Metabolomics of urinary tract infection: a multiplatform approach

Tiziana Pacchiarotta

ISBN: 978-94-6182-433-2
This thesis has been partially funded by the Stichting Tropische Geneeskunde – Leiden.

Printing: OffPage, Amsterdam

Metabolomics of urinary tract infection: a multiplatform approach

PROEFSCHRIFT

ter verkrijging van de graad van Doctor aan de Universiteit Leiden, op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker, volgens besluit van het College voor Promoties te verdedigen op dinsdag 20 mei 2014, klokke 13.45 uur

Tiziana Pacchiarotta

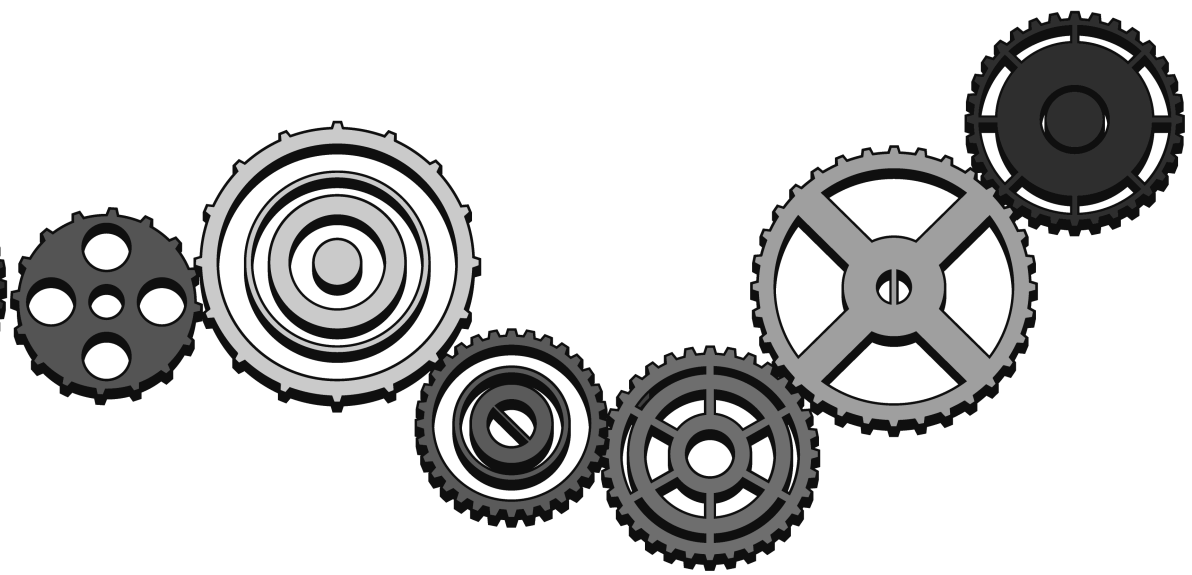
geboren te Avezzano, Italy in 1981

PROMOTIECOMMISSIE

Promotores:	Prof.dr. A.M. Deelder
	Prof.dr. J.T. van Dissel
Copromotor:	Dr. M. Mayboroda
Overige leden:	Dr. W. Dunn, *University of Birmingham, UK*
	Prof.dr. R.P.H. Bisschoff, *Rijksuniversiteit Groningen, Groningen*
	Prof.dr. J.W. de Fijter
	Dr. L.G. Visser
	Prof.dr. Ir. J.A.P. Willems van Dijk

Table of contents

Introduction



CLINICAL METABOLOMICS: AN OVERVIEW

Metabolomics is an example of a modern multidisciplinary research area. It has its roots in the technological revolutions of the sixties, the impressive breakthrough of the computer industry of the seventies and the re-birth of the holistic views in biology triggered by such a major undertaking as the human genome project. Metabolomics is a post-genomic discipline aiming at the study of metabolites: the end points and the intermediate products of the metabolism. Historically, the understanding of intracellular metabolic pathways and fluxes in model (mostly unicellular) organisms has been a dominant trend in the field [1]. However, clinical metabolomics, which has emerged as a spin-off of the 'mainstream application', focuses on the research of the metabolic regulation of the entire human organism and aims at the development of novel diagnostic/prognostic tools, prediction of patients responses to the treatment, exploration of the human metabolic individuality *et cetera.* Metabolomics of model systems and clinical metabolomics use the same analytical strategies but they differ with regard to experimental design and data treatment/interpretation.

Clinical metabolomics relies on the analysis of body fluids as the main source of information. The metabolic composition of the various body fluids provides the most accurate description of the individual physiological phenotype and as such is an essential element in the development of the concept of personalized medicine, the prediction of the susceptibility of an individual to a disease, the prediction of a state of disease or/and the outcome of a treatment. In a way, the most immediate goal of the clinical metabolomics is to reach the level of the technical confidence for being accepted as an independent measure in epidemiological studies.

Indeed, classical epidemiology operates within a limited space of clinical measures/readouts while metabolomics-based 'molecular epidemiology' is contributing by bringing in a comprehensive description of the chemical phenotype. Thus, instead of linking pre-existing knowledge to a certain level of the biological regulation (genes, gene expression protein expression), metabolomics offers a transverse approach to the problem evaluating the global outcome without taking into account the effect of any single variable.

At the time when metabolomics was just an emerging area, many studies were logically focused on revealing the potential of various analytical techniques and of the data processing. At the current state-of-art, metabolomics studies do not only provide class separation (e.g. diseased versus controls, response to a therapy and/or disease progression) but also a critical identification of those metabolites involved in the separation and their role in a meaningful biological context which can be used for further validation.

The metabolomics approach has been extensively used in the profiling of cancer phenotypes with the main intention to provide a valuable tool for highly needed alternative, earlier and molecule-based detection methods. A big portion of the present studies is dedicated to rapid diagnosis and monitoring of therapy. Probably, breast cancer represents the best example of metabolomics applied to diagnostic purposes: this tumor type showed at tissue level an elevated phosphocholine level, low glycerophosphocholine and low glucose compared with healthy tissue or benign tumors which allowed differentiation with a sensitivity of the 83% and a specificity of 100% [2]. Furthermore, it was calculated that magnetic resonance spectroscopic imaging analysis performed on choline-positive tissues could have prevented a biopsy in 66% of the cases [3]. Metabolite profiling has also contributed to the research of early diagnostic markers in lung cancer which, according to the World Health Organization, kills more than any cancer – a situation that is estimated

to continue till 2030. For this pathological condition, standard screening methods such as chest radiography and sputum histology have not increased the survival rate [4]. However, it has been proved that high risk cytochrome p450 genotypes may accelerate the catabolism of naturally occurring volatile organic compounds (VOCs) in the breath of lung cancer patients. Metabolite-based detection techniques have pointed at alkanes, alkenes and benzene derivatives in the classification of lung cancer patients' breath [5-7]. The additional investigation of body fluids showed an increased concentration of two aldehydes –hexanal and heptanal– in blood [8].

Among all the clinical conditions, infectious diseases deserve special attention: the metabolome of an infected organism, in fact, represents not only the metabolic phenotype of the host and/or the pathogen but it also depicts their cross-talk. Furthermore, for a comprehensive view of the metabolic pattern, multiple factors have to be taken into account such as the risk of exposure and the host's individual susceptibility to a given infection. Moreover, the word 'pathogen' encompasses a variety of infectious agents with very diverse invasion properties which are able to activate different pathways of the host immune system. The bulk of the literature dedicated to the metabolomics of infectious diseases concerns animal studies. For example, explorative studies of parasitic infections have been dedicated to the understanding of the pathogen metabolism and/or the host response after inducing the infection. Explicative examples are the most widespread parasite infections, namely malaria and schistosomiasis. For malaria, it has been shown that the dominant change in the urinary metabolic profile may provide a sensitive diagnostic tool for *Plasmodium* infection and for the evaluation of the malaria progression in mice [9]. A particular relevant finding is the specific conversion of arginine to ornithine induced by the parasite [10]. Tissue material has also provided interesting data for the understanding of the metabolic pathways. In fact, most tissues obtained from *S. mansoni*-infected mice were characterized by high levels of amino acids (leucine, isoleucine, lysine, glutamine and asparagine) and high levels of membrane phospholipid metabolites (glycerophosphoryl choline and phosphoryl choline). Additionally, low levels of energy-related metabolites (lipids, glucose and glycogen) were observed in ileum, spleen and liver samples of infected mice. These findings show that a *S. mansoni* infection causes a clear disruption of the metabolism in some tissues, which are in accordance with the previously reported indications in a biofluid (urine) [11].

METABOLOMICS OF THE URINARY TRACT INFECTION (UTI)

Despite the encouraging results achieved in animal model and *in vitro* studies, real clinical metabolomics studies in the field of infectious diseases are still lacking. Although large population studies are nowadays a matter of fact, the step from an ideal and controlled situation represented by a 'synthetic' study to human situation still appears tedious. In this respect, urinary tract infection (UTI) represents a real life situation in a clinical context and the comprehensive study of the UTI-induced changes in the metabolic pattern may offer the possibility to extensively investigate such a complex infection in man. This clinically well-defined disease represents a good model for the development of an analytically based scoring system of disease severity or/and evaluation of the effectiveness of a treatment. Urinary tract infection (UTI) is an inflammatory response to the presence of single or multiple species of pathogens in the urinary tract, with a high incidence of *Escherichia coli*, *Pseudonomas aeruginosa*, *Klebsialla pneumoniae*, *Enterobacter sp.*, *Acinetobacter sp.*, and *Proteus mirabilis*.

Despite the efficacy of the medical care and antibiotic treatment, UTI still contributes significantly to the annual morbidity and mortality rate in Western countries.

In a first approximation, UTI can be grouped into acute uncomplicated UTI (non-pregnant young women with cystitis but no evidence of urological abnormalities) and complicated UTI (UTI in males, all febrile UTI syndromes and those with urological abnormalities). In this respect, fever reflects the presence of a tissue invasive disease such as pyelonephritis, prostatitis or the urosepsis syndrome. The clinical management of UTI is relatively straightforward, but obtaining guidance for the duration of the treatment or for scaling of the morbidity inflicted to, for example, prostate or kidney is not such a simple task.

The identity of the causal pathogen is confirmed by a urine culture test. This typically needs 24-48 hours before the results are available. An alternative diagnostic method routinely used in clinical laboratories is a nitrite dipstick test: a fast and cheap method but with a high incidence of false-negative results, especially in the case of a Gram-positive bacteria pathogen.

An early and accurate diagnostic method would be valuable for a correct and prompt medical treatment, especially in view of the fact that epidemiological data suggest that a poorly diagnosed UTI or a recurrent UTI due to antibiotic-resistant pathogens is associated with a higher mortality rate [12, 13]. To this end, the scientific literature presents a growing range of NMR- and MS-based methods for fast identification of the pathogen [14]. It has been shown that uropathogens present a specific metabolite pattern which under controlled conditions can be used for accurate identification within six hours. For instance, *in vitro* studies have revealed specific metabolic 'signatures' for the different uropathogens: *E. coli* produces lactose from the lactate metabolism, *K. pneumoniae* produces 1,3 propanediol from the glycerol metabolism, *P. aeruginosa* produces 6-hydroxy nicotinic acid from the nicotinic acid metabolism and *P. mirabilis* produces 4-methylthio-2- oxobutyric acid from the methionine metabolism.

It has become evident that a molecule-based approach has practical value for the development of novel diagnostic tools. Furthermore, metabolomics creates new possibilities for the investigation of the host-pathogen cross-talk. This new aspect might help in more accurately scaling the progression of morbidity, a subject which has been largely ignored till this moment, and can help in the evaluation of the risk of the recurrent UTI. In fact, from a clinical point of view, recurrent and complicated UTI are more challenging problems since they can lead to more serious and life-threatening problems till organ failures.

The modeling of the metabolic phenotype has matured and improved over the years and is, nowadays, influencing translational medicine through the screening of a large cohorts of individuals with the ambition to deliver a classification in many different scenarios. An obvious application is the investigation of metabolic diseases often caused by a dysregulation in the metabolism of small molecules. It has been shown how a multiplatform approach to the metabolome-wide analyses of diabetes leads to a more sophisticated classification of the disease and the evaluation of the disease progression. Key observations could be linked to a series of metabolic plasma perturbations linked to kidney dysfunctions, lipid metabolism and gut microflora [15]. One of the first examples of a metabolome-wide association study has analysed risk factors (diet and blood pressure) in relation to exploratory spectroscopic data allowing the identification of good candidates for the metabolic signature of disease risk in different populations [16].

On the basis of the findings for metabolic diseases, a multiplatform explorative study could be expected to reveal the uropathogen expression and, even further, might uncover the response of the host and the complex signaling between the host and the (uro)pathogen. The so defined molecular pattern might lead to a deeper knowledge of the metabolome of UTI patients and thus to a more accurate disease classification and better understanding of the severity of the morbidity.

ANALYTICAL TECHNOLOGIES IN METABOLOMICS

The last decades of 20th century are often called the 'years of biology', a time when biomedical sciences made an enormous progress. Analytical and technical solutions, enabling large scale analysis of the 'biopolymers' (DNA, RNA and proteins), have played an important part in this phenomenon. The term 'biopolymer' implies the biological origin of a molecule and the limited number of monomeric units it consists of. A restricted number of the building blocks confines the chemical diversity of the oligomers (the analysis of which constitutes the basis of the analytical workflows) and at the same time one of the major analytical challenge, for example, in protein analysis remains the dynamic range. For metabolomics, dynamic range is just one of the challenges. The diversity of the physico-chemical properties of the metabolites is another one. Indeed, unlike the terms polynucleotides and polypeptides, the term 'metabolites' does not refer to a particular chemical family; even more, the most logical definition of metabolites is not a chemical but a functional one – the end products and intermediates of metabolism. Thus, metabolites belong to different chemical families and have a diverse size and polarity; their 'apparent simplicity' hides a concrete obstacle for the standardization of the analysis methods. Moreover, none of the existing analytical techniques provides a coverage of the entire metabolic space and therefore the analysis of the metabolome demands a multiplatform approach.

For metabolite profiling purposes, Nuclear Magnetic Resonance spectroscopy (NMR) and Mass Spectrometry (MS) are the main techniques applied.

Since the early 80's, NMR has been the first choice in metabolomics studies aiming at the analysis of body fluids [17, 18]. NMR is a highly robust and reproducible technology based on the detection of the resonance frequency of the atoms' nuclei when these are exposed to a magnetic field. Due to the high abundance of the ^{1}H nucleus, ^{1}H-NMR has been widely employed for metabolome profiling: any hydrogen-containing molecule in body fluids and tissues is potentially detectable. NMR is a non-destructive technique, can be largely automated and requires minimal sample preparation, all perfect characteristics for implementation in large scale clinical studies. In addition, according to its detection modality high-resolution NMR offers simultaneously structural elucidation and absolute quantification (at sub-micromolar range).

Despite the incomparable high throughput and the power of structural analysis, NMR only (un)covers a fraction of the metabolome, and thus mass spectrometry which may involve multiple ionization techniques, mass analyzers and hyphenation solutions represents a strong complementary technology. In mass spectrometry, molecules are detected on the basis of their mass over charge ratio (*m/z*) after an ionization process. Over the last decade, MS has become one of the key analytical methods in the biomedical sciences, primarily due to its potential to measure hundreds of metabolites over a wide dynamic range with a sensitivity that is several orders of magnitude higher than NMR. As a drawback, it is a

destructive technique, the metabolite profile is more susceptible to variability and results are platform-dependent. Recent improvements in achievable mass resolution and mass accuracy have enormously facilitated the task of structural elucidation of metabolites, with a mass accuracy often lower than a few ppm.

MS techniques show their best potential when used in the classical set-up of hyphenation with a separation technique which brings both benefits and complications. An MS analysis is more time consuming than NMR because of a number of reasons, such as the need of a more extensive sample preparation, a longer analysis-time and a more difficult interpretation of the resultant spectra. However, the hyphenation also provides extra information on the compounds' polarity, size or charge and reduces the complexity of the metabolic profile. The most commonly used separation techniques in combination with MS encountered in literature are liquid chromatography (LC-MS), gas chromatography (GC-MS) and capillary electrophoresis (CE-MS).

Of these, probably, the most applied hyphenation for profiling purposes is LC-MS due to its high versatility and the capability to analyze a wide variety of compounds. In addition there is a number of different columns, organic modifiers and technical solutions in order to maximize the chromatographic separation and MS ionization/detection with regard to a single compound, a certain matrix, or a specific compound family. For instance, the 'straight forward approach' encompasses the use of reversed phase columns that poorly retain polar compounds; to overcome this issue, hydrophilic interaction chromatography offers a solution. The advent of ultra high pressure LC systems (UPLC) in combination with MS analyzers has dramatically improved the chromatographic resolution, the peak capacity, and has reduced the limits of detection down to the picograms level while it also allows a better high-throughput capability than classic HPLC systems. Moreover, this type of platform allows coverage of a greater number of polar metabolites which increases its applicability and makes LC-MS an even more powerful technique for biomolecular research. Apart from improvement at the LC level, electrospray ionization (ESI) has significantly contributed to the success of LC-MS in its role of the perfect hyphenated technique for the analysis of the body fluids. For example, urine can be directly injected into LC-MS system and a plasma sample can be analyzed after minimal sample processing (protein precipitation).

GC is one of the oldest separation techniques and the first of the three separation techniques mentioned above which was used in a metabolite profiling study [19]. Despite its 'age', it still remains the only chromatographic technique able to separate, with high reproducibility, hundreds of compounds in one single run. GC-MS allows the simultaneous detection of a wide variety of analytes, although a derivatization step is often required which may be time-consuming and costly. The traditional sources for GC-MS are equipped with a vacuum stage interface – namely electronical impact (EI) and chemical ionization (CI). Under these conditions, EI fragmentation patterns have shown to be reproducible and as a result a number of open source and commercial spectral libraries are available which can be used for structural assignment. Nevertheless, the occurrence of extensive fragmentation can make the identification of unknown compounds in complex samples a challenging task. CI provides a softer ionization but the resultant spectra are strongly dependent on the experimental conditions (gas reagent, gas pressure) which makes the data less suitable for library search. The structural assignment of the compounds is, by its nature, the slowest element in an exploratory 'omics' workflow and there is always a place for new technology. Atmospheric Pressure Chemical Ionization (APCI) appears as a natural choice for a GC

platform. GC-APCI has recently been reintroduced for profiling aims after decades of oblivion. Reviving this technique has required a general optimization and definition of the analytical parameters as is shown in chapter 4 of this thesis. Moreover, one of the factors which seriously prevents GC-APCI from becoming a routine platform for explorative studies is the lack of a spectral such as extensively available for the more traditional sources. The first public spectra library for the APCI source is now available (http://metams.lumc.nl) (chapter 5) including a wide number of compounds (primary and secondary metabolites) most commonly encountered in the analysis of body fluids.

As indicated above, any single analytical detection modality brings its specific pros and cons, and thus the use of a multiplatform strategy to undertake an explorative metabolomics study of a clinical disease seems the most logical approach. UTI is almost an ideal test to prove how a metabolomics approach can offer a more comprehensive understanding of this complex clinical entity. The various analytical techniques have each worked as a magnifying lens providing a view on the metabolic signature and allowing a different perspective on the same pathological problem. The measurement of different panels of metabolites is the consequence of the complex status of the infection which represents a complex system where the host and the invader 'fight' their battle releasing specific chemical signals which may or may not be detected.

REFERENCES

1] Raamsdonk LM, Teusink B, Broadhurst D, Zhang NS, Hayes A, Walsh MC, et al. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nature Biotechnology. 2001;19:45-50.

[2] Bathen TF, Jensen LR, Sitter B, Fjoesne HE, Halgunset J, Axelson DE, et al. MR-determined metabolic phenotype of breast cancer in prediction of lymphatic spread, grade, and hormone status. Breast Cancer Research and Treatment. 2007;104:181-9.

[3] Jacobs MA, Barker PB, Bottomley PA, Bhujwalla Z, Bluemke DA. Proton magnetic resonance spectroscopic imaging of human breast cancer: A preliminary study. J Magn Reson Imaging. 2004;19:68-75.

[4] Humphrey LL, Teutsch S, Johnson M. Lung cancer screening with sputum cytologic examination, chest radiography, and computed tomography: An update for the US Preventive Services Task Force. Annals of Internal Medicine. 2004;140:740-53.

[5] Phillips M, Altorki N, Austin JHM, Cameron RB, Cataneo RN, Greenberg J, et al. Prediction of lung cancer using volatile biomarkers in breath. Cancer biomarkers : section A of Disease markers. 2007;3:95-109.

[6] Phillips M, Gleeson K, Hughes JMB, Greenberg J, Cataneo RN, Baker L, et al. Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. Lancet. 1999;353:1930-3.

[7] Poli D, Carbognani P, Corradi M, Goldoni M, Acampa O, Balbi B, et al. Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study. Respiratory Research. 2005;6.

[8] Deng CH, Li N, Zhang XM. Development of headspace solid-phase microextraction with on-fiber derivatization for determination of hexanal and heptanal in human blood. J Chromatogr B. 2004;813:47-52.

[9] Li JV, Wang Y, Saric J, Nicholson JK, Dirnhofer S, Singer BH, et al. Global metabolic responses of NMRI mice to an experimental Plasmodium berghei infection. Journal of Proteome Research. 2008;7:3948-56.

[10] Olszewski KL, Morrisey JM, Wilinski D, Burns JM, Vaidya AB, Rabinowitz JD, et al. Host-Parasite Interactions Revealed by Plasmodium falciparum Metabolomics. Cell Host & Microbe. 2009;5:191-9.

[11] Li JV, Holmes E, Saric J, Keiser J, Dirnhofer S, Utzinger J, et al. Metabolic profiling of a Schistosoma

mansoni infection in mouse tissues using magic angle spinning-nuclear magnetic resonance spectroscopy. International Journal for Parasitology. 2009;39:547-58.

[12] Foxman B. The epidemiology of urinary tract infection. Nature Reviews Urology. 2010;7:653-60.

[13] Dielubanza EJ, Schaeffer AJ. Urinary Tract Infections in Women. Medical Clinics of North America. 2011;95:27-+.

[14] Gupta A, Bansal N, Houston B. Metabolomics of urinary tract infection: a new uroscope in town. Expert Review of Molecular Diagnostics. 2012;12:361-9.

[15] Suhre K, Meisinger C, Doring A, Altmaier E, Belcredi P, Gieger C, et al. Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. PloS one. 2010;5:e13953-e.

[16] Holmes E, Loo RL, Stamler J, Bictash M, Yap IKS, Chan Q, et al. Human metabolic phenotype diversity and its association with diet and blood pressure. Nature. 2008;453:396-U50.

[17] Nicholson JK, Oflynn MP, Sadler PJ, Macleod AF, Juul SM, Sonksen PH. PROTON-NUCLEAR-MAGNETIC-RESONANCE STUDIES OF SERUM, PLASMA AND URINE FROM FASTING NORMAL AND DIABETIC SUBJECTS. Biochemical Journal. 1984;217:365-75.

[18] Bell JD, Brown JC, Sadler PJ. NMR studies of body fluids. NMR in biomedicine. 1989;2:246-56.

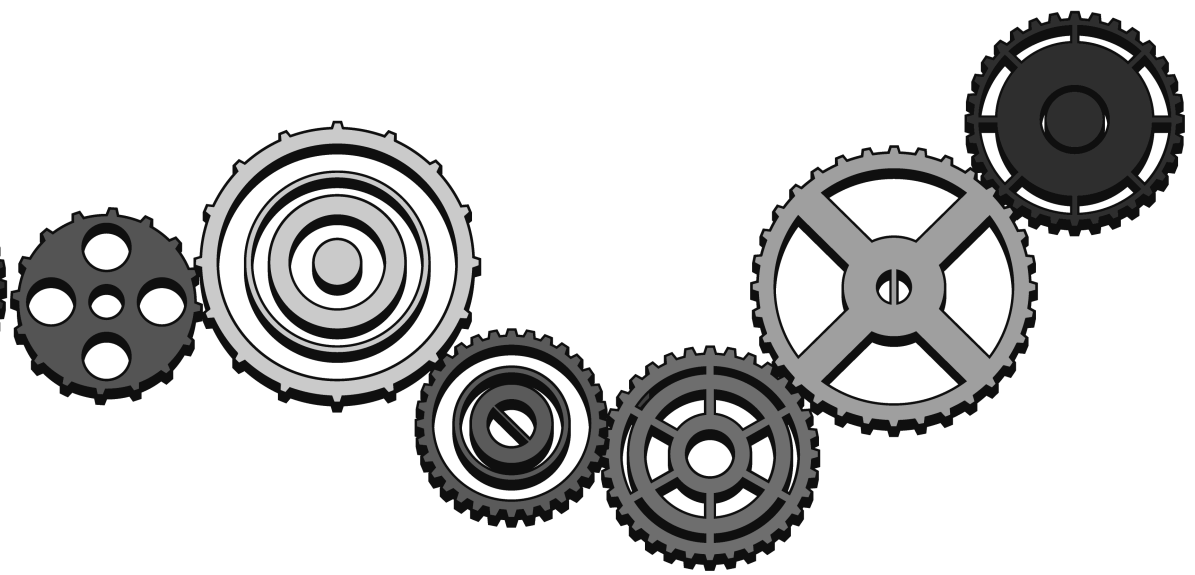
[19] Pauling L, Robinson AB, Teranish.R, Cary P. QUANTITATIVE ANALYSIS OF URINE VAPOR AND BREATH BY GAS-LIQUID PARTITION CHROMATOGRAPHY. Proceedings of the National Academy of Sciences of the United States of America. 1971;68:2374-&.

Chapter 1

Metabolomics investigation of human infections

Tiziana Pacchiarotta, André M. Deelder, Oleg A. Mayboroda

Bioanalysis 2012, 4(8) 2714-2722



ABSTRACT

Metabolomics has a special place among other 'omics' disciplines (genomics, transcriptomics and proteomics) as it describes the most dynamic level of biological regulation and, as such, provides the most direct reflection of a physiological status of an organism. Quick development of the analytical technologies in the first place –MS and NMR– has enabled the metabolomics analysis of such complex biological phenomena as host-pathogen interactions in the development of infection. In this review, an overview of the metabolomics studies of the infectious diseases carried out on human material is provided. The relevant papers on the metabolomics of human infectious diseases are comprehensively summarized in a table, including, for example, information on the study design, number of subjects, employed technology and metabolic discriminator. Future considerations, such as importance of the time-resolved study designs and the embedment of metabolomics in the large-scale epidemiological studies are discussed.

BACKGROUND

Recent advances in technology have strongly stimulated the development of the exploratory study models known otherwise as 'omics', for example, genomics, proteomics and metabolomics. The latter one refers to a study of the metabolome –the 'complete' set of the metabolites (small-molecules such as metabolic intermediates, amino acids or hormones) to be found within a biological sample [1, 2]. As the metabolome provides the most direct reflection of a physiological status or an organism, metabolomics is increasingly integrated in clinical studies as an essential phenotypic measure. A good overview of the technological basis and current achievements of metabolomics can be found in a number of excellent reviews that have been published over the last decade [3, 4]. Here, however, we will concentrate only on the progress in the field of metabolomics related to studies on of infectious diseases.

There are several reasons why metabolomics of infectious diseases requires special attention. First, every infectious disease is a process of host-pathogen interaction; thus with regard to metabolomics, the metabolic signature of an infected host represents the reciprocal actions of the host and the pathogen or their metabolic cross-talk [5]. Second, although the risk of exposure to a pathogen and the environment of a host play the most important part, genetic susceptibility of the host to an infection and consequently the host's metabolic individuality should not be underestimated. Finally the enormous diversity of pathogens should be mentioned. The evolutionary distance between pathogens is considerable; their 'strategies' for invasion of the host and evasion of the host's immune system differ extremely, which makes any straightforward generalization questionable.

Such a diversity complicates a review as well, since our perception of the progress in the field will be affected by the perspective taken. One could, for example, build an overview grouping the available material as traditional, emerging or re-emerging infections, or use a biological classification of the pathogens: bacterial pathogens, viruses, protozoa, helminths. Alternatively, one could consider a problem from the perspective of the type of the infection - acute or chronic. Assuming that the 'metabolic costs' and mechanisms mobilized by the host during the acute response are different from those used in the chronic phase, an overview of the metabolomics of human infections from a such point of view appears to be the most logical one.

METABOLOMICS OF THE INFECTIONS IN THE ACUTE PHASE

Every pathogen has to survive the acute-phase response of the host, but the development of the acute phase into a life-threatening condition happens not so often. Many pathogens are capable of subverting the initial reaction of the host and progress to a chronic phase. Consequently, here we try to address only those acute infections that are known as distinct clinical entities, such as, for example, urinary tract infection (UTI), which is indeed the most common bacterial infection leading to considerable health care expenditures. In fact the term 'urinary tract infection' encompasses a variety of clinical syndromes with the common denominator of a positive urine culture (i.e., significant bacteriuria > 10^3 CFU/ml) [6]. However, the presence of bacteria in urine is essential but not sufficient for the diagnosis of UTI and the correct assessment of the disease relies on a coordinated effort of clinicians and microbiologists. Within the limited number of metabolomics studies that have been performed on clinical material, UTI appears to be the most common topic.

The technical aspects, namely the assessment of the feasibility of analytical methods to

be applied, play an important role in those studies, regardless of whether an exotic technology is introduced [7] or a standard metabolomics technology, for example LC-MS, is used. A recent study of Lv *et al.* represents a good example of this technology-driven approach. The study is dedicated to the development of an LC-MS based method for metabolic profiling of UTI samples [8]: it includes optimization of the protocol, testing of columns with different particle sizes and both positive and negative ionization modes. The technology validation was followed by the analysis of a limited selection of samples. The authors thus identified a number of metabolites being specific for UTI (Table 1). However, it might be argued that far reaching biological conclusions, based on analysis of only a few samples, may need a critical re-evaluation. The work of Pacchiarotta *et al.* is another example of the use of the LC-MS technology for studying metabolomics of UTI [9]. Taking advantage of a prospective observational multicenter cohort study, the authors have selected a group of patients with culture-confirmed febrile Escherichia coli UTI. The study design included the control samples and samples of UTI patients collected at baseline (t=0) as well as patient samples collected after the antibiotic treatment (t=30). Thus, the authors had the possibility to not only compare UTI patients with symptom-free controls but also to observe how the metabolic phenotype of the patients reverts after the treatment. In doing so, they identified unique *O*-glycosylated fragments of the human fibrinogen alpha 1-chain and showed that presence of those fragments in urine strongly correlated with UTI symptoms.

Slupsky *et al.* explored another clinically important infection, namely acute pneumococcal pneumonia [10]. Pneumonia, like UTI, is a complex clinical entity, where an infection of the lower respiratory tract caused by bacteria or viruses is the common feature. The pneumococcal pneumonia caused by *Streptococcus pneumonia* is the main community-acquired form of the disease. To dissect the metabolic signature ('metabotype' according to authors' terminology) specific for S. pneumonia infection, a combination of targeted quantitative NMR profiling (61 metabolites) and a rather sophisticated study design was used. Anticipating that an acute infection represents a situation of metabolic stress for the host, the authors took a number of steps to ensure the specificity of their findings: the selection of patients included several additional 'control' groups of patients such as patients with non-infectious metabolic stress (cardio-vascular failure), patients with lung infections caused by other pathogens, fasting individuals and so on. Furthermore, a small longitudinal study on a sub-selection of eight samples was undertaken to monitor the evolution the 'metabotypes' of the patients. The approach used by the authors in the study design should serve as an example for future studies on metabolomics of infectious diseases. Still, one might ask whether an exploratory analysis of the entire NMR spectra could reveal more interesting compounds than those which were included in the list.

There are number of clinically relevant pathogens which present a threat to the human host during the acute phase of the infection. At first glance, it appears that many of those have been well studied and some (e.g., *Salmonella sp.*) have even been 'developed' into the standard model systems. However, the bulk of the publications address the metabolism of the pathogens in vitro or in animal models and the real clinical metabolomics studies of such important pathogens as *Vibrio cholerae* or even the otherwise well studied *Salmonella* are still awaiting their time.

Acute viral infections, such as influenza, represent a serious threat to the global health system. The previous pandemics (1918, 1957 and 1968) have clearly demonstrated the scope of the problem [11]. The 'efficiency' of influenza virus is based on the genetic re-assortment

and the effect of antigenic drift –a gradual change of the viral antigenic pattern due to frequent mutations [12]. Thus, even if a human host will acquire immunity to the invading strain, every new strain, originated as a result of the antigenic drift, can evade the host immune system. The value of the metabolomics studies in global control and management of influenza is far from being self-evident, which probably explains the lack of such reports in literature. There are, however, some publications addressing the impact of the virus on the metabolism of the hosts' cells. For example, Lin *et al.* have explored the metabolic pattern of human cell lines infected with influenza A virus [13]. They indicated that the metabolic signature might be associated with differentiated stages of the cells, probably due to fatty acids and cholesterol biosynthesis. Another possible niche for metabolomics study in a context of influenza infection could be an exploration of the 'collateral damage' effects like influenza-associated encephalopathy characterized by abrupt onsets of unconsciousness. The last case was addressed by Kavashima *et al.* who have studied CSF samples from six children affected by influenza-associated encephalopathy [14]. Using a high-field mass spectrometry (7 Tesla APEX III FTMS, Bruker Daltonics) authors have identified a number of signals specific for influenza-associated encephalopathy. However, neither statistical validation of the results, nor structural assignment of the signals, was presented.

METABOLOMICS OF THE INFECTIONS IN THE CHRONIC PHASE

If the question 'what do viruses and helminths have in common?' will ever be asked, a possible answer could be that these groups of organisms are 'masters of subversion'. To formulate it in more formal language, many viruses and helminths have the exceptional ability to develop long-lasting infections and coexist with the host for years or even decades. Hepatitis C virus (HCV) is a good example: in early infection the virus presents little to no damage to the host, however in the chronic infection it leads to progressive loss of liver function in 70% of patients [15]. Liver biopsy is still considered an essential element of HCV management and, thus, it comes as no surprise that one of the first applications of a metabolomics approach to this infection was an attempt to find an alternative way for the estimation of HCV-related morbidity. Zhang *et al.* have used targeted profiling of amino acids to reassess an old association between liver fibrosis and Fisher's ratio: the molar ratio of branched amino acids to aromatic ones. [16, 17]. A chronic Hepatitis B infection also results in the losses of liver function and an increase in the risk of hepatocellular carcinoma. A strong correlation between this morbidity and the levels of free amino acids in serum of HVB infected patients has been shown by Yu *et al.* [18].

An overview of the work carried out on HIV would require a separate review. There is a large body of literature dedicated to the studies of antiviral drugs and the metabolism of human CD4+ T-cells infected with the virus. Furthermore, an in-depth analysis of the metabolomics of HIV should discuss a fraction of the literature dedicated to the fundamentals of the human metabolism in the context of HIV infection and effects of the nutritional status of the host. These studies seldom use the 'metabolomics rhetoric' but they are essential for understanding the long-term effects of the infection. 'Classical' large-scale metabolomics studies on HIV are still lacking. However an original approach to the problem was taken by Ghannoum *et al.* who have evaluated oral wash samples as potential material for HIV diagnostics and monitoring of antiviral treatment [19]. Using a combination of

LC- and GC-MS, the authors have generated an overview of the metabolic composition of the oral cavity. The identity of all discussed metabolites has been verified by a custom built library of analytical standards. Next to the comparison between a control group and HIV patients a table summarizing the metabolic differences amongst HIV patients (naïve and exposed to the antiretroviral therapy) was presented. However, a closer examination of the reported p-values shows that the differences within the HIV group (naïve vs. antiretroviral therapy patients) are apparently more significant than between HIV and control groups. An interesting fact, which unfortunately was not discussed by the authors. While the biological system assumptions of the authors might need an independent validation on a larger set of samples, the idea to use the metabolome of the oral cavity for monitoring of the status of a viral infection opens many interesting possibilities. Oral wash has indeed a number of advantages: it is safer to handle than blood, it allows noninvasive sampling, the volume of the sample can be controlled and finally it is always available, which makes it an almost ideal sample for the 'time-resolved' studies [19].

A long-lasting coexistence with the host is typical for the survival of many parasitic helminths species. Some of these helminths, for example trematodes of the genus Schistosoma, can survive for decades. Schistosomiasis is after malaria the second most socio-economically devastating parasitic infection. According to the World Health Organization (WHO) there are approximately 200 million people chronically infected and 600 million at risk of the infection (www.who.int). Unlike malaria, where a large-scale clinical metabolomics study is still lacking, schistosomiasis has already attracted the attention of the metabolomics community. Two large NMR studies have been published within 4 years [20, 21]. Both studies were based on material collected in endemic areas: Ivory Coast [20] and in Uganda [21]. Despite many factors (geographical location, nutritional habits, co-infections) that make a direct comparison of the two studies difficult, both groups have reported that the age of the participants is a dominant source of the variance the data overshadowing all other factors including infection status and parasitic load. Anticipating this effect, Balog *et al*. have analysed their cohort by splitting it in two age-matched sub-groups; the first one included participants of 7-15 years old and the second of 20-40 years old. The list of the metabolic discriminators presented by the authors includes no surprising entries, but the data overlap significantly with results obtained on animal models [20, 22].

It is evident that viruses and helminths are not the only organisms causing chronic infections; bacterial pathogens, such as mycobacteria, can also be very efficient in subversion of the host defence and in long-time survival. The best example is *Mycobacterium tuberculosis* - the causative agent of the one of the most widespread of all human infectious diseases - tuberculosis (TB). The impact of the TB on the human population is so profound that the WHO has declared the state of a Global Emergency in 1993. With the development of multi-drug resistant tuberculosis, the decision of the WHO can hardly be considered as an overreaction. The value of the metabolomics for a systemic approach to the TB problem has been recognised already for some time and has even been addressed in several reviews [23, 24]. However, no metabolomics study on human material has been published so far and a study on a mouse model remains our closest approximation [25]. Leprosy, caused by Mycobacterium leprae, is another important mycobacterial infection. In comparison to TB, leprosy might appear to be a relatively minor problem but more than 200000 new cases worldwide were reported in 2010 alone [26] and this disease still represents a peculiar case when physical disability often enhanced by the social stigma associated with it [27]. The

work by Al-Moubarak *et al*. on metabolomics of *Mycobacterium leprae* [28] is to the best of our knowledge the only metabolomics study of a Mycobacterium infection performed on human material. The authors employed UPLC-MS for their study and used serum samples from a sample bank of an on-going project on the molecular epidemiology of leprosy. This technologically solid study has nevertheless one small weakness with respect to the sample selection. By comparing patients with high and low bacterial indices (BI), and by using a random sample selection from the database they ended up with two very unbalanced groups. The group of high BI consisted of only males (only one female) while the same time group of low BI was mixed. Thus, it remains an open question to what extent the differences reported by the authors can be attributed to the differences in the infectious status.

FUTURE PERSPECTIVES

The current overview covers only a fraction of the literature on the metabolomics of infectious diseases, namely studies performed on human material. At a first glance, the progress in this field might appear insignificant in comparison to such areas as metabolic syndrome, cardiovascular disorders or cancer [29-31]; only a limited number of studies have appeared and the data on some common infectious diseases, such as TB or malaria remain unavailable. Furthermore, the sets of specific metabolites observed often consist of a selection of the usual suspects (carboxylic acids, amino acids, sugars; for more details see Table 1) leading to the quite common conclusion about the involvement of the energy metabolism in the host reaction to an infection.

On the other hand, the link between the regulation of the host energy metabolism and the response to an infection has been known for decades and one might argue that the ability of a new technology to expose this well-established physiological phenomenon should be considered as a positive fact which provides a solid foundation for future investigations [32-34]. With regard to the usual suspects, one could emphasize the fact that, in general, not one or two specific metabolites but a specific panel of metabolites or even the metabolic pattern/profile of an organism as a whole should serve as a basis for clinically relevant conclusions. However, to identify those metabolic patterns, to reveal their specificity and to show their feasibility for clinical research, the static case-control study design will have to be substituted by a longitudinal one or any other type of design where time-dependent phenomena can be explored. So far, only a few studies (Slupsky *et al.* [10], Pacchiarotta *et al.* [9] –bacterial pathogens– and Balog *et al.* [21] –parasitic infection) have employed such designs.

Although, metabolomics of the infectious diseases is still in its early days, we are confident in its future. In a way, it is, probably, the most systemic part of the systems biology: it studies a unique phenomenon of the interaction of a pathogenic organism with human super-organism [35]. At the moment, it is difficult to foresee the way this field will evolve. Yet we might speculate that in the case of acute infections, a practical approach will not be the development of the diagnostics, which is usually established in clinical laboratories, but, for instance, the development of the analytically based scoring systems for the scaling of morbidity, or estimation of the response to treatment. The latter aspect is becoming increasingly important as the list of antibiotic-resistant strains is becoming longer and longer. With regard to chronic infections, an exploration of the metabolic trajectories of the infections unfolding in time is probably the most logical way to achieve the practical results. Finally, to address difficult cases - TB, HVC or HIV - the metabolomics should be integrated

in the large-scale epidemiological projects as a standard phenotypic measure.

Table 1. Overview of the works available on human material divided by acute and chronic infections.

Pathogen	Biofluids	Technique	Metabolite discriminator	Number of subjects	Study approach *	Publication year [Ref. Number]
Streptococcus Penumoniae	Urine	NMR	TCA cycle intermediates, branched amino acids, creatinine, taurine, myo-inositol	641	1, 2	2009 [10]
E. Coli	Urine	CE-MS	Phenylalanine, glutamic acid	18	1	2008 [7]
E. Coli, Gram positive	Urine	UFLC-MS	TCA cycle, terpenoid backbone biosynthesis, amino sugars and nucleotide sugars metabolism, arachidonic acid and steroid hormone biosynthesis	17	1	2011 [8]
E. Coli	Urine	UPLC-MS	C-terminal glycopeptide of the fibrinogen alpha chain 1	117	1, 2	2012 [9]
Klebisella pmeumoniae	Urine/Cell media	NMR	Glycerol metabolism	614	1	2006 [36]
E. coli, K. pneumoniae, P. aeruginiosa, Pr. mirabilis	Urine/Cell media	NMR	Lactate metabolism, methionine metabolism, nicotinic acid	617	1	2009 [37]
E. coli, K. pneumoniae, P. aeruginiosa, Pr. mirabilis	Urine	SELDI-ToF, NMR	N/A	80	1	2003 [38]
Plasmodium falciparum	Urine	NMR	Lipids, lactate	39	1	1988 [39]
Plasmodium falciparum	Red blood cells	NMR	Gamma aminobutyric acid (GABA), TCA intermediates, glutathione			2009 [40]
Influenza virus	CSF	Direct Infusion MS	N/A	6	1	2006 [14]
Hepatitis C virus	Urine	NMR	N/A	66	1	2010 [41]
Hepatitis C virus	Plasma	HPLC-MS	Branched and Aromatic Amino acids	53	1	2006 [17]
Hepatitis B virus	Serum	GC-MS	Branched and Aromatic Amino acids	47	1	2006 [18]
Hepatitis B virus	Plasma	GC/MS	Acetic acid, sorbitol, D-lactic acid, hexanoic acid, 1-nephtalenamine, butanoic acid, phosphoric acid, D-glucitol, glucose	40	1	2009 [42]
Hepatitis B virus	Serum	LC-MS	Lysophosphatidyl cholines, bile acids, hypoxanthine stearamide	59	1	2011 [43]
Hepatitis B virus	Plasma	UPLC-MS	Lysophosphatidyl cholines, primary fatty acid amides, comjugate bile acids	83	1, 2	2011 [44]
HIV virus	Oral wash samples	GC-MS, LC MS	Phenylalanine:tyrosine metabolism, aminosugars metabolism, purine metabolism, fructose mannose galactose, starch and sucrose metabolism, TCA cycle	24	1	2011 [19]
HIV virus	Plasma	HPLC-fluorescence	Phenylalanine/Tyrosine	107	2	2010 [45]
HIV virus	Plasma	NMR	Lipids, Glucose	63	1	2006 [46]
Mycobacterium Leprae	Serum	UPLC-MS	Polyunsaturated Fatty Acids, phospholipidis	23	1	2011 [28]
Schistosoma mansoni	Urine	NMR	TCA cycle, liver function, gut microflora	447	1, 2	2011 [21]
Schistoma mansoni	Urine	NMR	N/A		1	2007 [20]
Onchocerciasis volvulus	Serum	LC-MS	Fatty acids, sterol lipids	73	1	2010 [47]

*1 = case control study, 2 = longitudinal study

REFERENCES

[1] Oliver SG, Winson MK, Kell DB, Baganz F. Systematic functional analysis of the yeast genome. Trends in Biotechnology. 1998;16:373-8.

[2] Fiehn O. Metabolomics - the link between genotypes and phenotypes. Plant Molecular Biology. 2002;48:155-71.

[3] Kell DB. Metabolomics and systems biology: making sense of the soup. Current Opinion in Microbiology. 2004;7:296-307.

[4] Holmes E. The evolution of metabolic profiling in parasitology. Parasitology. 2010;137:1437-49.

[5] Kafsack BFC, Llinas M. Eating at the Table of Another: Metabolomics of Host-Parasite Interactions. Cell Host & Microbe. 2010;7:90-9.

[6] Wilson ML, Gaido L. Laboratory diagnosis of urinary tract infections in adult patients. Clin Infect Dis. 2004;38:1150-8.

[7] Ramautar R, Mayboroda OA, Derks RJE, van Nieuwkoop C, van Dissel JT, Sornsen GW, et al. Capillary electrophoresis-time of flight-mass spectrometry using noncovalently bilayer-coated capillaries for the analysis of amino acids in human urine. Electrophoresis. 2008;29:2714-22.

[8] Lv H, Hung CS, Chaturvedi KS, Hooton TM, Henderson JP. Development of an integrated metabolomic profiling approach for infectious diseases research. Analyst. 2011.

[9] Pacchiarotta T, Hensbergen PJ, Wuhrer M, van Nieuwkoop C, Nevedomskaya E, Derks RJE, et al. Fibrinogen alpha chain O-glycopeptides as possible markers of urinary tract infection. Journal of Proteomics. 2012;75:1067-73.

[10] Slupsky CM, Rankin KN, Fu H, Chang D, Rowe BH, Charles PGP, et al. Pneumococcal Pneumonia: Potential for Diagnosis through a Urinary Metabolic Profile. Journal of Proteome Research. 2009;8:5550-8.

[11] Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. Evolution and ecology of Influenza-A viruses. Microbiological Reviews. 1992;56:152-79.

[12] Carrat F, Flahault A. Influenza vaccine: The challenge of antigenic drift. Vaccine. 2007;25:6852-62.

[13] Lin S, Liu N, Yang Z, Song W, Wang P, Chen H, et al. GC/MS-based metabolomics reveals fatty acid biosynthesis and cholesterol metabolism in cell lines infected with influenza A virus. Talanta. 2010;83:262-8.

[14] Kawashima H, Oguchi M, Ioi H, Amaha M, Yamanaka G, Kashiwagi Y, et al. Primary biomarkers in cerebral spinal fluid obtained from patients with influenza-associated encephalopathy analyzed by metabolomics. International Journal of Neuroscience. 2006;116:927-36.

[15] Schuppan D, Krebs A, Bauer M, Hahn EG. Hepatitis C and liver fibrosis. Cell Death Differ. 2003;10:S59-S67.

[16] Fischer JE, Rosen HM, Ebeid AM, James JH, Keane JM, Soeters PB. Effect of normalization of plasma amino-acids on hepatic encephalopathy in man. Surgery. 1976;80:77-91.

[17] Zhang QW, Takahashi M, Noguchi Y, Sugimoto T, Kimura T, Okumura A, et al. Plasma amino acid profiles applied for diagnosis of advanced liver fibrosis in patients with chronic hepatitis C infection. Hepatol Res. 2006;34:170-7.

[18] Yu K, Sheng GP, Sheng JF, Chen YM, Xu W, Liu XL, et al. A metabonomic investigation on the biochemical perturbation in liver failure patients caused by hepatitis b virus. Journal of Proteome Research. 2007;6:2413-9.

[19] Ghannoum Mahmoud A. MPK, Jurevic Richard J.,Retuerto Mauricio, Brown Robert E., Sikaroodi Masoumeh, Webster-Cyriaque Jennifer, Gillevet Patrick M. Metabolomics Reveals Differential Levels of Oral Metabolites in HIV-Infected Patients: Toward Novel Diagnostic Targets. OMICS: A Journal of Integrative Biology. 2011.

[20] Singer BH, Utzinger J, Ryff CD, Wang Y, Holmes E. Chapter 11 - Exploiting the Potential of Metabonomics in Large Population Studies: Three Venues. In: John CL, Jeremy KN, Elaine HolmesA2 - John C. Lindon JKN, Elaine H, editors. The Handbook of Metabonomics and Metabolomics. Amsterdam: Elsevier Science

B.V.; 2007. p. 289-325.

[21] Balog CIA, Meissner A, Goraler S, Bladergroen MR, Vennervald BJ, Mayboroda OA, et al. Metabonomic investigation of human Schistosoma mansoni infection. Mol Biosyst. 2011;7:1473-80.

[22] Wang YL, Utzinger J, Xiao SH, Xue J, Nicholson JK, Tanner M, et al. System level metabolic effects of a Schistosoma japonicum infection in the Syrian hamster. Mol Biochem Parasitol. 2006;146:1-9.

[23] Sintchenko V, Iredell JR, Gilbert GL. Opinion - Pathogen profiling for disease management and surveillance. Nat Rev Microbiol. 2007;5:464-70.

[24] Forst CV. Host-pathogen systems biology. Drug Discovery Today. 2006;11:220-7.

[25] Shin J-H, Yang J-Y, Jeon B-Y, Yoon YJ, Cho S-N, Kang Y-H, et al. ^{1}H NMR-based Metabolomic Profiling in Mice Infected with Mycobacterium tuberculosis. Journal of Proteome Research. 2011;10:2238-47.

[26] Weekly epidemiological record. 2011. p. 389-400.

[27] Rafferty J. Curing the stigma of leprosy. Lepr Rev. 2005;76:119-26.

[28] Al-Mubarak R, Vander Heiden J, Broeckling CD, Balagon M, Brennan PJ, Vissa VD. Serum metabolomics reveals higher levels of polyunsaturated Fatty acids in lepromatous leprosy: potential markers for susceptibility and pathogenesis. Plos Neglect Trop Dis. 2011;5:e1303.

[29] Kind T, Tolstikov V, Fiehn O, Weiss RH. A comprehensive urinary metabolomic approach for identifying kidney cancer. Analytical Biochemistry. 2007;363:185-95.

[30] Lawton KA, Berger A, Mitchell M, Milgram KE, Evans AM, Guo L, et al. Analysis of the adult human plasma metabolome. Pharmacogenomics. 2008;9:383-97.

[31] Griffin JL, Atherton H, Shockcor J, Atzori L. Metabolomics as a tool for cardiac research. Nat Rev Cardiol. 2011;8:630-43.

[32] White RH, Frayn KN, Little RA, Threlfall CJ, Stoner HB, Irving MH. Hormonal and metabolic resposnse to glucose-infusion in sepsis studied by the hyperglycemic glucose clamp technique. Journal of Parenteral and Enteral Nutrition. 1987;11:345-53.

[33] Shaw JHF, Klein S, Wolfe RR. Assesment of alanine, urea, and glucose interrelationships in normal subjects and in patients with sepsis with stable isotopic tracers. Surgery. 1985;97:557-68.

[34] Carpentier YA DD, Siderova VS, Richelle M. Exogenous lipids and hepatic function. In: Kinney JM TH, editor. In Organ Metabolism and Nutrition: Ideas for Future Critical Care. New York: Raven Press, Ltd; 1994. p. 349-67.

[35] Nicholson JK, Holmes E, Wilson ID. Gut microorganisms, mammalian metabolism and personalized health care. Nat Rev Microbiol. 2005;3:431-8.

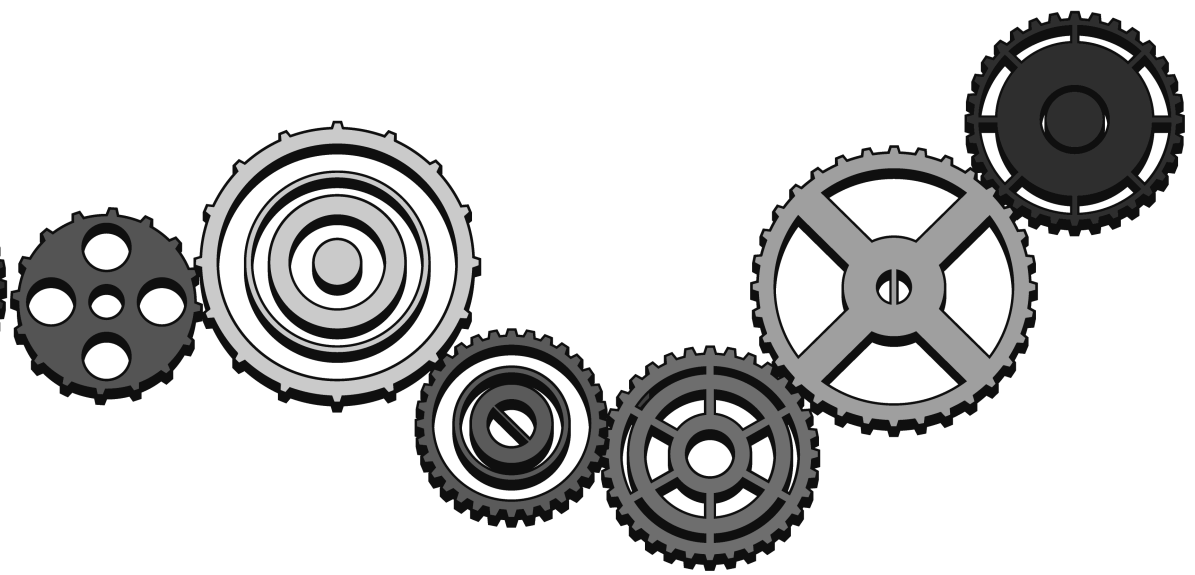
[36] Gupta A, Dwivedi M, Gowda GAN, Mahdi AA, Jain A, Ayyagari A, et al. H-1 NMR spectroscopy in the diagnosis of Klebsiella pneumoniae-induced urinary tract infection. NMR Biomed. 2006;19:1055-61.

Chapter 2

^{1}H NMR-based metabolic profiling of urinary tract infection: combining multiple statistical models and clinical data

Ekaterina Nevedomskaya, Tiziana Pacchiarotta, Artem Artemov, Axel Meissner, Cees van Nieuwkoop, Jaap T. van Dissel, Oleg A. Mayboroda1, André M. Deelder

Metabolomics 2012 8(6) 1227–1235



ABSTRACT

Urinary tract infection (UTI) encompasses a variety of clinical syndromes ranging from mild to life-threatening conditions. As such, it represents an interesting model for the development of an analytically based scoring system of disease severity and/or host response. Here we test the feasibility of this concept using ^{1}H NMR based metabolomics as the analytical platform. Using an exhaustively clinically characterized cohort and taking advantage of the multi-level study design, which opens possibilities for case–control and longitudinal modeling, we were able to identify molecular discriminators that characterize UTI patients. Among those discriminators a number of compounds (e.g. acetate, trimethylamine and others) showed association with the degree of bacterial contamination of urine, whereas others, such as, for instance, scyllo-inositol and para-aminohippuric acid, are more likely to be the markers of morbidity.

INTRODUCTION

Despite the progress made in understanding the mechanistic basis of many diseases in the last century, medicine is still essentially 'more an art than a science' [1]. Discovery of biochemical markers is thought to improve this, making diagnosis more reliable, specific and sensitive. These markers are sought at different levels of biological processes with the help of genomics, transcriptomics, metabolomics and other 'omics' approaches. Of these, metabolomics focuses on the analysis of metabolites present in biological fluids. Metabolites are end-points of all the biochemical processes of the organism and thus their collection – the metabolome is the closest approximation of the physiological phenotype and as such has a great potential for uncovering the biology underlying diseases and providing valuable markers of pathology [2,3].

The biological interpretation of results from metabolomics studies is rather complex and still in an early phase of development [4]. The human body is a 'super-organism' that unites its own network of interconnected tissues and organs with multiple colonies of microorganisms [5]. Interpretation of changes in concentration of metabolites found in biological fluids can readily be performed based on the underlying metabolic pathway; however, it is not always possible to link the observed change in systemic metabolite concentrations to a specific tissue or organ [6]. Especially in the case of disruption of highly abundant metabolites, e.g. from energy or amino acid metabolism, additional information would be required in order to interpret the data in respect to the tissue of origin. In addition, a change of such metabolites does not always improve the knowledge about the underlying cellular mechanisms and biology. A way to facilitate the interpretation of clinical metabolomics data is to integrate a plethora of available clinical parameters and to utilize a multilevel study design that should provide the opportunity to access the various levels of biological processes.

One of the examples of a complex and heterogeneous clinical entity, for which current diagnostic methods are not straightforward, is Urinary Tract Infection (UTI) [7]. Clinical manifestation of UTI can cover the range from mild cystitis to advanced pyelonephritis potentially leading to urosepsis and kidney failure. Physical symptoms may vary from patient to patient and be similar to a number of other diseases, mainly of infectious origin. Thus, the presence of bacteria and lymphocytes in urine can not be considered as a sole common denominator for UTI and even if it was, the criterion for the colony count is variable and anyway considered insensitive [8]. The correct and timely diagnosis relies on effective joint work of clinicians and microbiologists [8]. All of this explains the considerable interest in providing new, specific and sensitive markers for UTI, as well as for the uropathogen involved. The focus of the available metabolomics studies on UTI in the literature has so far been on the identification of pathogens: in the work of Gupta *et al.* a beautiful method with the use of ^{1}H NMR was proposed [9–11]. However, regrettably the method is not quantitative nor does it provide any information about the localization of the infection within the urinary tract, morbidity and preferred strategy of treatment

In the current study we investigated possibilities of using urinary metabolic profiles to monitor the health state of UTI patients, the degree of infection and the recovery process of UTI patients in the context of febrile, complicated UTI. We used a selection of samples from an exhaustively characterized cohort, with multiple urine samples available per individual and with the main pathogen identified as *Escherichia coli*, which is the most common pathogen for UTI. Samples from a group of age- and gender- matched UTI symptom-free subjects were included as control. The longitudinal design allowed studying various

biological processes: not only the difference between the patients and controls, but also the recovery process, using each patient as its own control.

MATERIALS AND METHODS

Samples

The study protocol was approved by the ethical committee of the Leiden University Medical Center and all included patients gave written informed consent.

Urine samples were collected at the Emergency Department and Primary Care Department. The sampling was carried out at several time points: the first urine samples were collected at the day of enrolment as baseline samples (t=0). Clean midstream-catch urine cultures were obtained and were analyzed using local standard microbiological methods. Three-four (t=4) and thirty days (t=30) after the day of enrolment, urine samples of the same patients were collected and new bacterial culture tests were performed (Supplementary Figure 1).

For the current study, from a database of about 700 subjects enrolled, 40 subjects, for which urine culture confirmed *E.coli*-positive complicated febrile urinary tract infection that recovered after antibiotic treatment, were selected. Samples from age- and gender- matched subjects with low bacterial culture in urine and without evidence of inflammatory diseases were used as controls (Table 1). A number of samples were missing, a few removed from the analysis due to either insufficient spectra quality or high glucose content (Supplementary Figure 1). In the end the study included four classes of samples originating from UTI symptom-free (N = 35) at day 0 (baseline control), UTI patients (N = 32) at day 0 (baseline), UTI patients (N= 29) at day 4 and UTI patients after recovery from infection (N = 37) at day 30 (Supplementary Figure 1).

Table 1. Characteristics of the studied patients and controls groups at baseline (t=0).

Characteristics	**UTI patients** n = 40	**Controls** n = 40	p
Age, years, median (sd)	59 (14.6)	58 (17.9)	0.9
Female, n (%)	22 (55)	22 (55)	1
Smoking, n (%)	5 (12)	5 (12)	1
Co-morbidity, n (%)			
Urinary tract disorder	4 (10)	4 (10)	1
Malignancy	4 (10)	1 (3)	0.17
Heart failure	5 (13)	3 (8)	0.46
Renal insufficiency	1 (4)	0 (0)	0.13
Diabetes mellitus	6 (15)	2 (5)	0.14
Immunocompromised	1 (3)	1 (3)	1
Urine dipstick results			
Nitrate	26/37 (75)*	0/37 (0)*	< 0.001
Leucocyte esterase	35/37 (95)*	5/37 (14)*	< 0.001

* 3 missing values

Sample preparation

Samples were thawed, transferred into 96 deep-well plates and centrifuged at 3000g for 15 minutes at 4°C to remove any precipitate. For sample preparation 520 µL urine were mixed with 60 µL of pH 7.0 phosphate buffer (1.5 M) in 100% D_2O containing 4 mM sodium 3-trimethylsilyl-tetradeuteriopropionate (TSP) and 2mM NaN3 in a 96 deep-well plate using a Gilson 215 liquid handler controlled by a Bruker Sample Track LIMS system (Bruker BioSpin, Karlsruhe, Germany).

NMR experiments and processing

^{1}H NMR data were collected using a Bruker 600 MHz AVANCE II spectrometer equipped with a 5 mm TCI cryogenic probehead and a z-gradient system; a Bruker BEST (Bruker Efficient Sample Transfer) system was used in combination with a 120 µL CryoFIT™ flow insert for sample transfer. One-dimensional (1D) ^{1}H NMR spectra were recorded at 300 K using the first increment of a NOESY pulse sequence [12] with presaturation (γB_1=50 Hz) during a relaxation delay of 4 s and a mixing time of 10 ms for efficient water suppression [13]. Eight scans of 65,536 points covering 12,335 Hz were recorded and zero filled to 65,536 complex points prior to Fourier transformation, an exponential window function was applied with a line-broadening factor of 1.0 Hz. The spectra were manually phase and baseline corrected and automatically referenced to the internal standard (TSP = 0.0 ppm). Phase offset artifacts of the residual water resonance were manually corrected using a polynomial of degree 5 least square fit filtering of the free induction decay (FID) [14]. In order to monitor proper filling of the NMR flow cell and for quality control 1D gradient profiles [15] along the z-axis were recorded for each sample prior and post data acquisition. Duration of 90 degree pulses were automatically calibrated for each individual sample using a homonuclear-gated nutation experiment [16] on the locked and shimmed samples after automatic tuning and matching of the probe head.

Statistical analysis.

Each spectrum was integrated (binned) using 0.014 ppm integral regions between 10 and 1 ppm, the residual water and urea region between 6 and 4.5 ppm was excluded, resulting in 550 data points used for the analysis. To account for any difference in concentration between the samples, each spectrum was normalized to a total area of 1. Absolute values were log-transforsmed. All pre-processing was done using in-house developed routines in R statistical environment (http://www.r-project.org/). Variables were centered and unit variance scaled prior to statistical analysis in SIMCA-P+ (version 12.0; Umetrics, Sweden) software package. For initial analysis and outlier detection, principal component analysis (PCA) was performed using 10 components. After the initial PCA analysis the following regions corresponding to paracetamol and its metabolites were excluded from the analysis: 7.5 – 6.75, 3.95 – 3.8, 3.7 – 3.45, 2.2 – 2.14 and 1.84-1.88 ppm according to Bales *et al.* {17]. For partial least squares-discriminant analysis (PLS-DA)[18] samples were categorized based on classes as defined by the study design. PLS model was built using 5 categories according to logarithm of bacterial count as a Y variable. Statistical models from supervised multivariate data analysis were validated by random permutation of the response variable and comparison of the goodness of fit (R^2Y and Q^2) [19–20]. For random permutation tests 100 models were calculated and the goodness of fit was compared with the original

model in a validation plot. Spectral regions responsible for the separation between classes in supervised models were identified based on the Variable Influence on Projection (VIP) values, which correspond to the importance of the variables (bins) for the model. The variables with a VIP value larger than 1.8 were considered significant and used for further analysis and identification of the responsible peak(s) within the spectrum. Prediction of class membership of samples by PLS-DA model was based on the predicted Y variable with the cut-off of 0.5.

For multilevel components analysis (MCA) using an in-house developed script in R as described by Jansen *et al.* [21] data were not log-transformed.

Univariate tests were performed to assess the statistical significance of the spectroscopic regions found using multivariate analysis: unpaired t-test was performed for the regions found as discriminating between UTI patients and controls by PLS-DA; ANOVA was performed on the regions that showed association with bacterial count in PLS; paired t-test was carried out on the regions identified in multilevel analysis. All the corresponding p-values were adjusted for multiple testing using Benjamini-Hochberg correction.

Identification of compounds of interest

Annotation of identified peaks was performed based on reference spectra from the Bruker Bioref database and in-house reference data. Confident identification was facilitated by the use of Statistical Total Correlation SpectroscopY method (STOCSY)[22].

Quantification of paracetamol

Quantification was performed by deconvolution and subsequent integration of paracetamol-glucuronide resonance at 5.10 ppm (d, 7.1 Hz) using an in-house developed automation routine. The absolute concentrations were calculated based on internal reference TSP. Values were not corrected for differential attenuation of the signals caused by relaxation during the mixing time and rapid-pulsing saturation effects.

RESULTS

The initial PCA on baseline samples revealed a trend in separation between UTI patients and controls in the scores plot of the first two principal components as shown in Figure 1A. However, we suspected that this could be due to over-the-counter (OTC) analgesics and antipyretics the patients might have taken prior to their visit to the hospital. Since paracetamol is one of the most commonly used OTC analgesic it was not surprising that we could find the major urinary metabolite of paracetamol, namely paracetamol-glucuronide, present in higher concentration in many of the urine samples from UTI patients, whereas it was present only in low concentrations in very few of the control samples (Supplementary Figure 2). The absolute concentration of paracetamol-glucuronide was used to stratify samples in the PCA plot: the direction of increase of paracetamol-glucuronide was found to match the direction of controls-patients separation (Figure 1B). The variables, corresponding to paracetamol and its metabolites, also dominated the loadings plot in the direction of separation between UTI patients and controls. As paracetamol is not an infection or morbidity marker, the further analysis was performed after the exclusion of the regions corresponding to the drug and its metabolites.

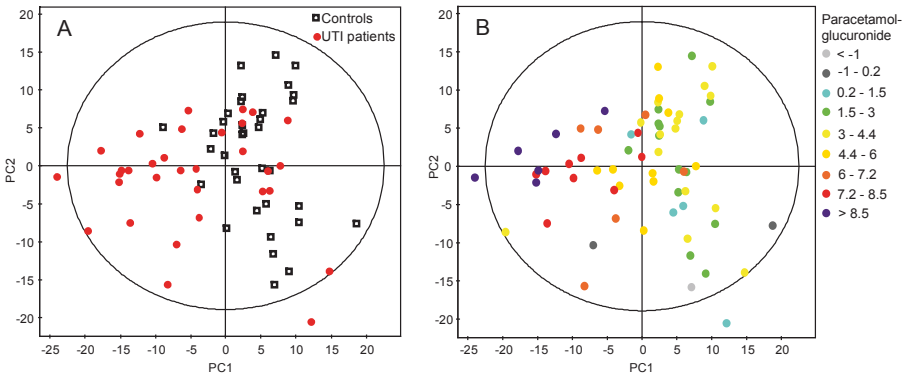


Figure 1. PCA scores plot of ^{1}H NMR data from controls and UTI patients urine samples at baseline, first two principal components covering 14.5 and 10.2% of variation respectively. (A) Colored according to controls (□) and UTI patients (●). (B) Colored according to the logarithm of absolute concentration of paracetamol-glucuronide.

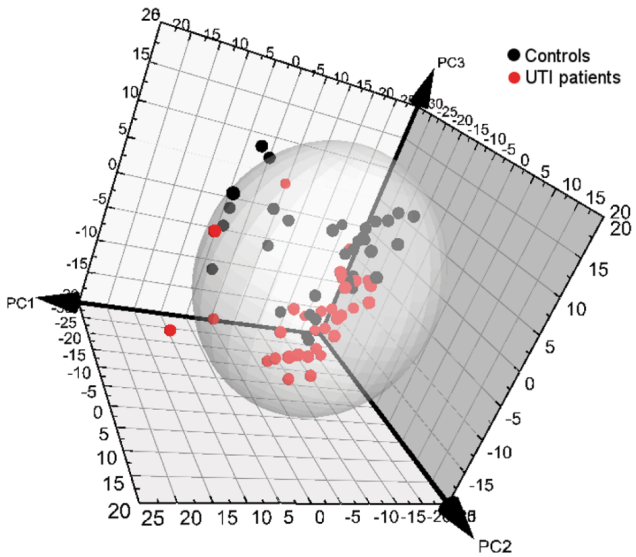


Figure 2. PCA scores plots of ^{1}H NMR data from controls (black) and UTI patients (red) urine samples at baseline after removal of the regions corresponding to paracetamol and its metabolites. First principal component covers 11.7%, second 11.2% and third 9.8% of variation.

The PCA analysis of the baseline samples after the removal of spectral regions of paracetamol and its metabolites did not show separation between UTI patients and controls within the scores plot of the first two principal components; however, a clear trend was identified along the third principal component (Figure 2), which means that inter-individual

variability is to a certain extent more prominent than the disease effect. No outliers were detected based on distance to the model (DModX).

In the next step a supervised PLS-DA model was built for t=0 using UTI class as a response variable. In the scores plot of the resulting model the two groups were well separated (Figure 3). Cumulative explained variance (R2Y) of 0.88 and cross validated predictive fraction (Q2) of 0.63 were calculated for the model; the model validation plot showed intercepts of the R2Y and Q2 regression lines with the vertical axis at 0.63 and –0.11, respectively, indicating a valid model. Molecular discriminators were identified based on relevant regions as identified by the corresponding VIP. A list of those regions, along with the p-values based on t-test (corrected for multiple testing), the direction of change and identities of the corresponding metabolites are summarized in Table 2.

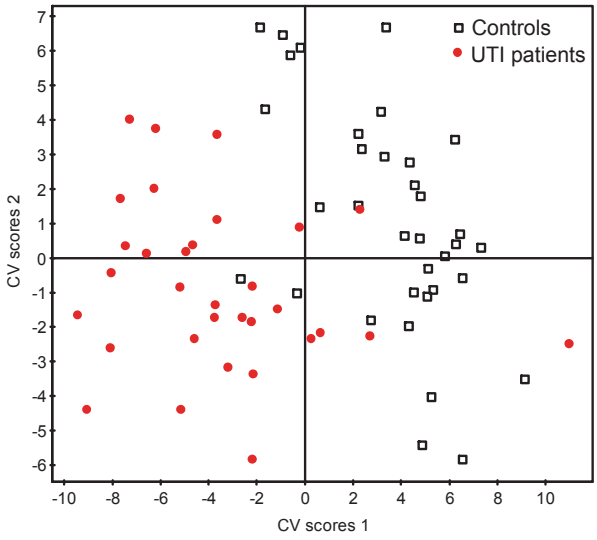


Figure 3. Cross-validated PLS-DA scores plot of urine ^{1}H NMR spectra of controls (□) and UTI patients at baseline (●), R2Y = 0.88, Q2 = 0.63.

Table 2. Spectroscopic regions that appear as influential in various statistical models and statistical significance of the corresponding univariate tests adjusted for multiple testing using Benjamini-Hochberg method.

		Controls vs. UTI patients		Bacteria concentration		Recovery from t=0 to t=30	
ppm region	Identity	t-test p-value	change	ANOVA p-value	change	paired t-test p-value	change
9.291 - 9.277	1-methylnicotinamide	<0.0001	-	<0.001	-		
9.277 - 9.264	1-methylnicotinamide	<0.01	-				
8.977 - 8.964	1-methylnicotinamide	<0.01	-				
4.491 - 4.477	1-methylnicotinamide	<0.01	-	<0.01	-		
1.941 - 1.927	Acetic acid	<0.01	+	<0.01	+		
1.927 - 1.914	Acetic acid	<0.0001	+	<0.0001	+		
3.196 - 3.182	Acetylcarnitine	<0.01	+				
2.568 - 2.555	Citric acid	<0.01	-				
2.541 - 2.527	Citric acid	<0.01	-				
4.082 - 4.068	Creatinine	0.03	-				
3.073 - 3.059	Creatinine	<0.01	-	0.07	-		
3.059 - 3.045	Creatinine	0.09	-				
7.709 - 7.696	Furoylglycine					<0.01	+
7.696 - 7.682	Furoylglycine	<0.01	-	<0.01	-		
3.959 - 3.946	Glycolic acid derivative	<0.001	-	<0.01	-	<0.0001	+
7.859 -7.846	Hippuric acid	<0.01	-	<0.01	-		
7.668 - 7.655	Hippuric acid	<0.001	-	<0.01	-		
7.655 - 7.641	Hippuric acid	0.01	-	0.02	-		
7.586 - 7.573	Hippuric acid	<0.01	-	0.05	-		
3.973 - 3.959	Hippuric acid	0.01	-	0.03	-		
8.555 - 8.541	Hippuric acid (amide)	<0.01	-				
8.541 - 8.527	Hippuric acid (amide)	<0.001	-	<0.01	-		
1.341 - 1.327	Lactic acid	<0.01	+	<0.01	+		
7.764 - 7.75	Para-aminohippuric					<0.001	+
3.332 - 3.318	Scyllo-inositol					<0.01	+
3.455 - 3.441	Taurine	<0.0001	+	<0.001	+	<0.0001	-
3.441 - 3.427	Taurine	<0.0001	+	<0.001	+	<0.0001	-
3.427 - 3.414	Taurine	<0.0001	+	<0.01	+		
3.264 - 3.250	Taurine	<0.001	+				
8.855 - 8.541	Trigonelline					0.01	+
4.45 - 4.436	Trigonelline					<0.01	+
2.896 -2.881	Trimethylamine	<0.0001	+	<0.0001	+		
8.486 - 8.473	Unknown					<0.01	+
7.968 - 7.955	Unknown					<0.001	+
7.75 - 7.736	Unknown					<0.01	+
7.518 - 7.505	Unknown			<0.01	+		
6.686 - 6.673	Unknown					<0.0001	+
6.509 - 6.496	Unknown					0.04	+
3.168 - 3.155	Unknown			<0.01	-		

1) two-group t-test for the healthy controls and UTI patients at baseline; positive direction of change corresponds to intensity of the region being higher in UTI patients compared to controls, negative – region intensity is lower in UTI patients compared to controls

2) ANOVA analysis for the number of bacteria present in urine; direction corresponds to the correlation to the number of bacteria: positive corresponds to the raise of the region intensity with the increase of the number of bacteria, negative - to the decrease of the region intensity with the increase of the number of bacteria

3) paired t-test for the UTI patients at baseline and 30 days; positive direction of change corresponds to intensity of the region being higher at 30 days compared to baseline, negative – region intensity is lower at 30 days compared to baseline

The advantage of PLS-based models is that they can easily be used to predict the class membership of new samples. Data of the UTI patients at t=4 were predicted using the two-class PLS-DA model that was built as described above. Of a total of 29 urine samples included in the prediction set, 19 (65.5%) were classified as controls, whereas 10 (34.5%) samples were classified as UTI (Figure 4). Besides using data from the 4-days time point as prediction set, we also performed a separate analysis for the 30-days time point (Figure 4). In this case, out of 37 samples collected, 32 (86.5%) were attributed to the group of controls and 5 (13.5%) were categorized as UTI.

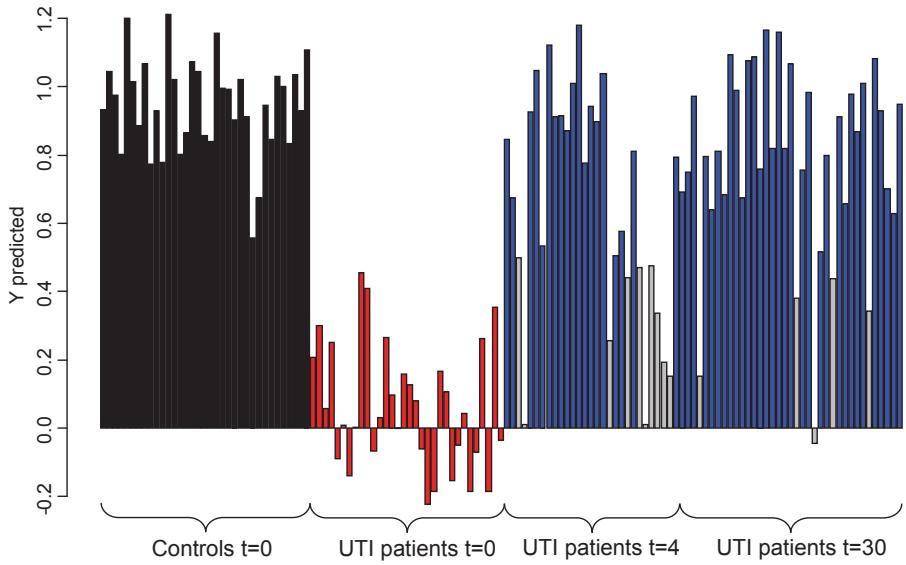


Figure 4. Predicted response value for two-class PLS-DA model based on controls (black bars) and UTI patients (red bars) at baseline: blue bars are the t=4 and t=30 classified as controls, grey are the t=4 and t=30 samples classified as UTI patients at t=0.

An important parameter characterizing UTI patients is the number of bacteria in urine; however, bacteria can also be present in urine of the individuals, who do not exhibit any symptoms of UTI [23]. We built a PLS regression model from NMR data of urine at baseline using the result of bacterial culture as response variable. Since bacterial count and UTI classification do not fully correlate we expected to obtain a slightly different model as compared to the model built based on UTI classification for this timepoint. Using 2 components a cumulative R2Y = 0.78 and Q2 = 0.44 were obtained and model validation showed intercepts of the R2Y and Q2Y regression lines with the vertical axis at 0.63 and -0.12, respectively, in the model validation plot. As can be seen from the PLS scores plot (Figure 5) the samples with the highest bacteria concentration in urine were very distinct from the rest forming a separate cluster, whereas the rest of the samples were overlapping. The spectral regions responsible for the correlation of the ^{1}H NMR data and bacterial count were chosen on the basis of the corresponding VIP. A list of those regions, along with the p-values derived from ANOVA (corrected for multiple testing), the direction of change and identities of the corresponding metabolites are summarized in Table 2.

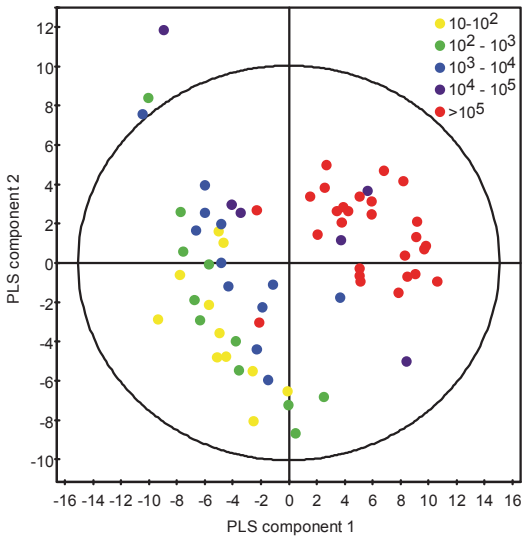


Figure 5. Scores plot of the PLS model of urine ^{1}H NMR spectra at baseline vs. the number of bacteria (CFU/mL) found in urine (R2Y = 0.78, Q2 = 0.44). Colored by the number of bacteria.

To better understand the process of patient recovery and to find the spectroscopic regions that correlate with this process, we took advantage of the longitudinal study design. One of the statistical methods suitable for such analysis is multilevel component analysis (MCA) that separates variation present in the data into two levels: between-individual and within-individual. We performed this analysis on the 29 patients for which both the data from the baseline and from the 30-days time point were available and concentrated on the within-individual information. This should best reflect the recovery from the baseline, when patients are diagnosed as infected, to 30 days, when they are considered UTI symptom-free. PCA scores plot of the first two principle components that cover 15.8 and 14.8% of the variation, respectively, showed good separation between baseline and t=30 time points (data not shown). The PLS-DA model of this data had high quality parameters (R2Y = 0.98, Q2 = 0.96 for four components), showed stability in permutation test (intercepts of the R2Y and Q2 regression lines with the vertical axis were at 0.42 and -0.6, respectively) and perfectly separated the two time points (data not shown). The NMR spectral regions responsible for the separation between baseline and the t=30 time point were identified based on VIP values. The underlying metabolites as well as the p-values from paired t-test (corrected for multiple testing) and the direction of change are summarized in Table 2.

DISCUSSIONS

UTI represents a complex clinical entity, for which diagnostics is not straightforward: there is no single test or criteria for it [7]. In the current paper we approach identification of metabolites that characterize UTI and its pathology with the use of ^{1}H NMR. We demonstrate how the use of clinical data and multiple samples per individual can enrich the biological interpretation of the findings. To reduce the heterogeneity typically posed by

UTI research, as a first attempt the smaller selection of UTI subjects from a bigger cohort was used, with similar diagnosis and with the major pathogen being *E. coli*. A set of matched controls was also available.

Unlike in animal experiments, in clinical research assigning people to certain groups is not always unconditional. The diagnosis of a disease can be fuzzy and defining the 'healthy' group is even more difficult, as there is hardly a definition of healthy. Thus, it may be very advantageous to supplement a traditional 'case-control' design with a more complex study design and the use of additional clinical data. When used without extra information, 'case-control' analysis might even be misleading. For example, the separation of the control and UTI groups was seen in the first two principal components of PCA; however, this discrimination was not disease-related, but the result of patients taking the antipyretic and analgesic drug paracetamol. An analysis strategy for such type of data is to identify all of the spectroscopic regions that contain signals from drug-related compounds and to exclude them prior to further analysis. However, it is not feasible to account for the whole range of the medication used and, more importantly within the context of clinical metabolomics studies in general, to account for drug-related shifts in metabolism, especially in the case of long-term treatment regimes of chronic conditions. It is essential to consider such effects when developing the study design in order to minimize or control such influences.

Samples from 4 days after admission, when the patients were still under therapy, but on the way to recovery, were used to check if the modeled differences were related to the effect of medication or not. The fact that the majority of those samples were classified as healthy by the model built on baseline samples is an indication that the model is not reflecting therapy/drug intake, but is indeed related to the clinical difference between the groups.

The samples from the 30-days time point, when UTI patients were symptom-free, could also be used to gain additional information on the performance of the model as well as to get insight into the underlying biology. When predicted using the PLS-DA model built on the baseline UTI infected and UTI symptom-free samples, most of the 30-days samples (86.5%) were projected to the control group. Those few, which were still predicted as infected UTI patients, may have another condition (as we do not know at this point how specific our model is) or have asymptomatic UTI. On the other hand, they can be healthy and be false positives, as the predictive ability of our model, estimated by cross-validation was 63%. Despite that, considering the prediction of 30-days samples as an independent statistical test for our model, it gives very satisfactory results.

Pair-wise analysis for baseline and 30-days samples from the same individuals was conducted in order to monitor the recovery process. It revealed a number of classifiers and improved their statistical significance. The identified metabolites overlapped with the compounds from the model discriminating healthy and UTI subjects, however a few of them were unique (para-aminohippuric acid, scyllo-inositol and a few unidentified compounds).

Besides the multilevel design, the advantage of the current study was the exhaustive clinical characterization of the patients. Among the variety of clinical parameters available, the number of bacteria in urine was of specific importance. We performed regression-based analysis of the relation between the ^{1}H NMR data and the bacterial load in urine as determined by bacterial culture. The classifiers that emerged from this analysis were to a certain extent overlapping with the classifiers derived from the discriminative model on baseline samples. This was no surprise, since UTI is generally characterized by the presence of bacteria in urine.

When comparing the lists of discriminators obtained from the different models (discriminating UTI patients from controls, modeling the recovery process and modeling the data against the degree of bacterial contamination of urine) it is evident that there is a large overlap which makes biological interpretation of the results feasible. For instance, some of the overlapping metabolites were already known from the literature to be related to the bacterial contamination of urine: acetate, lactate and trimethylamine [9]. Others, if they were found only in the comparative analysis of the two groups, could be attributed based on previous studies to certain phenomena. Hippuric acid, for example, is often associated with the gut microflora [24] and taurine with liver toxicity [25]. However, our findings suggest that they are also associated with the bacterial contamination of urine, which obviously does not mean that they are not related to the mentioned physiological processes as well, but that a complex network of interconnected factors is involved. The metabolites that appear to be related to the recovery process might be considered as potential morbidity markers. One of them, para-aminohippuric acid, is a well-established diagnostic marker for renal plasma flow and glomerular filtration [26]. The recovery from the complicated, tissue-invasive UTI is associated with the resumption of the kidneys' function, so the positive change in para-aminohippuric acid corroborates our assumption that some of the markers discovered in the paired analysis are the markers of morbidity.

CONCLUSION

In the current paper we used a metabolomics approach to profile UTI, which is on the one hand one of the most common infectious diseases among the adults, and on the other hand a disease that still lacks markers of morbidity. Using ^{1}H NMR profiles of urine we generated various statistical models: (a) discriminating UTI patients and control subjects, (b) following the recovery process of UTI patients and (c) associating urine metabolic content with bacterial contamination. The discriminative model was able to classify most of the independent samples correctly according to their diagnosis, which indicates its high predictive ability. Comparing the sets of molecules derived from different analyses, we concluded that some of the compounds (e.g. trimethylamine and acetate) can be attributed to the effect of bacterial contamination of urine; others (e.g. para-aminohippuric acid, scyllo-inositol) can be considered markers of morbidity.

ACKNOWLEDGMENT

The authors would like to thank Sibel Go¨raler M.Sc. for the analytical work.

REFERENCES

[1] Woodcock J. The prospects for "personalized medicine" in drug development and drug therapy. Clinical Pharmacology & Therapeutics. 2007; 81:164-169.

[2] Lindon JC., Holmes E, Bollard ME, Stanley EG, Nicholson,J.K. Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. Biomarkers. 2004;9:1-31.

[3] Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. Cell. 2008 134:714-717.

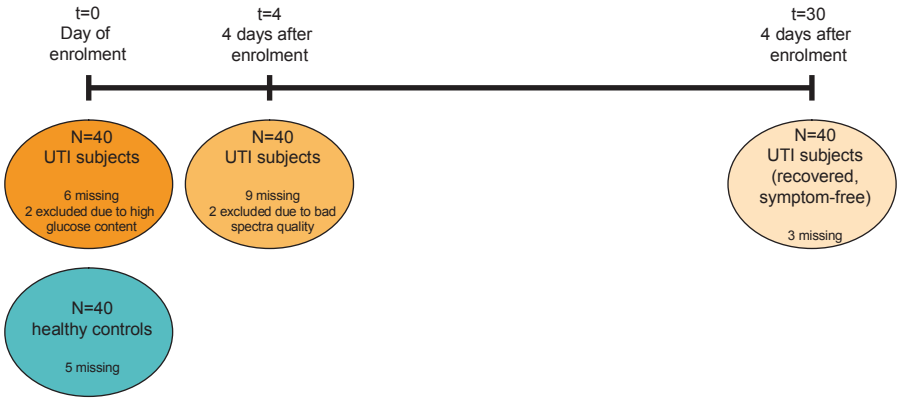
[4] Mendes P, Camacho D, de la Fuente A. Modelling and simulation for metabolomics data analysis. Biochem. Soc. Trans. 2005; 33:1427-1429.

[5] Nicholson JK, Holmes E, Wilson ID. Gut microorganisms, mammalian metabolism and personalized health care. Nat. Rev. Microbiol. 2005; 3:431-438.

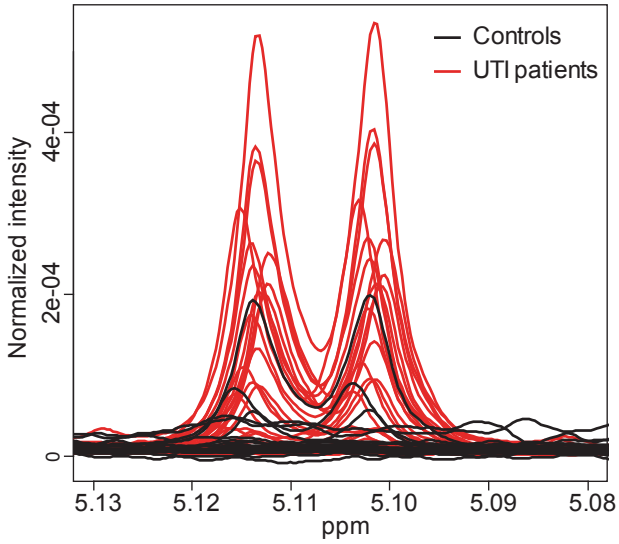
[6] Adourian A, Jennings E, Balasubramanian R, Hines WM, Damian D, Plasterer TN, Clish CB, Stroobant P, McBurney R, Verheij ER et al. Correlation network analysis for data integration and biomarker selection. Molecular

Biosystems. 2008; 4:249-259.
[7] Wilson ML, Gaido L. Laboratory diagnosis of urinary tract infections in adult patients. Clinical Infectious Diseases. 2004; 38:1150-1158.
[8] Johnson JR. Laboratory diagnosis of urinary tract infections in adult patients. Clinical Infectious Diseases. 2004; 39:873.
[9] Gupta A, Dwivedi M, Mahdi AA, Gowda GA, Khetrapal CL, Bhandari M. 1H-nuclear magnetic resonance spectroscopy for identifying and quantifying common uropathogens: a metabolic approach to the urinary tract infection. BJU. Int. 2009;104:236-244.
[10] Gupta A, Dwivedi M, Nagana Gowda GA, Ayyagari A, Mahdi AA, Bhandari M, Khetrapal CL. (1)H NMR spectroscopy in the diagnosis of Pseudomonas aeruginosa-induced urinary tract infection. NMR Biomed. 2005; 18:293-299.
[11] Gupta A, Dwivedi M, Gowda GA, Mahdi AA, Jain A, Ayyagari A, Roy R, Bhandari M, Khetrapal CL. 1H NMR spectroscopy in the diagnosis of Klebsiella pneumoniae-induced urinary tract infection. NMR Biomed. 2006;19:1055-1061.
[12] Kumar A, Ernst RR, Wuthrich K. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. Biochem. Biophys. Res. Commun. 1980; 95:1-6.
[13] Price WS. Water signal suppression in NMR spectroscopy. Annual Reports on Nmr Spectroscopy, 1999; 38:289-354.
[14] Coron A, Vanhamme L, Antoine JP, Van HP, Van HS. The filtering approach to solvent peak suppression in MRS: a critical review. J. Magn Reson. 2001; 152:26-40.
[15] Vanzijl PCM, Sukumar S, Johnson M O, Webb P, Hurd RE. Optimized Shimming for High-Resolution Nmr Using 3-Dimensional Image-Based Field-Mapping. Journal of Magnetic Resonance. 1994; Series A 111:203-207.
[16] Wu PS, Otting G. Rapid pulse length determination in high-resolution NMR. J. Magn Reson. 2005; 176:115-119.
[17] Bales JR, Nicholson JK, Sadler PJ. Two-dimensional proton nuclear magnetic resonance "maps" of acetaminophen metabolites in human urine. Clin. Chem. 1985; 31:757-762.
[18] Nocairi H, Qannari EM, Vigneau E, Bertrand D. Discrimination on latent components with respect to patterns. Application to multicollinear data. Computational Statistics & Data Analysis. 2005; 48:139-147.
[19] Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, van Velzen EJJ, van Duijnhoven JPM, van Dorsten FA. Assessment of PLSDA cross validation. Metabolomics. 2008; 4:81-89.
[20] Lindgren F, Hansen B, Karcher,W, Sjostrom M, Eriksson L. Model validation by permutation tests: Applications to variable selection. Journal of Chemometrics. 1996 10:521-532.
[21] Jansen JJ, Hoefsloot HCJ, van der Greef J, Timmerman ME, Smilde AK. Multilevel component analysis of time-resolved metabolic fingerprinting data. Analytica Chimica Acta. 2005 530:173-183.
[22] Cloarec O, Dumas ME, Craig A, Barton RH, Trygg J, Hudson J, Blancher C, Gauguier D, Lindon JC, Holmes E et al. Statistical total correlation spectroscopy: An exploratory approach for latent biomarker identification from metabolic H-1 NMR data sets. Analytical Chemistry. 2005; 77:1282-1289.
[23] Lin K, Fajardo K. Screening for asymptomatic bacteriuria in adults: evidence for the U.S. Preventive Services Task Force reaffirmation recommendation statement. Ann. Intern. Med. 2008; 149:W20-W24.
[24] Swann J, Wang Y, Abecia L, Costabile A, Tuohy K, Gibson G, Roberts D, Sidaway J, Jones H, Wilson ID et al. Gut microbiome modulates the toxicity of hydrazine: a metabonomic study. Mol. Biosyst. 2009; 5:351-355.
[25] Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica. 1999; 29:1181-1189.
[26] REUBI FC. Glomerular filtration rate, renal blood flow and blood viscosity during and after diabetic coma. Circ. Res. 1953; 1:410-413.

SUPPORTING INFORMATION



Supplementary Figure 1. Design of the study.



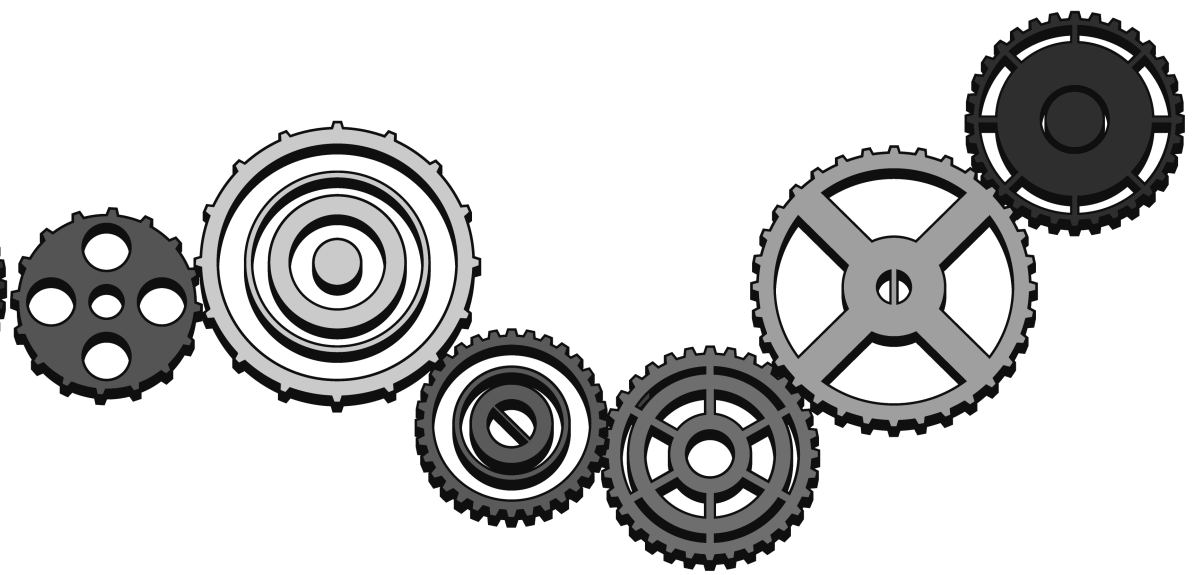
Supplementary Figure 2. Paracetamol-glucuronide doublet in baseline samples colored by controls (black) and UTI patients (red).

Chapter 3

Fibrinogen alpha chain *O*-glycopeptides as possible markers of urinary tract infection

Tiziana Pacchiarotta, Paul J. Hensbergen, Manfred Wuhrer, Cees van Nieuwkoop, Ekaterina Nevedomskaya, Rico J. Derks, Bart Schoenmaker, Carolien A. M. Koeleman, Jaap van Dissel, André M. Deelder, Oleg A. Mayboroda

Journal of Proteomics 2012, 75(3) 1067–1073



ABSTRACT

Urinary tract infection (UTI) is the most common bacterial infection leading to substantial morbidity and considerable health care expenditures across all ages. Here we present an exploratory UPLC–MS study of human urine in the context of febrile, complicated urinary tract infection aimed to reveal and identify possible markers of a host response on infection. A UPLC–MS based workflow, taking advantage of Ultra High Resolution (UHR) Qq-ToF-MS, and multivariate data handling were applied to a carefully selected group of 39 subjects with culture-confirmed febrile Escherichia coli UTI. Using a combination of unsupervised and supervised multivariate modeling we have pinpointed a number of peptides specific for UTI. An unequivocal structural identification of these peptides, as O-glycosylated fragments of the human fibrinogen alpha 1 chain, required MS^2 and MS^3 experiments on two different MS platforms: ESI-UHR-Qq-ToF and ESI-ion trap, a blast search and, finally, confirmation was achieved by matching experimental tandem mass spectra with those of custom synthesized candidate-peptides.

In conclusion, exploiting non-targeted UPLC-MS based approach for the investigation of UTI related changes in urine, we have identified and structurally characterized unique *O*-glycopeptides, which are, to our knowledge, the first demonstration of *O*-glycosylation of human fibrinogen alpha 1-chain.

INTRODUCTION

Urinary tract infection (UTI) is the most common bacterial infection leading to substantial morbidity and considerable health care expenditures across all ages [1]. The term 'urinary tract infection' encompasses a variety of clinical syndromes with the common denominator of a positive urine culture (i.e., significant bacteriuria ≥ 10^3 CFU/ml) [2]. Pathogens most commonly associated with UTI are *Enterobacteriacaea* such as *Escherichia coli*, *Klebsiella pneumoniae*, *Proteus sp.* and *Staphylococcus saprophyticus*. In a first approximation, UTI can be grouped into acute uncomplicated UTI (non-pregnant young women with cystitis but no evidence of urological abnormalities) and complicated UTI (UTI in males, all febrile UTI syndromes and those with urological abnormalities) [3, 4]. In this respect, fever reflects the presence of a tissue invasive disease such as pyelonephritis, prostatitis or the urosepsis syndrome. A clinical management of UTI is rather straightforward and implementation of modern analytical technologies such as mass spectrometry for diagnostic purposes will have little impact on the clinical decision making. However, mass spectrometry might provide new, highly needed input in scaling the morbidity inflicted to, for example, prostate or kidney or obtaining guidance for duration of the treatment.

So far, the most successful application of mass spectrometry in the context of UTI was the identification of a pathogen. Ferreira *et al* [5] have shown that matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-ToF MS) in combination with pattern recognition algorithms can provide a correct pathogen identity in more than 90% of cases with bacteriuria cut-off at 10^5 CFU/ml. In general, MALDI-ToF MS has quickly evolved as a method of pathogen identification and took its place in microbiological laboratories alongside with more classical techniques [6]. However, for an unbiased exploratory study, which is needed for a first evaluation of the host response to UTI, a hyphenated technique as, for example, liquid chromatography–mass spectrometry (LC–MS) is probably a better choice. Those techniques cover a broad range of chemical entities and most importantly a very special group of compounds such as small molecules and naturally occurring urinary peptides. The clinical relevance of the last ones was proven in a large scale study of Good *et al* [7]. Here, we present a first exploratory LC–MS study of human urine in the context of febrile, complicated urinary tract infection. Using a carefully matched selection of patients with infection due to *E. coli*, we have evaluated the applicability of a reverse phase (RP)UPLC-ToF-MS approach for UTI in a clinical study and identified a glycopeptide representing a unique proteolytic fragment of the fibrinogen alpha 1 chain as a possible marker of active UTI.

EXPERIMENTAL SESSION

Chemicals

Ethanol, methanol (HPLC grade) and formic acid were purchased from Sigma-Aldrich (Zwijndrecht, The Netherlands). Sodium hydroxide was purchased from Fluka (Zwijndrecht, The Netherlands). Peptides were synthesized at the LUMC-facility as described previously [8].

Clinical samples

Urine samples were collected in a prospective observational multicenter cohort study

of eight emergency departments (ED) of seven hospitals and 35 affiliating primary health care providers. From January 2004 through November 2008, patients who presented with a diagnosis of febrile UTI, were considered for enrolment in the study. The study was approved by the local ethics committees and all included patients gave written informed consent. Inclusion criteria and exclusion criteria have been described in detail elsewhere [9,10]; in short, patients of 18 years or older, with fever and presenting at least one symptom of UTI (dysuria, frequency, of urination, perineal pain, flank pain or costovertebral tenderness) and a positive nitrite dipstick test or a positive leukocyte esterase dipstick test were enrolled. Exclusion criteria were current treatment for urolithiasis or hydronephrosis, pregnancy, hemo- or peritoneal dialysis, a history of kidney transplantation or known presence of polycystic kidney disease. Details with respect to empiric therapy and follow-up are provided elsewhere [11]. Urine samples were collected at enrolment as baseline samples (t=0). After three (t=3) and thirty days (t=30) urine samples of the same patients were collected for follow-up. Clean midstream-catch urines were collected and stored, within 2 h after the collection, at -80° C until further analysis. Bacterial cultures were obtained at all time points and were analyzed using standard microbiological methods. A positive urine culture was defined as bacterial growth of over 10^3 colony-forming units per ml^2.

For the current study, from a database of about 700 subjects enrolled, a group of 39 subjects with culture-confirmed *E. coli* in the urine was selected. Samples from volunteers without UTI or symptoms of infection were enrolled as negative control (a negative bacterial culture; no evidence of inflammatory disease) and in this study used as references. Baseline characteristics of the samples are presented in Table 1. The study design includes the control samples and samples of UTI patients collected at baseline (t=0) as well as patient samples collected after the antibiotic treatment (t=30).

Table 1. Baseline characteristics. Data are presented as n (%) unless otherwise stated. IQR: interquartile range.

Characteristics	Cases n = 39	Controls n = 39	*p*
Age, years, median [IQR]	59 [42-70]	58 [49-67]	0.900
Male sex	18 (46)	18 (46)	1.000
Smoking	5 (13)	5 (13)	1.000
Co-morbidity			
Urinary tract disorder	4 (10)	4 (10)	1.000
Malignancy	4 (10)	1 (3)	0.166
Heart failure	5 (13)	3 (8)	0.456
Renal insufficiency	1 (3)	0 (0)	0.134
Diabetes mellitus	6 (15)	2 (5)	0.136
Immunocompromised	1 (3)	1 (3)	1.000
Urine dipstick results			
Nitrate	26/36 (72)*	0/36 (0)*	< 0.001
Leucocyte esterase	35/36 (97)*	5/36 (14)*	< 0.001

*: 3 missing values

Sample processing

Samples were prepared for non-targeted analysis of small molecules and natively occurring urinary peptides. Protein precipitation was carried out by adding 75 µl of cold ethanol to 25 µl of urine followed by 20 min incubation on ice. Subsequently, the samples

were centrifuged at 3660 g for 10 min and the supernatant was collected and evaporated until dryness overnight. The samples were then reconstituted with 100 µl of water and 10 µl was used for injection. The injection scheme was randomized and it included quality control samples (fresh voided urine samples prepared in the same way as the clinical samples) to ensure the robustness of the workflow and to evaluate the analytical variability. The total number of 117 clinical samples was organized in 4 sequences (4 injection days); two quality control samples were injected at the beginning, at the end and after every six biological samples. In total, 165 samples (48 quality controls plus 117 clinical samples) were analyzed by UPLC-ESI-Qq-TOF MS (Ultra High Resolution ToF, maXis, Bruker Daltonics, Bremen, Germany). The UPLC (Ultimate 3000 RS tandem LC system, Dionex, Amsterdam, The Netherlands) was equipped with a pre-column (Acclaim 120 C18, 5 µm, 120 Å, and 2.1 x 10 mm,) and two analytical columns (Acclaim RSLC 120 C18, 2.2 µm, 120 Å, and 2.1 x 100 mm) working alternatively to speed up the acquisition series. The UPLC flow was set at 400 µl/min and the mobile phases were water + 0.1 % formic acid v/v (Phase A) and methanol + 0.1% formic acid v/v (Phase B). The gradient was as follows: 1 min 0 % phase B, then in 1 min to 10% phase B, held for 1 min at 10% phase B, and subsequently in 6.5 min to 100% phase B and held for 3 min at 100% phase B. Before each chromatographic run, a calibrant solution of sodium formate was injected in flow injection mode.

The mass spectrometer was operated in positive ionization mode and acquired data in the mass range from *m/z* 50 to 1000 with a spectra rate of 1 Hz. The capillary was set at 4500 V, the end plate offset at -500 V, the nebulizer gas at 2 bars and the dry gas at 8 L/min at 200°C.

For compound identifications MS^2 and MS^3 experiments were performed using ESI-ultra high resolution UHR-Qq-ToF and ESI-ion trap MS (HCTultra, Bruker Daltonics). The data were acquired in a wider scan range from *m/z* 50 to 2000 in auto MS/MS mode and by using an inclusion list of precursor ions of interest (the ones relevant after statistical data analysis). Collision energies were the following: *m/z* 500 (singly charged) 50 V, *m/z* 500 (doubly charged) 30 V and *m/z* 1500 (singly charged) 35 V, and *m/z* 500 (doubly charged) 25 V. All the *m/z* values within this interval were fragmented with interpolated values of collision energy.

MS^3 experiments were performed using a reverse phase LC-ESI-ion trap MS system as described previously [12] with slight modifications. For electrospray (1100–1250 V), stainless steel capillaries with an inner diameter of 30 mm (from Proxeon, Odense, Denmark) were used. The solvent was evaporated at 170°C employing a nitrogen stream of 7 L/min and ions from *m/z* 500 to 1800 were registered. Automatic fragment ion analysis was enabled resulting in MS/MS spectra of the most abundant peaks. The two most intense ions of each MS/MS spectrum were subjected to an additional ion isolation/fragmentation cycle, resulting in MS^3 fragment ion spectra.

Data analysis

The LC–MS data files were exported as mzXML files and aligned by using the in-house developed alignment algorithm msalign2 tool [13] (http://www.ms-utils.org/msalign2/); peak picking was performed using XCMS package (The Scripps Research Institute, La Jolla,USA) using the default settings except for bandwidth (bw parameter) for grouping of features set to 10 [14].

The generated data matrix was imported to the SIMCA-P 12.0 software package (Umetrics, Umeå, Sweden). The data were mean centered and unit variance-scaled prior to statistical analysis. The validity and the degree of overfitting of the PLS-DA models were checked using a 200 permutations test.

To identify metabolites of interest, rational chemical formulas were generated based on internally calibrated monoisotopic masses within 10 mDa mass error, using the SmartFormula tool within the DataAnalysis software package (Bruker Daltonics).

RESULTS

Considering the fact that the presented study is the first exploratory LC–MS study of human urine in the context of febrile urinary tract infection, a general overview of the analytical consistency of the method was the starting point of the data analysis. To evaluate the effect of analytical variability, we performed Principal Component Analysis (PCA) using the whole data set of clinical and quality control samples (165 samples). Figure 1 shows the PCA score plot for the first two principal components. No analytical trends, such as day of injection, injection sequence or influence of the column, were represented in the first three principal components, which altogether cover 46% of the total variation. Thus, the analytical variability within our data set had negligible influence on the data matrix.

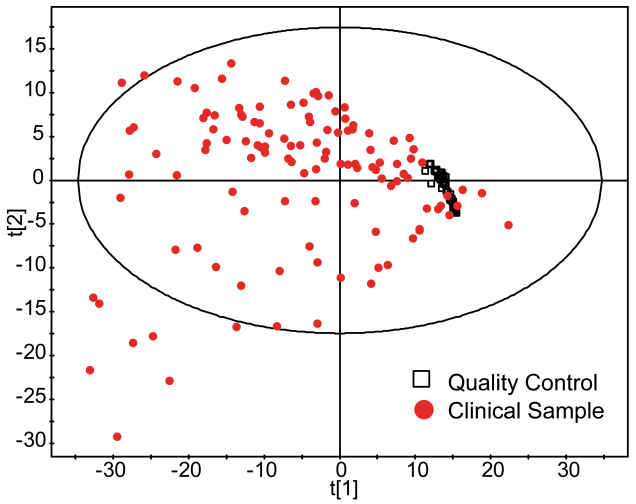


Figure 1. Scores plot of the PCA analysis of the entire data set (including quality control and clinical samples).

Subsequently, we removed the quality control samples from the model and we focused the data analysis on clinical samples. We observed a clear separation of UTI patients and UTI free subjects (Controls and t=30 days) in the PCA model (Supplementary Figure 1A). A two-class PLS-DA model built on baseline samples (controls and patients at t=0 days) had satisfactory values of goodness of fit and prediction ability (Supplementary Figure 1 B). The analysis of Variables Importance on the Projection values (VIP) higher than 1.5 revealed that the separation was related to the presence of acetaminophen metabolites in the patients

affected by febrile UTI. Taking advantage of the intrinsic properties of the UHR-Qq-ToF instrument such as the high mass accuracy and the high-confidence resolution of isotopic distribution we could identify glucuronide ($C_{14}H_{17}NO_8$), acetaminophen mercapturate ($C_{11}H_{14}N_2O_4S$) and N-acetylacetaminophen mercapturate ($C_{13}H_{16}N_2O_5S$) (Supplementary Figure 1C, Supplementary Table1). These data were confirmed by our internal database in which the use of fever reducers was reported for those subjects.

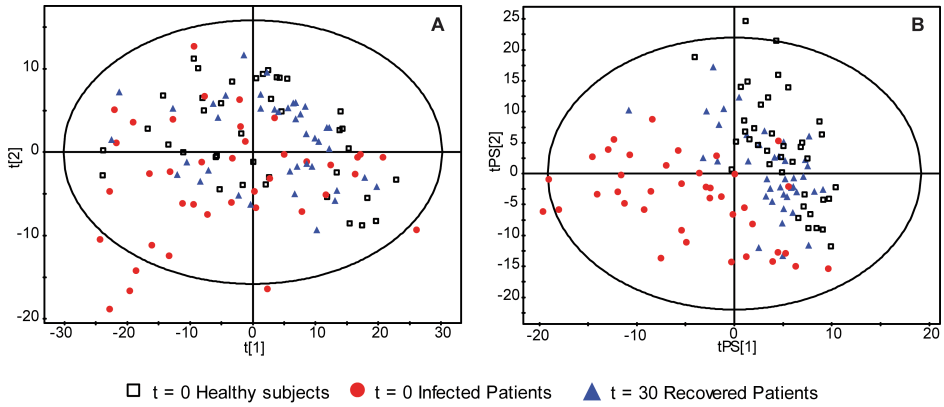


Figure 2. Statistical models of the LC-Qq-ToF-MS data. A) PCA and B) PLS-DA with Q2 (cum)=0.520 and R2Y (cum)=0.647. Samples are coloured according to the disease status.

It is evident that the metabolites of acetaminophen introduce a bias in the data concealing the effects relevant for the study. Therefore, we removed all variables related to acetaminophen metabolites from our dataset and re-analyzed the data. The PCA model built on this neatened data matrix did not show such a clear separation between infected and uninfected patients as in the previous model (Figure 2A). However, a visual trend for such a separation was evident. The goodness of fit and prediction values for the new two-class PLS-DA model built on baseline samples (t=0 days and controls) were lower (Figure 2B), but for human studies still within the statistically acceptable limits. The 30 days after enrolment samples were used to test the predictive power of a model: 34 of 39 samples were predicted as recovered patients, clustering with the group of not infected subjects. An overview of the VIP values > 1.5 revealed 40 features. After closer examination, half of them could be organized in three groups according to their retention times (Table 2).

Table 2 List of major classifiers expressed as *m/z* with their retention times and VIP>1.5 in bracket.

Group	Retention time (s)	*m/z* (VIP Value)
1	58.3	747.274(1.79) 675.239(1.81) 676.248(1.81) 672.174(2.21) 657.232(1.55) 292.103(1.67)
2	382.4	875.361 (1.78) 876.365(1.78) 877.37(1.77) 584.266(2.12) 375.211(1.74) 366.139(2.07) 292.103(2.11) 274.093(2.12)
3	410.3	816.86(2.16) 817.865(1.69)

Within the MS spectra of groups 1 and 2, we observed ions corresponding to some putative in source fragmentation of glycan moieties. Classifier 1 appeared to be a free sugar moiety composed of a sialic acid, one hexose and one N-acetylhexosamine (Figure 3A). Classifier 2 (Figure 3B) contains a similar glycan moiety most probably on a peptide backbone of 218.127 Da which can represent either the dipeptide threonine–valine or the dipeptide serine–(iso)leucine.

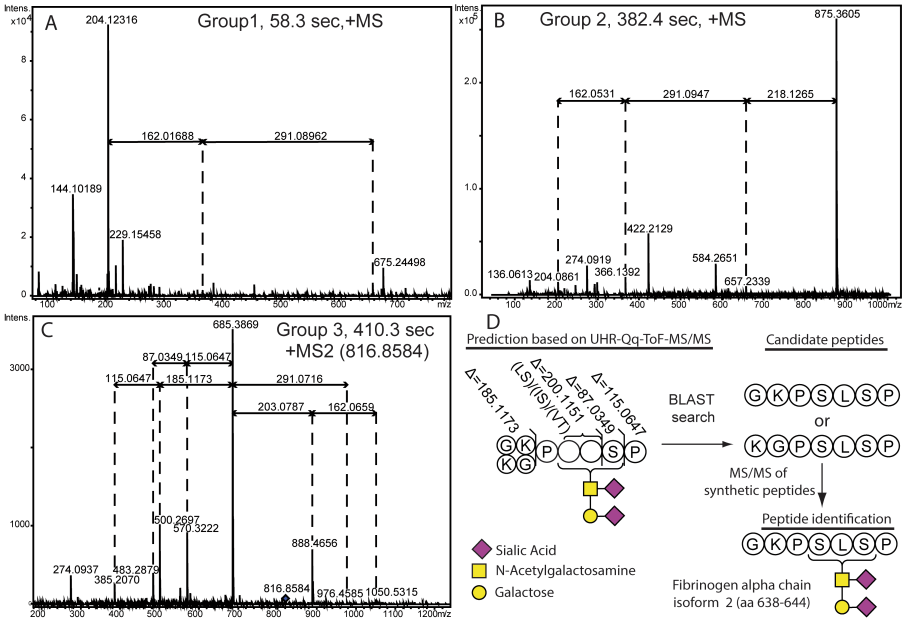


Figure 3. Mass spectrometric characterization of the major classifiers. MS spectra of the classifiers from group 1 (**A**) and 2 (**B**). MS^2 spectrum of *m/z* 816.86 belonging to group 3 (**C**) with the work flow for its identification (**D**) combining information from the fragmentation spectrum, BLAST searches and data from synthetic peptides.

To identify classifier 3, we performed MS/MS experiments using two different mass analyzers: UHR-Qq-ToF and ion trap to take advantage of the characteristics of each respective analyzer. We started the structural assignment form the largest identified precursor. MS/MS data were acquired selecting as precursor the double charged ion at *m/z* 816.86 (Table 2, Group 3) on the UHR-Qq-ToF (Figure 3C). The identification workflow is shown in Figure 3D. The fragmentation pattern revealed characteristics of a glycopeptide based on the sequential losses of two sialic acids (NeuAc; *m/z* 1050.5315), one hexose (Hex; *m/z* 888.4656) and one *N*-acetylhexosamine (HexNAc; *m/z* 685.3869). According to the literature [15-18] this most probably reflects a typical core-1 type O-glycan structure composed of $Gal_1GalNAc_1NeuAc_2$. The peptide backbone was represented by the ion at *m/z* 685.3869. Apparently, some peptide backbone cleavages also occurred, resulting in the ions at *m/z* 570.3222 and 483.2897. Using the mass accuracy of the UHR-Qq-ToF-analyzer, we assigned these ions as b-type ions generated after the sequential loss of a terminal proline (exact mass difference 115.0633 Da) and a serine, and not as y-type ions resulting from a

terminal aspartic acid (exact mass difference 115.0269 Da) and a serine. The loss of proline was also observed after the initial loss of 185.1173, resulting in an internal fragment ion. In this case, the accurate mass points towards the loss of the dipeptide glycine–lysine (GK/KG, exact mass 185.1164). To further corroborate that the above mentioned ions were indeed the result of peptide backbone fragmentation, an MS^3 spectrum from the ion at *m/z* 685.4 was recorded using ESI-ion trap MS (Supplementary Figure 2A). Assuming that the loss of the glycine–lysine dipeptide was the result of a cleavage N-terminal of a proline, we predicted that the peptide backbone should have the sequence (GK/KG)P(200.1151 Da)SP. The total mass of the two remaining amino acids (200.1151 Dalton) matches with a dipeptide consisting of (iso)leucine–serine or threonine–valine (which have the same elemental composition with an exact mass of 200.1161 Dalton). Therefore, we performed BLAST searches against the human NCBI database, taking all the above possibilities into account. This resulted in two positive identifications; one (GKPSLSP) corresponding to a peptide from the C-terminus of fibrinogen alpha 1chain (GI 11761629) and one (KGPSLSP) to an unnamed protein product (GI 22760358). No positive matches were found when identical BLAST searches were performed against the *E. coli* database. To discriminate between the two above mentioned possibilities we synthesized both peptides, performed MS/MS experiments on the ion trap and compared the fragmentation patterns with the MS^3 spectrum that we obtained from the urinary glycopeptide (Supplementary Figure 2B and C). This unambiguously showed that the glycopeptide we have identified corresponded to the C-terminal peptide from fibrinogen alpha 1 chain (GKPSLSP). At this stage we can relate group 2 (see above) as putative derivative of this same peptide containing the dipeptide SL/LS carrying the glycan moiety. Finally, using integrated peak areas for the ion at *m/z* 816.86 we constructed a box plot, which clearly demonstrated the differences between the patients with an active UTI status, the recovered group (t=30) and the control group (Figure 4).

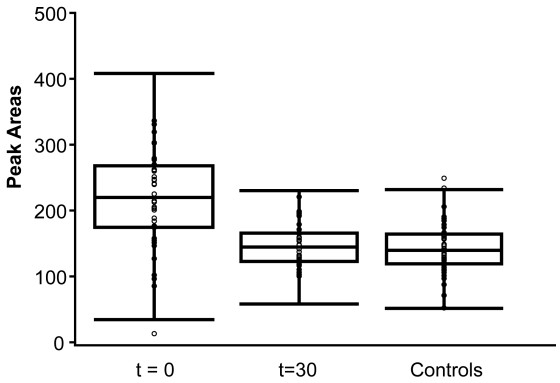


Figure 4. Box plot displaying the differences of peak area related to *m/z* 816.86 in the model classes.

DISCUSSION

Here we present the first exploratory UPLC-MS based study of UTI. As clinical entity UTI includes a number of syndromes in which the only common feature is a positive urine culture (bacteriuria). Several pathogens are most commonly associated with UTI:

Enterobacteriacaea such as *E. coli, K. pneumoniae, Proteus sp.* and *S. saprophyticus*. Thus, to minimize sample variability we have selected for the current study only the patients with culture-confirmed febrile *E. coli* UTI (Table 1).

Using a combination of unsupervised and supervised multivariate modeling we have selected forty most significant variables according to their VIP values. Approximately half of them could be grouped into three chromatographic peaks. A close examination of the spectra revealed the possible structural similarity of these three peaks with the peak at 410 s containing the most complete structure. The identification of this peak (*m/z* 816.86) as a product of proteolytic degradation of one of the most abundant serum proteins (human fibrinogen alpha chain isoform 2) may not seem encouraging at the first glance. The increased release of fibrinogen fragments in urine as a response to an infection or infection related kidney damage has little novelty. It was reported for the first time in the seventies and confirmed many times afterwards [7, 19-22]. Yet, the fact that this C-terminal peptide of fibrinogen alpha chain is carrying an O-glycan turns it into a rather unusual finding. Human fibrinogen alpha 1-chain was considered to be free of glycosylation, and only the alpha 2 isoform which is a splice variant with an elongated C-terminus has been shown to be *N*-glycosylated upon expression in monkey cells (COS cells) [23]. The proposed glycan structure corresponds to a T-antigen — core 1 type disaccharide Galβ1-3GalNAcα1 — which may carry 0, 1, or 2 sialic acid moieties. Such (partially) sialylated T-antigen structures are commonly found on various glycoproteins such as, for example, apolipoproteins [15-18]. There is only a single report by L'Hôte *et al*. which indicated a similar modification on porcine fibrinogen alpha chain on the basis of analysis with a combination of lectins [24]. However, neither the structure of the glycan nor the site of glycosylation was identified. In our case a combination of the data sets generated with the help of two mass analyzers and an additional confirmation using a synthetic peptide helped us to identify the exact sequence of the peptide backbone, although the question whether serine 641 or serine 643 carry the glycan moiety remains open.

Thus, our findings lead to a modification of the existing concept of fibrinogen structure. As an essential part of one of the most important physiological defense systems in the human organism, fibrinogen has been studied rather well. Nevertheless till now the alpha 1 chain was considered to be free of glycosylation. One might argue that the bulk of the work was done on the hepatic fibrinogen, which represents approximately 70% percent of total fibrinogen [25]. The peripheral tissues are capable to produce the 'localized' isoforms of fibrinogen as well; however, data on their structure and physiological significance remain scarce. An unusual *O*-glycosylation might be interpreted as an indication of the extra-hepatic origin of the fragments. However our data do not provide evidence for tracing of the physiological origin of this fragment and the question whether it is derived from hepatic or peripheral fibrinogen remains open.

It is well known that in healthy individuals the loss of fibrinogen through the coagulation cascade accounts only for 2% or 3% of total plasma fibrinogen [26]. Consequently, the peptides described in our study might well be filtrated from blood or proteolytically converted in the lower urinary tract or even be a result of urinary fibrinolytic activity. It has been shown that granulocytes possess a set of neutral proteases capable to use fibrinogen as a substrate [27]. In view of the involvement of granulocytes in antibacterial host defense the appearance of such proteolytic fragment in urine might then be considered as an indication of an active UTI. Regardless of the exact physiological interpretation, which at the moment

can only be speculative, a simple comparison of peak areas in each of three groups shows the difference between the group of patients with clear UTI symptoms (t=0) and the groups of the symptom free patients (controls and t=30).

In conclusion, using non-targeted exploratory UPLC–MS based approach for the investigation of UTI related changes in urine, we have identified and structurally characterized a unique C-terminal glycopeptide of the human fibrinogen alpha-chain, which to our knowledge is the first demonstration of glycosylation of alpha-chain. Of course, clinical significance of this finding should be evaluated on the larger cohort. Indeed, on the basis of the presented results we are now analyzing Leiden prospective UTI cohort of 700 samples.

REFERENCES

[1] Foxman B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. Am J Med. 2002;113 Suppl 1A:5S-13S.

[2] Wilson ML, Gaido L. Laboratory diagnosis of urinary tract infections in adult patients. Clin Infect Dis. 2004;38:1150-8.

[3] Lee JBL, Neild GH. Urinary tract infection. Medicine (Baltimore). 2007;35:423-8.

[4] Funfstuck R, Ott U, Naber KG. The interaction of urinary tract infection and renal insufficiency. Int J Antimicrob Agents. 2006;28 Suppl 1:S72-7.

[5] Ferreira L, Sanchez-Juanes F, Gonzalez-Avila M, Cembrero-Fucinos D, Herrero-Hernandez A, Gonzalez-Buitrago JM, et al. Direct identification of urinary tract pathogens from urine samples by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2010;48:2110-5.

[6] Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis. 2009;49:543-51.

[7] Good DM, Zurbig P, Argiles A, Bauer HW, Behrens G, Coon JJ, et al. Naturally occurring human urinary peptides for use in diagnosis of chronic kidney disease. Mol Cell Proteomics. 2010;9:2424-37.

[8] Hiemstra HS, Duinkerken G, Benckhuijsen WE, Amons R, de Vries RR, Roep BO, et al. The identification of CD4+ T cell epitopes with dedicated synthetic peptide libraries. Proceedings of the National Academy of Sciences of the United States of America. 1997;94:10313-8.

[9] van der Starre WE, van Nieuwkoop C, Paltansing S, van't Wout JW, Groeneveld GH, Becker MJ, et al. Risk factors for fluoroquinolone-resistant Escherichia coli in adults with community-onset febrile urinary tract infection. J Antimicrob Chemother. 2011;66:650-6.

[10] van Nieuwkoop C, Bonten TN, Wout JW, Becker MJ, Groeneveld GH, Jansen CL, et al. Risk factors for bacteremia with uropathogen not cultured from urine in adults with febrile urinary tract infection. Clin Infect Dis. 2010;50:e69-72.

[11] van Nieuwkoop C, Bonten TN, van't Wout JW, Kuijper EJ, Groeneveld GH, Becker MJ, et al. Procalcitonin reflects bacteremia and bacterial load in urosepsis syndrome: a prospective observational study. Crit Care. 2010;14:R206.

[12] Wang J, Balog CI, Stavenhagen K, Koeleman CA, Scherer HU, Selman MH, et al. Fc-glycosylation of IgG1 is modulated by B-cell stimuli. Mol Cell Proteomics. 2011;10:M110 004655.

[13] Nevedomskaya E, Derks R, Deelder AM, Mayboroda OA, Palmblad M. Alignment of capillary electrophoresis-mass spectrometry datasets using accurate mass information. Anal Bioanal Chem. 2009;395:2527-33.

[14] Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Analytical chemistry.

2006;78:779-87.

[15] Nilsson J, Ruetschi U, Halim A, Hesse C, Carlsohn E, Brinkmalm G, et al. Enrichment of glycopeptides for glycan structure and attachment site identification. Nat Methods. 2009;6:809-11.

[16] Wopereis S, Grunewald S, Morava E, Penzien JM, Briones P, Garcia-Silva MT, et al. Apolipoprotein C-III isofocusing in the diagnosis of genetic defects in O-glycan biosynthesis. Clinical chemistry. 2003;49:1839-45.

[17] Vaith P, Assmann G, Uhlenbruck G. Characterization of the oligosaccharide side chain of apolipoprotein C-III from human plasma very low density lipoproteins. Biochimica et biophysica acta. 1978;541:234-40.

[18] Gornik O, Lauc G. Glycosylation of serum proteins in inflammatory diseases. Disease markers. 2008;25:267-78.

[19] Fananapazir L, Eccleston M, Edmond E, Anderton JL. Raised urinary fibrin-degradation products, complement, and IgG during an influenza-like illness. Lancet. 1977;1:933-5.

[20] Kwaan HC, Barlow GH. Nature and biological activities of degradation products of fibrinogen and fibrin. Annu Rev Med. 1973;24:335-44.

[21] Blainey JD, Terry JM. Urinary fibrinogen degradation products and differential protein clearances in renal disease. Clin Nephrol. 1979;11:140-1.

[22] Metzger J, Kirsch T, Schiffer E, Ulger P, Mentes E, Brand K, et al. Urinary excretion of twenty peptides forms an early and accurate diagnostic pattern of acute kidney injury. Kidney Int. 2010;78:1252-62.

[23] Fu Y, Zhang JZ, Redman CM, Grieninger G. Formation of the human fibrinogen subclass fib420: disulfide bonds and glycosylation in its unique (alphaE chain) domains. Blood. 1998;92:3302-8.

[24] L'Hote C, Berger S, Karamanos Y. O-glycosylation of fibrinogen from different mammalian species as revealed by the binding of Escherichia coli biotinylated lectins. Thrombosis and haemostasis. 1996;76:710-4.

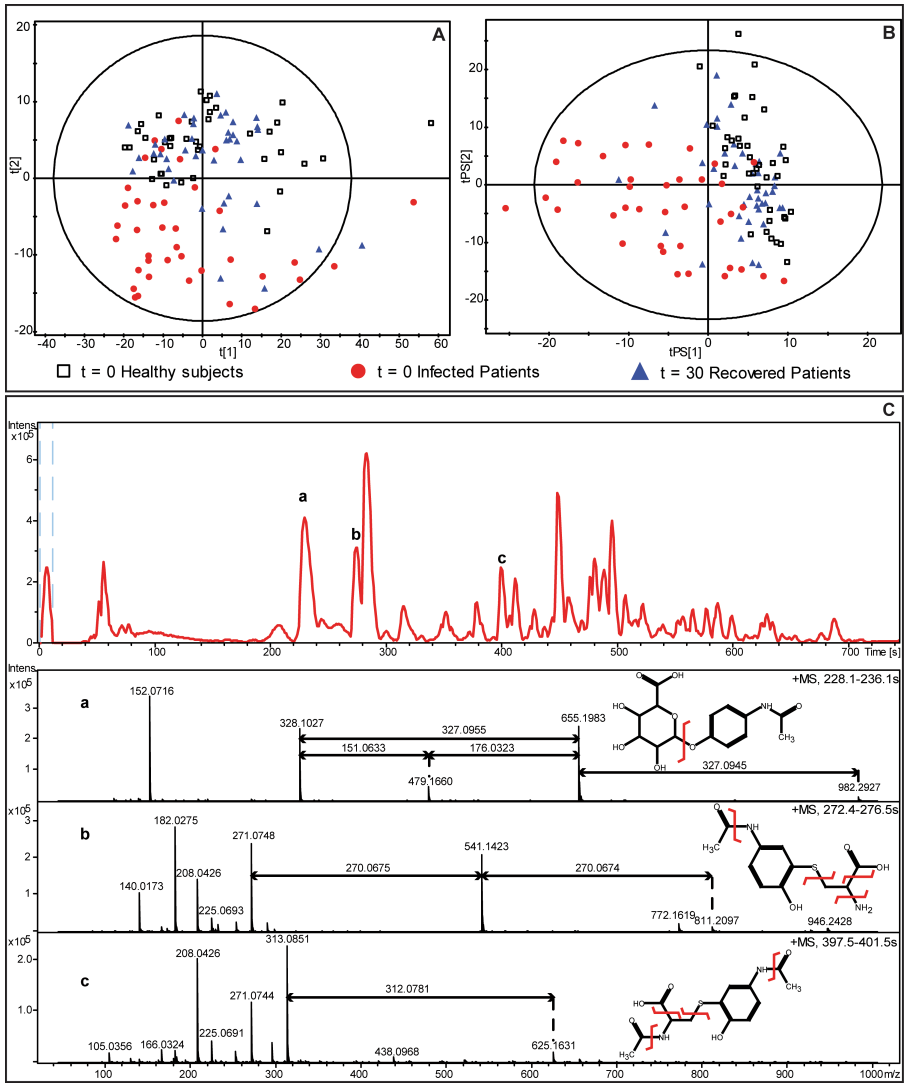
[25] Weisel JW. Fibrinogen and fibrin. Adv Protein Chem. 2005;70:247-99.

[26] Nossel HL. Radioimmunoassay of fibrinopeptides in relation to intravascular coagulation and thrombosis. N Engl J Med. 1976;295:428-32.

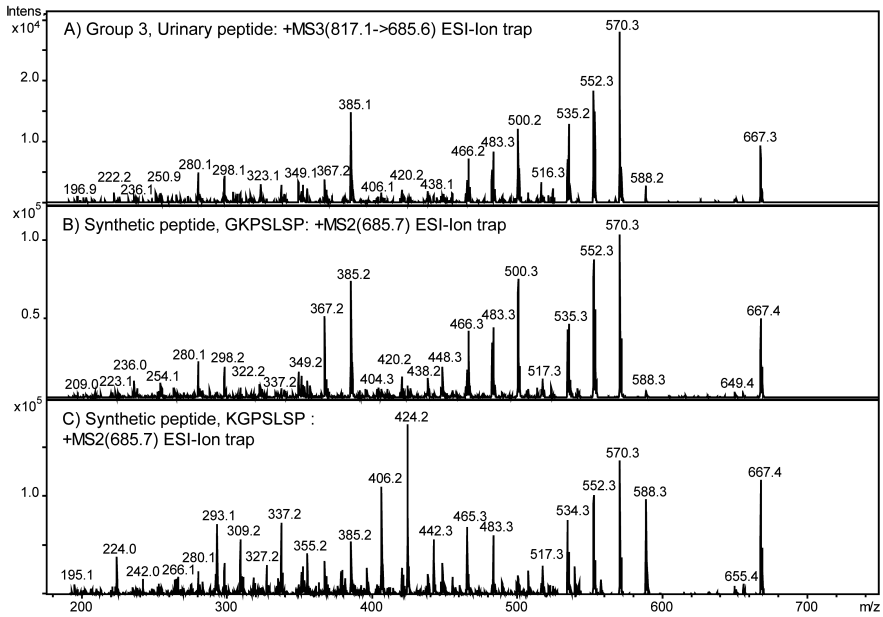
[27] Gramse M, Bingenheimer C, Schmidt W, Egbring R, Havemann K. Degradation products of fibrinogen by elastase-like neutral protease from human granulocytes. Characterization and effects on blood coagulation in vitro. The Journal of clinical investigation. 1978;61:1027-33.

Supplementary Table 1. Assignment of acetaminophen metabolites fragments in urine of UTI patients.

Retention time (s)	Compound		Experimental *m/z*	Theoretical *m/z*	Error (mDalton)
230.1	Acetaminophen glucuronide $C_{14}H_{16}NO_8$	$[3M+H]^+$	982.2927	982.2935	-0.8
		$[2M+H]^+$	655.1983	655.1981	+0.2
		$[2M+H\text{-Glucoronide}]^+$	479.1660	479.1660	0.0
		$[M+H]^+$	328.1027	328.1027	0.0
		$[M+H\text{-Glucuronide}]^+$	152.0716	152.0706	+1.0
374.5	Acetaminophen mercapturate $C_{11}H_{14}N_2O_4S$	$[3M+H]^+$	811.2097	811.2096	+0.1
		$[2M+H]^+$	541.1423	541.1413	+1.0
		$[M+H]^+$	271.0748	271.0747	+0.1
		$[M+H\text{-}OH]^+$	254.0457	254.0720	-2.63
		$[M+H\text{-}OH\text{-}CO]^+$	225.0693	225.0692	+0.1
		$[M+H\text{-}OH\text{-}CO\text{-}NH_3]^+$	208.0426	208.0427	-0.1
		$[M+H\text{-}OH\text{-}CO\text{-}NH_3\text{-}C_2H_2]^+$	182.0275	182.0270	+0.5
		$[M+H\text{-}OH\text{-}CO\text{-}NH_3\text{-}C_2H_2\text{-}C_2H_2O]^+$	140.0173	140.0165	+0.8
399.5	N-acetylacetaminophen mercapturate $C_{13}H_{16}N_2O_5S$	$[2M+H]^+$			
		$[M+H]^+$	625.1631	625.1633	-0.2
		$[M+H\text{-}H_2O]^+$	313.0851	313.0852	-0.2
		$[M+H\text{-}C_2H_2O]^+$	295.0745	295.0747	-0.2
		$[M+H\text{-}H_2O\text{-}C_2H_2O]^+$	271.0744	271.0747	-0.3
		$[M+H\text{-}H_2O\text{-}C_2H_2O\text{-}CO]^+$	253.0639	253.0641	-0.2
		$[M+H\text{-}H_2O\text{-}C_2H_2O\text{-}CO\text{-}NH_3]^+$	225.0691	225.0692	-0.1
			208.0426	208.0427	-0.1
		$[M+H\text{-}H_2O\text{-}C_2H_2O\text{-}CO\text{-}NH_3\text{-}C_2H_2O]^+$	166.0324	166.0321	+0.3



Supplementary Figure 1. **Models of LC-Qq-ToF-MS data**: A) PCA and B) PLS-DA, two class model built using baseline patients (t=0) and control (Q2 (cum) = 0.598 and R2Y (cum) = 0.686) with t=30 group used as a validation set. Samples are colored according to the disease status. C) A representative base peak chromatogram of urine sample collected at t=0; the peaks marked as a, b and c correspond to acetaminophen metabolites. In details: a) $[C14H17NO8+H]^+$, b) $[C11H14N2O4S+H]^+$ and c) $[C13H16N2O5S+H]^+$.



Supplementary Figure 2. Peptides MS^n spectra. A: Group 3, MS3 of urinary glycopeptide on ESI-Ion trap. B and C: MS^2 on ESI-Ion trap at m/z 685.7 of two synthetic peptides. The spectrum B resembles better the urinary peptide fragmentation showed in the spectrum A.

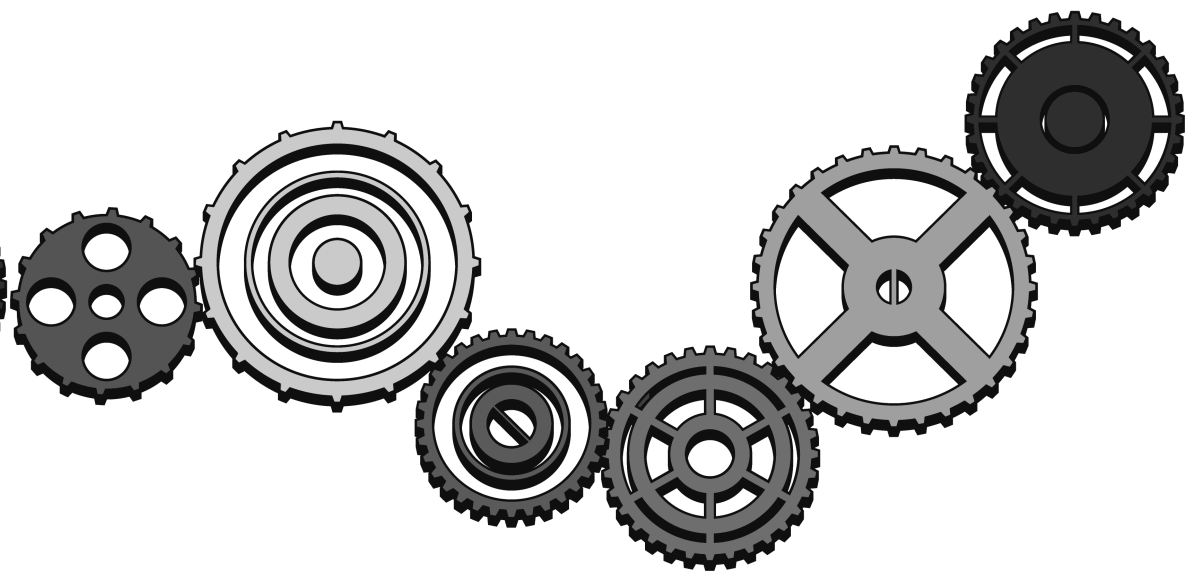
Chapter 4

Evaluation of GC-APCI/MS and GC-FID as a complementary platform

Tiziana Pacchiarotta. Ekaterina Nevedomskaya. Alegria Carrasco-Pancorbo.

André M. Deelder and Oleg A. Mayboroda

Journal of biomolecular techniques 2010; 21(4): 204–213



ABSTRACT

With a development of the metabolomics field, complementary cross-platform approaches started to attract attention, as none of the contemporary analytical methods had the capacity to cover the entire space of the human metabolome. In the current manuscript, we have evaluated an online coupling of gas chromatography (GC) – mass spectrometry (MS) and flame ionization detector (FID) as way of cross-detector analysis. The possible value of this combination was recognized from the very first days of GC-MS history but was never explored in detail. We have compared the basic analytical parameters of both detectors, such as limit of detection (LOD) and limit of quantification, with intra- and interday reproducibility. We show that for the majority of the tested compounds, MS detector demonstrates lower LOD. At the same time, FID appeared to be more robust, showing lower relative standard deviations (RSDs) for intra- and interday reproducibility. We conclude that the gain of this dual detector acquisition appears to be most evident for complex biological samples, where wide dynamic range and predictable response of FID are useful for an initial quantitative overview of sample composition and estimation of molar proportions of different metabolites. MS provides reliable, structural information and superior, at least in the case of atmospheric pressure chemical ionization, sensitivity. Taken together, both detectors represent a flexible tool for explorative studies and if supported by a powerful data-processing algorithm, would appear to be useful in any metabolic profiling study.

INTRODUCTION

The concept of metabolomics as a global overview of an organism's metabolites emerged in the early 1970s [1, 2], and from the beginning, it was evident that its technical realization demands complementary methodologies, as none of the contemporary analytical methods had the capacity to cover the entire space of the human metabolome. Despite the booming development of this field, 40 years later, we are still facing the same problem.

A classical way to expand the set of analyzable compounds –coming from different chemical families and with different molecular weights (MWs)– and their linear dynamic range is the use of multidimensional chromatography. In routine analysis, multidimensional chromatography is usually divided in heart-cut and comprehensive approaches [3]. The latter one is of primary importance for metabolic profiling, as it implies the separation of the complete sample in both chromatographic dimensions. A combination of two or more stationary phases leads to the exhaustive fractionation of the sample and eventually helps to use the dynamic range of the detector, usually a mass spectrometer, more efficiently. Multidimensional comprehensive chromatography is a relatively time-consuming method, and the depth of the coverage is proportional to the analysis time. When expensive, high-end mass spectrometry (MS) instruments are used as detectors, this can be a disadvantage. In addition, multidimensional comprehensive chromatography is, in practice, a single detector method, and regardless of the fractionation quality, the physical properties of the detector are the factors that 'constrain' the analysis. In the case of MS, those factors are the ionization technique and the type of mass analyzer. An alternative way to extend the coverage of the metabolome is a cross-platform analysis. The essence of cross-platform analysis is the acquisition of data using instruments with entirely different physics of detectors, followed by post–acquisition data fusion. A classical example is a combination of NMR spectroscopy and MS. NMR is, without reservation, the most powerful method of modern metabolomics [4-6]. With regard to metabolic profiling of body fluids, ^{1}H-NMR, which provides a fingerprint of proton-containing species, has become a routine tool. Moreover, NMR is probably the only analytical method where quantitative and structural information is 'embedded' in the physics of the detector. Still, a coanalysis of NMR with MS data expands the analytical domains of both methods and opens new ways of data interpretation [7]. For example, it has been shown that application of correlation-based algorithms, such as statistical heterospectroscopy (SHY) can be a useful way to dissect and structurally resolve xenobiotic metabolites in human body fluids [8].

The combination of NMR and MS is undoubtedly a powerful platform, but it requires costly instrumentation (NMR) and specialized expertise. In the current manuscript, we propose a way of cross-detector analysis, which can be practiced in almost any analytical laboratory: a combination of gas chromatography-mass spectrometry (GC-MS) and flame ionization detector (FID). GC is, without doubt, one of the most important and widely applied techniques in modern analytical chemistry, and FID has often been the first choice for GC routine analysis of complex biological samples. Indeed, since its introduction in 1958, FID has quickly become the most popular detector for GC and not without a reason: FID responds practically to all organic compounds, it is resistant to small fluctuations of the gas flow, it is insensitive to gas impurities, and most importantly, the FID response is very predictable, obeying the rule of equal carbon response [9]. The additional value of a combination of MS and FID has been recognized already in the early days of GC-MS history [10, 11]. Recently, the usefulness of FID as an auxiliary detector in NMR or/and MS

metabolic profiling studies has been demonstrated once more [12]. However, in practice, the combination of GC-MS and GC-FID has never developed into a routine analytical method, and there are a number of explanations for this. One of these is that for routine analytical applications, the benefits were not clearly obvious, and at the same time, technical issues with the synhronization of FID and vacuum-stage MS were considerable [11, 12]. With the reintroduction of Atmospheric Pressure (AP) sources for GC-MS [13, 14], the technical problems preventing optimal use of a combination of MS and FID have been largely overcome. In the current study, we demonstrate not only the simplicity of the coupling but also the possible complementarities of both detectors, in particular, for explorative studies about complex biological samples.

MATERIALS AND METHODS

A standard solution of 17 amino acids at 1 mM each in 1 M HCl was purchased from Sigma–Aldrich (Zwijndrecht, Nederland). Dopamine hydrochloride and phenyl-glycine (Phe-Gly) hydrate were from Fluka (UK). Sarcosine, theophylline, caffeine, nortriptyline hydrochloride, hippuric acid, 4-O-methyldopamine hydrochloride, benzoic acid, uric acid, and 5-hydroxyindole-3-acetic acid were acquired from Sig ma-Aldrich. Stock standard solutions of the 28 compounds were prepared in methanol at a concentration of 200 µM. N-methyl-N-trimethylsilyltrifluoroacetamide plus 1% trimethylchlorosilane (MSTFA + 1% TMS) were used as derivatization reagents (Pierce, Rockford, IL, USA). These reagents were used from freshly opened 1 ml bottles. Methoxyamine hydrochloride was purchased from Supelco (Park Bellefonte, PA, USA). Methanol (HPLC grade) and acetonitrile (GC grade) were acquired from Sigma-Aldrich, and pyridine (>99%, ultra-pure GC grade) was from Fluka. 4-Chlorophenol, decylamine, methyl caprate, prometryn, aldrin, and endosulfan α were purchased from Sigma-Aldrich. Stock solutions of these analytes were prepared in acetonitrile at a concentration of 10 mM. Reserpine was purchased from Sigma-Aldrich.

Sample preparation.

Aliquots of the standard mix were evaporated and derivatized by adding 100 µl methoxyamine (60 min, 40°C) and subsequently, 50 µl MSTFA (30 min, 40°C). Cerebrospinal fluid (CSF) samples were taken by lumbar puncture. The study was approved by the Ethical Committee of the Leiden University Medical Center (The Netherlands). Samples were processed within 1 h, centrifuged at 300 g to remove cells, aliquoted, and stored at - 80°C until use. The CSF samples were prepared as described previously [15]: the proteins were precipitated with cold methanol, and then the supernatant was evaporated under a gentle stream of nitrogen. The derivatization was carried out with methoxyamine and MSTFA + 1% TMCS.

GC-MS-FID Analysis

The samples (1 µl) were applied by splitless injection with a programmable Agilent CTC PAL multipurpose sampler (Agilent, Atlanta, GA, USA) into an Agilent 7890A GC (Agilent, Palo Alto, CA, USA). The chromatograph was equipped with a single HP-5-MS column (30 m, 0.25 mm ID, 0.25 µm film thickness) and a column flow splitter with deactivated capillary (0.6 m, 0.25 mm ID, 0.25µm film thickness), which allow a

simultaneous acquisition with MS and FID detectors. For sample injection, a septumless programmed temperature vaporizing (PTV) injector was used. For each analysis, the purge time of the PTV injector was set to 60 s at a purge flow rate of 20 ml/min and an equilibration time of 1 min. Helium was used as carrier gas at a constant flow rate of 1 ml/min through the column.

For the underivatized compounds, the PTV worked with the following temperature program: 60°C held for 0.1 min, and then the temperature was increased to 250°C at a rate of 500°C/min. The column temperature was first kept at 40°C for 0.1 min and then raised to 280°C at 30°C/min and held for 10 min. For the biological samples, the following temperature program was used: 90°C held for 0.1 min, followed by the temperature increase to 250°C at 500°C/min. The column temperature was first kept at 70°C for 5 min and then raised to 280°C at 5°C/min and held for 10 min.

The FID was operated at 300°C; hydrogen and air flows were set at 40 ml/min and 370 ml/min, respectively. The transfer line to the mass spectrometer was kept at 300°C. In the AP chemical ionization (APCI) source operating in positive mode, temperature and flow rate of the dry gas (nitrogen) were 300°C and 2.00 l/min, respectively. The APCI vaporizer temperature was 300°C; the pressure of the nebulizer gas (nitrogen) was set to 3.5 Bar, and the voltage of the corona discharge needle was + 4000 nA. Capillary voltage was set to –2000 V and the end-plate offset to –500 V. The maXis (Bruker Daltonics, Bremen, Germany), an orthogonal-accelerated TOF-MS, was used as a MS detector. The polarity of the APCI interface and all of the parameters of the TOF-MS detector were optimized with infusion of reserpine. Spectra were acquired in scan mode from mass-to-charge ratio (*m/z*) 100 to *m/z* 1000 with 1Hertz (Hz) frequency. External calibration was Performed using a tune mix.

The linear dynamic range for the mix of six underivatized compounds was estimated for a concentration range from 0.05 to 10,000 μM. For the standard mix, a concentration range, 0.5–50 μM was used. The two calibration lines (APCI-TOF and FID) were used for calculation of limit of detection (LOD) and limit of quantification (LOQ; Tables 1 and 2 in Supplementary Data); these were calculated on the basis of a signal-to-noise ratio of 3 for LOD and 10 for LOQ. The background noise was estimated from the peak baseline close to the analyte peak. Each calibration point is the result of three independent injections. The intra– and inter–day repeatability test was performed using standard mix as a reference. Results of the test were expressed as relative SD (RSD). Four independent injections were carried out at 25 μM concentration level. The interday repeatability was calculated after 24 and 48 h. The accurate mass data of the molecular ions were processed with DataAnalysis 4.0 software (Bruker Daltonics). A list of possible elemental formulas was generated by using the SmartFormula module of the DataAnalysis 4.0. It uses a CHNO algorithm, which provides standard functionalities, such as minimum/maximum elemental range, electron configuration, and ring-plus double bond equivalents, as well as a sophisticated comparison of the theoretical with the measured isotope pattern (σ value).

RESULTS

To illustrate the complementarity of APCI-MS and FID, we used a mix of six organic compounds: 4-chlorophenol, decylamine, methyl caprate, prometryn, aldrin, and endosulfan- α. The selection of these compounds is not random; it includes aliphatic and cyclic structures, with and without chlorine or sulfur. A comparison of the FID

chromatogram and the base peak chromatogram (BPC) shows (Figure. 1) that the response of the MS detector is more compound-dependent than FID. For example, the peaks at 5.3 and 9.2 min (namely, 4-chlorophenol and aldrin) are quite intense in FID but are nearly absent in the BPC generated by the MS detector.

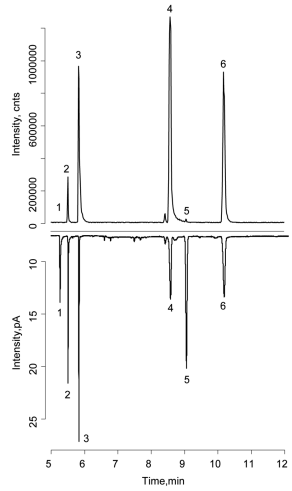


Figure 1. An example of comparison of the chromatograms recorded with GC-APCI-TOF MS (upper) and FID (lower) for six underivatized compounds (100 μM each): 1, 4-chlorophenol; 2, decylamine; 3, methyl caprate; 4, prometryn; 5, aldrin; 6, endosulfan α.

Table 1. Analytical parameters for underivatized compounds. The gray lines refer to GC-APCI-TOF MS detector while the white lines are referred to FID detector. r^2, Linear regression.

Compound	Calibration Curve	r^2	Linear Dynamic Range (μM)	LOD (μM)	LOQ (μM)	Retention time (min)
4'Chlorophenol	y = 583.54x - 54192	0.9951	50-5000	40.0	133.3	5.27
	y = 0.1639x - 14.011	0.9934	10-5000	8.33	27.78	5.27
Decylamine	y = 5317.7x + 96980	0.9933	10-2500	4.29	14.29	5.51
	y = 0.2578x - 27.218	0.9972	25-5000	12.00	40.0	5.50
Methyl Caprate	y = 11382x + 1E+06	0.9932	0.5-5000	0.38	1.25	5.84
	y = 0.1682x + 12.262	0.996	0.5-10000	0.59	1.95	5.83
Prometryn	y = 55477x + 36030	0.9941	0.05-500	0.06	0.20	8.58
	y = 0.2207x - 1.2977	0.9972	1-1000	0.85	2.82	8.56
Aldrin	y = 279.97x - 3776.1	0.9974	5-1000	3.75	12.50	9.07
	y = 0.4799x - 7.9728	0.9960	1-1000	1.58	5.26	9.05
Endosulfan α	y = 42543x - 97395	0.9907	0.05-100	0.03	0.09	10.19
	y = 0.2981x - 4.2042	0.9982	5-1000	1.64	5.48	10.16

Table 1 summarizes the analytical parameters for both detectors. Calibration lines, LODs, and LOQs were calculated, plotting peak areas as a function of concentration [extracted ion chromatograms (EICs) were used for calculations of peak areas in MS]. According to the correlation coefficients, each detector demonstrates a high degree of linearity

(r^2 > 0.991 in all cases). LOD and LOQ are significantly lower for GC–APCI-TOF-MS than GC-FID, except for 4-chlorophenol and aldrin. On the other hand, the linear range of detection is wider in the case of FID. It is important to mention that the retention times of a given compound are almost identical for both detectors (FID and APCI-MS). Thus, using this relatively simple example, we have shown that there are substantial differences in the response factors and basic analytical parameters (LOD and LOQ) between FID and APCI-MS detectors. A combination of six organic compounds, even when selected specifically with regard to chemical diversity, does not reflect the complexity of biological samples. Besides, a large proportion of body fluid metabolites is nonvolatile polar compounds and needs a derivatization before GC analysis. Therefore, as a next step, we analyzed a standard mix of compounds typically reported as components of body fluids. This mix included compounds belonging to different chemical families: amines, amino acids, organic acids, alcohols, and xanthines. All chemical species were selected with the specific aim to cover a wide range of polarities and MWs, mimicking as closely as possible a real life situation. Figure 2 shows typical chromatograms of FID and MS detectors.

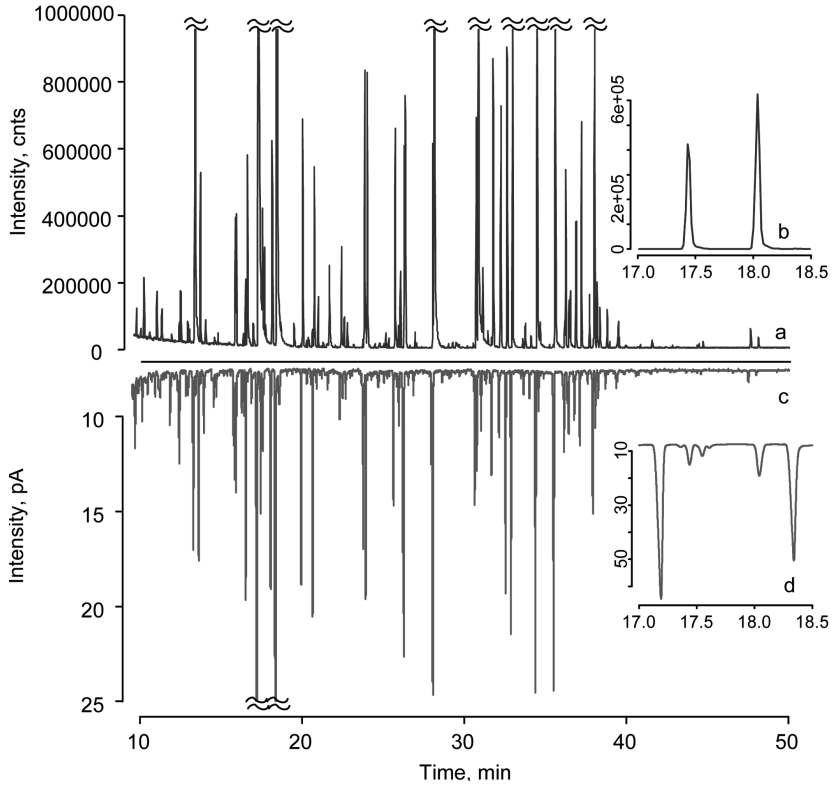


Figure 2. Representative chromatograms of the standard mix (25 μM): APCI-TOF-MS (BPC; upper) and FID (lower). Zoom-in: Ile (17.6 min) and Leu (18.2 min).

Table 2. Standard mix analytical parameters. The gray lines refer to GC-APCI-TOF MS; the

white lines refer to FID

Compounds	Retention Time (min)	Calibration Curve	r^2	LOD (µM)	LOQ (µM)	Repeatability intra-day	Repeatability inter-day (24h)	Repeatability inter-day (48h)
Alanine+2TMS+H	12.39	y = 13490x - 45721	0.9686	0.19	0.63	7%	18%	23%
	12.40	y = 0.2716x + 2.6521	0.9756	0.24	0.79	5%	14%	14%
Sarcosine+2TMS+H	13.41	y = 6145.7x - 32450	0.9665	3.00	6.47	17%	18%	26%
	13.41	y = 0.1097x + 0.0803	0.9537	8.00	26.88	6%	7%	11%
Proline+1TMS+H	14.53	y = 5473.2x - 59848	0.9943	21.58	71.94	16%	16%	21%
	14.55	y = 0.2041x - 0.5898	0.9969	18.29	60.98	3%	6%	8%
Valine+2TMS+H	15.83	y = 21344x - 73007	0.9199	0.10	0.33	9%	20%	35%
	15.83	y = 0.2403x + 0.0935	0.945	0.78	2.60	1%	3%	5%
Benzoic Acid + 1 TMS+H	16.51	y = 895.29x - 5812.6	0.9998	10.00	33.33	18%	18%	18%
	16.50	y = 0.5802x - 1.3997	0.9887	6.66	22.20	2%	6%	5%
Isoleucine+2TMS+H	17.42	y = 23962x - 80202	0.9911	0.34	1.11	19%	24%	25%
	17.44	y = 0.3432x - 1.0109	0.9928	5.00	16.60	2%	3%	6%
Leucine+2TMS+H	18.02	y = 30845x - 120980	0.8989	0.13	0.42	26%	17%	16%
	18.03	y = 0.6134x - 2.2667	0.9736	1.71	5.71	2%	4%	10%
Glycine+3TMS+H	18.34	y = 3430.7x + 2601.9	0.9747	0.06	0.20	15%	39%	37%
	18.34	y = 0.3445x + 1.1661	0.9997	2.70	9.90	14%	30%	27%
Serine+3TMS+H	19.92	y = 34064x - 87840	0.9373	0.07	0.22	12%	13%	19%
	19.92	y = 0.4867x - 1.18	0.9279	0.67	2.22	3%	3%	9%
Threonine+3TMS+H	20.62	y = 33259x - 58986	0.9442	0.05	0.16	13%	11%	20%
	20.63	y = 0.5272x - 1.0441	0.9241	1.55	5.15	2%	4%	7%
Methionine+2TMS+H	23.75	y = 44225x - 113892	0.9622	0.12	0.40	6%	17%	30%
	23.76	y = 0.4499x + 1.2307	0.9839	3.33	11.11	5%	6%	12%
Aspartic acid+3TMS+H	23.92	y = 36461x - 15137	0.9972	0.07	0.24	13%	24%	28%
	23.91	y = 0.5083x - 0.7629	0.9784	3.00	10.00	4%	6%	19%
Glutamic acid+3TMS+H	26.20	y = 29834x - 49222	0.9957	0.12	0.38	12%	21%	32%
	26.19	y = 0.4277x - 0.5318	0.9933	3.50	11.68	9%	11%	12%
Phenylalanine+2TMS+H	26.26	y = 49218x - 25330	0.9974	0.13	0.44	16%	18%	20%
	26.25	y = 0.7029x - 1.1104	0.9924	3.50	11.68	4%	5%	14%
Phenyl-Gly+H	28.06	y = 119260x - 165927	0.9972	0.06	0.18	13%	22%	34%
	28.04	y = 0.8171x - 0.7237	0.9956	0.35	1.18	5%	5%	8%
Hyppuric Acid+1TMS+H	30.66	y = 11673x - 26449	0.9918	0.26	0.86	16%	15%	16%
	30.64	y = 0.3526x - 0.5933	0.9835	4.77	15.92	6%	13%	11%
Caffeine+H	30.78	y = 114348x - 170144	0.9962	0.23	0.77	5%	10%	9%
	30.76	y = 0.3152x + 0.0298	0.9934	1.66	5.55	2%	6%	7%
Theophylline+1TMS+H	32.12	y = 45565x - 62580	0.9952	0.33	1.11	6%	16%	26%
	32.13	y = 0.2126x - 0.3059	0.9875	2.43	8.09	2%	37%	45%

Lysine+4TMS+H	32.53	y = 56140x - 14749	0.9948	0.11	0.36	20%	27%	35%
	32.53	y = 0.2126x - 0.3059	0.9875	0.75	2.50	7%	11%	11%
Tyrosine+3TMS+H	32.86	y = 75274x - 46887	0.9963	0.09	0.83	12%	21%	23%
	32.86	y = 0.6914x - 0.0326	0.9974	0.59	1.95	6%	8%	6%
4- Methyldopamine hydrochlor+3 Si+H	34.40	y = 91126x + 94865	0.9816	0.03	0.10	10%	16%	18%
	34.38	y = 0.8979x - 0.4778	0.9992	0.70	2.34	2%	6%	5%
Dopamine hydrochlor+4 TMS+H	35.53	y = 87513x + 30795	0.9786	0.05	0.18	13%	14%	21%
	35.50	y = 1.0494x - 1.0955	0.9945	1.20	4.00	1%	4%	4%
Uric Acid + 4 TMS+ H	36.15	y = 29816x - 35130	0.9984	0.14	0.45	12%	28%	38%
	36.15	y = 0.2464x + 0.1531	0.9999	0.94	3.13	6%	10%	34%
5-Hydroxyindole-3-acetic + 3 TMS+ H	37.90	y = 67994x - 66186	0.995	0.08	0.25	11%	22%	18%
	37.91	y = 0.4126x - 0.1815	0.9981	0.58	1.92	4%	5%	13%
Nortriptyline hydrochlor+H	38.32	y = 8649.4x + 5316.9	0.995	0.57	1.89	10%	16%	25%
	38.26	y = 0.0882x + 0.2164	0.9907	3.33	11.11	8%	27%	39%

Table 2 summarizes analytical data for compounds of the mix (only the most stable silylation form of a compound is shown). In agreement with previous data (Table 1), retention times for both detectors were found to be identical, and the MS detector showed a lower LOD (< 3 μM for MS and < 8 μM for FID), except for proline and benzoic acid (Table 2). The maximum intraday deviations were 9% for FID and 26% for MS; the maximum interday deviations (after 48 h) were 39% for FID and 38% for MS. The RSD after 48 h for both detectors indicates degradation in the samples once they had been derivatized. The linearity of response was found to be equally good for FID and MS ($r^2 > 0.95$). Even when the correlation factor was not ideal ($r^2 > 0.95$), as in the case of valine, serine, and threonine, this deviation affects both detectors, which indicates a problem of derivatization, a thorny step in the analytical procedure, rather than instability of the detector. The detailed analysis of the chromatograms presented in Figure 2 shows how differently both detectors respond to certain compounds despite their superficial similarity. For example, Ile and Leu are 'classical' positional isomers, which have the same atomic composition, number of single and double bonds, and consequently, the same masses. The peak areas observed for these compounds give a good approximation of the relative detector response. Calculating the peak area ratios for FID and MS, we obtain the following numbers: Leu FID/Ile FID = 0.52; Leu MS/ Ile MS = 0.69. Thus, there is a significant difference of the ratios in the Leu/Ile peak areas between FID and MS detectors. The effect is also reflected in the angular coefficients of the calibration lines: Ile has a higher angular coefficient than Leu (Table 2). This increase was found to be larger for FID (+ 44%) than for APCI-TOF (+ 22%). Analysis of the compound mass spectra (Figure 3) provides a possible explanation of this effect. Leu and Ile are represented by doubly silylated forms. In GC-APCI MS, the parent ion is the dominant feature of the spectra, unlike in conventional GC-electronic impact (EI) spectra, where the parent is a minor peak or even absent [15]. Main fragments, for Leu as well as for Ile, are the result of a loss of neutral groups such as $Si(CH_3)_3OH$ (*m/z* 90) or $Si(CH_3)_3OHCO$ (*m/z* 118), leading to formation of *m/z* 186.1 and *m/z* 158.1 fragments, respectively. However, Ile has an additional intense fragment at *m/z* 260.1. This fragment is the result of a loss of

the methyl group, most probably at position 5; the remaining tertiary carbon at position 3 stabilizes the positive charge with formation of the *m/z* 260.1 fragment. Anadditional loss of $Si(CH_3)_3OH$ (*m/z* 90) leads o the formation of a fragment with *m/z* 170.1 (Figure 3).

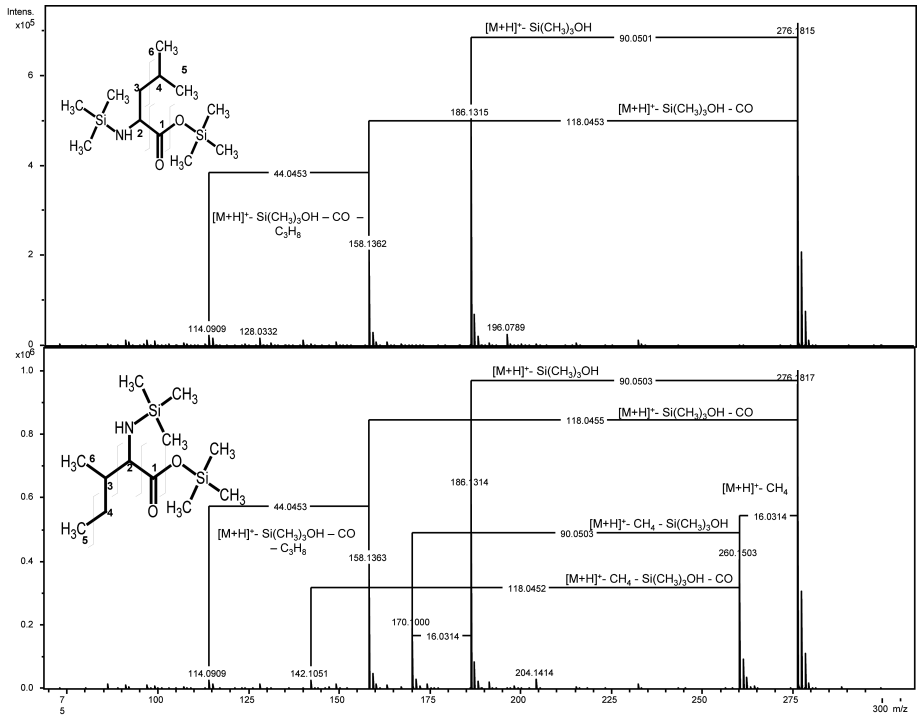


Figure 3. APCI-TOF spectra for Leucine (upper) and Isoleucine (lower).

Next, we used human CSF to test the applicability of our approach for biological material. Figure 4A shows typical BPC and FID chromatograms. With both detectors, very complex chromatograms were recorded. A number of molecular features extracted from MS data and overview of compounds detected in human CSF with GC-APCI-MS were reported elsewhere [15]. At first glance, there is no serious dissimilarity between the chromatograms. However, as the representative examples (Figure 4B and C) show, there is a number of differences between FID and APCI-MS. Figure 4B shows a few abundant compounds that saturate the MS detector. It is a common situation for biological fluids with their enormous concentration range. It is evident that realistic quantitative estimation based on MS data is impossible. FID, on the contrary, is still far from its saturation point within its linear range and therefore, can provide accurate quantitative information. Figure 4C illustrates a different situation: in the case of co-eluting or poorly resolved peaks, MS plays an essential role in the interpretation of the results.

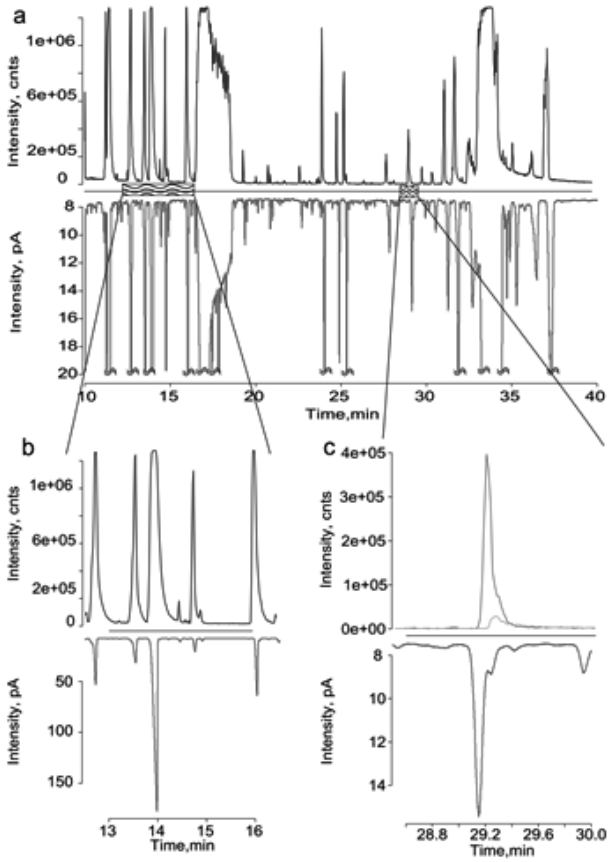


Figure 4. A) Representative chromatograms of CSF sample: APCI-TOF-MS (BPC upper); FID chromatogram (lower). (B) Zoom-in showing examples of peaks saturating the MS detector but remaining within linear dynamic range of FID. (C)Zoom-in showing an example of EIC used to resolve co–eluting compounds.

DISCUSSIONS

GC-MC at AP has a peculiar history. The first experiments with corona discharges have shown the potential of ionization processes at AP [16]. The first instrumental designs of GC-APCI/MS were described in the early 1970s, but for a number of reasons, vacuum-stage instruments took over the market. For decades, GC-APCI/MS remained an exotic application.

The combination of GC-MS and GC-FID also is an old idea, with roots going back to the 1960s [10, 11]. Horning [2] demonstrated the feasibility of the combination GC-MS/GC-FID in one of his first reports about APCI-GC. His source design is regularly quoted in reviews about APCI-MS, but his idea to use MS and FID as parallel detectors

remained unnoticed [17, 18]. One of the common arguments against the routine use of the combination of FID and MS is the lack of complementarity [19]. It has been postulated, for example, that MS and FID yield a similar chromatogram, and response factors for the majority of organic compounds are comparable [19]. This statement, however, is based on the comparison of FID signals, which were obtained with classical GC-MS ionization techniques: EI and CI. As we have shown here, the difference between APCI-MS and FID detectors becomes evident even with a simple mix of six organic compounds. Thus, the statement of equal response factors for FID and MS would appear not to be applicable to GC-APCI-TOF/MS. However, the difference in response factors and basic analytical parameters (LOD and LOQ) is not entirely surprising, considering the different nature of the physics of both detectors. Besides, the important question is not whether APCI-MS and FID produce different responses but whether those differences are complementary and whether the combination of both detectors might thus be beneficial for analysis of complex biological samples.

Our next example, a mix representing the most common body fluid metabolites (standard mix), is an attempt to address this question. Our 'standard mix' includes compounds from several chemical families: amines, amino acids, organic acids, alcohols, and xanthines. It is certainly a much more complex sample than the previous one and at a first glance, much more diverse. However, the derivatization procedure introduces a common bias for all metabolites. In addition, there are not so many heteroatoms, such as, for example, chloride. A simple visual inspection of the trace of the chromatogram is not sufficient for the comparison of detectors, but the analysis of the basic analytical parameters reveals a few interesting features. For example, the MS detector consistently demonstrates lower LOD and LOQ values, and the intraday repeatability is significantly better for FID (RSD between 1% and 9%, as compared with 6–26% for MS). Moreover, a comparison of the data acquired 24 h and 48 h after derivatization indicates that there is a common trend for both detectors toward higher RSDs, which is, most probably, a side-effect of derivatization instability. Considering that the effects of derivatization on the analytical performance of a given method can be circumvented by, for example, an in-line derivatization approach and that the MS detector shows better LOD/LOQ, one of the possible preliminary conclusions could be that the combination of FID and APCI-MS as parallel detectors is indeed of little practical use. However, our analysis of the response of two detectors to such classical positional isomers as Leu and Ile proves the opposite, or at least, makes the last statement questionable. The difference in Leu/Ile ratios observed with the two detectors is the result of 'sensitivities' of the detector to different proper ties of compounds: although FID responds mainly to the total carbon number, the response of MS is affected by the structural differences between the analytes. The tertiary carbon of Ile affects the fragment formation during the in-source/in-funnel fragmentation process, which leads to the differences in compound spectra and eventually changes the response of the detector. On the other hand, FID is also not an ideal 'carbon counter', and its response can be affected by the presence of such heteroatoms as chlorine or sulfur. Nevertheless, there are not that many physiologically relevant metabolites containing a large number of chlorine atoms, and if there are any, their presence is easily detected by unique isotopic patterns.

Our next example was human CSF, which plays a key role in the mechanical and immunological protection of the brain, the maintenance of its homeostasis, and metabolism and as such, is an important source of analytical information. We have used it to outline

areas where a parallel data collection by two detectors (FID and APCI-MS) could be useful for the interpretation of GC results. It would thus appear that the combination of detectors has at least two possible advantages. Firstly, the wide dynamic range of FID can be useful for the exact quantitative estimation of compounds present in biological samples at concentrations close to saturation of the MS detector, and those quantitative results are free of any ionization and instrumental bias. Secondly, the TOF mass analyzer provides high-quality structural information, which is essential for the data interpretation in metabolomics studies. Moreover, the advantages of APCI mass spectra for structural identification of compounds become more evident if we compare them with spectra of traditional GC-MS sources, such as EI and CI. The two latter ionization techniques produce spectra, where the quasi-molecular ion is low-abundant or even absent. This can be a serious disadvantage for the analysis of biological material, where a significant part of compounds is usually unknown. With an APCI source, the quasi-molecular ion ($[M+H]^+$) is a dominant one, which opens the possibility for structural characterization of unknown compounds using the combination of accurate mass and isotopic pattern. In addition, an ionization process resulting in the formation of a stable quasi-molecular ion and the technical characteristics of modern TOF instruments makes it possible to use such acquisition modes as MS^2 or multiple reaction monitoring, which are common in LC-MS but seldom used in GC-MS.

Thus, the power of MS as a tool for the structural elucidation of metabolites generally needs no additional confirmation, but quantification remains 'the weakest link'. The quantitative response of a mass spectrometer cannot be attributed to a single physical phenomenon: it is the product of hyphenation, ionization technique, efficiency of ion optics, and the type of mass analyzer. In other words, MS is a selective detector. In routine work, dealing with the measurements of known compounds, this selectivity is not a problem: the response of the MS detector can be perfectly linear for every single compound. However, in explorative studies, where the relative abundance of compounds within the complete metabolite profile is as important as the absolute quantity, the mass spectrometer may introduce an undesired bias. On the contrary, the quantitative response of the FID detector is free from ionization bias and those biases introduced by the type of mass analyzer or the instrumental design of a mass spectrometer. Consequently, the FID gives a better overall quantitative representation of complex biological samples. As a result, a parallel detection for GC-FID, GC-APCI, might offer a quick way to obtain simultaneously the quantitative and structural information in metabolic profiling studies.

There are, of course, a few technical issues needed to be resolved to turn parallel data acquisition with GC-FID/GC-MS into an efficient method. Those are: the different acquisition rates (few Hz for MS and routine 50 Hz for FID), the lack of a proper algorithm for data fusion, the development of an optimal strategy for treatment of over-lapping peaks, and so on. However, the main purpose of this manuscript was to draw attention to the elegant but forgotten idea of the simultaneous use of FID and MS, rather than resolve all of the technical problems of this combination. With the reintroduction of APCI sources for GC in their modern, user-friendly form, the parallel acquisition of FID and MS chromatograms has become a simple instrumental option that requires no additional technical modification of the chromatograph. The gain of this dual detector acquisition would appear to be most evident for the analysis of complex biological samples. FID is the only detector capable of handling concentration ranges typical for body fluids. Its wide dynamic range and predictable response are useful for an initial quantitative overview of sample composition

and the estimation of molar proportions of the different metabolites. MS provides reliable, structural information and superior, at least in the case of APCI ionization, sensitivity. Taken together, the use of both detectors represents a flexible tool for explorative studies and if supported by a powerful data-processing algorithm, would appear to be useful in any metabolic profiling study.

REFERENCES

[1] Pauling L, Robinson AB, Teranish.R, Cary P. QUANTITATIVE ANALYSIS OF URINE VAPOR AND BREATH BY GAS-LIQUID PARTITION CHROMATOGRAPHY. Proceedings of the National Academy of Sciences of the United States of America. 1971;68:2374-&.

[2] Horning EC, Horning MG. METABOLIC PROFILES - GAS-PHASE METHODS FOR ANALYSIS OF METABOLITES. Clinical Chemistry. 1971;17:802-&.

[3] Francois I, Sandra K, Sandra P. Comprehensive liquid chromatography: Fundamental aspects and practical considerations-A review. Analytica Chimica Acta. 2009;641:14-31.

[4] Nicholson JK, Wilson ID. HIGH-RESOLUTION PROTON MAGNETIC-RESONANCE SPECTROSCOPY OF BIOLOGICAL-FLUIDS. Progress in Nuclear Magnetic Resonance Spectroscopy. 1989;21:449-501.

[5] Lindon JC, Nicholson JK, Holmes E, Everett JR. Metabonomics: Metabolic processes studied by NMR spectroscopy of biofluids. Concepts in Magnetic Resonance. 2000;12:289-320.

[6] Bollard ME, Stanley EG, Lindon JC, Nicholson JK, Holmes E. NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. Nmr in Biomedicine. 2005;18:143-62.

[7] Crockford DJ, Holmes E, Lindon JC, Plumb RS, Zirah S, Bruce SJ, et al. Statistical heterospectroscopy, an approach to the integrated analysis of NMR and UPLC-MS data sets: Application in metabonomic toxicology studies. Analytical Chemistry. 2006;78:363-71.

[8] Crockford DJ, Maher AD, Ahmadi KR, Barrett A, Plumb RS, Wilson ID, et al. H-1 NMR and UPLC-MSE Statistical Heterospectroscopy: Characterization of drug metabolites (xenometabolome) in epidemiological studies. Analytical Chemistry. 2008;80:6835-44.

[9] Holm T. Aspects of the mechanism of the flame ionization detector. Journal of Chromatography A. 1999;842:221-7.

[10] Roeraade J, Enzell CR. A COMBINATION OF MASS SPECTROMETER AND FLAME IONISATION DETECTOR FOR OPEN TUBULAR COLUMN CHROMATOGRAPHY. Acta Chemica Scandinavica. 1968;22:2380-&.

[11] Bruner F, Ciccioli P, Zelli S. IMPROVED DOUBLE DETECTION GAS CHROMATOGRAPH MASS SPECTROMETER INTERFACE FOR ANALYSIS OF COMPLEX ORGANIC MIXTURES. Analytical Chemistry. 1973;45:1002-6.

[12] Fancy S-A, Beckonert O, Darbon G, Yabsley W, Walley R, Baker D, et al. Gas chromatography/flame ionisation detection mass spectrometry for the detection of endogenous urine metabolites for metabonomic studies and its, use as a complementary tool to nuclear magnetic resonance spectroscopy. Rapid Communications in Mass Spectrometry. 2006;20:2271-80.

[13] McEwen CN, McKay RG. A combination atmospheric pressure LC/MS : GC/MS ion source: Advantages of dual AP-LC/MS : GC/MS instrumentation. Journal of the American Society for Mass Spectrometry. 2005;16:1730-8.

[14] Schiewek R, Lorenz M, Giese R, Brockmann K, Benter T, Gaeb S, et al. Development of a multipurpose ion source for LC-MS and GC-API MS. Analytical and Bioanalytical Chemistry. 2008;392:87-96.

[15] Carrasco-Pancorbo A, Nevedomskaya E, Arthen-Engeland T, Zey T, Zurek G, Baessmann C, et al.

Gas Chromatography/Atmospheric Pressure Chemical Ionization-Time of Flight Mass Spectrometry: Analytical Validation and Applicability to Metabolic Profiling. Analytical Chemistry. 2009;81:10071-9.

[16] Hayhurst AN, Sugden TM. MASS SPECTROMETRY OF FLAMES. Proceedings of the Royal Society of London Series a-Mathematical and Physical Sciences. 1966;293:36-&.

[17] Byrdwell WC. Atmospheric pressure chemical ionization mass spectrometry for analysis of lipids. Lipids. 2001;36:327-46.

[18] Bruins AP. MASS-SPECTROMETRY WITH ION SOURCES OPERATING AT ATMOSPHERIC-PRESSURE. Mass Spectrometry Reviews. 1991;10:53-77.

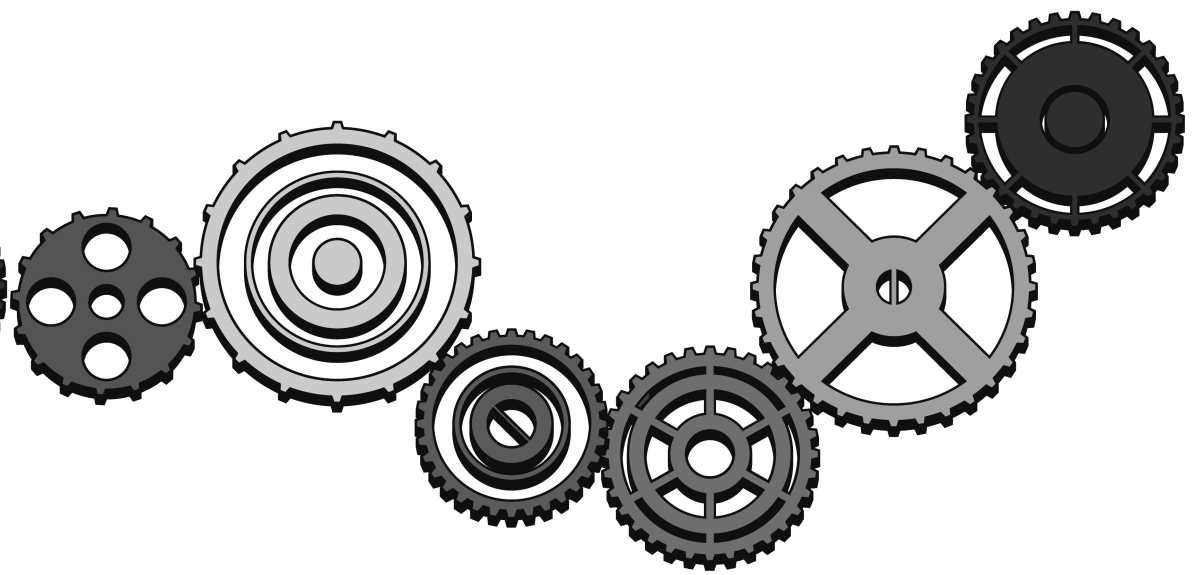
[19] H.J H. Handbook of GC/MS. Fundamentals and Applications. 2ed.: Wiley-VCH; 2008.

Chapter 5

On-line spectral library for gas chromatography/atmospheric pressure chemical ionization-time of flight mass spectrometry

Tiziana Pacchiarotta, Rico J.E. Derks, Elena Hurtado-Fernandez, Pauline van Bezooijen, Alex Henneman1, Ralf Schiewek, Alberto Fernández-Gutiérrez, Alegría Carrasco-Pancorbo, André M. Deelder, Oleg A. Mayboroda

Bionalysis 2013; 5(13): 1515-1525



ABSTRACT

Invented more than three decades ago by Horning, Gas Chromatography Mass Spectrometry under atmospheric pressure (GC-APCI-MS) has only recently emerged from years of obscurity. However, the general acceptance of GC-APCI-MS is certainly constrained by the lack of spectral libraries, which make the traditional GC-MS approaches so powerful. Here we present a concept of a GC-APCI-QqToF spectral library. The library is web-based, fully searchable and at moment includes spectra of 150 compounds from the most common chemical families. The fragmentation pattern of some chemical families is explained and a protocol for *de novo* identification has been provided in order to facilitate the identification of unknown compounds. A library for GC-APCI-QqToF is now publicly available online.

INTRODUCTION

The article by Linus Pauling on the quantitative analysis of urine vapour [1] is often referred to as the first metabolomics publication. The concept outlined in this article resembles closely the modern 'metabolomics' even though this term was introduced more than twenty years later. The work of Horning [2] on analytical methods for the study of steroid hormone metabolites is usually quoted in special literature as a seminal work on steroid analysis. Yet, to the best of our knowledge, in this manuscript the term 'profiling' was introduced for the first time as a description of multicomponent analysis of body fluids. Both manuscripts are milestones in the development of today's metabolomics and both exploited the same analytical platform – GC. 40 years later, GC still remains the core method of metabolomics and its position is hardly threatened by the progress of LC [3-8].

The enduring success of GC can be attributed to many factors, of which the unmatched chromatographic resolution and the very fact of being the first separation technique used with a mass spectrometer as a detector are probably the most important. Most of the commercial GC–MS use vacuum stage ionization sources: Electron Impact (EI) and Chemical Ionization (CI). EI is a hard ionization technique, which implies an extensive fragmentation of the analytes. The fragmentation patterns are highly reproducible and the possibility of using commercial and open source 70eV EI spectral libraries makes GC/EI-MS a powerful analytical technique. However, due to the extensive fragmentation the molecular ion is seldom present in a spectrum. This could be a disadvantage for the analysis of complex spectra populated by structurally similar analytes. Here, CI might offer a solution. Indeed, being a softer ionization technique, CI preserves the precursor ions but the results depend strongly on the reagent gas or reagent gas pressure used and, consequently, CI spectra are less suitable for a library search.

An atmospheric pressure ionization source appears to be a logical next step and atmospheric pressure CI (APCI) was pioneered by the same group, which introduced the term "profiling" [9]. Developed in the early seventies the GC-APCI was subsequently abandoned for almost thirty years till the development of the MS instrumentation created an opportunity for its reintroduction. First, McEwen *et al.* [10] introduced a multipurpose source, modifying a commercially available source for LC. Later on, in 2008, Schiewek *et al.* [11] demonstrated the feasibility of GC-MS at atmospheric pressure for APCI and atmospheric pressure laser ionization. To date, there have been a number of manuscripts published, ranging from technical papers [12-14] to applications such as the analysis of pharmaceutical residuals [12], food extracts [15, 16], bio-fluids [17, 18], the metabolome of *E. coli* [19] and sediment phthalates [20].

However, despite a growing number of applications, GC-APCI is still far from being a routine technique. One of the factors which seriously downscales the practical value of the GC-APCI-MS is the lack of spectral libraries such as available for EI and CI sources (e.g. NIST, ADMIS, GOLM) [21, 22]. Here we present the first attempt to create a spectral library for GC-APCI-quadrupole-quadrupole (Qq) ToF keeping in mind metabolic profiling of body fluids as a prime application. To this end, we have selected a set of compounds belonging to those chemical families which are most relevant for the analysis of the biological fluids and/or food specimens, and have created a library of reference MS and MS/MS spectra. Moreover, anticipating the well-known problem of 'unknown' compounds we also provide the guidelines for *de novo* identification.

Finally, keeping in mind the importance of web-based tools for the scientific community

we have presented our library as a publicly available database resource [101].

MATERIALS AND METHODS

Chemicals

All chemicals at analytical grade were purchased from Sigma or Fluka (Zwijndrecht, Netherlands). N-methyl-N-trimethylsilyltrifluoroacetamide with 1% trimethylchlorosilane (MSTFA + 1% TMCS) and N,O - Bis(trimethylsilyl) trifluoroacetamide with 1% trimethylchlorosilane (BSTFA + 1% TMCS) were applied as derivatization reagents (Pierce; Rockford. IL. USA); they were used from freshly opened 1 ml bottles. Methoxyamine hydrochloride and the mixes of saturated alkanes (C7-C30 and C7-C40) were purchased from Supelco (Zwijndrecht, Netherlands). APCI tune mix (containing purine, hexamethoxyphosphazine, hexakis(2,2-difluoroethoxy)phosphazine and hexakis(1H, 1H, 3H-tetrafluoropropoxy)phosphazine) and pyridine (99%. ultra-pure GC grade) were acquired from Fluka.

Derivatization

All compounds (with exception of some thermo labile ones such as histidine, guanine and arginine) were derivatized on-line with a double step silylation. The standards were prepared from the stock solutions to obtain a final concentration of 10, 25 or 50 ng/µl (depending on the molar response of the compound).

First, 40 µl of a methoxyamine solution (20 mg/ml in pyridine) was added to the dried standards and this was then incubated for 40 min at 40 °C; then 50 µl of MSTFA was added and incubated at 40 °C for 30 minutes.

For thermolabile compounds, the methoxyamination step was carried out at room temperature for 16 h with further addition of 50 µl of MSTFA, and an incubation step at room temperature for 1 h.

The flavonoids require no methoxyamination, consequently this step was omitted for these compounds. The derivatization was performed with the sole addition 40 µl of pyridine and 50 µl of BSTFA and further incubation at 60 °C for 1 h.

To all the compounds, 10 µl of hydrocarbons solution C7-30 or C7-40 were added to reach a final concentration of 25 µg/mg.

GC-APCI-MS-Flame ionization detector analysis

The samples (1 µl) were applied by splitless injection with a programmable MultiPurpose Sampler MPS2 (Gerstel, Mülheim an der Ruhr, Germany) into an Agilent 7890A GC (Agilent. Palo Alto. USA) equipped with one single column (HP-5-MS column 30 m. 0.25 mm ID. 0.25 µm film thickness) and two detectors running in parallel; a column flow splitter is required with deactivated capillaries (0.6 m. 0.25 mm ID. 0 µm film thickness) which allows the simultaneous acquisition with both MS and FID detectors (Figure 1). For sample injection a septumless PTV injector was used. For each analysis the purge time of the PTV injector was set to 60s at a purge flow rate of 20 ml/min and an equilibration time of 1 min. Helium was used as carrier gas at a constant flow rate of 1 ml/min through the column.

The hydrocarbons solution C7-C30 was spiked to the compounds at a concentration of 25 µg/mg and injected using the following temperature program: 40 °C held for 1 minute,

10 ºC/min till 310 ºC held for 2 minutes.

The compounds requiring the addition of the hydrocarbons solution C7-C40, because of their late elution, at a concentration of 25 μg/mg were injected with the following temperature program: 40 ºC held for 1 minute, 7 ºC/min till 310 ºC held for 15 minutes.

The transfer line to the mass spectrometer was kept at 300 ºC. The APCI source was operated in positive mode. The temperature and the flow rate of the dry gas (nitrogen) were 300°C and 2.00 l/min. respectively. The APCI vaporizer temperature was 300 °C; the pressure of the nebulizer gas (nitrogen) was set to 3.5 bar and the voltage of the corona discharge needle was +2000 nA. Capillary voltage was set to -1000V and the end-plate offset to -1000 V. The source and detector parameters have been set in order to achieve a minimal resolution of 40000.

A maXis 4G (Bruker Daltonics, Bremen, Germany), an orthogonal-accelerated Qq-ToF mass spectrometer, was used as mass analyzer. The spectra were acquired in a mass range from *m/z* 100 to *m/z* 1000 with 1 Hz acquisition rate. Before each chromatographic run, a tune mix solution was used as external calibrant and injected in Flow Injection mode.

MS/MS experiments were carried out selecting the precursor ion and applying a different collision energy for each analyte to achieve the best possible fragmentation.

The FID flows and temperature were set as follows: front detector 300 °C, hydrogen flow: 40 ml/min and air flow: 370 ml/min. This detector was used for the calculation of the Kovats' indices, embedded database information for each compound, relative to the specific column in use. For the calculation, the equation relative to the temperature programmed chromatography was applied [23].

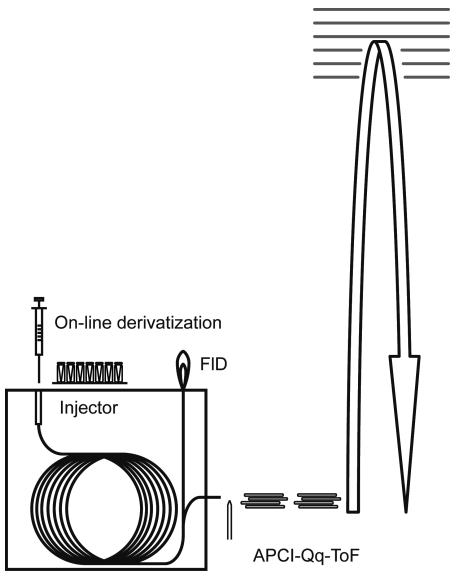


Figure 1. The setup that enables the use of two detectors (flame ionization detection and atmospheric pressure chemical ionization–quadrupolequadrupole-ToF) in parallel, the flow is split at the end. APCI–Qq-ToF: Atmospheric pressure chemical ionization–quadrupole-quadrupole-ToF; FID: Flame ionization detection.

Serum samples

Sera samples from healthy females volunteers were combined into one pool. The samples were prepared according to a reference method with minor modifications [24]. Briefly, 50 µl of serum was extracted with a 450 µl solution of methanol and water (8:1 v/v). The supernatant was dried and reconstituted with 30 µl of methoxyamine in pyridine (20 mg/ml) and then incubated for 1 hour at 40 ºC. Subsequently we added 30 µl of MSTFA and incubated at 40 ºC for 30 minutes.

Library management

The GC-APCI–MS metabolite database is designed to be freely accessible for the public over the internet [101]. It is built exclusively from open source software, or software that will soon enter the open domain. Debian GNU/Linux 6.0.1 is used as operating system running an Apache web server (version 5.3.3) with PHP 5.2 as scripting language for retrieving the information from a MySQL (version 5.1.49-3) database.

The architecture of the database consists of several tables containing the information about the metabolite, MS spectra and MS^2 spectra.

The metabolite database is accessible through a website, facilitating several general search options such as metabolite name, molecular formula or KEGG number. Moreover, a specific MS spectra search is also allowed using the measured *m/z* value with a set mass tolerance. The results are returned sorted by *m/z* error. As an extra option, the search can be expanded to seek MS fragments. The users can search for MS^2 spectra entering the fragments and their relative intensities with the option to take the precursor ion into account. In addition, for each compound the retention indices calculated for the used column are provided.

For future contributors an XML structure is being developed for the upload of data into the data base.

RESULTS AND DISCUSSIONS

Spectra interpretation

An important advantage of soft ionisation techniques such as APCI is the preserved signal of a precursor ion. Consequently, a combination of soft ionisation and modern high resolution ToF instruments creates a perfect platform for the analysis of highly complex samples, for instance body fluids, food specimens or even sediment samples. The last group has been addressed by David *et al.* in their report on the analysis of high molecular weight phthalates [20]. The parent ion is usually present as protonated species (for positive ionization mode). This fact has been carefully investigated by Portolés *et al.*, who analyzed the ionization process of a large selection of pesticides and reported that for the 90% of the analysed compounds the prominent ionized species was the $[M+H]^+$ [25]. However, as Bristow *et al.* have shown one should not exclude a possibility of odd-electron ions formation [12].

In the Table 1 we have listed the common losses we observed in the MS or MS^n spectra. Some of them are internal molecule fragments; others are related to the presence of the derivatization reagent. Some losses (CH_3, O, CO) are rather nonspecific, other, however, can provide structural information. For instance, the neutral loss TMSOH (mass 90.0501

Da) is characteristic for the hydroxyl group – the most reactive towards silylation functional group. The consecutive losses of TMSOH give us an indication of the number of OH groups in the molecules: representative examples of this phenomenon are the alcohol sugars, such as mannitol or sorbitol.

The loss of 118.0450 Da is characteristic for the carboxylic groups. The presence of an amino group can be corroborated by the release of the TMS group with theoretical mass of 72.0395 Da. Occasionally, we observe a radical molecular ion $M^{\cdot+}$ with the consequent loss of TMSO (89.0423 Da) and additional fragments related to the TMS group with theoretical mass of 73.0474 or 74.0551 Da, depending on the stability of the molecule.

Table 1. Common losses.

Common neutral losses	Theoretical mass
CH_3	15.0235
O	15.9949
NH_3	17.0265
H_2O	18.0106
C_2H_2	26.0157
CHN	27.0109
CO	27.9949
CH_3N	29.0265
CH_2O	30.0106
CH_4O	32.0262
C_3H_2	38.0157
C_2HN	39.0109
C_2H_3N	41.0265
C_2H_2O	42.0106
CO_2	43.9898
C_2H_4O	44.0262
CH_2O_2	46.0055
C_2H_5NO	59.0371
C_3H_8Si [TMS]	72.0395
C_3H_9Si	73.0474
$C_3H_{10}Si$	74.0552
C_4H_5NO	83.0371
$C_3H_9SiCH_3$	88.0708
$C_3H_9SiNH_2$	89.0661
C_3H_8SiOH [TMSOH]	90.0501
$C_3H_9SiNCH_2$	101.0661
$C_4H_{10}SiO$	102.0495
$C_3H_9SiOH + CO$ [TMSOH+CO]	118.0450

The compounds most commonly detected in body fluids represent the core of our selection. Hence, we have included in our database amino acids, sugars, organic acids, fatty acids, xanthines, etc. Additionally, we have added a selection of flavonoids, phenols and carotenoids as compounds frequently encountered in the bioavailability experiments. The website interface facilitates several search options, such as name, molecular formula or KEGG number. Since the database is GC-oriented, the Kovats' index of each metabolite is stored and searchable. A MS search can be performed with adjustable mass tolerance (default 0.05). The *m/z* value of the molecular or protonated ion of a derivatized compound is often used as a search parameter. If metabolites have multiple derivatization forms, a separate MS spectrum is stored for each of them. The MS spectrum selected in the search is highlighted. Furthermore, in order to unequivocally identify the compounds one is interested in, the theoretical isotopic distribution of the molecular/protonated ion is also provided.

The MS mode includes fragments related to the in-source or in-funnel fragmentation: a search option for these fragments (with the same tolerance) is available. The output results are sorted by *m/z* error and they are shown separately. When a result is selected all the information and MS spectrum of the metabolite are displayed.

An MS^2 search can be performed on multiple fragments with a set mass tolerance threshold (default: 0.05) and the molecular/protonated ion information can be included for a more accurate search. The output results are sorted by *m/z* error and the MS^2 spectrum displays relative intensities of the fragments. If multiple MS^2 spectra are given for a metabolite the selected MS^2 spectrum from the search is highlighted.

Overview of compound classes

Amino acids. The presence of two functional groups with a different reactivity towards the silylation reagent is a characteristic feature of amino acids. The carboxylic group is more reactive than the amino group: the compounds with lower silylation substitution always indicate the presence of the TMS on the hydroxyl group of the molecule and the compounds with higher substitution include derivatization of the amino group as well. As a result, for most of the amino acid more than one chromatographic peak was observed.

The amino acids spectra always showed the loss of TMSOH (90.0501 Da) and TMSOH+CO (118.0450 Da) corresponding to the carboxylic group and the loss of TMS (72.0395 Da) corresponding to the amino group. For all detected silylation forms of the amino acids their relative Kovats' indices are included in the database.

Fatty acids. A $[M+H]^+$ ion corresponding to the derivatized form of the carboxylic group was the main feature of the MS- spectrum of fatty acids (Figure 2A). MS^2 spectra, however, revealed some peculiarities (Figure 2B). In addition to the classical losses of TMSOH at *m/z* 239.2363 and its cluster series with consecutive losses of CH_2 (14.0157 Da), we have observed two additional fragments from the protonated molecule corresponding to the loss of CH_4 at *m/z* 313.2545 and its substitution with a H_2O molecule at *m/z* 331.2650. In order to identify the reaction site, we used the deuterated standard of the palmitic acid d-31 of which we report the MS and MS/MS spectra in Figure 2C and 2D respectively. MS/MS experiments were carried out with an isolation window of 2 Da. The figure 2D shows the protonated molecule *m/z* 360.4827 ($C_{19}HD_{31}O_2Si^+$), the loss of a CH_4 group at *m/z* 344.4493 and a substitution with a H_2O molecule confirmed by the exact mass of 18.0106 Da and the isotopic distribution at *m/z* 362.4596. The data shows no involvement of a deuterium atom, which leads to the conclusion that the observed effect is related to the silylation group. The exact location of the reaction site might require an experiment with isotopically labelled derivatization reagent.

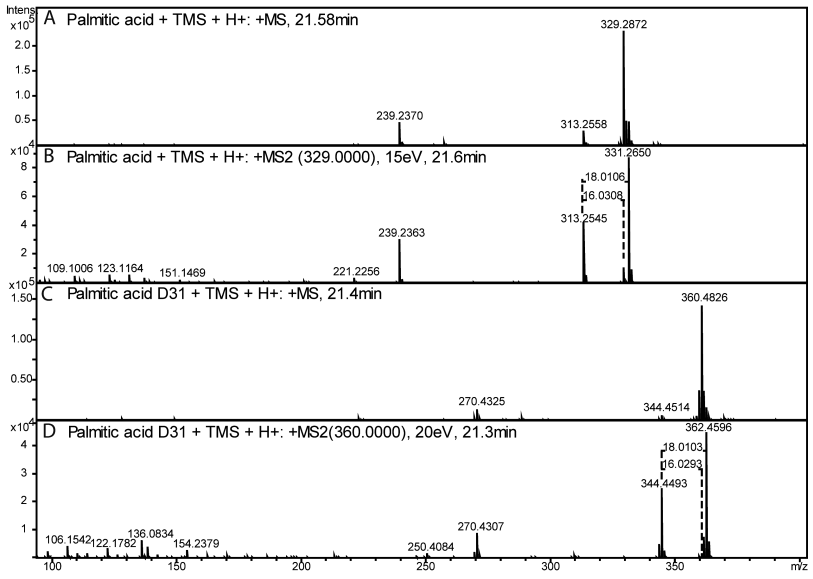


Figure 2. MS and MS^2 spectra of palmitic acid and palmitic acid d-31. The MS2 spectra of (A & B) palmitic acid and (C & D) palmitic acid d-31 show a substitution reaction of CH4 with a H_2O molecule.

In the analysis of unsaturated fatty acids, regardless of the unsaturation factor, we could detect the radical fragment corresponding to the TMS backbone after alpha-cleavage ($C_5H_{11}O_2Si^{+}$) with the theoretical *m/z* 131.0523. Moreover, in both mono and poly unsaturated series of C16 and C18, the fragment $C_{10}H_{15}^{+}$ with theoretical *m/z* 135.1168 indicates the presence of a double bond. Unfortunately, at the moment we cannot provide a sound strategy for the location of the position of the double bond. Vrkoslav and Cvacka have recently proposed a strategy for the localization of the double bond position for the methyl esters of fatty acids using APCI [26]; it remains, however, to be tested whether their approach has more universal applicability.

Sugars. The classical derivatization strategy for sugars consists of two steps, namely oximation and silylation. During the oximation the keto group is blocked and the formation of enolic forms is inhibited, preventing the cyclization. This reduces the possible isomeric forms of a single sugar to just α and β anomers of the sugars' linear form. Hence, all the sugars are represented as protonated molecule (molecular mass with the addition of TMS groups on the active hydrogens) and the CH_3N group on the carboxylic oxygen.

The ratio between the *E-* and *Z-* forms is different for aldohexoses (e.g. glucose) or ketohexoses (e.g. fructose). Our results are consistent with a previous report by Funcke *et al.*[27] : the *E-* form is dominating for the aldohexoses and for the ketohexoses the ratio between the two forms is almost equal (Figure 3 A and B). In the MS spectra instead (not shown here), the main discriminative feature between the two anomers is the ratio between the protonated ion at *m/z* 570 and the ion at *m/z* 480 corresponding to the fragment $[M+H-TMSOH]^{+}$. For the aldohexose glucose the *m/z* 480 presents almost the same intensity of the protonated ion in the *E-* form but only the 50% of intensity for the *Z-*form. For

the ketohexose fructose, the *m/z* 480 presents a low abundance or is absent in both of the anomers. The ratio of the anomeric forms can be discriminative feature for standards, however, in a 'real life' situation (with presence of co-elution, low abundance and high noise), it provides no unequivocal identification and needs to be correlated to the retention indices provided in the database. MS^2 spectra also contain signals characteristic for aldo- and keto-hexoses. Figure 3 C - F show the representative MS^2 spectra of an aldohexose (C and E) and a ketohexose (D and F): the spectra are clearly different for the two species due to the relative abundance of the specific fragments at *m/z* 307.1599 ($C_{12}H_{31}O_3Si_3^+$) and 217.1090 ($C_9H_{21}O_2Si_2^+$) for aldohexoses, and at *m/z* 319.1558 ($C_{13}H_{31}O_3Si_3^+$) and 205.1065 ($C_8H_{21}O_2Si_2^+$) for ketohexoses.

Similar fragmentation patterns have previously been described by Medeiros *et al.* [28]; however the authors assigned the effect to the cycle forms, pyrano or furano rings, which is not applicable in our case. Another explanation of the fragmentation has been suggested by MacLoed *et al.* [29], who used ^{13}C incorporation experiments to explain the origin of the fragment ions. Our data appear to be in agreement with their model implying fragment formation from the straight-chain derivatives.

Analysis of the disaccharides provides a good example of the differences between reducing (e.g. lactose) and non-reducing sugars (e.g. sucrose). The first differences are noticeable in the chromatographic behaviour of the compounds: lactose, for instance, shows two peaks corresponding to the two anomers, while sucrose, which has no open chain form, is represented by a single peak. Mass spectra of these compounds are apparently different. The spectrum of lactose is characterised by presence of $[M+H]^+$ ion at *m/z* 948.4639, the first glucose moiety ion at *m/z* 570 and the second glucose moiety after the glycosidic break and the subsequent loss of a TMSOH group at *m/z* 361.1665 (Figure 3G). However, in contrast to previously published results [30] the intensity of the *m/z* 538 ion (corresponding to the open chain hexose ion after glycosidic break and the loss of a methyl group) was too low to be included in the list of the discriminative fragments. A distinct feature of non-reducing sugar (sucrose) spectra is a dominant fragment at *m/z* 361.1688 resulting from cleavage of the glycosidic bond and low intensity signal of the single moiety at *m/z* 539.2525 (Figure 3H).

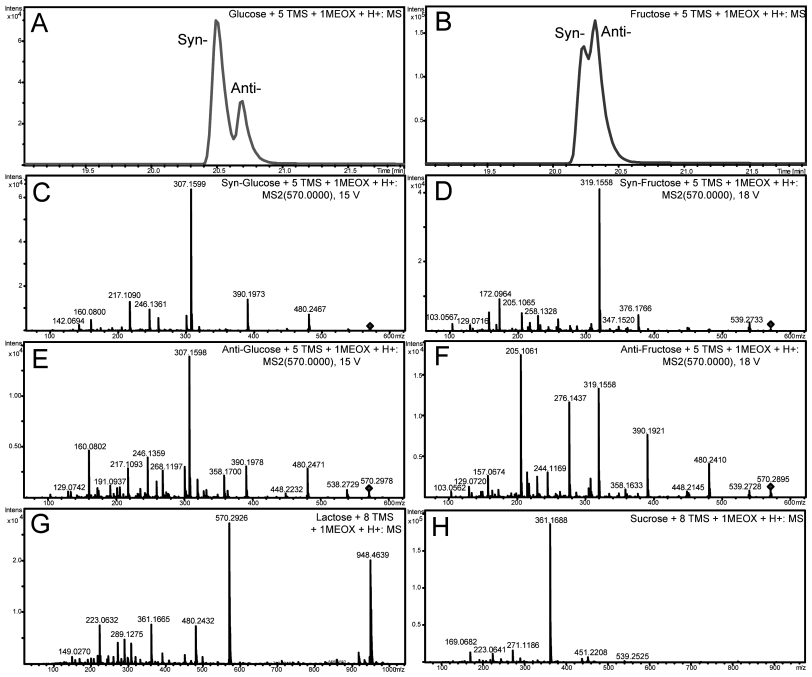


Figure 3.Chromatograms and spectra of sugars. Base peak chromatogram of (A) glucose and (B) fructose. (C & E) MS^2 of synand anti-forms of glucose. (D & F) MS^2 of syn– and anti–forms of fructose. (G) MS spectra example of reducing sugar – lactose and (H) MS spectra example of nonreducing sugar – sucrose.

The six carbons alcohol sugars do not have oxime forms so the fragmentation resulted in subsequential losses of TMSOH giving always at least one of these ions *m/z* 307.1576 ($C_{12}H_{31}O_3Si_3^+$), 217.1075 ($C_9H_{21}O_2Si_2^+$) or 205.1075 ($C_8H_{21}O_2Si_2^+$).

Organic acids. This group of compounds consists of relatively small molecules and only few of them have nitrogen atoms in their structure. Thus, the loss of the silylanol group is usually the only constant feature of their spectra. However, for such compounds as benzoic or hippuric acid an additional fragment ion $C_7H_5O^+$ (105.0335 Da), corresponding to the benzoyl ion stabilized by the aromatic ring, is detected.

Phenols .To date, only the simple phenols, phenolic acids, lignans and some flavonoids are included in the database. As far as phenolic acids are concerned, we included hydroxycinnamic acid and hydroxybenzoic acid. In the first group, we can mention *o*-, m- and *p*-coumaric acids, ferulic, sinapic and chlorogenic acids, and so on. The MS pattern for this kind of analytes is characterized by a predominant ion signal corresponding to the complete silylated form of each compound, although the hydroxycinnamic acid spectra always show the loss of CH_4 (16.0313 Da) and the loss of TMSOH (90.0501 Da). For isomers such as *o*-, *m*- and *p*-coumaric acids the ions observed in the MS spectra are exactly the same, although their relative intensities differ. *o*-Coumaric acid seems to be the less stable of the three isomers, showing the highest intensity of the ion resulting from the loss of the TMSOH group. This fact has to be with the position of the functional groups and is

confirmed by MS/MS analysis which provides an unequivocal identification of the structural isomers. The compounds included in the group of hydroxybenzoic acids are gallic, vanillic, protocatechuic, gentisic, 4-hydroxybenzoic and α-, β-, and γ-resorcylic acids. The protonated molecule of the fully silylated compounds is still the predominant ion in the MS spectra and the fragmentation pattern is characterized by the losses of CH_4 and TMSOH. In almost all spectra of the hydroxybenzoic acids, it is possible to observe the loss of CO_2 (43.9898 Da) as well. α-, β- , and γ-resorcylic acids show the same fragmentation pattern, however γ -resorcylic acid appears to be the least stable and vulnerable to the in-source fragmentation. Homovanillic acid is the only compound belonging to this family and it gives a major ion that results from the loss of 118.045006 Da (C_3H_9SiOH + CO).

Of the flavonoids, some flavanols, flavones and flavanones were included. Their spectra are characterized by the loss of CH_4 and additional fragments related to the TMS group with mass 73.0474 or 74.0552 Da, according to the molecule stability. In the case of flavanols (catechin and epicatechin) the loss of TMSOH is more commonly observed than for the other flavonoids.

We have also studied the APCI behaviour of hydroxytyrosol and tyrosol, which are simple phenols. They are represented by the radical $[M + x\,TMS]^+$ instead of the protonated molecule $[M+ x\,TMS+H]^+$ and they showed the loss of C_3H_9SiO. Pinoresinol, which belongs to the lignans, showed as the main signal in the MS spectra the loss of 18.010565 (H_2O). In its MS/MS spectrum, we observe the fragment 209, which is attributable to the stable substituted tropylium ion structure.

Vitamins. Some vitamins suitable for a GC analysis such as nicotinic acid, pantothenic acid, pyridoxine, ascorbic acid, etc. were included in our database as well. Their analysis revealed that all the compounds showed the complete silylation of their active hydrogen and a very high resistance to the in-source fragmentation process. For the four mentioned compounds we observed as predominant signal in the MS spectra the protonated molecule $[M + x\,TMS+H]^+$, and in every case we could observe losses of TMSOH, although piridoxine proved to be more sensitive to the in-source fragmentation process, since the loss of 90 was more easily observed in its spectrum.

De novo identification of unknown compounds

Currently our library contains APCI-MS, MS^2 spectra and Kovats' indices for 150 compounds. Thus, this is only the primer of a library, which should then create a framework for a further growth. Routine analysis of body fluids will undoubtedly lead to a quick growth of the number of entries. One would need, however, a set of rules or a standard procedure to be followed in order to make a valid assignment. Below, we outline the procedure currently used for the identification of the unknowns.

- Spectra calibration – an external calibration with tune mix is carried out. Additionally, an internal calibration using the losses of siloxanes from the column as lock-masses could be performed;
- Identification of the common losses (Table 1) to estimate a number of possible elemental compositions;
- Calculation of the theoretical elemental compositions, simulation of the isotopic distribution and fitting it to the experimental data. Possible candidates can be listed and confirmed by further experimental analysis;

- MS^n experiments - Match the MS^n spectra with the candidates and their empirical formulae and finally define the searched compound;
- Validation on analytical standard, if applicable, this step would provide the highest confidence for the identification.

An importance of an established routine for the identification of unknown becomes evident once we stepping into analysis of the complex biological samples such as biofluids. Figure 4 shows a representative chromatogram of a human serum sample. Using a feature finding algorithm provided by the MS vendor (Find Molecular Features; Bruker Daltonics) we have found 439 molecular features (the solvent segment was excluded). There is no algorithm for the conversion of this number to the number of the real metabolites but our experience shows that a 50% cut-off could be applied as a rule of the thumb. Even then we could expect twice as more metabolites than in a recent overview of the human serum metabolome [31]. At the moment our library provides an unequivocal identity only for a limited number of the metabolites (Table 2), a fact which once more emphasizes importance of a continuous updated library.

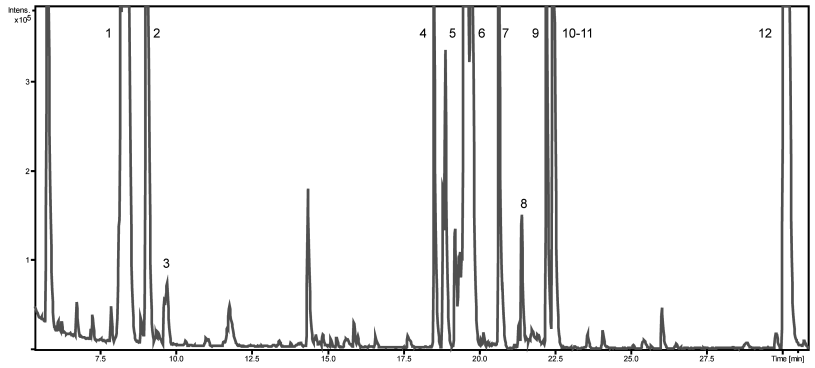


Figure 4. Representative GC-APCI-MS chromatogram of a serum sample and compounds identified in the database and listed in Table 2.

Table 2. Compounds identified in serum samples.

#	Compound	*m/z*	Error [mDa]	Error [ppm]	Kovats Index *	Kovats Index Error (%) **
1	Urea+ 2 TMS + H^+	205.1181	0.6	2.74	1287.5	3.19
2	Orthophosphoric acid + 3 TMS + H^+	315.1028	0.8	2.68	1287.6	0.29
3	Glycine + 3TMS+H^+	292.1584	-0.5	1.63	1316.0	-0.15
4	Citric Acid + 4 TMS + H^+	481.1915	0.9	1.88	1870.3	1.51
5	Syn-Glucose + 5 TMS + MEOX + H^+	570.2917	-0.1	-1.58	1934.2	-0.47
6	Anti-Glucose + 5 TMS + MEOX + H^+	570.2961	-0.7	-1.14	1952.5	-0.17
7	Palmitic acid + 1 TMS + H^+	329.2870	0.1	0.22	2052.00	0.16
8	Myo-Inositol+ 6 TMS + H^+	613.3056	2.2	3.66	2193.6	2.89
9	Linoleic acid + 1 TMS + H^+	353.2876	0.5	1.29	2216.6	-0.06
10	Oleic acid + 1 TMS + H^+	355.3029	-0.2	-0.65	2223.18	0.05
11	Stearic acid + 1 TMS + H^+	357.3183	-0.4	-1.11	2263.0	0.75
12	Cholesterol + TMS + H^+	459.4018	-0.1	-0.2	3195.6	0.07

*Calculated in the serum samples; ** Relative to the value reported in the database.

CONCLUSION

Here we present the concept of a GC-APCI-QqToF-MS library. The database is web-based and includes a number of spectra of compounds from the most common chemical families (amino acids, sugars, fatty acids, organic acids, phenols, vitamins and so on) [101]. All data included in the database were acquired using a modern UHR-ToF instrument, which ensures the high quality of the generated MS and MS/MS spectra. The database is fully searchable and includes retention indices and all possible silylation forms for a given compound. The fragmentation pattern of some chemical families is illustrated and a protocol for *de novo* identification has been provided in order to facilitate the determination of unknown compounds.

FUTURE PERSPECTIVES

Currently our library includes GC-APCI-QqToF-MS, MS^2 spectra and Kovats' indices for 150 compounds, which is only a fraction of metabolites one might encounter during metabolomics studies. The practical value of the library will certainly improve if it includes data obtained with the instruments of other vendors and/or using different mass analysers such as for example, triple quadrupole [14]. We are planning to update this library continuously extending a range of biological samples used and trying to extend a range of the mass analysers for the data acquisition.

ACKNOWLEDGMENT

This work was supported by the Stichting Tropische Geneeskunde – Leiden. The authors would like to thank Ing. Rob Marissen for technical support.

FINANCIAL & COMPETING INTERESTS DISCLOSURE

This work was supported by the Stichting Tropische Geneeskunde, Leiden, The Netherlands.

REFERENCES

[1] Pauling L, Robinson AB, Teranish.R, Cary P. QUANTITATIVE ANALYSIS OF URINE VAPOR AND BREATH BY GAS-LIQUID PARTITION CHROMATOGRAPHY. Proceedings of the National Academy of Sciences of the United States of America. 1971;68:2374-&.

[2] Horning MG, Moss AM, Horning EC. FORMATION AND GAS-LIQUID CHROMATOGRAPHIC BEHAVIOR OF ISOMETRIC STEROID KETONE METHOXIME DERIVATIVES. Analytical Biochemistry. 1968;22:284-&.

[3] Gullberg J, Jonsson P, Nordstrom A, Sjostrom M, Moritz T. Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of Arabidopsis thaliana samples in metabolomic studies with gas chromatography/mass spectrometry. Analytical Biochemistry. 2004;331:283-95.

[4] Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR. Gas chromatography mass spectrometry-based metabolite profiling in plants. Nature Protocols. 2006;1:387-96.

[5] Jonsson P, Gullberg J, Nordstrom A, Kusano M, Kowalczyk M, Sjostrom M, et al. A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. Analytical Chemistry. 2004;76:1738-45.

[6] Michell AW, Mosedale D, Grainger DJ, Barker RA. Metabolomic analysis of urine and serum in Parkinson's disease. Metabolomics. 2008;4:191-201.

[7] Kind T, Tolstikov V, Fiehn O, Weiss RH. A comprehensive urinary metabolomic approach for identifying kidney cancer. Analytical Biochemistry. 2007;363:185-95.

[8] Nishiumi S, Shinohara M, Ikeda A, Yoshie T, Hatano N, Kakuyama S, et al. Serum metabolomics as a novel diagnostic approach for pancreatic cancer. Metabolomics. 2010;6:518-28.

[9] Horning EC, Horning MG. METABOLIC PROFILES - GAS-PHASE METHODS FOR ANALYSIS OF METABOLITES. Clinical Chemistry. 1971;17:802-&.

[10] McEwen CN, McKay RG. A combination atmospheric pressure LC/MS : GC/MS ion source: Advantages of dual AP-LC/MS : GC/MS instrumentation. Journal of the American Society for Mass Spectrometry. 2005;16:1730-8.

[11] Schiewek R, Lorenz M, Giese R, Brockmann K, Benter T, Gaeb S, et al. Development of a multipurpose ion source for LC-MS and GC-API MS. Analytical and Bioanalytical Chemistry. 2008;392:87-96.

[12] Bristow T, Harrison M, Sims M. The application of gas chromatography/atmospheric pressure chemical ionisation time-of-flight mass spectrometry to impurity identification in Pharmaceutical Development. Rapid Communications in Mass Spectrometry. 2010;24:1673-81.

[13] Portoles T, Sancho JV, Hernandez F, Newton A, Hancock P. Potential of atmospheric pressure chemical ionization source in GC-QTOF MS for pesticide residue analysis. Journal of Mass Spectrometry. 2010;45:926-36.

[14] Portoles T, Mol JGJ, Sancho JV, Hernandez F. Advantages of Atmospheric Pressure Chemical Ionization in Gas Chromatography Tandem Mass Spectrometry: Pyrethroid Insecticides as a Case Study. Analytical Chemistry. 2012;84:9802-10.

[15] Garcia-Villalba R, Pacchiarotta T, Carrasco-Pancorbo A, Segura-Carretero A, Fernandez-Gutierrez A, Deelder AM, et al. Gas chromatography-atmospheric pressure chemical ionization-time of flight mass spectrometry for profiling of phenolic compounds in extra virgin olive oil. Journal of Chromatography A. 2011;1218:959-71.

[16] Portolés T, Cherta L, Beltran J, Hernández F. Improved gas chromatography–tandem mass spectrometry determination of pesticide residues making use of atmospheric pressure chemical ionization. Journal of Chromatography A. 2012;1260:183-92.

[17] Carrasco-Pancorbo A, Nevedomskaya E, Arthen-Engeland T, Zey T, Zurek G, Baessmann C, et al. Gas Chromatography/Atmospheric Pressure Chemical Ionization-Time of Flight Mass Spectrometry: Analytical Validation and Applicability to Metabolic Profiling. Analytical Chemistry. 2009;81:10071-9.

[18] Pacchiarotta T, Nevedomskaya E, Carrasco-Pancorbo A, Deelder AM, Mayboroda OA. Evaluation of GC-APCI/MS and GC-FID as a complementary platform. Journal of biomolecular techniques : JBT. 2010;21:205-13.

[19] Wachsrnuth CJ, Almstetter MF, Waldhier MC, Gruber MA, Nurnberger N, Oefner PJ, et al. Performance Evaluation of Gas Chromatography-Atmospheric Pressure Chemical Ionization-Time-of-Flight Mass Spectrometry for Metabolic Fingerprinting and Profiling. Analytical Chemistry. 2011;83:7514-22.

[20] David F, Sandra P, Hancock P. Determining High-Molecular-Weight Phthalates in Sediments using GC-APCI-ToF-MS. Lc Gc Europe. 2011;24:16-+.

[21] Halket JM, Przyborowska A, Stein SE, Mallard WG, Down S, Chalmers RA. Deconvolution gas chromatography mass spectrometry of urinary organic acids - Potential for pattern recognition and automated identification of metabolic disorders. Rapid Communications in Mass Spectrometry. 1999;13:279-84.

[22] Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmuller E, et al. GMD@CSB.DB: the Golm Metabolome Database. Bioinformatics. 2005;21:1635-8.

[23] Kovats E. GAS-CHROMATOGRAPHISCHE CHARAKTERISIERUNG ORGANISCHER VERBINDUNGEN .1. RETENTIONSINDICES ALIPHATISCHER HALOGENIDE, ALKOHOLE, ALDEHYDE UND KETONE. Helvetica Chimica Acta. 1958;41:1915-32.

[24] Jiye A, Trygg J, Gullberg J, Johansson AI, Jonsson P, Antti H, et al. Extraction and GC/MS analysis of the human blood plasma metabolome. Analytical Chemistry. 2005;77:8086-94.

[25] Portoles T, Pitarch E, Lopez FJ, Hernandez F, Niessen WMA. Use of soft and hard ionization techniques

for elucidation of unknown compounds by gas chromatography/time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry. 2011;25:1589-99.

[26] Vrkoslav V, Cvačka J. Identification of the double-bond position in fatty acid methyl esters by liquid chromatography/atmospheric pressure chemical ionisation mass spectrometry. Journal of Chromatography A. 2012;1259:244-50.

[27] Funcke W, Vonsonntag C. SYN AND ANTI FORMS OF SOME MONOSACCHARIDE O-METHYL OXIMES - C-13-NMR AND GLC STUDY. Carbohydrate Research. 1979;69:247-51.

[28] Medeiros PM, Simoneit BRT. Analysis of sugars in environmental samples by gas chromatography-mass spectrometry. Journal of Chromatography A. 2007;1141:271-8.

[29] MacLeod JK, Flanigan IL, Williams JF, Collins JG. Mass spectrometric studies of the path of carbon in photosynthesis: positional isotopic analysis of C-13-labelled C-4 to C-7 sugar phosphates. Journal of Mass Spectrometry. 2001;36:500-8.

[30] Fuzfai Z, Boldizsar I, Molnar-Perl I. Characteristic fragmentation patterns of the trimethylsilyl and trimethylsilyl-oxime derivatives of various saccharides as obtained by gas chromatography coupled to ion-trap mass spectrometry. Journal of Chromatography A. 2008;1177:183-9.

[31] Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, et al. The Human Serum Metabolome. Plos One. 2011;6.

WEBSITE

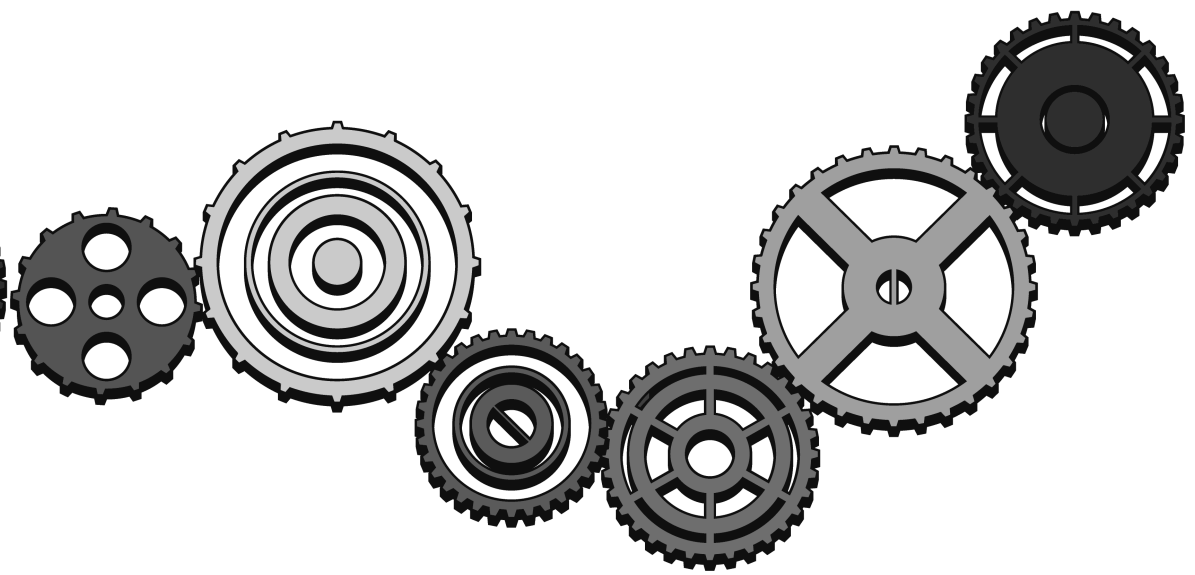
[101] GC-APCI–MS metabolite database. http://metams.lumc.nl

Chapter 6

Exploratory analysis of urinary tract infection using a novel GC-APCI-MS platform

Tiziana Pacchiarotta, Rico Derks, Ekaterina Nevedomskaya, Willize van der Starre, Jaap van Dissel, André Deelder, Oleg Mayboroda

Manuscript in preparation



ABSTRACT

Urinary tract infection (UTI) is among the most common bacterial infections worldwide. As such, it is an interesting model for the application of the modern integrative '- omics' technologies and metabolomics in particular. This work is part of a bigger project which has the aim to provide a comprehensive overview of UTI-induced changes in urine. Here we present the first explorative metabolomics study of UTI using a recently described GC-APCI-MS platform. This instrumental set up allows the combination of high resolution mass spectroscopy with the robustness and dynamic range of an FID detector, and opens new possibilities for cross-platform data analysis, especially in combination with NMR and LC-MS. Using well-defined clinical cohort, and taking advantage of a multi-level study design, a number of molecular discriminators were identified as specific for UTI patients, such as uremic solutes (myo-inositol and citrate) and isomers of hydroxyhippuric acid.

INTRODUCTION

UTI is among the most common bacterial infections worldwide. Clinical presentation of UTI is broad: from a mild cystitis to severe pyelonephritis, which can quickly evolve into life-threatening conditions such as urosepsis and multiple organ failure [1, 2]. Thus, with a positive bacterial urine culture as common denominator, the term 'urinary tract infection' includes a variety of clinical syndromes, [3]. To study the physiological alterations going along with such complex clinical phenomena, modern integrative '-omics' technologies, and metabolomics in particular, are ideally suited. Indeed, metabolomics approaches have already been applied to analyse urines from UTI patients, but those studies were either a kind of proof of principle studies [4, 5] or were dedicated to the development of a metabolomics based routine for uropathogen identification [6]. Each of these studies advocated a specific technological solution but it is well known that the use of different analytical platforms often results in complementary data. Our aim is to create a comprehensive overview of UTI-induced changes in the metabolic patterns of patients' urine by applying several explorative methodologies on a carefully matched selection of patients from a multicentre prospective observational study of UTI [7]. So far, well-established techniques such as liquid chromatography-mass spectrometry (LC-MS) and nuclear magnetic resonance (NMR) were used by us [8, 9]

GC-APCI-MS has only recently been included into the toolbox for metabolite analysis. The detection at atmospheric pressure rather than with a vacuum stage (e.g chemical ionization and electronic impact) has demonstrated its potential in a wide variety of fields ranging from pharmaceutical applications [10], to studies of bacterial metabolism [11], food [12, 13] and environmental analysis [14]. In this manuscript, we present the results obtained with a recently described GC platform equipped with two detectors running in parallel - namely an atmospheric pressure chemical ionization source - mass spectrometer (GC-APCI-MS) and a flame ionization detector (FID). We applied this platform to the investigation of changes in metabolic profiles of urinary tract infection (UTI) patients which may assist in monitoring the infectious status of the patients.

MATERIALS AND METHODS

Chemicals

Methanol, pyridine, iso-propanol, toluene (all at analytical grade) and urease were purchased from Sigma (Zwijndrecht, the Netherlands). N-methyl-N-trimethylsilyltrifluoroacetamide with 1% trimethylchlorosilane (MSTFA + 1% TMCS) was used as a derivatization reagent (Pierce; Rockford. IL. USA) and taken from freshly opened 1 ml bottles. Methoxyamine hydrochloride and the mixtures of saturated alkanes (C7-C30 and C7-C40) were purchased from Supelco (Zwijndrecht, the Netherlands). APCI tune mix (containing purine, hexamethoxyphosphazine, hexakis (2,2-difluoroethoxy) phosphazine and hexakis (1H,1H,3H-tetrafluoropropoxy)phosphazine) and pyridine (99%. ultra-pure GC grade) were acquired from Fluka.

Clinical samples

Urine samples were collected during a prospective observational multicenter cohort study in which eight emergency departments (ED) of seven hospitals and 35 affiliating

primary health care centers participated. From January 2004 till November 2008, patients who presented with a diagnosis of febrile UTI, were considered for enrolment in the study. The study was approved by the local ethics committees and all patients gave written informed consent. Inclusion criteria and exclusion criteria have been described in detail elsewhere [15, 16]. In short, patients of 18 years or older, with fever and presenting at least one symptom of UTI (dysuria, increased frequency of urination, perineal pain, flank pain or costovertebral tenderness) and a positive nitrite dipstick test or a positive leukocyte esterase dipstick test were enrolled. Exclusion criteria were current treatment for urolithiasis or hydronephrosis, pregnancy, hemo- or peritoneal dialysis, a history of kidney transplantation or known presence of polycystic kidney disease. Details with respect to empiric therapy and follow-up are provided elsewhere [7]. Urine samples were collected at enrolment as baseline samples (t=0). After three and thirty days (t=30) urine samples of the same patients were collected for follow-up. At all the time points, clean midstream-catch urine samples were cultured and analyzed using standard microbiological methods. A positive urine culture was defined as bacterial growth of over 103 colony-forming units per ml. Within 2 hours after collection all samples were stored at -80° C until further analysis.

For the current study, a group of 67 subjects with culture-confirmed bacteria in the urine was selected from a database of approximately 700 subjects. Samples from volunteers without UTI or symptoms of another infection were taken as controls. Baseline characteristics of the samples are presented in the Table 1. The study design includes the control samples and samples of UTI patients collected at baseline (t=0 days) as well as patient follow up samples collected after the antibiotic treatment (t=30 days).

Table 1. Baseline characteristics. Data are presented as n (%) unless otherwise stated. IQR: interquartile range.

Characteristic	Controls n=53	Cases n=67
Age, years, median [IQR]	59[47–70]	68[54–78]
Male	25 (47)	32 (47)
Smoking	8 (1 missing value)	11
Uropathogen (1 missing value)		
E.coli	2	48
Haemolytic Streptococcus Group B		2
Enterococcus faecalis		3
Klebisella pneumoniae		4
Pseudomonas aeruginosa		2
Staphylococcus aureus		1
Staphylococcus saprophyticus		2
Klebsiella oxytoca		1
Aerococcus urinae		1
acinetobacter species		1
Candida spcies	1	
Co-morbidity		
Urinary tract disorder	13	16
Malignancy	4	10
Renal insufficiency	3	4
Heart failure	8	7
Diabetes mellitus	0	0
Urine dipstick results		
Nitrites	4 (5 missing values)	38 (8 missing values)
Leucocyte esterase	13 (3 missing values)	62 (3 missing values)

Sample preparation

Urine samples were prepared according to a reference method with minor modifications [17]. Briefly, 100 µl of urine were incubated with 20 Units of urease in 100 µl of water for 45 minutes at 37 ºC to deplete the excess of urea. Protein precipitation was carried out by adding 300 µl of cold methanol and by incubating 10 min on ice. After centrifugation at 19,600g for 10 min, 100 µl of the supernatant was transferred to a GC/MS vial and evaporated to dryness. All the samples were stored at -80 ºC until the moment of the GC analysis.

The samples were derivatized online with a double step reaction. Firstly, we added 50 µl of methoxyamine (20 mg/ml in pyridine) and incubated for 1 hour at 60 ºC. Secondly, we added 50 µl of MSTFA and incubated at 60 ºC for 40 minutes. Subsequently, we added 100 µl of toluene.

Hundred ul of each urine sample were pooled, aliquoted and then used as quality control which were prepared with the same protocol.

GC-APCI-MS-FID analysis

The samples (1 µl) were applied by splitless injection with a programmable MultiPurpose Sampler MPS2 (Gerstel, Mülheim an der Ruhr, Germany) into an Agilent 7890A GC (Agilent. Palo Alto. USA) equipped with one single column (HP-5-MS column 30 m. 0.25 mm ID. 0.25 µm film thickness) and two detectors running in parallel; a column flow splitter with deactivated capillaries (0.6 m. 0.25 mm ID. 0 µm film thickness) allowed the simultaneous acquisition with both the MS and FID detector. For sample injection, a septumless PTV injector was used. For each analysis the purge time of the PTV injector was set to 60s at a purge flow rate of 20 ml/min and an equilibration time of 1 min. Helium was used as carrier gas at a constant flow rate of 1 ml/min through the column.

Urine samples were injected with the following temperature program: 70 ºC held for 1 minute, the temperature increased till 280 ºC at 5.5 ºC/min and held for 10 minutes.

When identification of compounds was required, we spiked derivatized urine samples with hydrocarbons solution C7-C40 at 25 µg/ml and they were injected with the following temperature program: 70 ºC held for 1 minute, 5.5 ºC/min till 315 ºC held for 15 minutes.

The transfer line to the mass spectrometer was kept at 300 ºC. The APCI source was operated in positive mode. The temperature and the flow rate of the dry gas (nitrogen) were 300°C and 2.00 l/min, respectively. The APCI vaporizer temperature was 300 °C; the pressure of the nebulizer gas (nitrogen) was set to 3.5 bar and the voltage of the corona discharge needle was +2000 nA. Capillary voltage was set to -1000V and the end-plate offset to -500 V. The source and detector parameters have been set in order to achieve a minimal resolution of 40000.

A maXis 4G (Bruker Daltonics, Bremen, Germany), an orthogonal-accelerated Qq-ToF mass spectrometer, was used as mass analyzer. The spectra were acquired in a mass range from *m/z* 100-1000 at an acquisition rate of 1Hz. Before each chromatographic run, a tune mix solution was used as external calibrant and injected in Flow Injection mode.

For compound identifications Auto MS^2 experiments were performed. Collision energies were the following: *m/z* 100 at 25 V, *m/z* 1000 at 50 V. All the *m/z* values within this interval were fragmented with interpolated values of collision energy.

The FID flows and temperature were set as follows: front detector 300 °C, hydrogen flow 40 ml/min and air flow 370 ml/min. This detector was used for the calculation of the

Kovats' indices in comparison with the one provided in our publicly available database [18]. For the calculation, the equation relative to the temperature programmed chromatography was applied [19].

Statistical analysis

The GC-MS data files were exported as mzXML files. The peak picking was performed using XCMS package (The Scripps Research Institute, La Jolla, USA) using the default settings except for the following parameter: prefilter=c(2,1000), scan range = c(300, 2920), noise level =500, band width=15, *m/z* width=0.05, minimal sample=41, minimal fraction=1 [20].

The generated data matrix was imported to the SIMCA-P 13.0 software package (Umetrics, Umeå, Sweden). The data were mean centered, unit variance-scaled and Log transformed prior to the statistical analysis. The validity and the degree of overfitting of the PLS-DA models were checked using a 200 permutations test.

To identify metabolites of interest, rational chemical formulas were generated based on internally calibrated monoisotopic masses within 2 mDa mass error, using the SmartFormula tool within the DataAnalysis software package (Bruker Daltonics).

RESULTS

GC-APCI-MS is a relatively recent addition to the analytical toolbox for metabolomics investigations. Despite the growing number of publications showing the feasibility of this new technique, there are no reports demonstrating how it performs in exploratory studies. Thus, as a first step of the data analysis, we performed Principal Component Analysis (PCA) on the entire data set which included clinical samples and quality control samples (197 samples in total). The PCA score plot revealed a trend in the first principle component; we were not able to assign this effect to known sources of the analytical variability such as the acquisition block, the order of injection or the influence of the column. However, the PCA model built on clinical samples only (Figure 1) resembled the pattern known from our previous work on the same set of samples [8, 9]. Thus, we may conclude that our data described UTI related changes in the metabolic composition of the urine adequately, despite the unknown trend influencing the first principal component of the PCA model.

In Figure 1 (A and B), the PCA analysis of the 156 clinical samples, including the controls, the baseline samples (UTI patients t=0 days) and the follow-up samples (UTI patients t=30 days), is shown. Although the first principal component is representing the main axis of the variance, a trend separating the UTI-free and UTI-subjects was identified along the second and the third components.

In the next step, we focused on the analysis of the baseline (UTI patients t=0 days) and control samples. A PCA analysis of 116 samples (Figure 2A) showed a weak trend in the clustering of the UTI samples (t=0 days) versus the controls. Subsequently, a two-class PLS-DA model built on the same set of samples demonstrated a rather low predictive power (Q2(cum)=0.265; Figure 2 B).

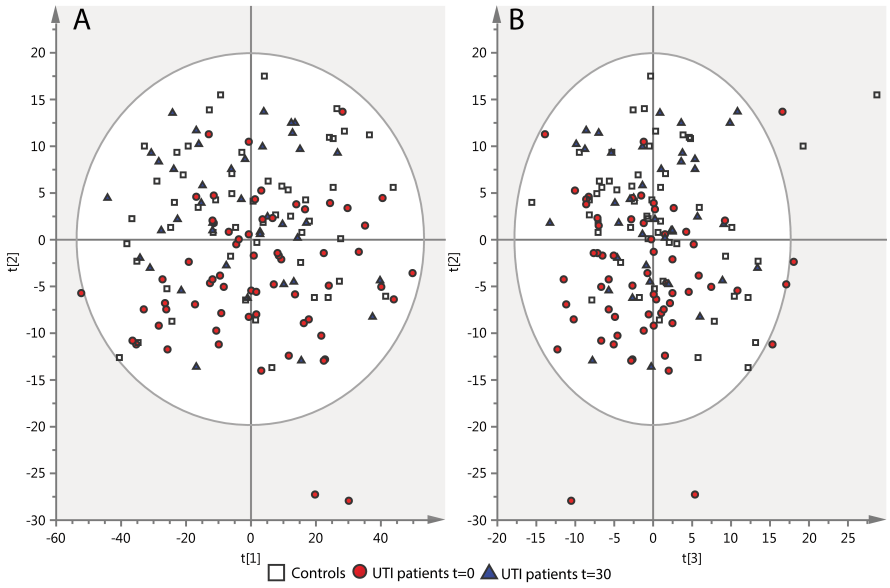


Figure 1. Clinical samples PCA model coloured by infectious status. The first and the second components display a major trend which cover more than 30% of the total variability (A). However, the second and the third component show a tendency in the cluster of the controls and UTI patients t=30 days against UTI patients t=0 day (B). R2X(cum)=0.642.

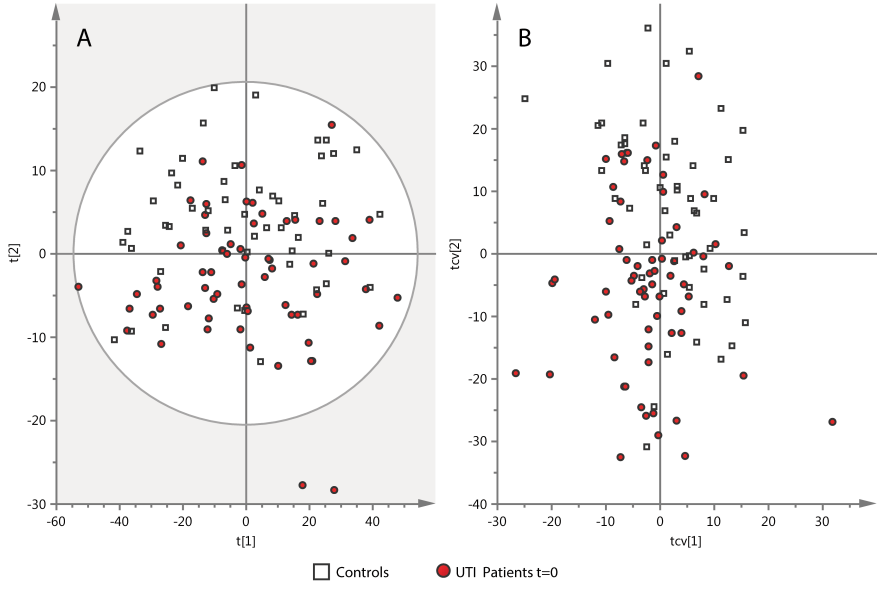


Figure 2. Baseline sample scatter plot, Controls versus UTI patients. A) PCA: R2X(cum)=0.622. B) Cross-validated PLS-DA: R2X(cum)=0.383, R2Y(cum)=0.444, Q2(cum)=0.265. The low statistical values express a poor separation of the controls against the UTI patients t=0 days.

Next, we took advantage of the longitudinal study design and built a two-class PLS-DA model using the time of the samples collection as class ID (t=0 and t=30 days). In figure 3 the PCA and PLS-DA models built on the longitudinal part of the dataset are depicted. As a result, the quality of the statistical models has clearly improved, displaying higher values for both the goodness of fit and goodness of prediction (Figure 3B).

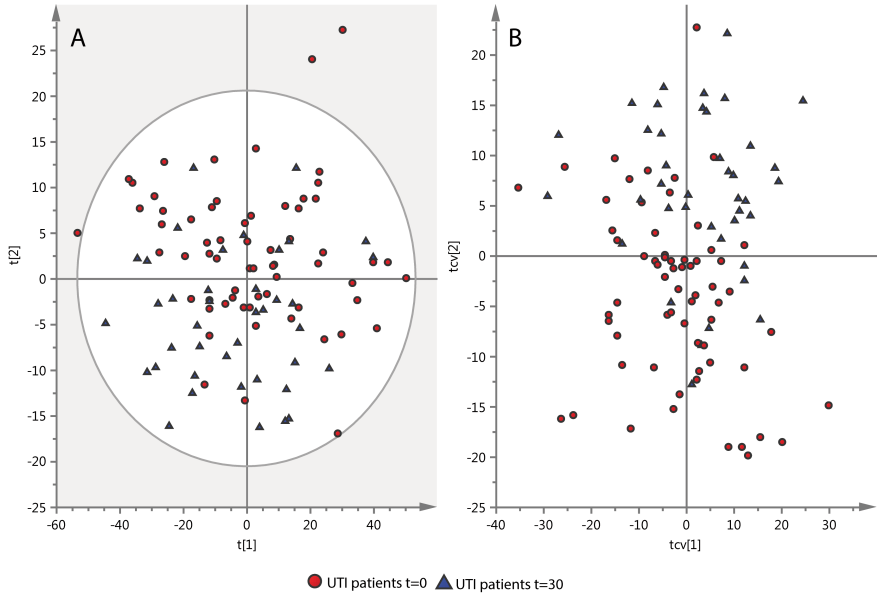


Figure 3. UTI patients scatter plot, UTI patients t=0 days versus UTI subjects t=30 days. A) PCA: R2X(cum)=0.604,. B) Cross-validated PLS-DA: R2X(cum)=0.389, R2Y(cum)=0.78, Q2(cum)=0.422. The improved values of the statistical models describe more adequately the infectious status displaying a clear trend of the UTI patients t=0 days against UTI patients t=30days.

To get an insight into the metabolic differences between the day of enrolment (t=0 days) and after the recovery (t=30 days), we focused on *m/z* signals with significant VIP values (> 1.5) within the PLS-DA model. As such, we selected 68 *m/z* values which could be grouped into 11 chromatographic peaks, when sorted by retention time. For the identification of the underlying compounds, we used two strategies. First, we checked whether the metabolites were present in our publicly available in-house developed database (http://metams.lumc.nl), using the spectra (MS and MS^2 spectra were compared by visual inspection) and the retention index estimated using FID (calculated errors for retention indices were kept below 1.5%). If the compounds were not present in the library, *de novo* identification, according to the protocol provided elsewhere [18], was applied. Table 2 summarizes the results of these analyses.

Two of the features with VIP value > 1.5 were identified as hydroxyhippuric acid. Hydroxyhippuric acid has three isomeric forms, 2-, 3- and 4-hydroxyhippuric acid. 2-Hydroxyhippuric acid is one of the main metabolites of salicylic acid; however, the use of aspirin was not reported in the metadata. Therefore, we reasoned that the observed signals correspond to 3- and 4-hydroxyhippuric acid. We also identified acetaminophen

glucuronide, one of the major metabolites of paracetamol; however, in comparison to our previous studies (LC-MS and NMR), its influence on the model appears to be less strong.

Table 2. Identification of the features with VIP values > 1.5

Compound	Experimental parent Ion *m/z*	Experimental Retention Index error (%)	Identification type
Unknown 1	307.1577		
Phosphoric acid (tri-TMS)	315.1025	0.72	In-house library
Unknown 2	287.1121		
Unknown 3	226.0892		
Unknown 4	287.1121		
Unknown 5	265.1435		
Citric Acid (tetra-TMS)	481.1929	0.29	In-house library
Myo-Inosiol (hexa-TMS)	613.3081	1.3	In-house library
3-Hydroxyhippuric Acid (di-TMS)	340.1391		*De novo*
4-Hydroxyhippuric acid (di-TMS)	340.1391		*De novo*
Acetaminophen Glucuronide (penta-TMS)	688.3006		*De novo*

One of special features of GC-MS at atmospheric pressure is the possibility to use a Flame Ionization Detector (FID) simultaneously with MS [21]. The FID is an extremely simple detector: burning of the carbon compounds in the hydrogen flame results in the production of ions and the detector's response is roughly proportional to the number of carbon atoms [22]. Due to an enormous dynamic range and a predictable response, the FID is often the detector of choice for quantitative analysis. Even though quantitative analysis was not part of this study, we sought to compare the FID responses with the quantitative data obtained for a number of metabolites in our previous study [9]. Figure 4 presents an example of such comparison for hippuric acid, a compound detected in both NMR and GC-MS studies. The response of the FID detector, expressed as area under the curve, correlates with the NMR data, while the correlation of the MS and NMR data was poor (data not shown).

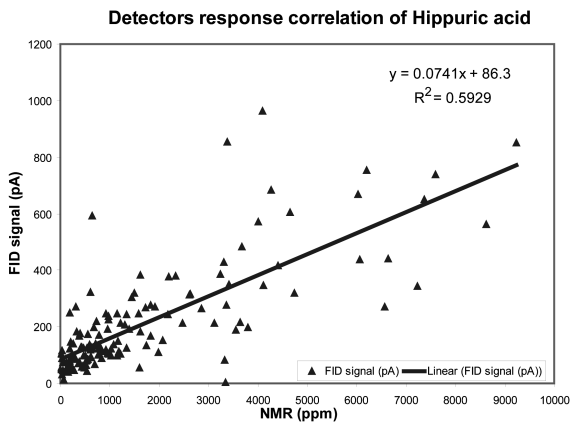


Figure 4. Hippuric acid response. The FID response, expressed as area under the curve, correlates with the quantitative NMR data. The deviations from the linearity might be explained with the presence of co-eluting peaks such as citric acid, which is present in high concentration in urine samples.

DISCUSSION

Here we present the first GC-APCI-MS-based explorative study of urines from UTI-patients. This work is part of a larger study in which we aim to create a comprehensive overview of UTI-induced changes in the metabolic patterns within the urine of these patients. The sample set used for the current study has previously been characterized with two other analytical platforms, namely NMR and LC-MS. Those two studies provided us with complementary pictures of the metabolic changes related to the UTI and the recovery process. NMR has revealed a set of metabolites indicative for the degree of the infection, while the LC-MS study resulted in the identification of O-glycosylated fragments of fibrinogen alpha chain as possible morbidity markers.

In comparison to the NMR and LC-MS studies, the GC-APCI-MS data set demonstrated some special features. Firstly, the influence of the exogenous compounds such as paracetamol metabolites, on the data matrix was much less pronounced than in both previous reports. Secondly, this data set contained a trend in the first principle component which we were not able to assign to known sources of analytical variability such as the acquisition block, the order of injection or the influence of the column. However, considering the number of principal components (in total 17) explaining the model and the fact that our PCA model on the GC-APCI-MS dataset closely resembled the pattern observed in our previous work, we attempted to explore the biologically relevant variance. To this end, we concentrated on the statistically strongest model, using the time of the samples collection as class ID (t=0 and t=30 days). In doing so, we observed 68 features with VIP values above 1.5, grouped into 11 chromatographic peaks. As Table 2 shows, we were able to provide structural identity only to a part of the classifiers but those which were identified are consistent with previous reports on UTI [9]. However, at the moment it is difficult to determine the degree of the overlap between GC-APCI-MS profiling and the other platforms used.

The features with the highest VIP values were identified as isomers of hydroxyhippuric acid - putatively 3- and 4- hydroxyhippuric acid. 3- and 4-Hydroxyhippuric acid are acyl glycines and usually seen as the products of microbial metabolism. However, hydroxyhippuric acid and especially 4- hydroxyhippuric acid is also implicated in the development of kidney failure [23] and was even proposed as a marker for monitoring the efficacy of dialysis [24]. More recently, 4-hydroxyhippuric acid has been indicated as a potent human erythrocytic Ca^{2+}- ATPase inhibitor in end-stage renal failure patients. [25]. With pyelonephritis being one of the frequent forms of complicated UTI, a link between hydroxyhippuric acid and kidney function provides an interesting topic for further study.

Citrate is another 'simple' metabolite with a complex functional network. Citrate is often solely considered as one of the intermediates of the TCA cycle. However, as a constituent of body fluids, it plays a different but equally important role in the regulation of homeostasis. The three carboxylic groups of a citrate molecule may form complexes with metal ions and/or other divalent cations; as such, citrate has been implicated in the control of renal stone formation by forming soluble complexes with calcium [26]. Since many years, citrate excretion in urine is used for the assessment of the renal function: for example, reduction of urinary citrate, which we observed for UTI patients at t=0 days, is an alerting signal for pathological changes in the renal metabolism. There are multiple reports showing the decrease of the urinary citrate in nephrolithiasis [27], glomerulonephritis [28] and more recently in infectious diseases [9, 29]. Thus, the results obtained with our GC-APCI-MS platform are in line with previous findings.

Even a brief overview of the metabolomics literature shows that the presence of inositol in urine, in its several isomeric forms, is an established fact. The picture seems a bit unclear, however, when it comes to the functional implication of urinary excretion of those compounds. A fraction of myo-inositol has a dietary origin, either as free inositol, phytic acid or inositol-containing phospholipids. Another and larger part is related to renal biosynthesis. Under normal conditions, up to 98% of myo-inositol is reabsorbed in the renal tubules. Consequently, increased concentrations of myo-inositol are considered as an indication of ongoing pathological changes in the renal system: increased concentrations of urinary myo-inositol were reported for patients with diabetes [30-32] and renal failure [33]. The increased levels of the myo-inositol observed in our study could also be interpreted as a consensus of acute pyelophritis.

In conclusion, this profiling platform combines high resolution mass spectroscopy data with the robustness and the wide dynamic range of the FID detector but also opens new possibilities for cross-platform data analysis, particularly in combination with such a well-established metabolomics platform as NMR.

CONCLUSION

Here, to the best of our knowledge, we have presented the first application of an explorative metabolomics study on a GC-APCI-MS platform. As part of a bigger project aimed at the comprehensive overview of UTI-induced changes in urine, this study points some physiological marker of the febrile UTI status. Although obviously no firm biological conclusions can yet be drawn, this study nevertheless has provided findings which may further on contribute to the understanding of the physiological mechanisms underlying the progression of UTI.

REFERENCES

[1] Pinson AG, Philbrick JT, Lindbeck GH, Schorling JB. Fever in the clinical diagnosis of acute pyelonephritis. American Journal of Emergency Medicine. 1997;15:148-51.

[2] Kunin CM. Definition of acute pyelonephritis vs the urosepsis syndrome. Archives of Internal Medicine. 2003;163:2393-.

[3] Wilson ML, Gaido L. Laboratory diagnosis of urinary tract infections in adult patients. Clin Infect Dis. 2004;38:1150-8.

[4] Lv H, Hung CS, Chaturvedi KS, Hooton TM, Henderson JP. Development of an integrated metabolomic profiling approach for infectious diseases research. Analyst. 2011;136:4752-63.

[5] Ramautar R, Mayboroda OA, Derks RJE, van Nieuwkoop C, van Dissel JT, Sornsen GW, et al. Capillary electrophoresis-time of flight-mass spectrometry using noncovalently bilayer-coated capillaries for the analysis of amino acids in human urine. Electrophoresis. 2008;29:2714-22.

[6] Gupta A, Bansal N, Houston B. Metabolomics of urinary tract infection: a new uroscope in town. Expert Review of Molecular Diagnostics. 2012;12:361-9.

[7] van Nieuwkoop C, Bonten TN, van't Wout JW, Kuijper EJ, Groeneveld GH, Becker MJ, et al. Procalcitonin reflects bacteremia and bacterial load in urosepsis syndrome: a prospective observational study. Critical Care. 2010;14.

[8] Pacchiarotta T, Hensbergen PJ, Wuhrer M, van Nieuwkoop C, Nevedomskaya E, Derks RJE, et al. Fibrinogen alpha chain O-glycopeptides as possible markers of urinary tract infection. Journal of Proteomics. 2012;75:1067-73.

[9] Nevedomskaya E, Pacchiarotta T, Artemov A, Meissner A, van Nieuwkoop C, van Dissel JT, et al. H-1 NMR-based metabolic profiling of urinary tract infection: combining multiple statistical models and clinical data. Metabolomics. 2012;8:1227-35.

[10] Bristow T, Harrison M, Sims M. The application of gas chromatography/atmospheric pressure chemical ionisation time-of-flight mass spectrometry to impurity identification in Pharmaceutical Development. Rapid Communications in Mass Spectrometry. 2010;24:1673-81.

[11] Wachsrnuth CJ, Almstetter MF, Waldhier MC, Gruber MA, Nurnberger N, Oefner PJ, et al. Performance Evaluation of Gas Chromatography-Atmospheric Pressure Chemical Ionization-Time-of-Flight Mass Spectrometry for Metabolic Fingerprinting and Profiling. Analytical Chemistry. 2011;83:7514-22.

[12] Portoles T, Cherta L, Beltran J, Hernandez F. Improved gas chromatography-tandem mass spectrometry determination of pesticide residues making use of atmospheric pressure chemical ionization. Journal of Chromatography A. 2012;1260:183-92.

[13] Portoles T, Mol JGJ, Sancho JV, Hernandez F. Advantages of Atmospheric Pressure Chemical Ionization in Gas Chromatography Tandem Mass Spectrometry: Pyrethroid Insecticides as a Case Study. Analytical Chemistry. 2012;84:9802-10.

[14] David F, Sandra P, Hancock P. Determining High-Molecular-Weight Phthalates in Sediments using GC-APCI-ToF-MS. Lc Gc Europe. 2011;24:16-+.

[15] van Nieuwkoop C, Bonten TN, van't Wout JW, Becker MJ, Groeneveld GH, Jansen CL, et al. Risk Factors for Bacteremia with Uropathogen Not Cultured from Urine in Adults with Febrile Urinary Tract Infection. Clinical Infectious Diseases. 2010;50:E69-E72.

[16] van der Starre WE, van Dissel JT, van Nieuwkoop C. Treatment duration of febrile urinary tract infections. Current infectious disease reports. 2011;13:571-8.

[17] Jiye A, Huang Q, Wang GJ, Zha WB, Yan B, Ren HC, et al. Global analysis of metabolites in rat and human urine based on gas chromatography/time-of-flight mass spectrometry. Anal Biochem. 2008;379:20-6.

[18] Pacchiarotta T, Derks RJE, Hurtado-Fernandez E, van Bezooijen P, Henneman A, Schiewek R, et al. Online spectral library for GC-atmospheric pressure chemical ionization–ToF MS. Bioanalysis. 2013;5:1515-25.

[19] Kovats E. GAS-CHROMATOGRAPHISCHE CHARAKTERISIERUNG ORGANISCHER VERBINDUNGEN .1. RETENTIONSINDICES ALIPHATISCHER HALOGENIDE, ALKOHOLE, ALDEHYDE UND KETONE. Helvetica Chimica Acta. 1958;41:1915-32.

[20] Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: Processing mass spectrometry data for metabolite profiling using Nonlinear peak alignment, matching, and identification. Analytical Chemistry. 2006;78:779-87.

[21] Pacchiarotta T, Nevedomskaya E, Carrasco-Pancorbo A, Deelder AM, Mayboroda OA. Evaluation of GC-APCI/MS and GC-FID as a complementary platform. Journal of biomolecular techniques : JBT. 2010;21:205-13.

[22] Holm T. Aspects of the mechanism of the flame ionization detector. Journal of Chromatography A. 1999;842:221-7.

[23] Nielsen VK. NERVE CONDUCTION. Acta Medica Scandinavica. 1975;197:14-9.

[24] Schoots AC, Devries P, Thiemann R, Hazejager WA, Visser SL, Oe PL. BIOCHEMICAL AND NEUROPHYSIOLOGICAL PARAMETERS IN HEMODIALYZED PATIENTS WITH CHRONIC RENAL-FAILURE. Clinica Chimica Acta. 1989;185:91-107.

[25] Jankowski J, Tepel M, Stephan N, van der Giet M, Breden V, Zidek W, et al. Characterization of p-hydroxy-hippuric acid as an inhibitor of Ca2+-ATPase in end-stage renal failure. Kidney international Supplement. 2001;78:S84-8.

[26] Rudman D, Dedonis JL, Fountain MT, Chandler JB, Gerron GG, Fleming GA, et al. HYPOCITRATURIA IN PATIENTS WITH GASTROINTESTINAL MALABSORPTION. New England Journal of Medicine. 1980;303:657-61.

[27] Angel Arrabal-Polo M, Arrabal-Martin M, Arias-Santiago S, Garrido-Gomez J, Poyatos-Andujar A, Zuluaga-Gomez A. Importance of citrate and the calcium : citrate ratio in patients with calcium renal lithiasis and severe lithogenesis. Bju International. 2013;111:622-7.

[28] Psihogios NG, Kalaitzidis RG, Dimou S, Seferiadis KI, Siamopoulos KC, Bairaktari ET. Evaluation of tubulointerstitial lesions' severity in patients with glomerulonephritides: An NMR-Based metabonomic study. Journal of Proteome Research. 2007;6:3760-70.

[29] Balog CIA, Meissner A, Goraler S, Bladergroen MR, Vennervald BJ, Mayboroda OA, et al. Metabonomic investigation of human Schistosoma mansoni infection. Molecular Biosystems. 2011;7:1473-80.

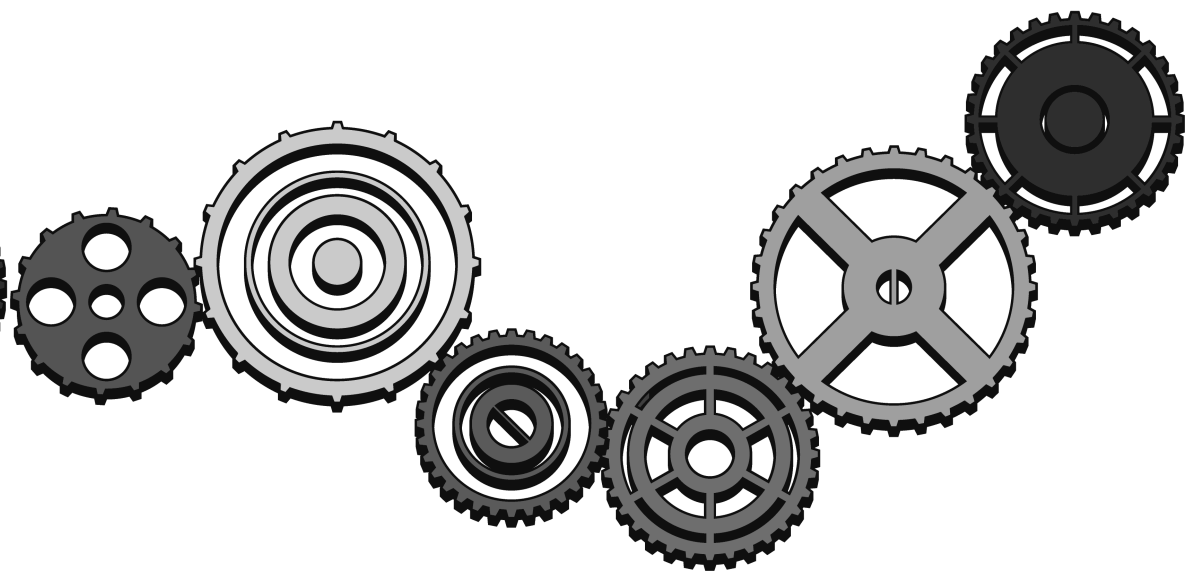
[30] Daughaday WH, Larner J, Houghton E. THE RENAL EXCRETION OF INOSITOL BY NORMAL AND DIABETIC RATS. Journal of Clinical Investigation. 1954;33:1075-80.

[31] Sarashina G, Yamakoshi M, Noritake M, Takahashi M, Kure M, Katsura Y, et al. A study of urinary myo-inositol as a sensitive marker of glucose intolerance. Clinica Chimica Acta. 2004;344:181-8.

[32] Hong JH, Jang HW, Kang YE, Lee JH, Kim KS, Kim HJ, et al. Urinary chiro- and myo-inositol levels as a biological marker for type 2 diabetes mellitus. Disease Markers. 2012;33:193-9.

[33] Niwa T, Yamamoto N, Maeda K, Yamada K, Ohki T, Mori M. GAS-CHROMATOGRAPHIC MASS-SPECTROMETRIC ANALYSIS OF POLYOLS IN URINE AND SERUM OF UREMIC PATIENTS - IDENTIFICATION OF NEW DEOXYALDITOLS AND INOSITOL ISOMERS. Journal of Chromatography. 1983;277:25-39.

General discussion and conclusion



Metabolomics is a technology driven discipline but unlike genomics or even proteomics there is not a single analytical platform which enables a comprehensive overview of the entire metabolic space. One can always find an argument in favor of (or against) a particular analytical solution but, regardless of the argumentation, any analytical technology provides only a slice of the metabolome. Thus, with this thesis one particular clinical problem has been studied, namely urinary tract infection, using several analytical approaches: this included well-established methods such as NMR and LC-MS, as well as a novel GC-APCI-MS platform. Each platform was used for the investigation of the urinary metabolite profile of a carefully matched selection of UTI patients and controls.

In chapter two we used a non-targeted profiling by NMR showing the number of metabolites such as acetate, lactate and trimethylamine which are correlated with number of bacteria (CFU/ml). However, more interesting results using the NMR technique were obtained when the longitudinal axis of the study design was used rather than the simpler case/control approach. A few compounds such as p-aminohippuric acid and scyllo-inositol were proposed as the markers of the recovery process. Those compounds could reflect the renal plasma flow and glomerular filtration efficiency and thus of the kidneys' functionality.

The NMR study in chapter two has once more demonstrated that an exploratory approach has some advantages of over the targeted. A non-targeted study gives the opportunity to discover the unknown associations in the data and uses, to the full extend, the multivariate nature on NMR and mass spectrometric data sets. For instance, the exploratory study performed with LC-MS (chapter 3) demonstrated a new and unexpected finding: using a combination of unsupervised and supervised multivariate modeling we discovered a fibrinogen alpha 1-chain peptide as potential morbidity marker. The release of proteolytic fragments of fibrinogen has earlier been reported in response to an infection or involved in cases of kidney disease is not exciting itself and is not new [1]. Recently, antimicrobial activity has also been attributed to the peptide released from fibrinogen [2]. The novelty of the findings reported here lies in the structure of the fibrinogen that carries a glycosylation which changes our understanding of the fibrinogen structure that was considered glycosylation free in this part of its chain. Certainly this observation needs further validation in a larger and more heterogeneous cohort of samples; nevertheless the increased release of the glycosylated fibrinogen alpha chain in febrile UTI patients has shown a reasonably good predictive value.

NMR and LC-MS are the most frequently used techniques in metabolomics; both can rely on broad range of data processing software solutions (commercial and open source), spectral libraries and data bases. However, a quest for novel analytical solutions is still ongoing since none of the existing methods matches chemical diversity and wide concentration range of the metabolites comprising human metabolome. At first glance GC-MS hardly can be presented as a novel method, this method was used for the investigation of metabolic profiles in body fluids long before the term metabolomics was introduced [3]. The modernization of such a technique flows into a variety of platforms which find applications in any type of metabolic profiling. Nonetheless, GC-MS under atmospheric pressure (GC-PCI-MS) is certainly a novel (for metabolomics) method.

Indeed, for GC-APCI-MS an extensive evaluation of the analytical performance is still needed. As indicated in the chapters 4 and 5, the combination of gas chromatography and mass spectrometry under atmospheric pressure is not entirely a new concept but rather a re-introduction of an old idea in a modern, user-friendly form. The value of this method for metabolomics was recognized soon after its reintroduction [4] and it allows the combination

of mass spectrometrer as detector with an addition, simple but nonetheless powerful detector - FID. Chapter 4 describes the detailed analytical evaluation of this idea and demonstrates the applicability of this platform to the study of the metabolite profile in biological samples. The dual detector set-up may be very convenient for the analysis of complex biological samples such as for example body fluids. On the one hand, GC-APCI-TOF-MS offers high-quality data that are essential for the structural assignment of the metabolites. Using the APCI source, in fact, the protonated molecule ($[M+H]^+$) is the dominant feature in the spectra, that provides the possibility for structural characterization of unknown compounds using the combination of the accurate mass, the isotopic pattern and the MS^n experiments. On the other hand, in explorative studies, the relative abundance of the observed compounds in the metabolic space covers a wide dynamic range which the mass spectrometer cannot cover. In this respect, FID is one of the few detectors capable of handling such wide concentration ranges and it can provide more reliable quantitative data.

Despite the growing number of studies which demonstrate its value, GC-APCI-MS is not yet considered as a routine platform. There are a few reasons for this but the lack of a spectral library certainly limits the practical value of GC-APCI-MS for metabolomics studies. The development of such library, as described in the chapter 5 of this thesis, might facilitate the identification of metabolites in body fluids. The spectral library presented here includes the majority of the compounds most commonly encountered in the analysis of body fluids, such as amino acids, sugars, organic acids and so on. The web-based library is fully searchable and publicly available. As any web-based application, it is planned to continuously update our database and the first practical applications are included in the chapters 5 and 6.

Finally, the application of GC-APCI-MS platform proved to be successful for the explorative study of the urinary metabolites in the context of UTI (Chapter 6). In this work, uremic solutes such as myo-inositol and citrate were found to play a role in the characterization of UTI patients as well as the altered concentration of two isomers of hydroxyhippuric acid, which is known as a potent human erythrocytic Ca^{+2} - ATPase inhibitor in end-stage renal disease. This might reflect the link between the tissue invasive process in febrile UTI, such as pyelonephritis, and functionality of the kidney.

The joint effort of these technologies has allowed a comprehensive overview of the complex sequence of the physiological events unfolding the process of an infection. The analytes detected with different techniques reflect the various aspects of the UTI such as the effect of the bacterial growth, the affected renal functionality and also the production of aberrant peptides.

In a longer and speculative perspective, this type of markers could potentially be used in routine clinical chemistry and for therapy decision making.

FUTURE PERSPECTIVES

The multiplatform approach, here reported in the UTI context, demonstrate the value of metabolomics as general profiling tool in a clinical environment, providing a better insight into altered metabolic phenotype and disease pathogenesis.

Despite the explosive growth of the field of metabolomics, 'clinical metabolomics' is still in its infancy, but with the analytical technologies and the computing power evolving rapidly it is to be expected that it will not take long before metabolomics will be fully integrated

into clinical and epidemiological studies. But can we foresee which analytical solution will become 'the standard' of metabolomics? If a question "What is the best analytical technique for metabolomics?" is going to be asked, no straightforward answer will be possible. The chance to find a molecule (or a set of molecules) with high clinical impact depends on the ability to efficiently combine the right technology with a meaningful study design. As shown for UTI, each analytical platform uncovers a slice of the biology of a certain status and the key of the success lies in the formulation of the proper biological question. From that point of view, one should choose the proper collection of samples and instrument according to the phenomenon of interest.

This thesis has shown the complexity of the UTI problem demonstrating how the metabolic patterns measured with different analytical instrumentation can reflect different physiological processes: bacterial infection (NMR), the host response (LC-MS) and the physiological status (GC-APCI-MS). Each of this aspect is a valuable point in the extensive understanding of the UTI as infectious disease and could be useful during the assessment of the disease severity or the decision making.

All this was possible due to the flexibility in the selection of the analytical instrumentation and an appropriate study design. From this experience, it is clear that with regard to generating markers with a potentially high clinical impact, the key of success does not lay in the chosen platform itself but what makes those markers suitable for practical application is the right biological question, a study design that describes adequately the phenomenon of interest, and the last but not the least practical considerations such as the number of samples and the platform robustness.

REFERENCES

[1] Metzger J, Kirsch T, Schiffer E, Ulger P, Mentes E, Brand K, et al. Urinary excretion of twenty peptides forms an early and accurate diagnostic pattern of acute kidney injury. Kidney International. 2010;78:1252-62.

[2] Pahlman LI, Morgelin M, Kasetty G, Olin AI, Schmidtchen A, Herwald H. Antimicrobial activity of fibrinogen and fibrinogen-derived peptides - a novel link between coagulation and innate immunity. Thrombosis and Haemostasis. 2013;109:930-9.

[3] Pauling L, Robinson AB, Teranish.R, Cary P. QUANTITATIVE ANALYSIS OF URINE VAPOR AND BREATH BY GAS-LIQUID PARTITION CHROMATOGRAPHY. Proceedings of the National Academy of Sciences of the United States of America. 1971;68:2374-&.

[4] Carrasco-Pancorbo A, Nevedomskaya E, Arthen-Engeland T, Zey T, Zurek G, Baessmann C, et al. Gas Chromatography/Atmospheric Pressure Chemical Ionization-Time of Flight Mass Spectrometry: Analytical Validation and Applicability to Metabolic Profiling. Analytical Chemistry. 2009;81:10071-9.

Appendices

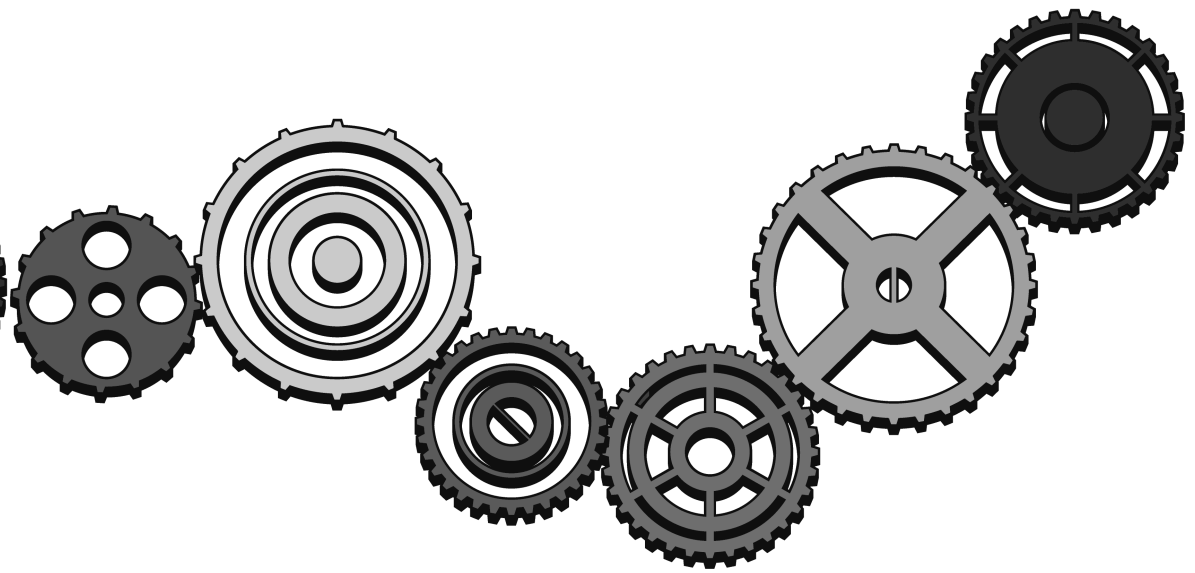
Summary

Nederlandse samenvatting

Acknowledgments

Curriculum vitae

List of publications



SUMMARY

Metabolomics is a rapidly evolving post-genomics discipline which focuses on the analysis of metabolites. The totality of an organism's metabolites –the metabolome– is considered to be the closest approximation of the phenotype of a given organism. Analysis of the metabolome and its changes in the complex transition between health and sickness may contribute to our understanding of human physiology and open new ways for diagnostics. However, an implementation of 'metabolomics thinking' in clinical research is still in its relatively early days. On the other hand metabolomics itself has to adopt to the requirements of clinical research. These two elements are essential in the process of convergence between clinical and analytical components which should lead to the recognition of metabolomics of body fluids as a fully accepted measure in medical research. This thesis represents a further step in this direction using a multiplatform metabolomics approach to investigate one of the most frequent infectious diseases– urinary tract infection (UTI).

Chapter 1 shows a detailed overview of the explorative metabolomics studies of human infections with the emphasis on metabolic adaptation to an acute or chronic infection.

The relevant literature has been comprehensively summarized including, for example, information on the study design, number of subjects, employed technology and metabolic discriminators. Future considerations, such as importance of the time-resolved study designs and the embedment of metabolomics in large-scale epidemiological studies are discussed.

This thesis provides an exhaustive exploratory study of the metabolic pattern of patients affected by urinary tract infection. UTI is a common infectious disease that encompasses a variety of clinical syndromes with a positive bacterial culture as common denominator. Here, this complex clinical entity was investigated with a multiplatform approach.

Chapter 2 describes the application of nuclear magnetic resonance (NMR) to the study of UTI-induced changes in the urinary profile. Firstly, multivariate data analysis adequately described the difference between cases and controls. The models built on the classical case/control design correctly predicted the outcome of the recovered patients in 86.5 % of the cases. However, this separation was partially biased by the presence of paracetamol metabolites in the urine samples. In this respect, more complex data analysis approaches provided more relevant information of the infectious status. For instance, pair–wise analysis of the UTI patients and their follow-up samples revealed a number of classifiers with improved statistical significance as well as the regression-based analysis of the relation between ^{1}H-NMR data and the bacterial load in urine. The lists of discriminators obtained from different models presented some overlap such as acetate, lactate, and trimethylamine, which are often associated with bacterial growth, but also showed some unique features such as para-aminohippuric acid, a well-established diagnostic marker for glomerular filtration.

Chapter 3 provides an additional perspective to the same set of samples by using another well-established technique in metabolomics namely liquid chromatography coupled with mass spectrometry (LC-MS). In this case, the combination of unsupervised and supervised multivariate modeling indicated an altered concentration of O-glycosylated proteolytic fragments of the human fibrinogen alpha 1–chain in UTI patients. The unequivocal identification of this peptide required MS^2 and MS^3 experiments on two different platforms –namely ESI-UHR-Qq-ToF and ESI-ion trap MS–, manual spectra interpretation for *de novo* sequencing, BLAST searches and, finally, a comparison of the experimental tandem

mass spectra with those of custom synthesized candidate peptides. The novelty of these findings lies in the observed modification of the existing concept of the fibrinogen structure which was considered, to the best of our knowledge, to be free of glycosylation in this part of its chain.

As was mentioned above, clinical metabolomics is still in its infancy and as such the field is open for new analytical solutions in order to provide markers with high clinical impact.

In this thesis, the application of one of the oldest separation techniques, gas chromatography (GC), in a renewed and more user friendly role, showed complementary results to those obtained with a NMR and LC-MS platform. GC in combination with an atmospheric pressure chemical ionization source (APCI) can provide new insight in the metabolic profile. This possible value was recognized from the very first day in the seventies when the technology was applied to the analysis of urinary organic acids but it was never investigated in detail. Moreover, the use of GC-APCI-MS in combination with one of the oldest GC detection methods, namely flame ionization (FID), allows a parallel acquisition.

Chapter 4 presents the extensive evaluation of the basic analytical parameters for both detectors such as limit of detections, limit of quantification, reproducibility intra– and inter– day, and linear dynamic range. As expected, FID appeared to be more robust over time, showing lower relative standard deviations while MS demonstrated higher sensitivity, reflected in lower limits of detection. The benefits of the dual detection was evident in the analysis of complex biological samples where the wide dynamic range and the predictable molar response of the FID may offer a quantitative overview of the sample composition. On the other hand, APCI–MS provides superior information for structural elucidation. The analysis of complex real samples such as cerebrospinal fluid showed that this combination represents a powerful tool for explorative studies.

Despite the advantage of such a built platform for explorative studies, real clinical applications are still lacking. There several reasons for this but the lack of a spectral library, that makes the application of the traditional sources so powerful, is one of them.

Chapter 5 reports the first web-based spectral library for GC-APCI-MS. This library is publicly available at http://metams.lumc.nl/ , fully searchable and facilitates many search options such as metabolite name, molecular formula and KEGG number. Moreover, a specific MS spectrum search is also allowed using the measured *m/z* value with a set mass tolerance. The entries include MS and MS^2 spectra, and retention indices calculated with the FID detector. At the moment, the library contains the metabolites mostly encountered in the analysis of body fluids (amino acids, sugars, fatty acids, vitamins) but, as any web-tool, it is planned to be continuously updated. Its application was tested in the analysis of complex samples (serum). However, in the analysis of biological samples one might encounter compounds which are not yet included in the library or unknown compounds. In order to facilitate the identification of such molecules, a protocol for *de novo* identification is also provided.

Finally, **Chapter 6** describes the first application of GC-APCI-MS in a clinical environment to investigate the UTI-induced change in the urinary metabolic profile. The multivariate statistical analysis allowed the clustering of the UTI patients against their follow

up which, on the basis of bacterial culture results, showed no longer sign of infection. The previous developed spectral library (Chapter 5) allowed the identification of the molecules driving this separation by matching both the spectra and the retention indices. Moreover, some of the compounds not present in the database could be identified with the *de novo* protocol. The list of molecular descriptors includes uremic solutes such as myo-inositol and citrate but also isomers of hydroxyhippuric acid, which might be descriptors of a certain physiological status since their presence has previously been related to the Ca^{2+}–ATPase inhibitory activity in end–stage renal failure patients. In addition, the robustness and dynamic range of the FID detector was used to investigate new possibilities for cross-platform data analysis, especially in combination with NMR and LC-MS.

This thesis shows the complexity of the UTI problem and the complementarity of different analytical tools which are able to measure different metabolic patterns, reflecting diverse physiological processes. Each of the used platforms added a unique perspective to the further understanding of the infection process. The assessment of the bacterial growth (NMR), of the host response (LC–MS) and of the physiological status (GC-APCI-MS) could eventually be useful during the assessment of the disease severity and/or decision making.

NEDERLANDSE SAMENVATTING

Metabolomics is een snel groeiende post-genomics tak van de wetenschap die zich bezighoudt met de analyse van metabolieten. Het totaal aan metabolieten in een organisme, het metaboloom, wordt gezien als de meest nauwkeurige representatie van het fenotype. De analyse van het metaboloom, en de veranderingen ervan bij de overgang van 'gezond' naar 'ziek', kan een bijdrage leveren aan het beter begrijpen van de humane fysiologie en zodoende mogelijk nieuwe diagnostische methoden verschaffen. Echter, de implementatie van het denken in termen van metabolomics in klinische studies is tot nu toe beperkt. Omgekeerd zullen metabolomics studies aangepast moeten worden aan de eisen die klinisch wetenschappelijk onderzoek met zich meebrengt. Het samengaan van deze klinische en analytische aspecten zijn de belangrijkste factoren voor de acceptatie van metabolomics van lichaamsvloeistoffen in het medisch onderzoek. Dit proefschrift beoogt aan dit proces een bijdrage te leveren door de toepassing van een multi-platform metabolomics benadering bij onderzoek aan urineweginfecties, één van de meest voorkomende infectieziekten.

Hoofdstuk 1 geeft een overzicht van de exploratieve metabolomics studies van humane infecties met de nadruk op de metabole veranderingen die gepaard gaan met acute en chronische infecties. Er wordt een uitgebreid overzicht gegeven van de relevante literatuur met inachtneming van informatie over studie opzet, aantal patiënten, gekozen technologie en metabole veranderingen. Toekomstperspectieven zoals het belang van longitudinale studies en de implementatie van metabolomics in grootschalige epidemiologische studies worden bediscussieerd.

Het proefschrift behandelt verder uitgebreide exploratieve analyses van de metabole patronen van patiënten met urineweginfecties (UTI). UTI is een veelvoorkomende infectieziekte met een verscheidenheid aan klinische verschijnselen met als gemeenschappelijk kenmerk een positieve bacteriële kweek. In deze studie is deze complexe klinische entiteit onderzocht met een verscheidenheid aan analytische platforms.

In **hoofdstuk 2** is UTI onderzocht door middel van NMR (nuclear magnetic resonance) analyse van urine met als doel de UTI-geïnduceerde veranderingen in kaart te brengen. Multivariate data analyse resulteerde in eerste instantie in een duidelijk onderscheid tussen patiënten en controles. Een model gebaseerd op deze eenvoudige patiënt-controle studie gaf in 86.5% van de gevallen een correcte voorspelling van de patiënten die hersteld waren. Echter, de aanwezigheid van paracetamol speelde hierin een belangrijke rol. Een uitgebreidere analyse gaf een meer relevante weergave van de infectiestatus. Een paarsgewijze analyse van urines van patiënten zowel tijdens de infectie als ook na herstel gaf een duidelijker en significanter onderscheid tussen de verschillende groepen weer. Dit gold ook voor de op regressieanalyse gebaseerde beschrijving van de relatie tussen de proton NMR data en de hoeveelheid bacteriën in de urine. Sommige metabolieten kwamen in meerdere analyses als onderscheidend naar boven, zoals bijvoorbeeld acetaat, lactaat en trimethylamine. Deze metabolieten worden vaak geassocieerd met groei van bacteriën. Daarnaast werden er ook model specifieke metabolieten aangetoond, zoals bijvoorbeeld para-aminohippurinezuur, een algemeen geaccepteerde diagnostisch marker voor glomulaire filtratie.

In **hoofdstuk 3** is dezelfde set aan samples geanalyseerd met een reeds algemeen geaccepteerde techniek in metabolomics, LC-MS (vloeistofchromatografie gekoppeld met massaspectrometrie). Uit deze dataset werd via een combinatie van zowel 'supervised' als 'unsupervised' multivariate analyse een verandering in de concentratie van een

O-geglycosyleerd peptide van de alpha-1 keten van fibrinogeen gevonden in urines van mensen met UTI. De identificatie van dit peptide werd mogelijk gemaakt door het gebruik van MS^2 en MS^3 experimenten op twee verschillende massaspectrometrisch platforms, ESI-UHR-Qq-ToF en ESI-ion trap, handmatige interpretatie van de verzamelde spectra en sequentie-specifieke zoekopdrachten in databases. Uiteindelijk vond de positieve identificatie plaats door middel van het vergelijken van de experimentele fragmentatiespectra met die van een, op basis van bovenstaande analyse, gesynthetiseerd peptide. Dit is naar ons inziens de eerste keer dat in dit deel van fibrinogeen glycosylering is aangetoond en mede daardoor is deze ontdekking zo opvallend.

Zoals boven beschreven is metabolomics een relatief jong vakgebied en dat biedt dus de mogelijkheid tot het vinden van nieuwe analytische oplossingen om te komen tot markers met een hoge klinische impact.

In dit proefschrift blijkt dat een 'zeer oude' scheidingstechniek zoals GC (gaschromatografie) in een nieuwe en meer gebruikersvriendelijkere opzet resultaten oplevert die complementair zijn aan die van NMR en LC-MS. Gekoppeld aan chemische ionisatie onder atmosferische druk (APCI) kan GC nieuwe inzichten geven in het metabole profiel. De kracht van deze combinatie werd al in de jaren zeventig onderkend in een studie gericht op de analyse van urinezuren, maar is nooit verder in detail onderzocht. Bovendien geeft de bijzondere combinatie van GC-APCI-MS en de oudste GC detectie methode, vlamionisatie (FID), de mogelijkheid tot parallelle data acquisitie (zie hieronder).

In **hoofdstuk 4** is voor beide detectiemethoden een uitgebreide evaluatie gedaan van de analytische parameters zoals gevoeligheid, reproduceerbaarheid (zowel binnen een dag als tussen verschillende dagen) en het lineaire dynamische gebied. Zoals verwacht bleek FID in de tijd een meer robuuste techniek en gaf als zodanig lagere standaarddeviaties terwijl MS gevoeliger was en dus een lagere detectielimiet had. Het voordeel van de combinatie van beide detectietechnieken was het meest duidelijk bij de analyse van complexe biologische monsters. In deze gevallen geven het hoge dynamische bereik en de voorspelde molaire respons van FID de mogelijkheid om een kwantitatief overzicht van de monstersamenstelling te geven. Aan de andere kant is APCI-MS essentieel voor de opheldering van de structurele kenmerken van de aanwezige componenten. Uit de analyse van een complex monster zoals cerebrospinale vloeistof werd duidelijk dat deze combinatie een krachtige methode is voor exploratieve studies.

Ondanks de mogelijkheden van het boven beschreven platform is de echte klinische toepassing beperkt . Eén van de redenen hiervoor is het gebrek aan databases met spectra, zoals die in het verleden met veel succes zijn toegepast bij andere manieren van ionisatie.

In **hoofdstuk 5** wordt de eerste web-based database met GC-APCI-MS spectra beschreven. Deze database is voor iedereen toegankelijk via http://metams.lumc.nl. In deze database kan gemakkelijk gezocht worden op bijvoorbeeld de naam van een metaboliet, de molecuulformule of een KEGG nummer. Daarnaast is er de mogelijkheid om met MS spectra te zoeken gebruikmakend van de *m/z*-waarde en een gegeven tolerantie. De database bestaat uit zowel MS als MS^2 spectra gecombineerd met de retentietijd indices zoals die uit de FID data zijn berekend. Op dit moment bevat de database gegevens van de belangrijkste metabolieten die in een lichaamsvloeistof aanwezig zijn (aminozuren, suikers, vetzuren, vitamines) maar het is de bedoeling dat deze, zoals iedere webtool, in de toekomst steeds wordt bijgewerkt. We hebben de database getest aan de hand van een analyse van serum monsters. Echter, de analyse van dit soort monsters geeft altijd een verscheidenheid aan

moleculen die nog niet in de database staan. Om de identificatie hiervan te vergemakkelijken hebben we een protocol voor *de novo* identificatie geïmplementeerd.

In **hoofdstuk 6**, tenslotte, wordt GC-APCI-MS voor het eerst in een klinisch georiënteerd onderzoek toegepast. Hiervoor zijn de veranderingen in het urineprofiel van patiënten met UTI geanalyseerd. De multivariate statistische analyse gaf een duidelijke clustering van UTI-monsters t.o.v. de urinemonsters van dezelfde mensen die, op basis van een bacteriële kweek, na verloop van tijd hersteld waren. De in hoofdstuk 5 beschreven database maakte het mogelijk om de moleculen die ten grondslag lagen aan deze clustering op basis van zowel de spectra als de retentietijd te identificeren. Daarnaast konden enkele nieuwe metabolieten worden opgehelderd middels het nieuwe *de novo* protocol. De lijst van moleculaire descriptoren omvat zowel myo-inositol en citraat maar ook isomeren van hydroxyhippurinezuur. De fysiologische relevatie van hydroxyhippurinezuur blijkt onder andere uit studies waarbij het de aanwezigheid hiervan gekoppeld werd aan de remming van de Ca^{2+}-ATPase tijdens de eindfase van patiënten met nierfalen.

In dit deel van de studie werden verder de robuustheid en het dynamisch bereik van de FID detector getest met betrekking tot nieuwe mogelijkheden voor cross-platform data analyse methoden, voornamelijk in combinatie met NMR en LC-MS.

Dit proefschrift laat de complexiteit van het UTI-vraagstuk zien en geeft de complementariteit weer van verschillende analytische methoden die metabole veranderingen, die ten grondslag liggen aan fysiologische processen, zichtbaar kunnen maken. Elk van de analytische platforms gaf een unieke bijdrage aan het beschrijven van UTI-gemedieerde metabole veranderingen tijdens het infectieproces. De bepaling van bacteriële groei (NMR), de gastheer respons (LC-MS) en de fysiologische status (GC-APCI-MS) kunnen in de toekomst mogelijk een rol spelen bij het vaststellen van de hevigheid van de infectie en de daaraan te koppelen therapeutische aanpak.

ACKNOWLEDGEMENTS

This PhD-time has flown by quickly as always happens when I enjoy what I do. This experience is far more than a challenging achievement in my scientific career, this experience represents the biggest opportunity I had for professional and personal growth. In these years there have been plenty of people who walked into or out of my life; I can certainly say that each of these people has contributed to this adventure and that I am pleased to acknowledge. I would like to start from the people who believed on me in the first place, André and Oleg.

André, I will never thank you enough for giving me the chance to join this group. I truly appreciate the wideness of your scientific understanding and, of course, I am amazed by everything you have built in your career. However what I mostly admire is your capability to be an efficient superior as well as an exquisite person.

Oleg, after so many years of working together I still confirm what I said in one of my first presentation in this department 'thanks for being Oleg'. You provided me guidance and support whenever I needed that, while preserving my scientific freedom and let me make my own mistakes. What I have learned from you is priceless and I will always carry your advice with me.

A huge thanks goes to all my colleagues over these years who are or have been part of this group. You all have taught me something.

Katja, we shared so much in these years. Not only long working day but also a lot of great moments. You are a special person who I deeply admire.

Alegría. Everything has started with you, we liked each other at the very first glance. Working with you is always a great pleasure and a delightful experience.

Bart and Rico, you certainly made this thesis possible. Thank you so much for being at my side any time I have needed this.

Paul, I am glad we have written a chapter of this thesis together. The best way of learning how to be a good scientist is working with one.

Kate, Sibel, and Erica, we did so many things together that you contributed to my social well-being.

I have been so fortunate that there would be many more people to mention though I do not have enough space for all. I am grateful to the incredible crew of colleagues who turned into special friends. The working day flows easier if you spend it with a nice companion. We had an incredible time together and I am glad I have shared the most important moments of my life with you.

Any person who lives far from the native country knows how important is to have close friends. I feel really fortunate because I have found more than that, I have found a real family. Irene, Pietro, Simona, Francesco, Valentina, Gloria, Raul, Lara, Rafael and Manu, you have a special place in my heart and I know that wherever life will bring us you will always be part of me. No word would really express how much I appreciate to have you in my life.

Although a thousand kilometers divide us, my dearest Italian friends, you also had a part in this thesis. You are my happy island when things here become too dense. It looks like yesterday that we were university mates and now that we are 'grown-up' we still do things with the same enthusiasm. You are my past and my future.

Alla mia famiglia, grazie per essermi stati sempre vicino soprattutto quando ne ho avuto piú bisgono. La lontananza divide in nostri giorni ma non i nostri cuori.

Thank you little Diego, your smile fills my days of joy. You changed my life and my

priorities teaching me that I cannot be a good scientist before being a good mother for you. You remind me every day that each person, scientist and not, should deliver to the next generation a better world than the one they found.

Simone, this is one thing more we have achieved in this long walk together that our lives are. None of this would be possible without you by my side and in my heart. During these years we have been fiancé, parents, friends, confidants, colleagues and much more. I can say with no reservation that you are my happiness.

Tiziana
Leiden, 2014

CURRICULUM VITAE

Tiziana Pacchiarotta was born on July 28, 1981, in Avezzano, Italy. In 2000, after completing her studies at a scientific high school, she continued her education in Chemistry at the University of Rome 'La Sapienza'. She obtained her degree in 2006 with a master thesis entitled 'Development of analytical methods based on LC-MS/MS to measure algal toxins in fish and lake water. Monitoring of Albano Lake'. Her first job was at the Institute of Atmospheric Pollution of the National Council of Research, Italy. She was employed for the analysis of organic micropollutants such as dioxins, PAH, PCB in industrial emissions. In 2008 she spent 10 months at the University of Almería as fellowship student under the supervision of Prof. Dr. A. R. Fernandez-Alba. The core of her project was the analysis of pesticides residues in food matrices. The same year, she was employed as PhD student under the supervision of Prof. Dr. A.M. Deelder and Dr. Oleg. A. Mayboroda at the Department of Parasitololgy, LUMC, the Netherlands, in a collaborative research project with Prof. Dr. Jaap van Dissel (Department of Infectious Diseases) founded by Stichting Tropische Geneeskunde – Leiden. That research work resulted in this thesis entitled 'Metabolomics of urinary tract infection: a multiplatform approach'. Since 2012 she is post-doc at the Centre for Proteomics and Metabolomics, LUMC. She is currently working with Dr O.A. Mayboroda and Prof Dr. J.W. de Fijter (Department of Nephrology) for the investigation of the urinary metabolic profile of patients affected by renal diseases such as polycystic kidney disease and after kidney transplantation.

LIST OF PUBLICATIONS

1. Hurtado-Fernandez E, **Pacchiarotta T**, Longueira-Suarez E, Mayboroda OA, Fernandez-Gutierrez A, Carrasco-Pancorbo A: Evaluation of gas chromatography-atmospheric pressure chemical ionization-mass spectrometry as an alternative to gas chromatography-electron ionization-mass spectrometry: Avocado fruit as example. *Journal of Chromatography A* 1313, 228-244 (**2013**).

2. **Pacchiarotta T**, Derks RJ, Hurtado-Fernandez E et al.: Online spectral library for GC-atmospheric pressure chemical ionization-ToF MS. *Bioanalysis* 5(12), 1515-1525 (2013).

3. **Pacchiarotta T**, Hensbergen PJ, Wuhrer M et al.: Fibrinogen alpha chain O-glycopeptides as possible markers of urinary tract infection. *Journal of Proteomics* 75(3), 1067-1073 (**2012**).

4. Nevedomskaya E, **Pacchiarotta T**, Artemov A et al.: H-1 NMR-based metabolic profiling of urinary tract infection: combining multiple statistical models and clinical data. *Metabolomics* 8(6), 1227-1235 (**2012**).

5. **Pacchiarotta T**, Deelder AM, Mayboroda OA: Metabolomic investigations of human infections. *Bioanalysis* 4(8), 919-925 (**2012**).

6. Hurtado-Fernandez E, **Pacchiarotta T**, Gomez-Romero M et al.: Ultra high performance liquid chromatography-time of flight mass spectrometry for analysis of avocado fruit metabolites: Method evaluation and applicability to the analysis of ripening degrees. *Journal of Chromatography A* 1218(42), 7723-7738 (**2011**).

7. Garcia-Villalba R, **Pacchiarotta T**, Carrasco-Pancorbo A et al.: Gas chromatography-atmospheric pressure chemical ionization-time of flight mass spectrometry for profiling of phenolic compounds in extra virgin olive oil. *Journal of Chromatography A* 1218(7), 959-971 (**2011**).

8. **Pacchiarotta T**, Nevedomskaya E, Carrasco-Pancorbo A, Deelder AM, Mayboroda OA: Evaluation of GC-APCI/MS and GC-FID as a complementary platform. *Journal of biomolecular techniques* 21(4), 205-213 (**2010**).

9. Bogialli S, D'ascenzo G, Di Corcia A, Innocenti G, Lagana A, **Pacchiarotta T**: Monitoring quinolone antibacterial residues in bovine tissues: extraction with hot water and liquid chromatography coupled to a single- or triple-quadrupole mass spectrometer. *Rapid Communications in Mass Spectrometry* 21(17), 2833-2842 (**2007**).