MAPPING THE CONSEQUENCES OF PHYSICAL EXERCISE AND NUTRITION ON HUMAN HEALTH

A PREDICTIVE METABOLOMICS APPROACH

Elin Chorell

Department of Chemistry
Umeå University, Sweden 2011
We believe that mere movement is life, and that the more velocity it has, the more it expresses vitality

- Rabindranath Tagore (1861-1941)
Abstract

Human health is a complex and wide-ranging subject far beyond nutrition and physical exercise. Still, these factors have a huge impact on global health by their ability to prevent diseases and thus promote health. Thus, to identify health risks and benefits, it is necessary to reveal the underlying mechanisms of nutrition and exercise, which in many cases follows a complex chain of events. As a consequence, current health research is generating massive amounts of data from anthropometric parameters, genes, proteins, small molecules (metabolites) et cetera, with the intent to understand these mechanisms. For the study of health responses, especially related to physical exercise and nutrition, alterations in small molecules (metabolites) are in most cases immediate and located close to the phenotypic level and could therefore provide early signs of metabolic imbalances. Since there are roughly as many different responses to exercise and nutrients as there are humans, this quest is highly multifaceted and will benefit from an interpretation of treatment effects on a general as well as on an individual level. This thesis involves the application of chemometric methods to the study of global metabolic reactions, i.e. metabolomics, in a strategy coined predictive metabolomics. Via the application of predictive metabolomics an extensive hypothesis-free biological interpretation has been carried out of metabolite patterns in blood, acquired using gas chromatography-mass spectrometry (GC-MS), related to physical exercise, nutrition and diet, all in the context of human health. In addition, the chemometrics methodology have computational benefits concerning the extraction of relevant information from information-rich data as well as for interpreting general treatment effects and individual responses, as exemplified throughout this work. Health concerns all lifestages, thus this thesis presents a strategic framework in combination with comprehensive interpretations of metabolite patterns throughout life. This includes a broad range of human studies revealing metabolic patterns related to the impact of physical exercise, macronutrient modulation and different fitness status in young healthy males, short and long term dietary treatments in overweight post menopausal women as well as metabolic responses related to probiotics treatment and early development in infants. As a result, the studies included in the thesis have revealed metabolic patterns potentially indicative of an anticyclobolic response to macronutrients in the early recovery phase following exercise. Moreover, moderate differences in the metabolome associated with cardiorespiratory fitness level were detected, which could be linked to variation in the inflammatory and antioxidative defense system. This work also highlighted mechanistic information that could be connected to dietary related weight loss in overweight and obese postmenopausal women in relation to short as well as long term dietary effects based on different macronutrient compositions. Finally, alterations were observed in metabolic profiles in relation to probiotics treatment in the second half of infancy, suggesting possible health benefits of probiotics supplementation at an early age.

Keywords: Metabolomics, physical exercise, cardiorespiratory fitness nutrition, high protein and fat diet, Lactobacillus F19, probiotics, GC-MS, plasma, chemometrics, multivariate analysis statistical experimental design, design of experiments
Contents

List of papers 7
Other papers by the author not appended to the thesis 8
Abbreviations 9
Notation 10

AIM 11

BACKGROUND 13
  Exercise, nutrition and health 13
  Metabolomics 16
  Multivariate data processing and analysis 18
  Predictive metabolomics 23

RESULTS 25
  Papers I and II 25
  Paper III 29
  Paper IV 33
  Paper V 40

CONCLUSION 45

ACKNOWLEDGEMENTS 48

POPULÄRVETENSKAPLIG SAMMANFATTNING 50

REFERENCES 52
List of papers

This thesis is based on the following papers, referred to in the text by the Roman numerals. In the list of papers Pohjanen E and Chorell E refers to the same person.


III. *Chorell E*, Moritz T, Svensson MB and Antti H, Physical fitness level is reflected by alteration in the human plasma metabolome. *Submitted manuscript 2011.*


*Reprinted with the permission from the American Chemical Society.*
Other papers by the author not appended to the thesis


VIII. Thysell E, **Chorell E.** Svensson M.B, Moritz T, Jonsson P and Antti H, Processing of mass spectrometry based metabolomics data for large scale screening studies and diagnostics, *Submitted 2011.*
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI</td>
<td>body mass index</td>
</tr>
<tr>
<td>CV</td>
<td>cross validation</td>
</tr>
<tr>
<td>CVD</td>
<td>cardiovascular disease</td>
</tr>
<tr>
<td>DA</td>
<td>discriminant analysis</td>
</tr>
<tr>
<td>DoE</td>
<td>design of experiments</td>
</tr>
<tr>
<td>GC</td>
<td>gas chromatography</td>
</tr>
<tr>
<td>HMCR</td>
<td>hierarchical multivariate curve resolution</td>
</tr>
<tr>
<td>HOMA</td>
<td>homeostasis model assessment</td>
</tr>
<tr>
<td>LC</td>
<td>liquid chromatography</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>MUFA</td>
<td>mono unsaturated fatty acid</td>
</tr>
<tr>
<td>MVA</td>
<td>multivariate data analysis</td>
</tr>
<tr>
<td>OSC</td>
<td>orthogonal signal correction</td>
</tr>
<tr>
<td>OPLS</td>
<td>orthogonal projections to latent structures</td>
</tr>
<tr>
<td>PCA</td>
<td>principal component analysis</td>
</tr>
<tr>
<td>PLS</td>
<td>partial least squares</td>
</tr>
<tr>
<td>PUFA</td>
<td>poly unsaturated fatty acid</td>
</tr>
<tr>
<td>SED</td>
<td>statistical experimental design</td>
</tr>
<tr>
<td>TOF</td>
<td>time of flight</td>
</tr>
<tr>
<td>VO$_2$max</td>
<td>maximum oxygen uptake</td>
</tr>
</tbody>
</table>
Notation

The following notations will be used throughout the thesis. Matrices are denoted by bold capital letters, e.g. $X$. Vectors are denoted by bold, lower case letters, e.g. $t$, and assumed to be column vectors if not stated otherwise.

- $E$: Residual matrix of predictor variables, $[N \times K]$
- $F$: Residual matrix of response variables, $[N \times M]$
- $P$: Matrix of loading vectors for $X$, $[K \times A]$
- $T$: Matrix of score vectors for $X$, $[N \times A]$
- $X$: Matrix of predictor variables, $[N \times K]$
- $Y$: Matrix of response variables, $[N \times M]$
- $p$: Loading vector for $X$, $[K \times 1]$
- $t$: Score vector for $X$, $[N \times 1]$
Aim

*What is food to one man may be fierce poison to others*
- Lucretius (c. 99-55 BC)

This work aims to investigate the intrinsic relationship between nutrients and its effect on human metabolism in relation to physical exercise and diet. The ambition is to answer parts of the complex question: To what extent can we influence our metabolic health status by diet and physical activity throughout life?

A more specific aim is to present and apply metabolomics for revealing metabolite patterns that can be used to generate new hypotheses and mechanistic information concerning the impact of nutrition and physical exercise on human metabolism.

The first part of this thesis will present a background to the area of exercise and nutrition highlighting health applications as well as applied metabolomics and bioinformatics methods. In the following part the results from the included papers are discussed and outlined. The third and last part consists of a collection of cited literature.
Background

Exercise, nutrition and health

The human body is astonishing. Thousands of complex processes act in concert with each other to keep the metabolic system in homeostasis. This balance, or equilibrium, is highly influenced by our own lifestyle, including factors such as physical activity, diet, smoking and alcohol consumption\textsuperscript{1-3}.

Health is defined by Merriam-Webster as the condition of being sound in body, mind or spirit and especially free from physical disease or pain\textsuperscript{4}. Compared to treatment of a specific disease, health regards treating and preventing all diseases. So where should we start? The Global burden of disease study from 2009 lists high blood pressure, high blood glucose, physical inactivity and obesity among the leading risk factors affecting human health globally. These factors are all related to increased risk of heart disease, diabetes and cancer and are also highly correlated to our current lifestyle, in particular regarding physical activity and nutrition\textsuperscript{5}.

Regular physical activity is widely considered to have favorable consequences on many aspects of health, such as prevention of insulin resistance and cardiovascular disease (CVD)\textsuperscript{6-8} as well as enhancing cell proliferation and neurogenesis\textsuperscript{7}. Several biological mechanisms associated with improved health are direct consequences of physical exercise including improved lipoprotein profile\textsuperscript{8}, and carbohydrate metabolism\textsuperscript{9} as well as lower blood pressure\textsuperscript{10,11} and weight loss\textsuperscript{12}. However, there are considerable individual differences in responsiveness to physical exercise, in particular regarding the above mentioned risk factors. This inter-variability still remains when subjects are exposed to
an exercise intensity adjusted to their own tolerance level.\textsuperscript{13} Thus, interpreting the mean response, as done in many conventional studies, can be somewhat misleading if not investigating the intra- and inter individual differences as well. In addition to the complexity of individual response to exercise there are also different metabolic effects related to the intensity and type of the exercise\textsuperscript{14,15}. Regardless of the apparent complexity in metabolic responses related to physical exercise and health, the majority of the information available is still conclusive, that an active lifestyle is beneficial for health.

The impact of nutrition, on the other hand, is even more complicated due to the addition of exogenous compounds. For this reason, different dietary recommendations are under constant debate in the scientific community\textsuperscript{16-18} as well as in the public media. Food is not only a source of nourishment it is also pleasure, fuel and highly linked to social behavior\textsuperscript{9}. However, inappropriate dietary choices can lead to metabolic imbalance such as obesity, type-2 diabetes, hypertension, inflammatory diseases and cardiovascular diseases\textsuperscript{19,20}. Today, these conditions are not only increasing globally, but also occurring earlier in life\textsuperscript{21} and even more rapidly in developing countries\textsuperscript{22}.

The so-called essential nutrients were discovered by excluding one nutrient at a time from the diet and considered essential if the host system failed to function as a consequence of this. In general, everybody has the same essential nutrients but humans respond very differently to nutrients depending on variations in lifestyle, genetics and phenotype. Thus, the paradigm that everyone should follow the same diet does not longer hold.\textsuperscript{23-24}

Nowadays, almost every country has its own national food recommendation to guide the population to a healthy diet (\textbf{Figure 1}).\textsuperscript{25-27} Interestingly, there has been a dramatic change in recommendations over the past couple of decades. However, the main foci have always been the same – a more or less restricted intake of fats. For example, in the 1980’s guidelines from United States Human Health Service (HHS) an overall reduction of fats was suggested, which in the 1990’s was changed to include $>30\%$ of dietary fats. In 2010, these recommendations were extended to include the distinction of different types of fats, such as \textit{trans} fatty acids, saturated and unsaturated fats\textsuperscript{28}. 
Besides the evolution of popular weight loss regimes, science has evolved with the understanding of the benefits of a well composed intestinal flora\textsuperscript{39-40}. This has emphasized a role for health promoting pro-, pre- and symbiotics also known as functional foods. Probiotics, ‘for life’ is currently used to define living microbial food of potential beneficial bacterial strains\textsuperscript{31-32}. One of the most frequently used strains is *lactobacilli*, the lactic acid bacteria. This bacterium has a long history and is considered as non-opportunistic bacteria, i.e. a bacterium with a low risk of infecting the host\textsuperscript{33}. Literature reports indicate possible anti-obesity\textsuperscript{34-35} and anti-allergic effects\textsuperscript{36-37} although the underlying molecular mechanisms have not yet been fully understood\textsuperscript{38-39}.

Nevertheless, there is still a world-wide increase in sedentary behaviour\textsuperscript{40} and obesity\textsuperscript{41}. So what is more important, keeping a normal weight or being physically active? Evidence states the importance of a physically active lifestyle to prevent senior disabilities in obese individuals\textsuperscript{42}. However, even though body mass index (BMI) and physical activity can be used as individual predictors for diabetes incidence\textsuperscript{43-44}, facts point towards a greater magnitude associated with BMI as compared to physical activity\textsuperscript{45}. Today, there are several anthropometrical measurements, in addition to BMI, which incorporate the distribution of weight, such as sagittal abdominal diameter\textsuperscript{46} and waist-hip ratio. This since abdominal adiposity has proved to be highly associated with disease development\textsuperscript{47}. Still, blood metabolites, such as triglycerides, insulin and cholesterol, are the most reliable markers indicative for insulin resistance\textsuperscript{48}. This could stem from the fact that these biomarkers incorporate the effects of physical activity and nutrition to a greater extent as compared to anthropometrical measurements\textsuperscript{49}.

Humans are diverse on so many levels: genetic, gender, age, hormonal cycle in different lifestages, et cetera. In addition, we are under constant influence of environmental factors

\textsuperscript{†} Reprinted with kind permission from the Swedish National Food Administration (*Livsmedelsverket*).
and, as mentioned before, in terms of lifestyle. These differences will contribute to an individual metabolic response to nutrients, physical exercise and the combination thereof. The underlying mechanisms responsible for individual differences in relation to nutritional intake and physical activity are very complex and today poorly understood. Therefore, methods assessing health in a way similar to diagnosis and prognosis of disease will be vital for the future. In this was it might be possible to capture the phenotypic diversity to fully understand the intrinsic relation between nutrients and metabolic responses. A new biological perspective is moving into health research guided by powerful bioinformatics. Bioinformatics provide tools for handling, processing and analyzing large and complex data sets. Such approaches have become a key player in the development of the field of functional genomics and more recently also systems biology. Via the emerging of systems biology, the focus has shifted from the more conventional analysis of studying isolated molecular events to instead consider the biological system as whole both structurally and dynamically. This includes a comprehensive and integrated analysis of genes (genomics), transcription factors (transcriptomics), protein (proteomics) and metabolites (metabolomics) (Figure 2).\(^{50-52}\)

![Building blocks of systems biology](image)

**Figure 2.** A schematic picture of the building blocks in systems biology and their inherent hierarchy: Genomics, mapping the entire genome (DNA), Transcriptomics, measurements of overall transcript expression (RNA), Proteomics, identifying, sequencing and characterizing the functional protein network (proteins), Metabolomics, the comprehensive analysis of the metabolome under specified conditions (metabolites).

In this chain of events, changes in the phenotype are of ultimate interest for the study of the impact of exercise and nutrition on health. In the light of that, alterations in the metabolome, located closest to the phenotypic level, could provide early signs as well as biochemical explanations of phenotypic variation\(^{53-54}\).

### Metabolomics

The overall aim of metabolomics is to identify and quantify all metabolites present in the metabolome of a specific biological system\(^{55}\). The human metabolome is defined as all low molecular weight metabolites present in a human sample. This includes not only the essential vitamins and amino acids, but the whole set of metabolites that can be expressed in humans under any condition\(^{56-57}\). This wide-ranging analysis is set to cover the vast chemical diversity of metabolites such as amino acids, fatty acids, polyols, carbohydrates, sterols and nucleosides (Figure 3).
Figure 3. A few examples of the physio-chemically diverse set of metabolites targeted by metabolomics analysis.

Many of these physio-chemically diverse metabolites fluctuate over time due to environmental factors, lifestyle, circadian rhythm and variability throughout the different lifestages. Consequently, the metabolome is in constant change, as compared to the genome, which is fixed for life except for when mutations occur. This means that a collected biological sample will only provide a snapshot of an organism’s metabolic state and that special consideration needs to be taken regarding study design, sample preparation and selection of platform for metabolite detection in order to capture the full set of metabolites.

Today there is no single analytical method that can detect all metabolites in a biological sample. Instead, numerous complementary methods are available. The metabolome can be characterized using vibrational spectroscopy platforms, e.g. Raman and Fourier transform IR (FT-IR) spectroscopy, nuclear magnetic resonance (NMR) spectroscopy based methods, e.g. proton (1H)-NMR, 2-dimensional 1H-NMR, magic angle spinning (MAS)-NMR or mass spectrometry based methods. The latter, often coupled to chromatography to improve compound separation, e.g. gas chromatography-mass spectrometry (GC-MS) and liquid chromatography-mass spectrometry (LC-MS). These platforms differ in sample preparation techniques, sensitivity, resolution, mass range, reproducibility, et cetera and should thus be used to complement each other. GC-MS, which is the preferred platform in this work, combines the advantage of a sensitive detection with the ability to identify metabolites through spectral library comparisons. Although being the most sensitive, the MS platforms still have some drawbacks concerning reproducibility and metabolite identification compared to NMR based methods in particular.
The grand vision of metabolomics is to be able to, from a blood, urine or other tissue sample, quantify and identify hundreds to thousands of metabolites that could be used to predict health status, disease risk, treatment response or the likelihood to suffer from drug side effects. This sci-fi vision is today starting to become a reality. As an example, in Sweden a nationwide newborn screening program is in place where blood from all newborns are collected, analyzed and stored. In this screening, the so-called phenylketonuria (PKU) testing, mass spectrometry based methods are today applied to search for biomarkers (metabolites) that could be used as indicators for over twenty different inborn errors of metabolism².

**Metabolomics in human nutrition and physical exercise**

Metabolomics applied in human nutrition, and even more so in relation to physical exercise, is still at its infancy. Studies of the metabolome are likely to bring further insight into the myriads of metabolic interactions caused by nutrients and exercise that are difficult to target by conventional analyses. Traditional studies into metabolic health is commonly conducted by measuring a few pre defined nutrients or metabolites. This will however only provide metabolic information limited to the specific hypothesis being tested. Instead, by complementing these methods with a more global investigation using metabolomics, an overview of the interaction pattern of ‘all’ metabolites without missing out on the accuracy provided by a more targeted analysis looking at specific metabolites or metabolite classes.

As stated before, the human metabolome is highly complex and confounded by numerous effects of different origins. Metabolomics as a strategy can thus be of great value to take on the challenge of characterizing the metabolome and the changes therein in areas concerning nutrition and health status³⁷. Current applications range from monitoring and detecting habitual dietary patterns³⁴-⁷⁵ to investigating nutritional effects³⁶ or developing personalized food³⁷-⁴⁰ and exercise regimes for the primary and secondary (rehabilitation) protection of disease development³¹. In addition, metabolomics can be used to investigate normal variation and fluctuations in the human metabolome³². Despite the progress of the methodology, there are still many challenges ahead for metabolomics, especially when it comes to the identification and interpretation of the vast amount of spectral data produced from metabolomics analyses³³. Some of these challenges have been addressed by the assembly and development of the human metabolome database (HMDB)³⁴.

**Multivariate data processing and analysis**

In GC-MS, molecules are anticipated to elute from the chromatographic step one at a time for the detection in the mass spectrometry. However, when performing global metabolite analysis, i.e. metabolomics, a large number of metabolites have similar properties and will hence travel at the same speed and co-elute. Thus, in order to obtain
pure mass spectral profiles and correct quantifications of detected compounds for further evaluation, additional data processing by mathematical curve resolution or deconvolution is usually required\textsuperscript{85-87}.

Hierarchical multivariate curve resolution

Recent advances in analytical technologies have enabled high-throughput metabolomics analysis of numerous types of biological samples. This simultaneous detection of metabolites using hyphenated methods such as GC-MS captures the characteristics of a biological sample in information-rich data tables. These data tables will then describe the detected compounds from an analyzed sample in two dimensions, one spectral and one chromatographic, where each detected compound’s chromatographic and spectral profile will depend on the time range at which it elutes from the chromatographic system. When analyzing samples containing hundreds to thousands of compounds this will produce highly overlapping profiles, making manual integration difficult. GC-MS analysis of multiple samples will give rise to a three dimensional data structure including the spectral, chromatographic and sample dimensions. Deconvolution, or curve resolution corresponds to a mathematical transformation or model of the raw GC-MS data, in which compounds that are not, or only partly, separated by chemical chromatography are further separated into their pure profiles \textit{in silico}. Throughout this work, hierarchical multivariate curve resolution (HMCR) has been applied for the deconvolution step, which is carried out for all samples simultaneously. After smoothing and filtering of mass channels, all samples are aligned by using one or several pure mass channels, preferably from internal standards. This is followed by a manual division of chromatograms into time windows, with borders set at local minima to avoid peak splitting. Each time window is then deconvoluted separately to create a data matrix where all samples are described by a common set of quantitative variables (chromatographic peaks) and with corresponding resolved mass spectra stored in a separate matrix for identification. Thereby, no peak matching is necessary and the output data is ready for sample comparisons based on the whole metabolite profile. The advantage of using a window wise deconvolution is that it takes care of some of the retention drift still remaining after the alignment. In addition, it speeds up the calculation process, thus making it more high-throughput. Since it allows for simultaneous deconvolution of multiple samples, HMCR makes it possible to resolve two mass spectra from completely overlapping compounds as long as they differ in intensity between samples.\textsuperscript{86-89}

HMCR also includes a predictive feature by which independent samples can be resolved by using processing parameters acquired from the HMCR training set, i.e. the original processing (\textbf{Figure 4}). This predictive deconvolution is much faster than the original processing making it a feasible option for the processing of many samples. The combined high-throughput and predictive capabilities of HMCR makes it a very interesting method for e.g. screening of large sample sets e.g. in metabolome-wide association studies.
In almost all MS-based metabolomics studies, the sample comparison analysis, following data acquisition and curve resolution, is often performed using multivariate statistical methods. Commonly used methods include principal component analysis (PCA)\textsuperscript{92-93}, hierarchical cluster analysis (HCA)\textsuperscript{94}, discriminant analysis\textsuperscript{95}, correlation network analysis\textsuperscript{96} and partial least squares based regression and discriminant analysis\textsuperscript{97-99}.

Chemometrics

A biofluid sample is only representative for the specific time point at that it was withdrawn from the host. Thus, it of utmost importance to sample at a time point representative for the underlying study aim. Otherwise the data will not be predictive or provide an answer to the study question. Thus, analyses concerning human health factors in general, and metabolomics in particular, require well designed studies from bedside (subjects) to bench (computer). By doing so, we can minimize bias from daily variations.
in human subjects, instruments and laborative work confounding the response and in that way maximize the information output.

The main objective of metabolomics is to quantify and identify all metabolite present in a biological sample. These data sets will for that reason include samples described by hundreds to thousands of metabolites, which often are highly correlated. Thus, the data analysis should preferable involve sample comparisons based on the whole set of detected metabolites simultaneous. It is therefore a requirement to use methods for data analysis, such as multivariate statistics\(^{100}\), that can handle this type of correlated data structures. Classical statistical methods such as ANOVA, multiple linear regression analysis (MLR) and Student’s t-test are in these cases less reliable since they all assume variable independence. However, saying that, the classical methods should still be seen as an important complement to the multivariate counterpart, e.g. in the detection and validation of single metabolic markers or for determining the strength of multivariate models or model components.

Chemometrics is a concept that, as the name suggest, has its origin in chemistry. The overall chemometric aim is to provide mathematical tools to extract represent and display relevant information from high-dimensional empirical data and is thus applicable to many contemporary research areas including metabolomics. This concept can be summarized into two parts: (i) study design and (ii) multivariate analysis\(^{101-104}\).

(i) The purpose of study design is to make data contain information relevant for the study aim. Statistical experimental design (SED), or design of experiments (DoE), provides means to plan and execute experiments with the aim to maximize the information output while keeping the number of experiments as low as possible. The methodology enables simultaneous investigation of several variables and their interactions with the option of extending the number of experiments in an organized fashion based on the initial results SED can preferably be used throughout the whole process from defining the study aim to the final extraction of information.\(^{105-106}\)

(ii) The multivariate analysis (MVA) part of chemometrics concerns the extraction of information from high-dimensional data. This is often performed in two steps. First, using unsupervised projection methods, such as principal components analysis (PCA), to extract the main features of the data\(^{3,107}\). Secondly, a more focused and supervised evaluation to decipher relationship between the data set and additional sample knowledge, for example, age, gender and diet. Partial least squares (PLS)\(^{99}\) and Orthogonal-PLS\(^{98}\) represent multivariate regression methods for relating information in two or more data tables. An important part of MVA is the projection of the data to low dimensional planes defined by latent variables, e.g. principal components. In this way simple maps of the complex interactions occurring in the data are generated, which largely facilitates interpretation and further evaluation.
PCA

It is highly desirable, within the -omics and systems biology fields, to visually and mathematically summarize the main features of information-rich data sets. As an example, complex genetic structures in human populations can literally be transformed into visual maps by using principal component analysis (PCA)\(^{108}\).

The data overview obtained by PCA displays the relationship between observations (e.g. biological samples) and variables (e.g. metabolites) in the original data matrix \(X\). This can be of high value in many steps of analyzing and interpreting metabolomics data, for example in data inspection to detect outliers\(^{109}\) as well as when selecting representative samples for model training and test sets\(^{110}\).

In brief, PCA is an unsupervised projection method comparable to an ordinary window where the three dimensional outside environment is projected down to a two dimensional picture. This picture can contain different aspects of information depending on from which direction you are viewing the environment. In practice, the multidimensional data table (e.g. hundreds of metabolites) is projected down to a few orthogonal principal components (PCs), also referred to as latent variables, containing the systematic variation in the data. A latent variable is defined as a variable that cannot be directly be measured or observed but concretized from other variables. In MVA, latent variables are calculated for the analysis of data structures by statistically relating the co-variation between observed variables to these new latent variables\(^{111}\). The number of PCs used to describe the variation in the data can be calculated in many ways. Commonly used methods are singular value decomposition (SVD)\(^{92}\) and the non iterative partial least squares (NIPALS) algorithm\(^{112}\). The latter has the advantage of handling missing values which often occurs in large metabolomics datasets and will therefore be used throughout this work. Irrespective of algorithm, PCA works to deflate the systematic variation in the original matrix \(X\), into the latent variable description \(TP^T\), i.e. scores \(T\) and loadings \(P\). The loadings correspond to the angle of which the PC is placed among the original variables (e.g. metabolites), and will therefore indicate which variables are important for the sample distribution displayed in the scores \(T\). Nonsystematic or random variation is not included in the model but is instead stored in the residual matrix \(E\).

\[
X = TP^T + E
\]

**PLS and OPLS**

In metabolomics studies, a more focused analysis of the data is often needed to answer the specific underlying study questions. This aim may for instance be to investigate dietary patterns in metabolite profiles\(^{25}\) or to find the largest metabolic differences between treatment groups in intervention studies\(^{113}\) as well as more complex maneuvers such as to relate metabolite information to protein and gene expression data\(^{114}\).
Methods providing such focused analyses are multivariate regression methods including partial least squares (PLS)\textsuperscript{99} and orthogonal projection to latent structures (OPLS)\textsuperscript{98}. These methods work to decipher the relationship between the original data matrix $X$, e.g. metabolite concentrations in blood samples, to additional sample data stored in a response matrix $Y$, e.g. treatment groups, sample collection time point or anthropometric variables. In general, PLS and OPLS will yield the same results regarding the predictive power of the model. OPLS can be seen as PLS with an integrated orthogonal signal correction filter (OSC), a method for identifying and removing the systematic variation in $X$ that are not related to $Y$, i.e. orthogonal or linearly independent\textsuperscript{102}. Thus, the advantage of using OPLS, as compared to PLS, lies in the interpretation of the data by allowing a separate inspection of the $Y$-predictive and $Y$-orthogonal variation\textsuperscript{115-116}. Still, despite the obvious advantage for interpretation, consideration regarding the number of orthogonal components to be calculated must be taken to minimize the risk for over fitting your data. The OPLS method is used throughout the papers of thesis as the preferred regression method, why PLS will not be further discussed.

In brief, OPLS is a supervised multivariate regression method based on the previous mentioned latent variables. The latent variables in OPLS correspond to the maximized squared covariance between the measured variables in $X$ and the response variable(s) in $Y$. In addition, OPLS have the ability to handle the problem with multicollinearities, which are present in metabolomics and similar datasets containing numerous of highly correlated variables. This is done by creating the projected score matrix $T$, which consist of new orthogonal variables. The score matrix $T$ summarizes the variation in $X$ and separates it into two parts: $T_0$, the systematic variation in $X$ related to $Y$ and $T_0$, the systematic variation in $X$ orthogonal (unrelated) to $Y$. $E$ is the residual matrix containing the non-systematic or random variation in the data.

$$X = T_p P_p^T + T_0 P_0^T + E$$

By altering the response matrix $Y$ to become a dummy matrix containing information regarding sample class, multivariate discriminant analysis (DA) can be carried out. OPLS-DA\textsuperscript{97} is today commonly used for sample classification within metabolomics, mainly due to the advantages regarding interpretation as described before\textsuperscript{117-118}.

**Predictive metabolomics**

Predictive metabolomics\textsuperscript{91,113} is a strategy for generating representative high quality metabolomics data for multiple sample comparison analysis, by applying the above described chemometric concept in metabolomics analyses. In metabolomics studies, representative data corresponds to a reliable characterization of the metabolome and all of its entities. Because of the constant flux and the interdependency of metabolites,
validation of study results is crucial for achieving reliable detection and interpretation of metabolite patterns as well as unique biomarkers indicative of specific metabolic states. Thus, as the name implies, the key feature of the predictive metabolomics approach lies in the ability to allow for prediction of independent samples throughout the workflow. This is enabled by the use of SED for study designs, standardized protocols for sample handling and chemical characterization as well as a predictive attribute of the applied data processing, e.g. curve resolution, and multivariate statistical methods (Figure 5). As a result, the predictive metabolomics approach allows for screening of large number of samples in a high-throughput manner without compromising on the data quality.

Another important issue for the metabolomics studies is dealing with the potential biases from sample treatments\textsuperscript{59,119}. Addition of internal standards for normalization is one step towards minimizing such biases, but will not cope with the unique variation that is related to each single metabolite or to the order of analysis. In the predictive metabolomics strategy chemometric methods such as SED and MVA methods is utilized to construct sample treatment batches, e.g. for metabolite extraction and derivatization, creating run order designs for chemical analysis and selecting training sets for curve resolution and statistical analyses\textsuperscript{110,120}. Thus, by recognizing and considering the obvious sources of biases these can be minimized and with that the information output from the data maximized.
Results

In this chapter, results from papers I-V are outlined and discussed. The common theme throughout all papers is the predictive metabolomics methodology. This strategy is here applied to investigate the impact of exercise and nutrition on human metabolism to promote the understanding of human health mechanisms. The discussed papers are presented as appendices.

Papers I and II describe and evaluate the results of a hypothesis free multivariate screening methodology for studying human exercise metabolism and post exercise nutrition response. Paper III follows-up findings in paper II by studying differences in plasma metabolite profiles associated with physical fitness level. In paper IV an investigation of short and long term effects of diet on the metabolome in post menopausal women is presented. Finally, in paper V the predictive metabolomics strategy is applied to study alterations in metabolic profiles linked to probiotics treatment in infants.

Papers I and II

Metabolomics analysis of physical exercise induced stress responses and macronutrient modulation

The plasma response following exercise involves multifaceted interactions between thousands of primary and secondary metabolites. This metabolic response gets even more complicated when considering the exogenous influence of nutrients. Conventional
approaches to study exercise metabolism often restrict the analysis to a few compounds and to a pre-defined hypothesis. This approach will limit the possibility to detect novel and unexpected mechanistic phenomena. Therefore, the multivariate screening approach presented and evaluated in papers I and II is designed to complement conventional methods. The aim with this approach is to increase the understanding of the integrative biology behind, and reveal new information related to, physical exercise and post exercise nutrition. A schematic picture of the study design used in papers I and II is presented in figure 6.

**Figure 6.** A schematic overview of the study design used in papers I and II. All subjects performed four matched exercise sessions on individualized workloads and were either given water, low-carb+protein, low-carb or high-carb beverage immediately after completed exercise. Blood samples were collected pre and post exercise (0min) as well as in the early recovery phase (15-90min).

**Results**

The predictive metabolomics strategy is specifically designed to allow for independent sample prediction regarding both HMCR processing and multivariate analysis. This is an important feature to ensure that the model system is reliable and suitable for further mechanistic evaluation. In papers I and II a validation sample set (test set) was predictively resolved using the HMCR settings obtained from the training set, i.e. resolving the same metabolite peaks in the validation set as in training set before predicting it into the existing multivariate models. Additionally, in paper II the GC-time of flight (TOF)/MS analyses of serum samples were performed in two different runs separated by eight months. This bias caused by the differences in time of analysis was effectively accounted for by normalization by means of added internal standards and by a within subject normalization using each subject’s 0min sample as an individual reference.

A reproducibility test was performed of the obtained serum profiles in paper I by comparing data from analytical replicates. This resulted in >80% correlation for 210 of the 420 resolved putative metabolites, as well as for 28 of the 34 metabolites highlighted as significantly altered in response to the acute effect of exercise as decided by OPLS-DA.
modeling and a permutation test. Prediction of analytical replicates and an independent sample set into the existing OPLS-DA model resulted in 100% accuracy with regard to class (pre vs. post exercise). The calculated OPLS-DA models in paper II revealed an average of 99% correct classification according to the fitted model and 79% correct class predictions of independent samples.

The post exercise macronutrient response in paper II revealed expected general effects on glucose, insulin and free fatty acids (FFA) suggesting a different metabolic response in relation to macronutrient composition in the ingested beverage. A global OPLS-DA model comparing the serum metabolome for all four macronutrient compositions showed a clear beverage composition dependent separation in the model scores (Figure 7A).

![Figure 7. A) Cross validated OPLS-DA scores (tcv[1-3]P) revealing clustering of subjects in relation to which macronutrient composition that were ingested in the early recovery period following physical exercise. B) OPLS-DA covariance loadings plot (w[1-2]) of the resulting metabolic pattern explaining the clustering in the two first components of the model scores.](image)

The corresponding model loadings (Figure 7B) highlighted expected as well as new interesting metabolite patterns. Numerous amino acids increased after ingesting protein, sugars increased after ingesting carbohydrates and fatty acids increased after ingesting water. A more specific finding was the suggested anti-catabolic effect in subjects ingesting the low-carbohydrate and protein beverage compared to the sole ingestion of carbohydrates or water. This effect was shown by a specific decrease in 3-methylhistidin, a muscle catabolism marker, and an increase in pseudouridine, a marker of whole-body turnover rate of tRNA and rRNA and thus likely accelerated protein synthesis.

Additional findings included the detection of differences in metabolite recovery profiles in relation to fitness status. This was done by comparing the top and bottom five subjects, when ingesting water after completed exercise, by means of their aerobic capacity (VO_{2\text{max}}) in an OPLS-DA model. Interestingly, when predicting the same low-fit subjects, after ingesting the carbohydrate and protein beverage with suggested anti-
catabolic effect, into the existing model they all moved towards a more high-fit metabolite profile. One subject presented a deviating metabolite profile at two time-points. Coincidentally, this subject also revealed an abnormal insulin response following the ingestion of carbohydrate and protein beverage at the same time-points (Figure 8).

![Figure 8](image)

**Figure 8. Left:** Cross validated OPLS-DA scores describing the separation of subjects with high and low fitness status (i.e. VO₂max) in samples taken in the early recovery period after exercise. Subjects with top five VO₂max values are plotted as blue dots while the five subjects with bottom VO₂max values are plotted as yellow dots. All model samples were collected after ingestion of water only. Corresponding low-fit samples, collected after ingesting low-carb+protein beverage, were predicted into the existing model and are displayed as yellow circles in the plot. The same subject/time point is connected with a dashed line. **Right:** Insulin concentration of the deviating subject from the OPLS-DA predictions when ingesting the low-carb+protein beverage (red line) compared to the average (grey line).

**Summary and conclusion**

It is of high importance to have a working methodology that allows for independent sample predictions into existing models (both in terms of curve resolution and sample comparison models) as well as to allow for additional analyses of samples at a later stage. This is emphasized and evaluated in papers I and II.

The results from papers I and II conclude that we have a working model system for detecting exercise and post exercise macronutrient responses in serum metabolite profiles as well as a potential model system for discovering pattern related to fitness status. In addition, the predictive metabolomics methodology allows interpretation on an individual subject levels to reveal low/slow and high/fast responders as well as for detecting deviating metabolite profiles. This is not possible when only interpreting univariate p-values where all subjects are averaged or ranked and not individually displayed. By utilizing the cross-validated scores, the model variability will be incorporated in the model interpretation to a greater extent as compared to using only
the fitted model scores. This becomes even more important for the interpretation and visualization of inter subject variation when an external prediction/test set is lacking due to a too small sample set.

Exercise works as a stressor which amplifies the metabolic response and when using individualized workloads all subjects are in theory stressed at their own level and to the same extent. Therefore, each subject can act as its own control. In **paper II** this was applied by utilizing the sample taken immediately after completed exercise (0min) as the individual reference sample for within subject normalization. As an effect of this normalization, samples analyzed eight months apart could still be used together in sample comparison modeling.

The results from **paper II** also suggest an improved anabolic-catabolic balance favored by the ingestion of a more complex composition of macronutrients containing both carbohydrates and proteins immediately after finished exercise compared to only carbohydrates or water. Even though this study was not designed to investigate differences in fitness levels, the results in **paper II** provided interesting findings in this area. In addition, findings in **paper II** include changes in the metabolite profile indicative of possible early signs of an impaired insulin function or resistance. Thus, we hypothesize that this observation in a non-risk population, consisting of healthy regularly training young males, suggests that individualized exercise in combination with macronutrient intake could be used as a model system to amplify metabolic responses which potentially could lead to an early detection of metabolic disorders. Although interesting, this results needs to be confirmed in a larger cohort in a specifically design study.

**Paper III**

Investigation of metabolite profiles associated with cardiorespiratory fitness in healthy subjects

In relation to previous findings in **paper II**, in terms of differences in blood metabolite profiles related to subjects fitness level, **paper III** aims to further investigate the systematic changes in blood metabolites related to moderate differences in aerobic capacity. An investigation of the metabolic consequences related to physical fitness status, here defined by VO₂max, revealed mechanistic information both related to disease susceptibility as well as to properties of a healthy phenotype. Thus, metabolic stressors such as individualized exercise and nutritional loads (carbohydrate, protein and fat) were utilized in **paper III** with the intent to amplify metabolic signals and potentially expose more subtle metabolic differences linked to fitness level.
Results

**Paper III** revealed a clear and statistical significant separation (p<2.57E-06) of metabolite profiles related to moderate differences in fitness level. These changes were observed in plasma samples collected in 1) the early response following a nutritional load (during resting), 2) after completed exercise with and without a nutritional load and 3) in a verification test, i.e. subjects performing a matched exercise session including a nutritional load. Validation samples from session 3 were used as an independent prediction set to test the robustness of the obtained discriminating metabolic profiles (Figure 9).

![Three Different Sessions Diagram]

**Figure 9.** A schematic overview of the collected blood samples included in **paper III.** 1) All subjects were given a nutritional load during resting. Blood samples were collected pre and every following 15min for a total of 90min. 2) An exercise session with individualized workloads was performed where subjects were given either water or a nutritional load during the recovery period. Blood samples were collected pre and immediately after completed exercise as well as every 15min for a total of 60min. 3) All subjects performed a matched exercise session for validation purposes. Blood samples were collected pre and immediately after completed exercise as well as every 15min for the following 45min.

Prior to chemical characterization of the collected blood samples, a design was carried out to create sample batches for sample preparation and GC-TOF/MS analysis. This was done to minimize the risk of sample preparation order interfering with the final results. The design was set up by using the PCA scores calculated from anthropometric measurements, e.g. BMI, VO2max, percent fat in tissue, age and cetera. Batches were selected to maximize the PC space as well as to include a balanced set of high and low-norm fit subjects and for subjects in session 2, in relation to nutrition intake (**Figure 10**). All samples from one subject and one session were included in the same batch. Corresponding samples from each subject were analyzed in close connection but in a randomized order on the GC-TOF/MS.
Figure 10. A PCA model based on anthropometric measurements from which the sample preparation and GC-TOF/MS batches were designed. **Left:** Cross validated PCA scores (tcv[1-2]) describing the inter subject variation where high fit subject are displayed as dots and low-norm as circles. Batches are in different colors and matched subjects are connected by a line. **Right:** PCA loadings (p[1-2]) describing the variation among the anthropometric variables from which the PCs were calculated.

As a validation, interpretation of the metabolic responses to the acute effects of a bout of exercise in paper III resulted in a comparative metabolic pattern to that seen in papers I and II. Interpretation of these patterns revealed signs of oxidative stress and muscle leakage among the subjects after completed exercise in terms of increased levels of hypoxanthine\(^{21}\) and taurine\(^{22}\). However, no differences were observed in metabolite profiles related to different fitness groups immediately after completed exercise on individualized workloads, i.e. comparing post exercise samples between fitness groups.

Interestingly, differences between the fitness groups were seen in pre samples at fasting rest, in samples taken after a nutrition load as well as in the early recovery phase from exercise with water or a nutritional load. Further interpretation of the discriminating metabolic patterns revealed an overall decrease in gamma-tocopherol (GT), a vitamine E isomer, and an increase in docosahexaenoic acid (DHA), a C22:6ω3-poly unsaturated fatty acid (PUFA) in the high fit subjects. In relation to this, levels of saturated- (C16:0) and mono unsaturated- (C18:1) (MUFA) and trans (C16:1) fatty acid decreased in the high fit group while alpha-tocopherol (AT), another vitamin E isomer increased. Prediction of independent samples from the matched validation session showed small intra-subject variations, thus indicating a robust and representative metabolic pattern obtained in the initial OPLS-DA model. Model scores and loadings from the OPLS-DA model comparing plasma metabolite profiles from high fit to low-norm subjects following exercise and nutritional load can be seen in Figure 11A and B.
Figure 11. OPLS-DA model comparing plasma metabolite profiles from high fit to low-norm subjects following exercise and nutritional load. A) Cross-validated OPLS-DA scores, where high fit subjects are displayed as dark green dots and low-norm as light green dots. The independent sample predictions (from a replicate session) are displayed as circles. B) OPLS-DA model loadings describing the metabolite pattern related to fitness level. Metabolites in the upper part of the plot are positively correlated, i.e. increased, in high fit subjects while metabolites in the bottom part are negatively correlated, i.e. decreased in high fit subjects. Identified metabolites are displayed in black, where squares denotes metabolites classified as amino acids, crosses denotes metabolites classified as fatty acids and dots denotes non-classified metabolites.

Summary and conclusion

Cardiovascular function, which is improved with exercise training, can confer antioxidant protection that is valuable at rest and in the metabolism of exogenous nutrients. These improvements in the antioxidant defense system, which could help to deal with inflammatory responses linked to disease development, is potentially reflected in blood metabolites.

In paper III, moderate differences in the metabolome related to cardiorespiratory fitness (VO₂max) in healthy young males, which were all regularly training, was detected. Interestingly, by using external stressors such as nutritional load and individualized exercise the small differences observed in a few specific metabolites in pre samples during fasting were amplified and extended to include differences in the whole metabolic recovery profile, i.e. the dynamic response to nutrients and/or exercise.

To summarize, the high fit subjects revealed a potentially more favorable blood lipid profile as compared to low-norm. This profile was specifically defined by lower levels of gamma-tocopherol (GT), a vitamin E isomer, and higher level of docosahexaenoic acid (DHA), a ω3-PUFA. GT has proven effective in quenching reactive nitrogen species and inhibit heat shock proteins²³, which is a part of the oxidative stress response to
exercise and thus increased in the contracting muscle per se\textsuperscript{124}. As a consequence, the lower levels in high fit subjects might be explained by an increased uptake in the muscle cells. The higher level of the ω3-PUFA, DHA, in combination with lower levels of ω-6-PUFA, saturated- monounsaturated- and \textit{trans} fatty acids in the high fit subjects could potentially be a result of a metabolism more prone to use lipids as fuels, which is a well described phenomena following regular exercise\textsuperscript{125}. Also the increased availability of DHA in high fit subjects suggests an improved metabolism by means of reducing pro inflammatory activities\textsuperscript{126}. To conclude, the high fit subjects might have obtained an increased cardiovascular inflammatory and antioxidant defense system, which could be more prone to deal with the inflammatory response following exercise and nutrition intake.

One can speculate that a more active lifestyle, as for the high fit subjects, will increase the awareness of their own nutrient intake as well as of other factors that influence health. Thus, further studies of metabolic effects linked to physical fitness level will require a more detailed description of the participant’s normal metabolic pattern. This could be achieved by combining conventional food questionnaires with metabolomics analysis of tissues to investigate the normal variation within groups of different fitness levels.

**Paper IV**

The impact of long and short term diet interventions on the metabolome in overweight and obese postmenopausal women

The epidemiological link between obesity (mainly central), insulin resistance, diabetes and cardiovascular disorders is very clear but there are still questions remaining regarding the underlying mechanisms. After the menopause, there is a clear rise in the incidence in diabetes mellitus as well as cardiovascular diseases. A contributing factor might be the reduced levels of estrogen that causes a change in the fat distribution towards a more central localisation\textsuperscript{127}, which is a marker of cardiovascular risk per se\textsuperscript{127}. Weight management is therefore crucial for this population. Notably, a diet consisting of high amounts of saturated fats and nutrients with high glycemic index\textsuperscript{128} has been suggested to increase the prevalence of obesity\textsuperscript{129}. In contrast, a diet comprising of high amounts of fats, mainly mono- and poly unsaturated fatty acids (MUFA and PUFA), and low amounts of carbohydrates may reduce weight in comparison with a low fat and high carbohydrate diet\textsuperscript{130}. In addition, a relatively high content of protein in the diet may per se contribute to weight loss\textsuperscript{131}.

The study in \textbf{paper IV} aimed to investigate the metabolic response to different dietary restrictions in overweight and obese postmenopausal women, from a short- and a long-term perspective. For this, a high protein and MUFA diet was compared to a low protein MUFA diet; the latter in accordance with the Nordic nutrition recommendations from 2004\textsuperscript{132}. 
Results

**Paper IV** includes two separate studies of metabolic responses to dietary intervention in overweight and obese postmenopausal women. A short term study over five weeks and a long term study over two years (*Figure 12*). In the short term study an investigation of the effects of a high protein and MUFA diet composed by a low amount of carbohydrates (30 E%), normal to high in proteins (30 E%) and high in fats (40 E%) mainly MUFA and PUFA and low amounts of saturated and *trans*-fatty acids was carried out. In the long term study participants were either assign to the high protein and MUFA diet or to a low protein and MUFA diet, consisting of higher amounts of carbohydrates (55 E%), lower in protein (15 E%) and in fats (30 E%).

**TWO DIFFERENT INTERVENTIONS**

1) Short term effect of a high protein and MUFA diet
   - Five weeks, 10 subjects

2) Long term effect of two different diets
   - Two years
   - 34 subjects on a high protein and MUFA diet
   - 32 subjects on a low protein and MUFA diet

---

**Figure 12.** A schematic overview of the collected blood samples in paper IV. 1) A short term dietary intervention study including ten overweight and obese postmenopausal women following a high protein and MUFA diet. Blood samples were collected pre intervention, after five weeks on the restricted diet as well as after an additional three months on a non restricted diet. 2) A long term dietary intervention study comparing two different diets, low versus a high protein and MUFA diet, in 66 overweight and obese postmenopausal women. Blood samples were collected pre intervention as well as after six and 24 months on the restricted diets.

**Short term dietary restriction (study1)**

The ten overweight and obese post menopausal women showed an average weight loss of 1 kg/week and an average decrease in waist by 6 cm after five weeks on the high protein and MUFA diet.

From the metabolomics analysis of plasma metabolite profiles, a robust OPLS-DA model was obtained with high predictability of analytical replicates, i.e. 93% model sensitivity (% correct predictions) and 100% model specificity (% correct model fit). This model revealed clear differences between samples collected pre and post five weeks on the high protein and MUFA diet (*Figure 13A*). Additional validation of the chemical characterization and data processing, i.e. GC-TOF/MS and HMR, resulted in >90% correlation (Pearson’s) of resolved metabolites and internal standards between analytical replicates.
Prediction of follow-up samples collected after three months on a non-restricted diet into the existing OPLS-DA model comparing pre and post dietary samples, revealed highly diverse individual responses. When comparing these responses to anthropometric parameters, such as continued weight loss, the four subjects that had moved even further in the ‘diet direction’ continued to lose weight while one subject that had moved in the opposite direction did not (Figure 13B).

After five weeks of high protein and MUFA diet, numerous blood lipids and amino acids were significantly altered (p<0.05) and seen as robust trends in the OPLS-DA loadings (Figure 14). More specifically, there were an increase in MUFAs and PUFAs, such as docosahexadecanoic acid, (DHA, C22:3ω6), arachidonic acid (AA, C20:4ω6), oleic acid (OA, C18:1) and palmitoleic acid (PA, C16:1) after five weeks of high protein and MUFA diet. In contrast, there was a decrease in the saturated fatty acid lauric acid (LauA, C12:0).

Several amino acids decreased after the high protein and MUFA diet, i.e. tryptophan, tyrosine, proline and asparagine, while the amino acid conjugate methionylcysteine (MeCys) increased. MeCys has been reported to have cholesterol lowering effects\(^ {33} \) and was here seen in combination with a significant decrease in cholesterol (Figure 14).
Figure 14. OPLS-DA loading (p[1]) describing the metabolic pattern responsible for the observed separation between plasma samples collected pre and post five weeks on a high protein and MUFA diet. Entities displayed in the upper part of the plot increased in concentration after the five week intervention while entities in the lower part decreased.

Long term dietary restriction (study 2)

Both diets resulted in a significant decrease (p<0.04) in clinical measurements of waist, hip, BMI and percent body fat over the whole dietary period of 24 months. The only difference in the clinical measurements between the two diets was a significant decrease (p=0.02) over time in high sensitive C-reactive protein (hsCRP) associated with the high protein and MUFA diet. This was not seen in relation to the low protein and MUFA diet.

As expected, the metabolic response to the long term dietary intervention was much more diverse on an individual level, as compared to the short term intervention study. Over the whole two years dietary period both diets revealed individuals with high and low metabolic response to diet (Figures 15A and B). The individuals showing high metabolic responses were also those who lowered their BMI and percent body fat the most. Consequently, the low responders had negligible changes in both BMI and weight loss as compared to the mean of each group. Further confirmation of high and low responders could be found in homeostasis model assessment indices (HOMA), calculated from fasting glucose and insulin levels, were the high responders lowered their values and low responders showed increased values.
Figure 15. OPLS-DA models comparing plasma metabolite profiles in overweight and obese postmenopausal women in plasma samples collected pre to samples collected after 24 months on a restricted diet A) High protein and MUFA diet. Cross-validated OPLS-DA scores (tcv[1]) revealing a clear separation between plasma samples collected pre (white) to those after 24 months (red) on the high protein and MUFA diet. B) Cross-validated OPLS-DA scores (tcv[1]) revealing a clear separation of plasma sample collected pre (white) to those after 24 months on low protein and MUFA diet (grey). A and B) The highest responder within each diet is denoted by an arrow, low responders by a pound sign and reverse responder by a dashed line.

For both diets, a general trend was seen pointing to a decrease in saturated fatty acids and ω6-PUFA over the whole 24 months intervention. More specifically, a cohesive decrease were seen in LauA (C12:0), palmitic acid (C16:0), stearic acid (C18:0), AA (C20:4ω6) and di-homo-gamma-linolenic acid (DGLA, C20:3ω6) (Figure 16).

Unique changes in the metabolic pattern of the respective diet group were seen as an increase in DHA (C22:6ω3) in relation to the high protein and MUFA diet and a decrease of indole-3-propionate in relation to the low protein and MUFA diet. In both cases no change was observed in the other group (Figure 16).

An opposite metabolic response in the two diet groups were observed in ethanolamine and cholesterol, which were decreased in relation to the high protein and MUFA diet while increased in the low protein and MUFA diet (Figure 16).
Figure 16. OPLS-DA correlation loadings (p[1]) from two different OPLS-DA models plotted against each other in a shared and unique structure (SUS)-plot. **Y-axis:** Low protein and MUFA diet; p[1] for the OPLS-DA model discriminating between samples collected pre to those after 24 months of restricted diet. **X-axis:** High protein and MUFA diet; p[1] for the OPLS-DA model discriminating between samples collected pre to those after 24 months of restricted diet. The joint and opposite metabolite changes for both diets are displayed along the diagonals (arrows), while diet specific changes are seen along the x-(high protein and MUFA) and y-axis (low protein and MUFA).

The dynamic metabolic response was different between the two diet groups. Subjects restricted to a high protein and MUFA diet revealed a significant metabolic response (p<1.6E-04) in the early stage, i.e. baseline to six months, as well as in the late stage, i.e. six to 24 months. However, for subjects on the low protein and MUFA diet a significant multivariate model (p=3.4E-08) was only obtained when including the whole dietary period, i.e. baseline to 24 months.

Moreover, there was a difference in metabolite responses in the early phase, i.e. baseline to six months, related to the starting point of the high protein and MUFA diet, seen in the OPLS-DA scores (Figure 17). The first group, started the diet in October 2007, showed the lowest overall response while diet group three, started the diet in March 2008, revealed the highest overall metabolic response. The second group starting their diet in November 2007 showed the most homogenous metabolic response to the diet. This pattern was not observed in the later phase, i.e. six to 24 months.
**Figure 17.** Cross-validated OPLS-DA scores (tcv[1]) displaying a clear separation between samples collected pre (white) to sample collected after six months (grey) on a high protein and MUFA diet. The individual metabolic response of each subject is highlighted by a dashed line. A solid line divides the plot into segments corresponding to group assignments. Diet group one started their diet in October 2007, diet group two in November 2007 and diet group three in March 2008.

**Short and long term response to a high protein and MUFA diet**

There were a cohesive increase in the ω3-PUFA docosahexadecanoic acid (DHA, C22:6ω3) and myo-inositol as well as a decrease in lauric acid (LauA, C12:0) related to the high protein and MUFA diet, which was also seen in the short term study. These alterations were not seen in relation to the low protein and MUFA diet. In addition, differences were seen in relation to intervention length of the high protein and MUFA diet. The long term diet was associated with additional decrease in saturated fatty acids while the short term study revealed a decrease in aromatic and branched chain amino acids (BCAA).

**Summary and conclusion**

**Paper IV** revealed metabolic differences associated with dietary related weight loss in overweight and obese postmenopausal women restricted to a low protein and MUFA diet or a high protein and MUFA diet. By utilizing the predictive metabolomics approach, individual as well as subgroup variations were detected within each diet group. The obtained metabolic patterns from individuals with high and low metabolic responses to dietary related weight loss were highly correlated to clinical measurements such as BMI and HOMA indices, a measurement of insulin sensitivity\(^{13}\). In addition, subgroup variations related to the starting point of the high protein and MUFA diet was also seen. Subjects starting their diet during late fall, i.e. October and November, had an overall lower metabolic response to diet as compared to those starting in early spring, i.e. March (Figure 17). In the short term study, the prediction of follow-up samples after a non-restricted diet into the existing OPLS-DA model did also reveal high and low responders that were highly consistent with clinical weight loss measurements. These results suggest
that the metabolomics data carries mechanistic information that could detect and potentially explain the complexity of individual dietary related weight loss.

After five weeks on the high protein and MUFA diet the overweight and obese postmenopausal women revealed a significant increase in PUFAs and MUFAs and a decrease in saturated fats and cholesterol. This indicates that the participants have complied to the given dietary restrictions as well as obtained a more favorable blood lipid profile for the prevention of coronary heart diseases. Whether this favorable blood lipid profile is related to the specific diet or the overall weight loss is hard to tell.

In the long term study both diets resulted in a significant weight loss. However, differences in plasma metabolites were seen when comparing the low protein and MUFA to the high protein and MUFA diet. From this, specific dietary related alterations were observed in ethanolamine, cholesterol, DHA and myo-inositol. Both ethanolamine and DHA are major components in hepatic phospholipids, and ethanolamine supplementation has, in combination with a high fat diet, proven to lower serum cholesterol. In view of this, metabolomics analysis in humans has suggested that metabolites linked to hepatic phospholipids could predict the development of heart disease. Thus, the observed lower levels of these circulating metabolites in subjects on a high protein and MUFA diet could be related to a lower risk of developing CVD.

The menopause confers a higher risk of impaired insulin signaling and thus higher risk of developing insulin resistance. Some of the action of insulin might involve blood metabolites such as inositol, which acts as a second messenger of insulin action. In view of this, supplementation of myo-inositol has shown to improve the insulin sensitivity in postmenopausal women with metabolic syndrome. In paper IV, the increase in plasma myo-inositol levels for subjects restricted to the high protein and MUFA diet were seen in combination with an improved lipid profile, i.e. reduction of circulating saturated fatty acids and an increase in ω3-PUFA DHA (C22:6ω3), as well as lower levels of fasting glucose and hsCRP, a marker for low-grade inflammation. Thus, this suggests that the mechanisms related to weight loss by means of different diets might differently impact factors related to insulin sensitivity and cardiovascular disease risk in overweight and obese postmenopausal women.

Paper V

The impact of feeding *Lactobacillus* F19 during weaning: A study of the plasma metabolome

Microbial exposure is necessary for the development of a normal immune function. Infancy seems to represent a critical phase for the initiation of the immune response that could lead to an increased tolerance or the development of allergies. In addition,
differences in gut microbial composition have even been demonstrated to precede the development of obesity42.

The administration of live microorganism in adequate amounts has in general proven to confer health benefits on the host32. So far, probiotics treatment in the early life has shown to be effective for the treatment of diarrhea143. Still, the overall effects as well as the underlying mechanisms of probiotics treatment in humans are non-conclusive and further investigation needs to be made to establish whether and how probiotics benefit health38,144.

**Paper V** aims to evaluate the plasma metabolic response of *Lactobacillus paracasei* ssp. *paracasei* strain F19 (LF19) treatment in weaning infants as compared to matched placebo-controls. This strain has shown to reduce body fat and change the lipoprotein profile in early life145. Thus, the predictive metabolomics approach was applied to investigate the overall metabolic effects in plasma related to probiotics treatments as compared to placebos in 13 month old infants. A schematic study description is presented in **figure 18**.

**Figure 18.** A schematic overview of the study design and blood sample collection in **paper V**. A double-blind study of *Lactobacillus* F19 (LF19) treated weaning infants compared to placebos. Blood samples were collected at 5½ months of age and at 13 months of age after a daily intake of LF19 treated cereals or placebo.

**Results**

No significant multivariate model could be obtained for the discrimination between metabolite profiles of LF19 treated infants to placebos at 13 months of age. Thus, further investigations of treatments effects were based on differences in dynamics, i.e. comparing changes in the metabolite pattern related to age obtained in separate OPLS-DA models of LF19 and placebo subject (**Figure 19**).
Despite the lack of a significant multivariate model, there were still unique metabolites differing between treatment groups that were not related to age. Compared to placebos, the LF19 treated group had significantly lower levels of the saturated fatty acid palmitic acid (C16:0) and the MUFA palmitoleic acid (C16:1), whereas the BCAA tryptophan and putrescine, a polyamine important for cell growth and proliferation\textsuperscript{16} increased.

The dynamic metabolic profiles, i.e. comparing samples collected at 5½ months of age to those collected at 13 months of age in an OPLS-DA model, were robust for both treatment groups. By plotting the variable weights from the treatment specific models against each other in a shared and unique structure (SUS)-plot\textsuperscript{16}, the general (joint) and unique metabolic responses for the two study groups were exposed (Figure 20). For both groups, an age related change in the amino acid pattern was detected. This was seen as an increase in pipelicolic acid, phenylalanine and aspartic acid and a decrease in hydroxyproline, glutamine, beta-alanine and cysteine with age. In addition, we detected a general decrease in myo-inositol and saturated fatty acid lauric acid (C12:0) in relation to age (Figure 20).

Interestingly, dynamic metabolic variation unique for the LF19 treated infants, as compared to the placebo group, could also be detected in the model comparison. These included a decrease in methylycysteine and cholesterol. In addition, there was an increase in the \( \omega6\)-PUFA linoleic acid (C18:2\( \omega6\)) in relation to age in the LF19 treated group while a decrease was detected in placebos. Moreover, the LF19 treated infants revealed a decrease with age in metabolites linked to the pentos phosphate pathway, i.e. erythrose, glutarate, D-galactono-1,4-lactone and glucopyranose. This metabolite pattern was instead increased in placebos (Figure 20).
Figure 20. Loadings (p[1]) from two different OPLS-DA models discriminating between samples collected at 5½ months and 13 months of age plotted against each other in a shared and unique structure (SUS)-plot. Y-axis: LF19 (probiotics) group. X-axis: placebo group. In the plot the general (joint) and opposite metabolite changes for both groups are displayed along the diagonals, while group specific (unique) changes are seen solely along the x- (placebo) and y-axis (LF19).

Summary and conclusion

The metabolomics approach revealed metabolic patterns in response to age development in early life. This was seen as general pattern irrespective of probiotic treatment as well as a specific pattern uniquely associated with LF19 supplementation. In paper V, the value of the multivariate analysis was exemplified when picking up general and unique metabolic patterns by combining models of the dynamic age development between the groups. In this way, a robust pattern of metabolites linked to the pentos phosphate pathway (PPP) was also detected. The PPP metabolites were decreased with age in LF19 treated infants, while increased in the placebo group. The main functions of the PPP is to meet the need for nicotinamide adenine dinucleotide phosphate (NADPH) for reductive biosynthesis and the formation of ribose, used in ATP, DNA and RNA\(^{147}\). The lower levels of PPP metabolites in LF19 treated infants could be a result of a lower demand for reductive agents for handling oxidative stress. In view of this, some lactobacilli strains has proven to possess antioxidative qualities and are able to decrease the accumulation of reactive oxygen species\(^{148,149}\).

Results from paper V did also detect a cohesive age-related change in both treatment groups in metabolites that could be linked to a more diversified gut microbiota, as expected during weaning. This was seen as an increase in picovaleric acid, which can act as a precursor to several useful microbial secondary metabolites\(^{150}\).

In relation to previous finding linking LF19 treatment in mice to a decreased fat storage, we saw in paper V that lower level of palmitoleic acid (C16:1) was associated with LF19 treatment. Palmitoleic acid is one of the major MUFAs in humans and is an important
product of endogenous lipogenesis. Elevated serum levels of this fatty acid have been correlated to adiposity indices and visceral adiposity in children\textsuperscript{151-152}.

Whether or not probiotics treatment by means of LF19 is beneficial for health in the developing child still remains to be answered. Still, results from paper V highlights several metabolites and metabolite pathways that could be of interest in further investigations of microbial influence on human metabolism.
Conclusion

"Quit worrying about your health. It’ll go away."
- Robert Orben, Speakers handbook of humor

Health research of today has become a molecular science. As a consequence of this, interactions between scientists crossing traditional disciplinary borders are a necessity for moving this scientific field further to be able to provide answers to the highly complex question being human health.

From a health perspective, the most important issue may not be whether we carry a few extra pounds of weight or not. Instead, the key might be if we as individuals are metabolically healthy. For some individuals, shedding weight will definitely be most beneficial for their health while for others a change of their diet or exercise habits could be just as vital. Thus, our predictive metabolomics strategy was in this work applied to evaluate the overall metabolic responses related to physical exercise, nutrients and diets both in a general context as well as on an individual level.

Predictive metabolomics, being a combination of metabolomics analyses and advanced bioinformatics based on chemometric methods, runs as a common theme throughout the studies in this thesis. The results clearly highlight how the methodology is useful in all steps from the planning of a metabolomics study to the evaluation of complex metabolic interactions in interpretable maps as well as for validation and verification of metabolite patterns in a completely predictive chain including both data processing and multivariate models. In addition, evidence is provided that the strategy is useful for detecting and interpreting individual metabolic responses, which suggests that it can be a
valuable contribution to the development of personalized approaches in disease treatment and prevention.

Since health is to prevent and treat all diseases it is of high importance to detect signs of an imbalanced metabolism at an early stage. We were thus utilizing stressors such as individualized exercise and nutrients with the aim of amplifying metabolic responses and thereby detect early subtle signs of an imbalanced metabolism in healthy phenotypes and risk populations.

As a result, this work includes the detection of a potential anti-catabolic pattern related to carbohydrate and protein ingestion as compared to the sole ingestion of carbohydrates or water in the early recovery phase from exercise. The metabolomics approach could also identify a young healthy regularly training male with possible signs of early insulin resistance. This was seen as a deviating metabolic as well as insulin response in relation to carbohydrate and protein ingestion. Although a very interesting result, it still needs to be confirmed in a larger cohort in a specifically designed study.

Furthermore, a favorable metabolic pattern in terms of blood lipids associated with cardiorespiratory fitness was detected. Higher fitness level was associated with an increase in the cardiovascular inflammatory and antioxidant defense system as compared to low or normally fit individuals. This defense system is not only important during and after physical activity but can also be beneficial under normal circumstances such as for handling environmental and nutritional stress.

Weight loss achieved by a high protein and MUFA diet as compared to a low protein and MUFA diet resulted in different plasma metabolic responses, when studied in overweight and obese postmenopausal women. Here, the metabolic patterns revealed dietary specific changes in metabolites that were related to the prediction of insulin sensitivity and cardiovascular disease. The multivariate analysis methods did also highlight individuals as well as subgroup variation within the diet groups. The high correlation between metabolic response and clinical parameters, such as BMI and HOMA indices, suggests that there is information in the metabolite data that could explain the complexity of individual responses to dietary related weight loss. In addition, the predictive metabolomics strategy enabled the detection of a within diet group variation that was related to the starting point of the dietary intervention.

The final study aimed to investigate whether probiotics supplementation during the second half of infancy would affect the plasma metabolite profile. From this, several metabolites and metabolite patterns were highlighted that could help to explain the complex mechanisms related to probiotics treatment. The probiotics supplementation was suggested to have an impact on both the accumulation of reactive oxygen species as well as lipogenesis. However, whether the observed metabolic response to probiotics will benefit health in a long term perspective needs to be addressed in further studies.
To conclude, this work exemplifies the use of our predictive metabolomics as a useful tool for studying metabolite patterns in healthy phenotypes, risk populations and infants. An important lesson from all included studies is that metabolite patterns are potentially more representative for explaining the metabolic interactions in a larger population as compared to only using one or a few biomarkers, which is the traditional way of addressing it. Thus, the use of several metabolites as biomarkers could potentially allow capturing the individual variation needed when progressing towards diagnostic tools and personalized healthcare. At present, health risks are in a transition stage where the population is aging and patterns of physical activity and food consumption are changing. Understanding the role of these risk factors is important for the development of clear and effective strategies for preventing disease and thus improving health. Hopefully the results and methodologies presented in this thesis can be of value in this development.
Acknowledgements

Tänk vad många som har hjälp mig genom mina år som doktorand på kemiska institutionen. Vilken resa! Jag vill rikta ett stort TACK till alla er som jag har fått äran att jobba med.

TACK, thanks, merci, kitos, gracias, kha, cheers, dziekuje...

Henrik Antti, min huvudhandledare som hela tiden låtit mig få växa med uppgiften, och givit mig en välbehövlig knuff i rätt riktning då jag som mest behövde det. Speciellt många knuffar blev utdelade i slutskeden, som möjliggjorde denna textsamling. Tack för att du har låtit mig axla uppgifter som jag inte trodde att jag skulle klara, det är sånt en riktig bra handledare gör.

Mina handledare Thomas Moritz och Michael Svensson som delat med sig av ovärderlig kunskap och vägledning. Jag har lärt mig otroligt mycket av er.

Ett speciellt tack är riktat till Tommy Ohlsson och Christina West med tillhörande forskargrupper, för att ni låtit en kemist som mig få delta i otroligt intressanta medicinska projekt och diskussioner. Tack även för er ovärderliga hjälp med korrekturläsning av avhandling.

Min nära arbetskamrater och vänner på kemin: Anna Wuolikainen, Elin Thysell, Pär Jonsson, Lina Mören, Tommy Öhman, Mattias Hedenström, Johan Trygg, Max Bylesjö, Hans Stenlund, Izabella Surowiec, Stephen Bruce, Fredrik Pettersson, Tommy Löfstedt
Jenny Önskog och Rasmus Madsen. Samt alla andra fantastiska medarbetare på kemiska institutionen.

Ett extra supertack till er som har hjälp mig när jag har satt på andra sidan atlanten utan internet när det blev dags för provtryckning: Henrik, Anna och Elin.

Till alla på UPSC där jag gjort allt mitt labjobb vill jag rikta ett stort tack till: Krister Lundgren, Inga-Britt Carlsson, Annika Johansson, Maria Ahnlund, Pernilla Linden och Margareta Zetherström. Utan er hjälp hade jag varit vilse.


I would also like to give a giant thank to my fellow researchers at WashU during my time as the deputy lab mascot (second in command after ‘the Hulk’): Jeff Henderson, Chia, Kaveri, Haitao, Robin and Jan. You all made me feel so welcome!

Thanks to my dear friends and extra family ‘the Pinkners’: Sarah, Jerry, Chloe, Jesse, Cody, Wesley and Coco who help us with everything from house hunting, educated us in all-American sports such as baseball and football, pizza making as well as saving Lovisa from her tired grumpy parents. You guys made my final thesis work possible and enjoyable. I really appreciate all you have done!


Och till sist, mitt allt, mina allra närmaste, mitt bästaste. Mina tjejer, ni som alltid håller mig i nuet, vare sig det är med en spark i revbenet från insidan, eller genom att få se livet genom ögonen på en 1.5 åring. Min man Erik som gör mig till den allra lyckligaste, starkaste och vackraste i världen.
Populärvetenskaplig sammanfattning

"Ett äpple om dagen är bra för magen"
-Din mamma


En ny bild har genom de senaste åren växt fram om hur kroppen programeras så att rätt gener börjar arbeta vid rätt tillfälle och på rätt ställe i kroppen. Tidigare har man trott att en levande organism är förprogrammerad sedan konceptionsögonblicket och att vår arvsmassa och våra gener därför har något väldigt robust. Nu vet vi att generna slås på och av genom ett mycket intrikat signalssystem. Detta system är mycket lättare att störa än själva arvsmassan och det är här vår livsstil kommer in i bilden. Dessa insikter har medfört vetenskapen om att det inte längre räcker att endast fokusera på isolerade delar av ett biologiskt system för att förstå helheten utan hela dynamiska systemet måste beaktas. Detta tankesätt och tillika forskningsfält fick namnet systemsbiologi. Systemsbiologi inbegriper en global kartläggning av alla gener (genomik), RNA (transkriptomik), proteiner (proteomik) samt metaboliter (metabolomik). De sistnämnda påverkas i stor utsträckning omedelbart av vår livsstil, t.ex. vad vi äter och hur fysiskt aktiva vi är.
Traditionella tillvägagångsställ för att utreda biologiska mekanismer utförs ofta genom att, baserat på en förutbestämd hypotes, mäta ett fåtal ämnen/parametrar och sedan dra slutsatser från dessa resultat. Detta begränsar dock möjligheten att hitta oväntade resoner och nya frågeställningar. Denna avhandling har som fokus att belysa de mer globala metabolab effekterna relaterat till hälsa. Målet är att genom en hypotesfri kartläggning av alla små molekyler, s.k. metaboliter, i blodet hos människor kunna öka förståelsen för grundläggande mekanismer som är kopplade till hälsosämre aktiviteter, såsom träningskrav och näringsintag. Tillämpandet av en strategi vid namn prediktiv metabolomik är den röda träd som präglar hela avhandlingen. Denna strategi består av en kombination av metabolomikanalyser utförda med hjälp av högkänsliga analytiska instrument och kemometriska beräkningsmetoder och tankesätt för design av studier samt datahantering och utvärdering. Prediktiv metabolomik utgör ett effektivt verktyg för datahantering och processning för att skapa representativa data för ett biologiskt system. Metoden möjliggör dessutom en övergripande tolkning av generella sätt som individspecifika metabolab responser i dessa system, något som också exemplifieras i samtliga presenterade studier i avhandlingen.

Hälsa är något som påverkar oss alla genom hela livet. Detta behandlas i denna avhandling genom att tolka metabolab responser hos människor, av båda könen, i diverse livsfaser, ur ett hälso- och nutriktivt perspektiv. Metabolab effekter relaterat till träningsnivåer och nutrienter i den tidiga återhämtningsfasen samt jämställdhet mellan olika träningsnivåer har studerats hos unga friska män. Vidare har lång- och kortvariga effekter av en diet med högt fettsäurehalten jämförts med effekterna av en kolhydratradiet hos överviktiga kvinnor som passerat klimakteriet (postmenopausal). Slutligen har metabolab effekter i blodet av probiotikabehandling hos spådbarn under deras andra levnadshalvår utvärderats.

Förhoppningen är att en grupp metabolab som samverkar, ett s.k. metabolitmönster, i mänskligt blod ska bidra med tidigare och mer tillförlitliga indikationer på metabolab störningar hos tillsynes friska individer. Genom att på detta sätt använda ett flertal parametrar kan individuella skillnader tillåtas och hanteras i en högre utsträckning, vilket är en svaghet i dagens sjukvård där endast ett fåtal hälsomarkörer, till exempel kolesterol och glukos, används.

Resultaten från de presenterade studierna visar på tydliga metabolab mönster kopplade till såväl skillnader i tränishäftning som positiva effekter av viktiminskning genom olika dieter samt potentiellt positiva hälsoeffekter av probiotikabehandling hos spådbarn. En detaljerad tolkning av dessa metabolab mönster kan ge oss kunskap om vad en hälsosam metabolism är och hur olika individer bör agera för att förbättra just sin metabolab status för en bättre hälsa.
References


22. Popkin, B.M. Part II. What is unique about the experience in lower- and middle-income less-industrialised countries compared with the very-high-income industrialised countries? The shift in stages of the nutrition transition in the developing world differs from past experiences! *Public Health Nutrition* **5**, 205-214 (2002).


72. Laboratory, P. Centre for Inherited Metabolic Diseases. (ed. Hospital, K.U.) (Stockholm, Sweden, 2011).


