LIST OF PUBLICATIONS

The following papers are claimed as part of the PhD thesis.

- **Stanstrup J**, Gerlich M, Dragsted LO, Neumann S.
  Metabolite Profiling and beyond: Approaches for the rapid processing and annotation of human blood serum mass spectrometry data.
  Anal. Bioanal. Chem. 2013.

- **Stanstrup J**, Rasmussen JE, Ritz C, Holmer-Jensen J, Hermansen K, Dragsted LO.
  Intakes of whey protein hydrolysate and whole whey proteins are discriminated by LC–MS.
  Metabolomics. 2013.

- **Stanstrup J**, Schou SS, Holmer-Jensen J, Hermansen K, Dragsted LO.
  Whey protein delays gastric emptying and suppresses plasma fatty acids and their metabolites compared to casein, gluten and cod protein.
  2013. (*submitted*)

During the PhD period contributions were made to the following additional papers.

- Jiang P, Dragsted LO, Barri T, **Stanstrup J**, Jensen ML, Wan JM, Sangild PT.
  Antibiotics markedly affect/alter the urinary and plasma metabolome of preterm pigs susceptible to necrotizing enterocolitis.
  (*in preparation*)

- Johansson A, Barri T, Ulmius M, **Stanstrup J**, Önning G, Dragsted LO.
  LC-QTOF/MS metabolomic profiles in human urine after a 5-week high dietary fiber intake.
  (*in preparation*)

- Jiang P, Dragsted LO, **Stanstrup J**, Thymann P, Sangild PT.
  Severe malnutrition alters the urinary and plasma metabolome of neonatal pigs.
  (*in preparation*)

ABSTRACT

Classical human intervention studies are typically carried out to prove a hypothesis or answer a specific question. Therefore the design of the experiment and the performed analyses are based only on a few parameters related to the specific effect to be investigated while a wealth of information contained in the collected samples is not even recorded.

In recent years another approach, *metabolomics*, has proved to be a very valuable technique in nutrition research. In the metabolomics approach an intervention is undertaken without pre-defining which variables will be monitored – on the contrary the aim is to acquire as much information as possible in an unbiased way. In the present Ph.D. thesis the LC-MS metabolomics approach was applied to human nutrition intervention studies.

The plasma and urine samples collected in connection with three human nutrition intervention studies were analyzed using a metabolomics approach for the purpose of finding biomarkers of milk-derived whey protein intake and investigating the effects of whey intake on the human metabolome. It was an additional aim to devise computer-assisted methods to rationalize the process of compound identification.

In the first study the metabolomics profiles were compared following high-fat meals containing either cod, gluten, casein or whey as the protein source, while in the second study whole whey, subfractions of whey (α-lactalbumin or caseinoglycomacropeptide) and whey hydrolysate were compared. Both studies were performed with obese non-diabetics and the last study was repeated with diabetics as well.

We demonstrated that intake of whey causes a decreased rate of gastric emptying compared to other protein sources. This is in contrast to previous findings suggesting that whey is cleared faster than other proteins.

Paradoxically, we also find disproportionately elevated levels and shorter T_{max} of some aromatic and branched-chain amino acids following the whey meal. This suggests that whey affects absorption of amino acids in a way independent from, or at least not wholly controlled by, gastric emptying.

In addition, we find that whey caused decreased levels of a number of fatty acids due to increased insulin levels, which in turn is likely induced by the exaggerated amino acid levels.

We found no differences between the subfractions however, except for those explained by their different amino acid compositions. However, the hydrolysate contained a number of cyclic dipeptides that may be causing the hypoglycaemic effects observed for the hydrolysate. In addition we found that the manufacturing process for the hydrolysate caused methionine oxidation products, which were metabolized endogenously to metabolites not observed previously in humans.

We did not succeed in finding highly specific exposure markers of whey as the effects were confined to modifying levels of endogenous metabolites. Whey hydrolysate, on the other hand, contained unusual cyclic dipeptides; they are however, unlikely to be whey specific but rather a result of the hydrolysis process. These results could not have been recognized using traditional hypothesis testing approaches.

We also found a number of markers of the cod meal. While most are likely also markers of meat intake, arsenobetaine may be a specific marker of recent salt water fish intake.

We did not find any difference between obese non-diabetics and diabetics in their responses to whey. This, however, might be due to the semi-quantitative and separate nature of the analyses of the two sample sets, not allowing direct comparison of the "true" plasma levels.

At present, the major bottleneck in metabolomics studies of this kind is compound identification. Therefore this thesis will also present and discuss state-of-the-art tools for computer-assisted compound identification, including: annotation of adducts and fragments, determination of the molecular ion, *in silico* fragmentation, retention time mapping between analytical systems and *de novo* retention time prediction. A pipeline combining these tools in a single workflow is described, and the potential impact in the field of metabolomics highlighted. These tools were applied in the reported metabolomics studies.

RESUMÉ (ABSTRACT IN DANISH)

Klassiske humane interventionsstudier bliver sædvanligvis udført for at bevise en hypotese eller for at svare på et specifikt spørgsmål. Derfor er studiedesignet og de udførte analyser baseret på nogle få parametre relateret til de specifikke effekter, der ønskes undersøgt, mens en stor mængde information indeholdt i de indsamlede prøver ikke engang bliver målt.

I de seneste år har en anden metode, kaldet *metabolomics*, vist sig at være en særdeles fordelagtig teknik i ernæringsforskning. I metabolomics-tilgangen bliver interventionsstudiet udført uden at predefinere, hvilke variable der skal måles – derimod er målet at opsamle så meget information som muligt på en ikke selektiv måde. I denne ph.d.-afhandling blev væskekromatografi-massespektrometri (LC-MS) metabolomics-tilgangen anvendt i forbindelse med humane ernærings-interventionsstudier.

Plasma- og urinprøverne, opsamlet i forbindelse med tre humane ernærings-interventionsstudier, blev analyseret med metabolomics-tilgangen med det formål at finde biomarkører for mælkeproteinet valle og for at undersøge påvirkningen af valle på det menneskelige metabolom. Det var yderligere et mål at udarbejde computerassisterede metoder til at rationalisere processen ledende til identifikation af stoffer.

I det første studie blev metabolomics-profiler efter indtag af et måltid med højt fedtindhold, der indeholdt torsk, gluten, kasein eller valle som proteinkilde, sammenlignet, mens valle i det andet studie blev sammenlignet med underinddelinger af valle (α-laktalbumin eller caseinoglycomacropeptid) og hydrolyseret valle. Begge studier blev udført med overvægtige ikke-diabetikere og det sidste studie blev derudover gentaget med diabetikere.

Vi viser, at indtag af valle forårsager en lavere ventrikeltømningshastighed sammenlignet med andre proteinkilder. Dette står i kontrast til tidligere undersøgelser, som viser, at valle tømmes hurtigere fra ventriklen end andre proteiner.

Paradoksalt i forhold til ovenstående så finder vi disproportionalt høje niveauer og kortere T_{max} af nogle aromatiske og forgrenede aminosyrer efter indtag af valle-måltidet. Dette indikerer, at absorptionen af aminosyrer påvirkes uafhængigt af, eller i det mindste ikke fuldstændigt kontrolleres af ventrikeltømningshastigheden.

Derudover finder vi, at valle nedsætter niveauet af et antal fedtsyrer på grund af øget insulinniveau, hvilket sandsynligvis er induceret af de forhøjede aminosyreniveauer.

Vi fandt ingen forskel mellem underinddelingerne af valle, men hydrolysatet indeholdt en række cykliske dipeptider, der måske er årsagen til den hypoglykæmiske effekt observeret for hydrolysatet. Tillige fandt vi, at produktionsprocessen for hydrolysatet førte til dannelse af oxidationsprodukter af methionin, som metaboliseres endogent til tidligere ukendte metabolitter.

Det lykkedes ikke at finde eksponeringsmarkører med høj specificitet for valle, da effekterne var begrænset til at ændre niveauerne af endogene metabolitter. Valle-hydrolysatet derimod indeholdt usædvanlige cykliske dipeptider; det er dog ikke sandsynligt, at de er valle-specifikke, men snarere et udtryk for hydrolyseringsprocessen. Disse resultater kunne ikke være opnået ved de traditionelle hypotese-testende tilgange.

Vi fandt desuden en række markører for torskemåltidet. De fleste er sandsynligvis også markører for kødindtag, men arsenobetain kan være en specifik markør for fiskeindtag.

Vi fandt ingen forskel mellem overvægtige ikke-diabetikere og diabetikere. Det kan dog skyldes, at analysen ikke blev udført på en måde, der tillod direkte sammenligning af plasmaniveauer.

På nuværende tidspunkt er identifikation af stofferne den største flaskehals i metabolomics studier af denne type. Derfor vil denne afhandling præsentere og diskutere nogle af de nyeste værktøjer til computerassisteret identifikation af stoffer, herunder: annotering af fragmenter, bestemmelse af molekylær-ionen, *in silico* fragmentering, retentionstids-overførsel mellem analytiske systemer og *de novo* retentionstidsforudsigelse. Et sammenhængende system, der kombinerer disse værktøjer, bliver beskrevet, og den potentielle indflydelse på metabolomics-feltet understreges. Dette system blev brugt i de rapporterede studier.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

In this chapter the context in which this thesis is placed will be presented together with its aim and outline.

1.1 BACKGROUND

During the last one hundred years the populations in most regions of the world have undergone dramatic changes in terms of wealth, health [1] and lifestyle. In the Western World this has led to societies of affluence. While this is arguably a positive development a strong increase in the prevalence of obesity [2], metabolic syndrome and diabetes [3] have become a consequence of availability of an unrestricted amount of food and a more sedentary lifestyle.

Obesity and related risk factors are today some of the leading causes of preventable death [4] and thus of immense humanitarian and economic importance. The development of obesity is complex with serious social and psychological dimensions affecting virtually all ages and socioeconomic groups [5] and is not caused by a high-fat diet and low exercise level alone. In fact it has not been proven in any convincing way that there is a strict dose-response relationship between intake of calories and body fat; the quality of the fat plays an important role as well [6]. Genetic disposition, pre- and postnatal factors as well as intake of nutrients, independent of their caloric value, may also play a role. The question is which.

In the quest to answer this question the qualitative and quantitative intake of protein is of special interest since it has been shown that it is possible to achieve greater weight loss on a low fat/high protein diet compared to a low fat/high carbohydrate diet and that protein has a higher calorie per calorie satiety power [7]. The mechanism is so far unknown but possible causes include inhibition of energy intake due to release of gut peptides, liver metabolism and/or direct effects of certain amino acids [7].

Furthermore, it has been demonstrated that specifically *whey* proteins have certain biological properties that might be beneficial in the treatment and prevention of the metabolic syndrome related to obesity and diabetes. Whey proteins are derived from milk where it constitutes approximately 20 % of total proteins in bovine milk. The remaining milk proteins are caseins. Casein is the protein precipitated when milk is curdled by rennet (or by acidification) and is the main constituent of cheese. Conversely, whey is the proteins that remain in solution and is considered a waste product of cheese production. Whey proteins are therefore inexpensive, but also of high nutritional value and serves as an excellent additive to improve food formulations. These consideration, together with the possibly health promoting effects of whey protein, have sparked intense research into the effects and biochemical actions of whey protein. In fact several hundred journal papers describing research related to whey have been published during the timespan of the Ph.D. studies described in this thesis.

Whey proteins improve fasting lipids and insulin levels in overweight and obese individuals following a period of whey supplementation [8] and reduces short-term food intake compared to other protein sources [9] and thus might be valuable in the prevention and treatment of the metabolic syndrome.

While fasting glucose and lipid levels are important indicators of homeostasis, the deleterious effects of the metabolic syndrome are perhaps even more strongly associated with postprandial hyperglycemia and hyperlipaemia.

Holmer-Jensen *et al.* recently demonstrated that whey also has beneficial effects immediately following intake of a high fat meal. Whey caused lower postprandial lipemia (plasma triglycerides and free fatty acids), lower blood glucose and higher insulin levels compared to supplementation with cod and gluten [10].

While whole whey has been extensively studied, the role of individual whey proteins and peptides has not been thoroughly explored. Bovine-derived whey protein consists of approximately 50-55 % β-lactoglobulin, 20-25 % α-lactalbumin, 10-15 % caseinoglycomacropeptide, 10-15 % immunoglobulins, 5-10 % albumin in addition to small amounts of lactoferrin and lactoperoxidase [11]. Caseinoglycomacropeptide is a hydrophilic glycopeptide released from κ-casein during cheese production (using rennin) [12].

To investigate whether the activity of whey protein could be attributed to one of these subfractions Holmer-Jensen *et al.* conducted a new meal study similar to the one above but comparing whey isolate to products with enhanced proportions of α-lactalbumin and caseinoglycomacropeptide, respectively. In addition, a whey hydrolysate product was included in the study. This study did not show any difference between whey isolate or any of the two subfractions, however, the whey hydrolysate induced a larger incremental area under the plasma insulin curve (iAUC) at 30 min.

In this thesis I report on the results from the analysis of plasma samples collected in conjunction with the two studies by Holmer-Jensen *et al.* described above using a *metabolomics* approach.

1.2 NUTRITIONAL METABOLOMICS

The traditional approach to investigating the effect of different nutrients, i.e. intervention studies, is both time-consuming and expensive. The conclusions you can draw from such studies are also inherently obscured by methodological limitations. It is for example well-known that weight conscious individuals systematically underreport energy intake [13] and thus reliable information is difficult to obtain through such studies. Alternatively the study design could include stricter control of food intake, however, this is uncomfortable to both subject and researcher in all but the shortest studies and accurate assessment of intake of specific nutrients is still difficult to ascertain.

For these reasons it could be advantageous if surrogate measures, so called biomarkers, for the exposure under investigation could substitute self-reporting of intake.

World Health Organization (WHO) has defined a biomarker [14] as

> *"any measurement reflecting an interaction between a biological system and an environmental agent, which may be chemical, physical or biological".*

In addition WHO distinguish between two classes of biomarkers namely "*biomarkers of exposure*" defined as

> *"an exogenous substance or its metabolite or the product of an interaction between a xenobiotic agent and some target molecule or cell that is measured in a compartment within an organism"*

and *"biomarkers of effect"* defined as

> *"a measurable biochemical, physiological, behavioral or other alteration within an organism that, depending upon the magnitude, can be recognized as associated with an established or possible health impairment or disease".*

In the studies described in this thesis the purpose was two-fold. We sought both to establish biomarkers of intake of different protein sources, especially whey, but also to examine the effect of whey intake on the *metabolome.*

The term *metabolome* was first introduced by Oliver *et al.* in 1998 [15] and is generally taken to mean the collection of all low molecular weight molecules (metabolites) present in biological system in a particular physiological or developmental state [16].

Traditionally the investigation of perturbations to the metabolome of biological systems has been hypothesis-driven. In a hypothesis-driven approach the metabolites (or macromolecules or micronutrients) are predefined. This allows highly accurate targeted analytical methods to be employed. However, the degree to which the results of such studies reflect the biologically most important alterations are constrained by the strength and appropriateness of the *a priori* hypothesis.

In contrast, new developments in analytical instrumentation and computer power have opened the door to investigate changes to the metabolome in a holistic data-driven fashion.

Describing the holistic analysis of changes to the metabolome Nicholson *et al.* introduced the term "*metabonomics*" in 1999 [17] as

> *"the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification"*

while Oliver Fiehn in 2001 [18] defined the term "*metabolomics*" as

> *"A comprehensive and quantitative analysis of all metabolites"*.

The difference between the terms is mostly philosophical and the terms are used interchangeably. In this thesis I will use the more general term *metabolomics,* even though the studies described fall under the sub-set of metabolomics studies designated by the *metabonomics* definition.

Metabolomics is ideally suited to study the biochemical response to a stimulus when no clear hypothesis of the expected response can be formed beforehand, as is the case with the studies described here. The major challenge of metabolomics is the massive amount of data gathered (Chapter 2) when attempting to characterize and quantify the whole metabolome. First the data gathered need to pre-processed into a form appropriate for statistical treatment (Chapter 3). However, the main bottleneck in metabolomics is that initially the molecular structures of the "metabolites" found to characterize a certain stimulus (intervention) response are

unknown. Therefore the structure of the compounds (metabolites) of interest need to be elucidated (Chapter 4) which is a time consuming process.

1.3 THE AIM OF THE THESIS

The aim of the projects described in this thesis was to establish biomarkers of whey intake and to investigate the effects of whey intake on the human metabolome. Because compound identification constitutes the major bottleneck in any metabolomics study it was also the aim to develop a pipeline for computer-assisted compound identification to more rationally achieve the above goals.

1.4 OUTLINE OF THE THESIS

In this thesis I will first introduce the analytical platform and the associated terms and characteristics of the data associated with a liquid-chromatography-mass-spectrometry metabolomics analysis (Chapter 2).

Next, in Chapter 3, data pre-processing and statistical analysis will be described briefly.

In Chapter 4 I will describe state-of-the-art methods for annotation of liquid-chromatography -mass-spectrometry (LC-MS) data and computer-assisted compound identification. Because the description of these methods are scattered between technical papers and software documentations I have compiled a comprehensive summary of the individual steps.

In Chapter 5 my three metabolomics studies on the investigation of the effect of whey protein on the metabolome will be summarized.

Finally concluding remarks and perspectives are offered in Chapter 6 and 7, respectively.