli", *Proc. Natl. Acad. Sci. U. S. A.,* **110** (3), 1130-5.

Kochanowski, K., Volkmer, B., Gerosa, L., R. Haverkorn van Rijsewijk, B., Schmidt, A., Heinemann, M. (2012) "Functioning of a metabolic flux sensor in *Escherichia coli*", *Proc. Natl. Acad. Sci. U. S. A.*

Koebmann, B.J., Westerhoff, H.V., Snoep, J.L., Nilsson, D., Jensen, P.R. (2002) *J. Bacteriol.,* **184** , 3909.

Koek, M.M., van der Kloet, F.,M., Kleemann, R., Kooistra, T., Verheij, E.R., Hankemeier, T. (2011) "Semi-automated non-target processing in GC × GC-MS metabolomics analysis: applicability for biomedical studies", *Metabolomics,* **7** (1), 1-14.

Kopka, J., (2006) "Current challenges and developments in GC-MS based metabolite profiling technology", *J. Biotechnol.,* **124** (1), 312-322.

Kotte, O., Zaugg, J.B., Heinemann, M. (2010) "Bacterial adaptation through distributed sensing of metabolic fluxes", *Mol. Syst. Biol.,* **6** .

Kresnowati, M.T., van Winden, W.A., Almering, M.J., ten Pierick, A., Ras, C., Knijnenburg, T.A., Daran-Lapujade, P., Pronk, J.T., Heijnen, J.J., Daran, J.M. (2006) *Mol. Syst. Biol.,* **2** , 49.

Langley, G.J., Herniman, J.M., Townell, M.S. (2007) "2B or not 2B, that is the question: further investigations into the use of pencil as a matrix for matrix-assisted laser desorption/ionisation", *Rapid Commun. Mass Spectrom.,* **21** (2), 180-190.

Lay, J.O., (2001) "MALDI-TOF mass spectrometry of bacteria*", *Mass Spectrom. Rev.,* **20** (4), 172-194.

Lee, M.V., Topper, S.E., Hubler, S.L., Hose, J., Wenger, C.D., Coon, J.J., Gasch, A.P. (2011) "A dynamic model of proteome changes reveals new roles for transcript alteration in yeast", *Molecular Systems Biology,* **7** (514), 514.

Lewis, W.G., Shen, Z., Finn, M.G., Siuzdak, G. (2003) "Desorption/ionization on silicon (DIOS) mass spectrometry: background and applications", *Int. J. Mass Spectrom.,* **226** (1), 107-116.

Li, Q., Li, F., Jia, L., Li, Y., Liu, Y., Yu, J., Fang, Q., Cao, A. (2006) "New asymmetric AB(n)-shaped amphiphilic poly(ethylene glycol)-b-[poly(l-lactide)](n) (n = 2, 4, 8) bridged with dendritic ester linkages: I. Syntheses and their characterization", *Biomacromolecules,* **7** (8), 2377-87.

Li, X., Lu, X., Tian, J., Gao, P., Kong, H., Xu, G. (2009) "Application of fuzzy c-means clustering in data analysis of metabolomics", *Anal. Chem.,* **81** (11), 4468-75.

Li, X., Wilm, M., Franz, T. (2005) "Silicone/graphite coating for on-target desalting and improved peptide mapping performance of matrix-assisted laser desorption/ionization-mass spectrometry targets in proteomic experiments", *Proteomics,* **5** (6), 1460-1471.

Liew, C., Pan, C., Tan, A., Ang, K., Yap, C. (2012) "QSAR classification of metabolic activation of chemicals into covalently reactive species", *Mol. Divers.,* **16** (2), 389-400.

Lindon, J.C. and Nicholson, J.K. (2008) "Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics", *Ann. Rev. Anal. Chem.,* **1**, 45-69.

Lisec, J., Schauer, N., Kopka, J., Willmitzer, L., Fernie, A.R. (2006) "Gas chromatography mass spectrometry-based metabolite profiling in plants", *Nat. Protoc.,* **1** (1), 387-96.

Lu, W., Clasquin, M.F., Melamud, E., Amador-Noguez, D., Caudy, A.A., Rabinowitz, J.D. (2010) "Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer", *Anal. Chem.,* **82** (8), 3212-21.

Luo, B., Groenke, K., Takors, R., Wandrey, C., Oldiges, M. (2007) "Simultaneous determination of multiple intracellular metabolites in glycolysis, pentose phosphate pathway and tricarboxylic acid cycle by liquid chromatography-mass spectrometry", *J. Chromatogr. A,* **1147** (2), 153-164.

Mahadevan, S., Shah, S.L., Marrie, T.J., Slupsky, C.M. (2008) "Analysis of metabolomic data using support vector machines", *Anal. Chem.,* **80** (19), 7562-70.

Maharjan, R.P. and Ferenci, T. (2003) *Anal. Biochem.,* **313** , 145.

Makarov, A., Denisov, E., Lange, O., Horning, S. (2006) "Dynamic range of mass accuracy in LTQ orbitrap hybrid mass spectrometer", *J. Am. Soc. Mass Spectrom.,* **17** (7),

977-982.

Mapelli, V., Olsson, L., Nielsen, J. (2008) "Metabolic footprinting in microbiology: methods and applications in functional genomics and biotechnology", *Trends Biotechnol.,* **26** (9), 490-497.

Martínez-Antonio, A., Velázquez-Ramírez, D.,A., Sánchez-Mondragón, J., Santillán, M. (2012) "Hierarchical dynamics of a transcription factors network in E. coli", *Mol. BioSystems,* **8** (11), 2932-6.

Mashego, M.R., van Gulik, W.,M., Heijnen, J.J. (2007) "Metabolome dynamic responses of Saccharomyces cerevisiae to simultaneous rapid perturbations in external electron acceptor and electron donor", *FEMS Yeast Res.,* **7** (1), 48-66.

Matsuda, F., Yonekura-Sakakibara, K., Niida, R., Kuromori, T., Shinozaki, K., Saito, K. (2009) "MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites", *Plant J.,* **57** (3), 555-77.

Mendes, P., Camacho, D., de, l.F., A. (2005) "Modelling and simulation for metabolomics data analysis", *Biochem. Soc. Trans.,* **33** (6), 1427-1429.

Miura, D., Fujimura, Y., Tachibana, H., Wariishi, H. (2010) "Highly Sensitive Matrix-Assisted Laser Desorption Ionization-Mass Spectrometry for High-Throughput Metabolic Profiling", *Anal. Chem.,* **82** (2), 498-504.

Miura, D., Fujimura, Y., Wariishi, H. (2012) "In situ metabolomic mass spectrometry imaging: Recent advances and difficulties", *Journal of Proteomics; Special Issue: Imaging Mass Spectrometry: A User's Guide to a New Technique for Biological and Biomedical Research,* **75** (16), 5052-5060.

Miura, D., Fujimura, Y., Yamato, M., Hyodo, F., Utsumi, H., Tachibana, H., Wariishi, H. (2010a) "Ultrahighly Sensitive *in Situ* Metabolomic Imaging for Visualizing

Spatiotemporal Metabolic Behaviors", *Anal. Chem.,* **82** (23), 9789-9796.

Miura, D., Tsuji, Y., Takahashi, K., Wariishi, H., Saito, K. (2010b) "A Strategy for the Determination of the Elemental Composition by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Based on Isotopic Peak Ratios", *Anal. Chem.,* **82** (13), 5887-5891.

Müller-Linow, M., Weckwerth, W., Hutt, M. (2007) "Consistency analysis of metabolic correlation networks", *BMC Syst. Biol.,* **1** (1), 44.

Mustafa, D.A.N., Burgers, P.C., Dekker, L.J., Charif, H., Titulaer, M.K., Smitt, P.A.E.S., Luider, T.M., Kros, J.M. (2007) "Identification of glioma neovascularization-related proteins by using MALDI-FTMS and nano-LC fractionation to microdissected tumor vessels", *Mol. Cell. Proteomics,* **6** (7), 1147-57.

Nicholson, G., Rantalainen, M., Maher, A.D., Li, J.V., Malmodin, D., Ahmadi, K.R., Faber, J.H., Hallgrímsdóttir, I.,B., Barrett, A., Toft, H., Krestyaninova, M., Viksna, J., Neogi, S.G., Dumas, M., Sarkans, U., The, M.C., Silverman, B.W., Donnelly, P., Nicholson, J.K., Allen, M., Zondervan, K.T., Lindon, J.C., Spector, T.D., McCarthy, M.I., Holmes, E., Baunsgaard, D., Holmes, C.C. (2011) "Human metabolic profiles are stably controlled by genetic and environmental variation", *Mol. Syst. Biol.,* **7** (1), 525-525.

Nikerel, I.E., van, W., Wouter A., Verheijen, P.J.T., Heijnen, J.J. (2009) "Model reduction and a priori kinetic parameter identifiability analysis using metabolome time series for metabolic reaction networks with linlog kinetics", *Metab. Eng.,* **11** (1), 20-30.

Nordström, A., O'Maille, G., Qin, C., Siuzdak, G. (2006) "Nonlinear data alignment for UPLC-MS and HPLC-MS based metabolomics: quantitative analysis of endogenous and exogenous metabolites in human serum", *Anal. Chem.,* **78** (10), 3289-95.

Norris, J.L., Cornett, D.S., Mobley, J.A., Andersson, M., Seeley, E.H., Chaurand, P., Caprioli, R.M. (2007)*Int. J. Mass Spectrom.,* **260** , 212.

Oldiges, M., Lütz, S., Pflug, S., Schroer, K., Stein, N., Wiendahl, C. (2007) "Metabolomics: current state and evolving methodologies and tools", *Appl. Microbiol. Biotechnol.,* **76** (3), 495-511.

Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., Palsson, B.O. (2011) "A comprehensive genome-scale reconstruction of Escherichia coli metabolism[mdash]2011", *Mol Syst Biol,* **7** .

Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I., van Oudenaarden, A. (2004) "Multistability in the lactose utilization network of Escherichia coli", *Nature,* **427** , 737-740.

Pasikanti, K.K., Ho, P.C., Chan, E.C. (2008)*J. Chromatogr. , B: Anal. Technol. Biomed. Life Sci.,* **871** , 202.

Patti, G.J., (2011) "Separation strategies for untargeted metabolomics", *Journal of separation science,* **34** (24), 3460-9.

Patti, G.J., Yanes, O., Siuzdak, G. (2012) "Innovation: Metabolomics: the apogee of the omics trilogy", *Nat. Rev. Mol. Cell Biol.,* **13** (4), 263-9.

Perkins, T.J. and Swain, P.S. (2009) "Strategies for cellular decision-making", *Mol. Syst. Biol,* **5** (326), 326-326.

Persson, S., Sönksen, C.P., Frigaard, N., Cox, R.P., Roepstorff, P., Miller, M. (2000) "Pigments and proteins in green bacterial chlorosomes studied by matrix-assisted laser desorption ionization mass spectrometry", *Eur. J. Biochem.,* **267** (2), 450-456.

Plumbridge, J., (2002) "Regulation of gene expression in the PTS in Escherichia coli: the role and interactions of Mlc", *Curr. Opin. Microbiol.,* **5** , 187-193.

Price, N.D., Reed, J.L., Palsson, B.Ø. (2004) "Genome-scale models of microbial cells: evaluating the consequences of constraints", *Nat. Rev. Microbiol.,* **2** (11), 886-97.

R Core Team (2012) *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Roberts Jr., J.M., (2000) "Correspondence analysis of two-mode network data", *Soc. Networks,* **22** (1), 65-72.

Rojas-Chertó, M., Kasper, P.T., Willighagen, E.L., Vreeken, R.J., Hankemeier, T., Reijmers, T.H. (2011) "Elemental composition determination based on MS(n)", *Bioinformatics,* **27** (17), 2376-83.

Rojas-Cherto, M., Peironcely, J.E., Kasper, P.T., van der Hooft, J.,J.J., de Vos, R.,C.H., Vreeken, R., Hankemeier, T., Reijmers, T. (2012) "Metabolite identification using automated comparison of high-resolution multistage mass spectral trees", *Anal. Chem.,* **84** (13), 5524-34.

Roseman, S., (1969) "The Transport of Carbohydrates by a Bacterial Phosphotransferase System", *J. Gen. Physiol.,* **54** (1), 138-184.

Sáez, M.J. and Lagunas, R. (1976) "Determination of intermediary metabolites in yeast. Critical examination of the effect of sampling conditions and recommendations for obtaining true levels", *Mol. Cell. Biochem.,* **13** (2), 73-78.

Saier, M.H., J., (1989) "Protein phosphorylation and allosteric control of inducer exclusion and catabolite repression by the bacterial phosphoenolpyruvate: sugar phosphotransferase system", *Microbiol. Mol. Biol. Rev.,* **53** (1), 109-120.

Schaefer, U., Boos, W., Takors, R., Weuster-Botz, D. (1999) "Automated Sampling Device for Monitoring Intracellular Metabolite Dynamics", *Anal. Biochem.,* **270** (1), 88-96.

Schäfer, J. and Strimmer, K. (2005) "An empirical Bayes approach to inferring large-scale

gene association networks", *Bioinformatics,* **21** (6), 754-64.

Schäfer, J., Opgen-Rhein, R., Strimmer., K. (2012) "GeneNet: Modeling and Inferring Gene Networks", .

Schellenberger, J., Que, R., Fleming, R.M.T., Thiele, I., Orth, J.D., Feist, A.M., Zielinski, D.C., Bordbar, A., Lewis, N.E., Rahmanian, S., Kang, J., Hyduke, D.R., Palsson, B.O. (2011) "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0", *Nat. Protocols,* **6** (9), 1290-1307.

Scheltema, R.A., Kamleh, A., Wildridge, D., Ebikeme, C., Watson, D.G., Barrett, M.P., Jansen, R.C., Breitling, R. (2008) "Increasing the mass accuracy of high-resolution LC-MS data using background ions: a case study on the LTQ-Orbitrap", *Proteomics,* **8** (22), 4647-56.

Schlag, E.W., Grotemeyer, J., Levine, R.D. (1992) "Do large molecules ionize?", *Chemical Physics Letters,* **190** (6), 521-527.

Schlegel, A., Danot, O., Richet, E., Ferenci, T., Boos, W. (2002) "The N terminus of the Escherichia coli transcription activator MalT is the domain of interaction with MalY", *J. Bacteriol.,* **184** , 3069-3077.

Schriemer, D.C. and Li, L. (1996) "Detection of High Molecular Weight Narrow Polydisperse Polymers up to 1.5 Million Daltons by MALDI Mass Spectrometry", *Anal. Chem.,* **68** (17), 2721-5.

Senko, M.W. and McLafferty, F.W. (1994) "Mass spectrometry of macromolecules: has its time now come?", *Annu. Rev. Biophys. Biomol. Struct.,* **23** , 763-85.

Seshasayee, A.S., Bertone, P., Fraser, G.M., Luscombe, N.M. (2006) "Transcriptional regulatory networks in bacteria: from input signals to output responses", *Curr. Opin. Microbiol.,* **9** , 511-519.

Shroff, R., Muck, A., Svatos, A. (2007a)*Rapid Commun. Mass Spectrom.,* **21** , 3295.

Shroff, R., Muck, A., Svatos, A. (2007b) "Analysis of low molecular weight acids by negative mode matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.", *Rapid Commun. Mass Spectrom.,* **21** (20), 3295-3300.

Siu, F. and Che, C. (2006) "Quantitative Structure−Activity (Affinity) Relationship (QSAR) Study on Protonation and Cationization of α-Amino Acids", *J Phys Chem A,* **110** (44), 12348-12354.

Slupsky, C.M., Rankin, K.N., Wagner, J., Fu, H., Chang, D., Weljie, A.M., Saude, E.J., Lix, B., Adamko, D.J., Shah, S., Greiner, R., Sykes, B.D., Marrie, T.J. (2007)*Anal. Chem.,* **79** , 6995.

Smith, C.A., O'Maille, G., Want, E.J., Qin, C., Trauger, S.A., Brandon, T.R., Custodio, D.E., Abagyan, R., Siuzdak, G. (2005) "METLIN: a metabolite mass spectral database", *Ther. Drug Monit.,* **27** (6), 747-751.

Smolinska, A., Blanchet, L., Buydens, L.M.C., Wijmenga, S.S. (2012) "NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review", *Anal. Chim. Acta,* **750** (0), 82-97.

Soga, T. and Heiger, D.N. (2000) "Amino Acid Analysis by Capillary Electrophoresis Electrospray Ionization Mass Spectrometry", *Anal. Chem.,* **72** (6), 1236-1241.

Soga, T., Igarashi, K., Ito, C., Mizobuchi, K., Zimmermann, H., Tomita, M. (2009) "Metabolomic profiling of anionic metabolites by capillary electrophoresis mass spectrometry", *Anal. Chem.,* **81** (15), 6165-74.

Soga, T., Ohashi, Y., Ueno, Y., Naraoka, H., Tomita, M., Nishioka, T. (2003) "Quantitative Metabolome Analysis Using Capillary Electrophoresis Mass Spectrometry", *Journal of Proteome Research,* **2** (5), 488-494.

Soga, T., Ueno, Y., Naraoka, H., Ohashi, Y., Tomita, M., Nishioka, T. (2002) "Simultaneous Determination of Anionic Intermediates for Bacillus s ubtilis Metabolic Pathways by Capillary Electrophoresis Electrospray Ionization Mass Spectrometry", *Anal. Chem.,* **74** (10), 2233-2239.

Sontag, E., Kiyatkin, A., Kholodenko, B.N. (2004) "Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data", *Bioinformatics,* **20** (12), 1877-86.

Sriyudthsak, K., Shiraishi, F., Hirai, M.Y. (2013) "Identification of a metabolic reaction network from time-series data of metabolite concentrations", *PloS one,* **8** (1), e51212-e51212.

Stahl, B., Steup, M., Karas, M., Hillenkamp, F. (1991) "Analysis of neutral oligosaccharides by matrix-assisted laser desorption ionization mass spectrometry", *Anal. Chem.,* **63** (14), 1463-1466.

Steuer, R., Kurths, J., Fiehn, O., Weckwerth, W. (2003) "Observing and interpreting correlations in metabolomic networks", *Bioinformatics,* **19** (8), 1019-1026.

Steuer, R., (2006) "Review: on the analysis and interpretation of correlations in metabolomic data", *Briefings in bioinformatics,* **7** (2), 151-158.

Strupat, K., Karas, M., Hillenkamp, F. (1991) "2,5-Dihydroxybenzoic acid: a new matrix for laser desorption—ionization mass spectrometry", *Int. J. Mass Spectrom. Ion Proc.,* **111** , 89-102.

Sugimoto, M., Wong, D.T., Hirayama, A., Soga, T., Tomita, M. (2010) "Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles", *Metabolomics,* **6** (1), 78-95.

Suhre, K., Shin, S., Petersen, A., Mohney, R.P., Meredith, D., Wägele, B., Altmaier, E.,

Deloukas, P., Erdmann, J., Grundberg, E., Hammond, C.J., de Angelis, M.H., Kastenmüller, G., Köttgen, A., Kronenberg, F., Mangino, M., Meisinger, C., Meitinger, T., Mewes, H., Milburn, M.V., Prehn, C., Raffler, J., Ried, J.S., Römisch-Margl, W., Samani, N.J., Small, K.S., Wichmann, H., Zhai, G., Illig, T., Spector, T.D., Adamski, J., Soranzo, N., Gieger, C. (2011) "Human metabolic individuality in biomedical and pharmaceutical research", *Nature,* **477** (7362), 54-60.

Sun, G., Yang, K., Zhao, Z., Guan, S., Han, X., Gross, R.W. (2007) *Anal. Chem.,* **79** , 6629.

Sun, J. and Chen, P. (2011) "A flow-injection mass spectrometry fingerprinting method for authentication and quality assessment of Scutellaria lateriflora-based dietary supplements", *Analytical and bioanalytical chemistry,* **401** (5), 1577-84.

Sunner, J., Dratz, E., Chen, Y. (1995) "Graphite surface-assisted laser desorption/ionization time-of-flight mass spectrometry of peptides and proteins from liquid solutions", *Anal. Chem.,* **67** (23), 4335-42.

Suzuki, Y., Suzuki, M., Ito, E., Goto-Inoue, N., Miseki, K., Iida, J., Yamazaki, Y., Yamada, M., Suzuki, A. (2006) "Convenient structural analysis of glycosphingolipids using MALDI-QIT-TOF mass spectrometry with increased laser power and cooling gas flow", *J. Biochem.,* **139** (4), 771-7.

Svatoš, A., (2010) "Mass spectrometric imaging of small molecules", *Trends Biotechnol.,* **28** (8), 425-434.

Szymanski, J., Jozefczuk, S., Nikoloski, Z., Selbig, J., Nikiforova, V., Catchpole, G., Willmitzer, L. (2009) "Stability of metabolic correlations under changing environmental conditions in Escherichia coli – a systems approach", *PloS one,* **4** (10), e7441.

Taylor, J., King, R.D., Altmann, T., Fiehn, O. (2002) "Application of metabolomics to plant

genotype discrimination using statistics and machine learning", *Bioinformatics,* **18**, S241-S248.

Theobald, U., Mailinger, W., Baltes, M., Rizzi, M., Reuss, M. (1997) *Biotechnol. Bioeng.,* **55** , 305.

Theobald, U., Mailinger, W., Reuss, M., Rizzi, M. (1993) "In Vivo Analysis of Glucose-Induced Fast Changes in Yeast Adenine Nucleotide Pool Applying a Rapid Sampling Technique", *Anal. Biochem.,* **214** (1), 31-37.

Theobald, U., Mailinger, W., Baltes, M., Rizzi, M., Reuss, M. (1997) "In vivo analysis of metabolic dynamics in Saccharomyces cerevisiae : I. Experimental observations", *Biotechnol. Bioeng.,* **55** (2), 305-316.

Todeschini, R., Consonni, V. (2000) *Handbook of Molecular Descriptors.* Weinheim: Wiley-VCH.

Todeschini, R., Lasagni, M., Marengo, E. (1994) "New molecular descriptors for 2D and 3D structures. Theory", *J. Chemometrics,* **8** (4), 263-272.

Tolstikov, V.V. and Fiehn, O. (2002) "Analysis of Highly Polar Compounds of Plant Origin: Combination of Hydrophilic Interaction Chromatography and Electrospray Ion Trap Mass Spectrometry", *Anal. Biochem.,* **301** (2), 298-307.

Trauger, S.A., Kalisak, E., Kalisiak, J., Morita, H., Weinberg, M.V., Menon, A.L., Poole, F.L., Adams, M.W.W., Siuzdak, G. (2008) "Correlating the transcriptome, proteome, and metabolome in the environmental adaptation of a hyperthermophile", *J. Proteome Res,* **7** (3), 1027-35.

Trimpin, S., Clemmer, D.E., McEwen, C.N. (2007) "Charge-remote fragmentation of lithiated fatty acids on a TOF-TOF instrument using matrix-ionization", *J. Am. Soc. Mass Spectrom.,* **18** (11), 1967-72.

Trygg, J. and Wold, S. (2002) "Orthogonal projections to latent structures (O-PLS)", *J. Chemometrics,* **16** (3), 119-128.

Tsugawa, H., Arita, M., Kanazawa, M., Ogiwara, A., Bamba, T., Fukusaki, E. (2013) "MRMPROBS: a data assessment and metabolite identification tool for large-scale multiple reaction monitoring based widely targeted metabolomics", *Anal. Chem.,* **85** (10), 5191-9.

Vaidyanathan, S., Gaskell, S., Goodacre, R. (2006) "Matrix-suppressed laser desorption/ionisation mass spectrometry and its suitability for metabolome analyses.", *Rapid Commun. Mass Spectrom.,* **20** (8), 1192-1198.

Vaidyanathan, S. and Goodacre, R. (2007a)*Rapid Commun. Mass Spectrom.,* **21** , 2072.

Vaidyanathan, S. and Goodacre, R. (2007b) "Quantitative detection of metabolites using matrix-assisted laser desorption/ionization mass spectrometry with 9-aminoacridine as the matrix.", *Rapid Commun. Mass Spectrom.,* **21** (13), 2072-2078.

Vaidyanathan, S., Jones, D., Broadhurst, D.I., Ellis, J., Jenkins, T., Dunn, W.B., Hayes, A., Burton, N., Oliver, S.G., Kell, D.B., Goodacre, R. (2005) "A laser desorption ionisation mass spectrometry approach for high throughput metabolomics", *Metabolomics,* **1** (3), 243-250.

van den Berg, R.,A., Hoefsloot, H.C.J., Westerhuis, J.A., Smilde, A.K., van der Werf, M.,J. (2006) "Centering, scaling, and transformations: improving the biological information content of metabolomics data", *BMC Genomics,* **7** (1), 142-142.

van der Greef, J., Martin, S., Juhasz, P., Adourian, A., Plasterer, T., Verheij, E.R., McBurney, R.N. (2007) "The art and practice of systems biology in medicine: mapping patterns of relationships", *J. Proteome Res.,* **6** (4), 1540-59.

van Kampen, J.J., Burgers, P.C., de Groot, R., Gruters, R.A., Luider, T.M. (2011)

"Biomedical application of MALDI mass spectrometry for small-molecule analysis", *Mass Spectrom. Rev.,* **30** (1), 101-20.

Venables, W.N., Ripley, B.D. (2002) *Modern Applied Statistics with S.* Fourth edn. New York: Springer.

Vermillion-Salsbury, R.L. and Hercules, D.M. (2002) "9-Aminoacridine as a matrix for negative mode matrix-assisted laser desorption/ionization", *Rapid Commun. Mass Spectrom.,* **16** (16), 1575-1581.

Vestal, M.L., Juhasz, P., Martin, S.A. (1995) "Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry", *Rapid Commun. Mass Spectrom.,* **9** (11), 1044-1050.

Vestal, M.L. and Campbell, J.M. (2005) "Tandem Time-of-Flight Mass Spectrometry", *Meth. Enzymol.,* **402**, 79-108.

Viant, M.R. and Sommer, U. (2012) "Mass spectrometry based environmental metabolomics: a primer and review", *Metabolomics,* **9**, 144-158.

Vidal, M., Cusick, M., Barabási, A. (2011) "Interactome Networks and Human Disease", *Cell,* **144** (6), 986-998.

Villas-Boas, S.G., Hoejer-Pedersen, J., Aakesson, M., Smedsgaard, J., Nielsen, J. (2005) "Global metabolite analysis of yeast: Evaluation of sample preparation methods", *Yeast,* **22** (14), 1155-1169.

Visser, D., van Zuylen, G.,A., van Dam, J.,C., Oudshoorn, A., Eman, M.R., Ras, C., van Gulik, W.,M., Frank, J., van Dedem, G.,W.K., Heijnen, J.J. (2002) "Rapid sampling for analysis of in vivo kinetics using the BioScope: a system for continuous-pulse experiments", *Biotechnol. Bioeng.,* **79** (6), 674-81.

Wagner, A. and Fell, D.A. (2001) "The small world inside large metabolic networks", *Proc.*

R. Soc. London, B, **268** (1478), 1803-1810.

Wang, Q., Wu, C., Chen, T., Chen, X., Zhao, X. (2006) "Integrating metabolomics into a systems biology framework to exploit metabolic complexity: strategies and applications in microorganisms", *Appl. Microbiol. Biotechnol.,* **70** (2), 151-161.

Want, E.J., O'Maille, G., Smith, C.A., Brandon, T.R., Uritboonthai, W., Qin, C., Trauger, S.A., Siuzdak, G. (2006) "Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry", *Anal. Chem.,* **78** (3), 743-52.

Waybright, T.J., Van, Q.N., Muschik, G.M., Conrads, T.P., Veenstra, T.D., Issaq, H.J. (2006) "LC‐MS in Metabonomics: Optimization of Experimental Conditions for the Analysis of Metabolites in Human Urine", *J. Liq. Chromatogr. Rel. Technol.,* **29** (17), 2475-2497.

Weckwerth, W. and Fiehn, O. (2002) "Can we discover novel pathways using metabolomic analysis?", *Curr. Opin. Biotechnol.,* **13** (2), 156-160.

Weckwerth, W. and Morgenthal, K. (2005) "Metabolomics: from pattern recognition to biological interpretation", *Drug Discov. Today,* **10** (22), 1551-1558.

Wei, J., Buriak, J.M., Siuzdak, G. (1999) "Desorption-ionization mass spectrometry on porous silicon", *Nature,* **399** (6733), 243-246.

Weibel, K.E., Mor, J., Fiechter, A. (1974) "Rapid sampling of yeast cells and automated assays of adenylate, citrate, pyruvate and glucose-6-phosphate pools", *Anal. Biochem.,* **58** (1), 208-216.

Welker, M. and Moore, E.R.B. (2011) "Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology", *Syst. Appl. Microbiol.,* **34** (1), 2-11.

Werhli, A.V., Grzegorczyk, M., Husmeier, D. (2006) "Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and bayesian networks", *Bioinformatics,* **22** (20), 2523-2531.

Werner, E., Croixmarie, V., Umbdenstock, T., Ezan, E., Chaminade, P., Tabet, J., Junot, C. (2008) "Mass spectrometry-based metabolomics: accelerating the characterization of discriminating signals by combining statistical correlations and ultrahigh resolution", *Anal. Chem.,* **80** (13), 4918-32.

Whitehouse, C.M., Dreyer, R.N., Yamashita, M., Fenn, J.B. (1985) "Electrospray interface for liquid chromatographs and mass spectrometers", *Anal. Chem.,* **57** (3), 675-679.

Wick, L.M., Quadroni, M., Egli, T. (2001) *Environ. Microbiol.,* **3** , 588.

Wikoff, W.R., Anfora, A.T., Liu, J., Schultz, P.G., Lesley, S.A., Peters, E.C., Siuzdak, G. (2009) "Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites", *Proc. Natl. Acad. Sci. U. S. A.,* **106** (10), 3698-703.

Winder, C.L., Dunn, W.B., Schuler, S., Broadhurst, D., Jarvis, R., Stephens, G.M., Goodacre, R. (2008) "Global metabolic profiling of Escherichia coli cultures: an evaluation of methods for quenching and extraction of intracellular metabolites", *Anal. Chem.,* **80** (8), 2939-48.

Wishart, D.S., (2011) "Advances in metabolite identification", *Bioanalysis,* **3** (15), 1769-82.

Wishart, D.S., (2009) "Computational strategies for metabolite identification in metabolomics", *Bioanalysis,* **1** (9), 1579-96.

Wishart, D.S., Knox, C., Guo, A.C., Eisner, R., Young, N., Gautam, B., Hau, D.D., Psychogios, N., Dong, E., Bouatra, S., Mandal, R., Sinelnikov, I., Xia, J., Jia, L., Cruz, J.A., Lim, E., Sobsey, C.A., Shrivastava, S., Huang, P., Liu, P., Fang, L., Peng, J., Fradette, R., Cheng, D., Tzur, D., Clements, M., Lewis, A., De Souza, A., Zuniga, A.,

Dawe, M., Xiong, Y., Clive, D., Greiner, R., Nazyrova, A., Shaykhutdinov, R., Li, L., Vogel, H.J., Forsythe, I. (2009) "HMDB: a knowledgebase for the human metabolome", *Nucleic Acids Res.,* **37** , D603-10.

Wittmann, C., Krömer, J.,O., Kiefer, P., Binz, T., Heinzle, E. (2004) "Impact of the cold shock phenomenon on quantification of intracellular metabolites in bacteria", *Anal. Biochem.,* **327** (1), 135-139.

Wold, S., Sjöström, M., Eriksson, L. (2001) "PLS-regression: a basic tool of chemometrics", *Chemometrics Intellig. Lab. Syst.,* **58** (2), 109-130.

Wu, G., (2009) "Amino acids: metabolism, functions, and nutrition", *Amino Acids,* **37** (1), 1-17.

Yanes, O., Clark, J., Wong, D.M., Patti, G.J., Sánchez-Ruiz, A., Benton, H.P., Trauger, S.A., Desponts, C., Ding, S., Siuzdak, G. (2010) "Metabolic oxidation regulates embryonic stem cell differentiation", *Nat. Chem. Biol.,* **6** (6), 411-7.

Yanes, O., Tautenhahn, R., Patti, G.J., Siuzdak, G. (2011) "Expanding Coverage of the Metabolome for Global Metabolite Profiling", *Anal. Chem.,* **83** (6), 2152-2161.

Yang, L., Bennett, R., Strum, J., Ellsworth, B.B., Hamilton, D., Tomlinson, M., Wolf, R.W., Housley, M., Roberts, B.A., Welsh, J., Jackson, B.J., Wood, S.G., Banka, C.L., Thulin, C.D., Linford, M.R. (2009) "Screening phosphatidylcholine biomarkers in mouse liver extracts from a hypercholesterolemia study using ESI-MS and chemometrics", *Anal. Bioanal. Chem.,* **393** (2), 643-54.

Yap, C.W., (2011) "PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints", *J. Comput. Chem.,* **32** (7), 1466-1474.

Yergey, A.L., Coorssen, J.R., Backlund, P.S., Blank, P.S., Humphrey, G.A., Zimmerberg, J., Campbell, J.M., Vestal, M.L. (2002) "De novo sequencing of peptides using

MALDI/TOF-TOF", *J. Am. Soc. Mass Spectrom.,* **13** (7), 784-91.

Yukihira, D., Miura, D., Saito, K., Takahashi, K., Wariishi, H. (2010) "MALDI-MS-Based High-Throughput Metabolite Analysis for Intracellular Metabolic Dynamics", *Anal. Chem.,* **82** (10), 4278-4282.

Zaima, N., Goto-Inoue, N., Hayasaka, T., Setou, M. (2010) "Application of imaging mass spectrometry for the analysis of Oryza sativa rice", *Rapid Commun. Mass Spectrom,* **24** (18), 2723-9.

Zhang, Q., Zou, H., Guo, Z., Zhang, Q., Chen, X., Ni, J. (2001) "Matrix-assisted laser desorption/ionization mass spectrometry using porous silicon and silica gel as matrix", *Rapid Commun. Mass Spectrom.,* **15** (3), 217-223.

Zhang, W.C., Shyh-Chang, N., Yang, H., Rai, A., Umashankar, S., Ma, S., Soh, B., Sun, L., Tai, B., Nga, M., Bhakoo, K., Jayapal, S., Nichane, M., Yu, Q., Ahmed, D., Tan, C., Sing, W., Tam, J., Thirugananam, A., Noghabi, M., Huei Pang, Y., Ang, H., Mitchell, W., Robson, P., Kaldis, P., Soo, R., Swarup, S., Lim, E., Lim, B. (2012) "Glycine Decarboxylase Activity Drives Non-Small Cell Lung Cancer Tumor-Initiating Cells and Tumorigenesis", *Cell,* **148** (1), 259-272.

Zhuang, K., Izallalen, M., Mouser, P., Richter, H., Risso, C., Mahadevan, R., Lovley, D.R. (2011) "Genome-scale dynamic modeling of the competition between Rhodoferax and Geobacter in anoxic subsurface environments", *The ISME journal,* **5** (2), 305-16.

Zou, H., Zhang, Q., Guo, Z., Guo, B., Zhang, Q., Chen, X. (2002) "A mass spectrometry based direct-binding assay for screening binding partners of proteins", *Angew. Chem., Int. Ed.,* **41** (4), 646-648.