Volume 3 | Issue 2 | December 2013 ISSN 2182-0287

Journal of Integrated OMICS

A methodological journal

JIOMICS

Journal of Integrated OMICS

Focus and Scope

Journal of Integrated OMICS, JIOMICS, provides a forum for the publication of original research papers, preliminary communications, technical notes and critical reviews in all branches of pure and applied "-omics", such as genomics, proteomics, lipidomics, metabolomics or metallomics. The manuscripts must address methodological development. Contributions are evaluated based on established guidelines, including the fundamental nature of the study, scientific novelty, and substantial improvement or advantage over existing technology or method. Original research papers on fundamental studies, and novel sensor and instrumentation development, are especially encouraged. It is expected that improvements will also be demonstrated within the context of (or with regard to) a specific biological question; ability to promote the analysis of molecular mechanisms is of particular interest. Novel or improved applications in areas such as clinical, medicinal and biological chemistry, environmental analysis, pharmacology and materials science and engineering are welcome.

Associated editors

Australia and New Zealand

Food Research Division, Bureau of Chemical Safety, Health Canada, Ottawa, Canada

Xuequn Chen

Department of Molecular & Integrative Physiology, University of Michigan, Ann Arbor, USA

Ye Fang

Biochemical Technologies, Science and Technology Division, Corning Incorporated, USA

Ying Qu

Microdialysis Experts Consultant Service, San Diego, USA

Ying Xu

Department of Biochemistry and Molecular Biology, Institute of Bioinformatics, University of Georgia, Life Sciences Building Athens, GA, USA

Preface

DOI: 10.5584/jiomics.v3i2.162

Dear Associated Editors, Dear colleagues and Dear friends,

I am proud to announce the launch of the second issue of the Journal JIOMICS corresponding to 2013.

In this issue Ge et al. have written a nice review dealing with environmental Omics, an emerging scientific field that links the OMICS areas with the environment. They point out that OMICS breakthroughs are empowering the fields of environmental toxicology, chemical toxicity characterization, and health risk assessment. It is expected that future environmental OMICS will focus more on real environmental issues and challenges such as the characterization of chemical mixture toxicity, the identification of environmental and health biomarkers, and the development of innovative environmental OMICS approaches and assays. García -Sevillano et al. present a step forward in her career working with little mammals as (bio)indicators of pollution. In this case they have characterized the biological response to metals in liver, brain, kidneys, lungs and plasma of the free-living mouse *Mus spretus* in polluted areas located in Doñana National Park (southwest Spain) and the surroundings, mainly affected by agriculture, mining and industry activities. Very interesting results are presented relating metals and proteins in this organism. Colgrave et al. publish an article about profiling of the membrane compartment of bovine testis cell populations. The authors describe that the proteomic profiles generated in the published work support and complement transcriptomics studies and reinforce the potential of proteomics in identifying and characterising the protein effectors of self-renewal and/or differentiation in stem cells. Allmer shows us how to the C-Terminal Amino Acid of a Peptide from MS/MS Data is determined using new software. As 90% of protein terminal peptides may not end with either arginine or lysine and may thus contain any of the other amino acids, an algorithm is presented which predicts the c-terminal amino acid to be arginine, lysine or any other. Halouska et al. show us, which are the best protocols for the NMR Analysis of Bacterial Metabolomes, considering factors as collection, handling, processing and analysis of metabolomics data. Gonçalves et al. give a new perspective on the expansion of the resistance to antibiotics in bacteria present in animals, in this case belonging to a captive program This study showed specimens of Iberian lynx acting as reservoirs of resistance genes. The authors highlight that in future (re)-introductions they could spread resistant bacteria throughout the environment. Mangapiane et al. explain us through an elegant work entailing an integrated proteomic and physiological approach, the adhesion mechanism of the probiotic *Lactobacillus reuteri* Lb2 BM DSM 16143 concluding that that the moonlighting proteome of this lactobacillus can contribute to adhesion processes.

Radhouani et al. present an interesting work related to the changes in the proteome of *Escherichia coli* stressed by antibiotics. The comparison of stressed vs non-stressed bacteria proteome revealed that the abundance of numerous protein species changed in the strain stressed by CTX compared to the non-stressed wild-type strain. Baudin et al. in an interesting work address the morphogenic changes in endothelial cells from human umbilical vein induced by artificial angiogenesis concluding that phorbol esters modify a number of proteins involved in multiple and intricate pathways for promoting a phenotype ensuring cell survival and cell migration for new vessels formation. Wang-Fat et al. use protein/peptide profiles to identify cyano-bacterial species. An interesting work related to environmental OMICS. The authors concluded that through this method the fast identification of harmful cyanobacteria is possible.

As you have read above, this issue brings an excellent number of manuscripts, providing a level of excellence to JIOMICS. I would like to kindly acknowledge all authors for their excellent work and for trust in JIOMICS, disseminating their research through this journal.

I take this opportunity to wish you Merry Christmas and a Happy New Year 2014 full of Success.

Yours truly,

On behalf of the Editors-in-Chief,

José-Luis Capelo-Martínez

REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal.

JOURNAL OF INTEGRATED OMICS

A methodological Journal

CONTENTS OF VOLUME 3 | ISSUE 2 | DECEMBER 2013

REVIEW ARTICLES

ORIGINAL ARTICLES

JOURNAL OF INTEGRATED OMICS
A METHODOLOGICAL JOURNAL
HTTP://WWW.JIOMICS.COM

REVIEW ARTICLE | DOI: 10.5584/jiomics.v3i2.141

Environmental OMICS: Current Status and Future Directions

Yue Ge*[1], Da-Zhi Wang[2], Jen-Fu Chiu[3], Susana Cristobal[4], David Sheehan[5], Frédéric Silvestre[6], Xianxuan Peng[7], Hui Li[7], Zhiyuan Gong[8], Siew Hong Lam[8], Hu Wentao[5], Hitoshi Iwahashi[9], Jianjun Liu[10], Nan Mei[11], Leming Shi[11], Maribel Bruno[1], Heidi Foth[12], Kevin Teichman[13]

[1]National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency; [2]State Key Laboratory of Marine Environmental Science, College of the Environment and Ecology, Xiamen University, China; [3]University of Hong Kong and Shantou University College of Medicine, China; [4]Linkoping University, Sweden; [5]Department of Biochemistry, University College Cork, Ireland; [6]Research Unit in Environmental and Evolutionary Biology, University of Namur, Belgium; [7]Center for Proteomics, State Key Laboratory of Bio-Control, School of Life Sciences, Sun Yat-Sen University, China; [8]Department of Biological Sciences, National University of Singapore, Singapore; [9]Department of Applied Biological Sciences, Gifu University, Japan; [10]Shenzhen Center for Disease Control and Prevention, China; [11]National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR 72079, USA; [12]Institute for Environmental Toxicology, Martin Luther University, Halle/Saale, Germany; [13]Office of Research and Development, U.S. Environmental Protection Agency, Washington D.C., USA.

Received: 05 May 2013 **Accepted:** 24 September 2013 **Available Online**: 29 September 2013

Abstract

Objectives: Applications of OMICS to high throughput studies of changes of genes, RNAs, proteins, metabolites, and their associated functions in cells or organisms exposed to environmental chemicals has led to the emergence of a very active research field: environmental OMICS. This developing field holds an important key for improving the scientific basis for understanding the potential impacts of environmental chemicals on both health and the environment. Here we describe the state of environmental OMICS with an emphasis on its recent accomplishments and its problems and potential solutions to facilitate the incorporation of OMICS into mainstream environmental and health research.

Data sources: We reviewed relevant and recently published studies on the applicability and usefulness of OMICS technologies to the identification of toxicity pathways, mechanisms, and biomarkers of environmental chemicals for environmental and health risk monitoring and assessment, including recent presentations and discussions on these issues at The First International Conference on Environmental OMICS (ICEO), held in Guangzhou, China during November 8-12, 2011. This paper summarizes our review.

Synthesis: Environmental OMICS aims to take advantage of powerful genomics, transcriptomics, proteomics, and metabolomics tools to identify novel toxicity pathways/signatures/biomarkers so as to better understand toxicity mechanisms/modes of action, to identify/categorize/prioritize/screen environmental chemicals, and to monitor and predict the risks associated with exposure to environmental chemicals on human health and the environment. To improve the field, some lessons learned from previous studies need to be summarized, a research agenda and guidelines for future studies need to be established, and a focus for the field needs to be developed.

Conclusions: OMICS technologies for identification of RNA, protein, and metabolic profiles and endpoints have already significantly improved our understanding of how environmental chemicals affect our ecosystem and human health. OMICS breakthroughs are empowering the fields of environmental toxicology, chemical toxicity characterization, and health risk assessment. However, environmental OMICS is still in the data generation and collection stage. Important data gaps in linking and/or integrating toxicity data with OMICS endpoints/profiles need to be filled to enable understanding of the potential impacts of chemicals on human health and the environment. It is expected that future environmental OMICS will focus more on real environmental issues and challenges such as the characterization of chemical mixture toxicity, the identification of environmental and health biomarkers, and the development of innovative environmental OMICS approaches and assays. These innovative approaches and assays will inform chemical toxicity testing and prediction, ecological and health risk monitoring and assessment, and natural resource utilization in ways that maintain human health and protects the environment in a sustainable manner.

Keywords: OMICS; Environmental; Health; Chemicals; Toxicology; Genomics; Proteomics; Metabolomics; Biomarkers; Metagenomics; Risk Assessment; Biomarker.

***Corresponding author:** Yue Ge, Ph.D. Integrated Systems Toxicology Division, National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711 USA; Tel: 919-541-2202; Fax: 919-541-0694; Email Address: Ge.yue@epa.gov

1. Introduction

The suffix "-omics" was applied to describing heterogeneous networks of objects by physicists and computer scientists who produced novel papers on scale-free network properties in biological systems in the 1990s [1]. It later became a specific word describing the science that was comprehensively embraced by the four disciplines of genomics, transcriptomics, proteomics, and metabolomics in the early 2000s [2]. OMICS has developed rapidly in recent decades, triggered by the improvements in genome decoding techniques and high-throughput technologies enabling profiling of mRNA, proteins, and metabolites. OMICS-based data collectively provide a snapshot picture of gene expression, protein expression, and metabolite pattern, which altogether can enable much deeper insight into how tested organisms cope with external stressors. Applications of OMICS technologies to environmental toxicology and health research resulted in the emergence of a new research field: environmental OMICS. Environmental OMICS is the applications of OMICS technologies including genomics, transcriptomics, proteomics, and metabolomics to better understand the environmental and genetic factors, toxicity mechanisms, and modes of action in response to both acute and chronic exposure to environmental chemicals and, in the long-term, development of diseases caused or influenced by these exposures (Figure 1). Environmental OMICS is still in the early stage of OMICS data collection and validation of the molecular profiles for identifying toxic mechanisms, toxicity signatures, biomarkers, and pathways after exposure to environmental chemicals.

Environmental OMICS research can be roughly divided into three categories. The first category focuses mainly on chemical toxicity and environmental monitoring enabling risk assessment. The second category focuses on adverse human health outcomes and environmental impacts, while the last category focuses on ecological functions and environmental adaptation. The primary goal of these research fields is similar, namely to identify molecular changes, especially changes at expression levels of mRNA, proteins, and metabolites, in cells or tissues exposed to environmental toxicants and to relate these molecular changes to ecological and health outcomes.

Environmental OMICS has been used to study toxicity mechanisms and short- and long-term effects of environmental chemicals on human health outcomes, to define the acceptable levels and potential impacts of environmental toxicants on sensitive target species and ecosystems, to pro-

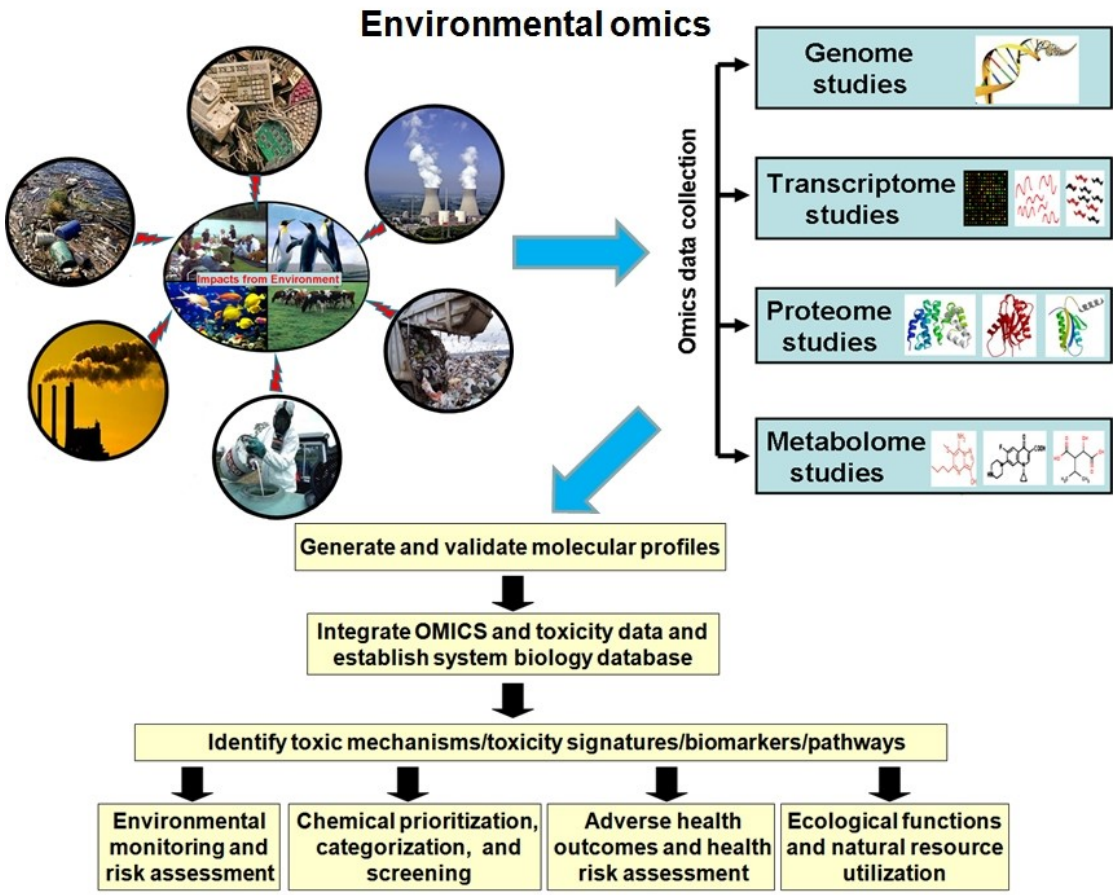


Figure 1. Environmental OMICS and its applications

vide insights into yet unsolved problems of environmental risk assessment such as chemical mixtures and combined effects of different environmental stressors, and to uncover unknown microbial communities and other natural resources. For example, genomic, proteomic, and metabolomic technologies have been used extensively to study the molecular mechanisms of how arsenic acts as a carcinogen [3, 4]. Also, the difference in gel electrophoresis (DIGE) proteomics technology has been used to decipher toxicity pathways and detoxification pathways of nanomaterials, protein-interacting network maps, biological response, potential toxicity, and detoxification pathways in titanium dioxide-treated BEAS-2B cells [5]. Proteomics was also applied to the investigation of differential proteomes of environmental bacteria for a better understanding of antibiotic and antibiotic-resistance mechanisms [6].

Genomic, proteomic, or metabolomic expression profiling is probably the most popular application of OMICS to environmental toxicology and health risk research. These and other OMICS studies help to reduce uncertainties associated with the ecological and health risk assessment process by deciphering toxicity mechanisms and modes of action [7-9], identifying biomarkers of exposure and toxicity [10], studying toxic effects and environmental diseases [11, 12], and facilitating cross-species extrapolation [13, 14]. A combination of genomic, proteomic, and bioinformatic approaches was also used to study toxic mechanisms of fungicides [15], and toxicity pathways of endocrine disrupting chemicals with different known or unknown toxic modes of action [16]. These integrated OMICS studies provided novel insights into toxic mechanisms and/or modes of action of environmental chemicals. Data from these experiments were used for human health risk assessment of the chemical, and established the basis of toxicity prediction approaches for species, endpoint, and chemical extrapolation.

In addition to the expression profiling studies, identification of modifications at the gene, protein, and metabolite levels is another important application of OMICS to environmental toxicology and human health research [17, 18, 19]. It is known that protein carbonylation and phosphorylation are among the major signal transduction pathways in cell biology. They are currently being studied to determine whether they are also the key to estimate toxicity pathways for environmental pollutants. For example, to understand the contributions of oxidative stress to toxicity, an integrated proteomics approach involving the identification of carbonylated proteins was utilized for the systematic measurement of protein oxidation in the livers of propiconazole-treated mice [17]. This study suggested a mode of propiconazole-induced toxicity in the mouse liver that primarily involves oxidative damage to cellular proteins. In another set of experiments, Enan and Matsumura [20] reported that the environmental chemical 2, 3, 7, 8-tetrachlorodibenzo-*p*-dioxin (TCDD) at very low concentrations was found to cause a rapid rise in protein phosphorylation activities in the extranuclear fraction of the adipose tissue from male guinea pigs. They are currently being investigated to determine whether they are also the key to estimate toxicity pathways for environmental pollutants, and have received comparatively more attention than other protein modifications in environmental toxicology research.

Table 1 is a summary of the environmental OMICS articles that were published from June 2011 to June 2012. In the past year, 611 genomic, 231 proteomic, and 84 metabolomics studies have been published based on a search of PubMed. Publications based on the chemicals, subjects, purposes, experimental systems, and goals that were used in these studies were also categorized to get an overview of the developing

Table 1. Summary of environmental OMICS publications from June, 2011 to June 2012

		Transcriptomics	Proteomics	Metabolomics	Systems Biology
No. of publications		611	231	84	598
% of publications	*Single chemicals*	70	98	86	96
	Chemical mixtures	30	2	14	4
	Health	72	52	46	38
	Ecosystem	28	48	54	62
	Mechanisms	30	34	8	8
	Pathway/biomarkers	70	66	92	92
	In vitro	56	42	40	48
	In vivo	44	58	60	52
	Environmental monitoring	80	52	54	58
	Health risk assessment	20	48	46	42

trends in environmental OMICS. The criteria to categorize the published literature were whether known contaminants were studied either as single chemicals or mixtures; whether risk factors for human health or environmental damage should be clarified further, and whether mechanisms (modes of action) and pathways were studied that are beyond the parameters for approval of hazardous properties. As shown in Table 1, genomics (611 publications) and proteomics (231 publications) are still the major "work horses" in environmental OMICS. Surprisingly, the number of environmental OMICS studies using a systems biology approach was ranked the second (598 publications), suggesting a systems biology approach is becoming increasingly important to environmental research. In addition, environmental OMICS studies focusing on toxicity pathways, biomarkers, and adverse health outcomes received more attention than studies focusing on either toxicity mechanisms or ecosystem toxicity. Interestingly, *in vitro* and *in vivo* experimental systems are almost equally used in the OMICS studies. However, OMICS studies focusing on chemical mixtures were much less numerous than studies focusing on individual chemicals. As for the techniques used in OMICS studies, it is worth mentioning that the percentage of the environmental proteomic studies using mass spectrometry-based approaches was almost the same as those using 2-D gel based approaches. This suggests that the previous dominance of the 2-D gel electrophoresis as a proteomic tool is waning.

While substantial progress has been made in certain areas of environmental OMICS, further improvements are still needed to facilitate the incorporation of OMICS into environmental toxicology and health research. There are a number of experimental design, sample collection, technical implementation, and data interpretation issues that impede the use of OMICS approaches in environmental and health research. In environmental OMICS research, many biological samples come from outbred individuals, populations, and communities. The field of environmental OMICS is thus a diverse and heterogeneous discipline, often involving a complex array of organisms and multifactorial experiments that can have an extremely large number of measurable parameters. Therefore, environmental OMICS strategies and approaches for the selection

and control of measurable parameters, as well as the identification and capture of essential information associated with environmental toxicity and health risks, should be established. Technical guidelines for standardization of OMICS research procedures are crucial, which need to cover sampling, technical data analysis, and interpretation of results, as well as the definition of cut off criteria, reference points, and normal values.

This paper provides an overview of recent environmental OMICS research including the presentations at The First International Conference on Environmental OMICS (ICEO) with an emphasis on the applications of genomics and proteomics technologies and methodologies to environmental science and health research. In the following discussions, we do not attempt to be exhaustive, but rather focus on some representative and traditional genomics and proteomics research topics such as OMICS characterizations of chemical toxicity, oxidative stress, and protein post-translational modifications. We also discuss some emerging research topics such as OMICS characterization of chemical mixtures and OMICS-based environmental monitoring and adaptation. Some comments on the challenges, potential solutions, and future directions in these research fields are also provided in this review. Our aim in writing this review is to stimulate interest in a bold, new, scientifically rigorous, and comprehensive environmental OMICS strategy for environmental toxicology and health research.

2. Environmental OMICS research issues

2.1 *OMICS-based vs. traditional environmental toxicology*

Traditional environmental toxicology and environmental health research mainly focus on the effects of environmental chemicals on the functions of various organ systems or toxicity phenotypes and modes of actions of the chemicals at cellular levels, usually relying on animal models for these studies [5, 21]. In addition to being labor-, time- and resource-intensive, this approach is primarily descriptive in nature and is low throughput and unable to characterize the full spectrum of targets and toxicity mechanisms for chemicals that affect multiple systems. To understand the toxicity mechanisms and/or modes of action, there is a need to understand the toxic processes at the molecular level. This will involve integrating genomic, proteomic, and metabolomic technologies for toxicology research.

Genomics was probably the first OMICS approach to be applied to mechanistic studies of toxicity, toxicity pathways, tumor biomarkers, and carcinogenicity. These studies involved environmental chemicals and their mixtures and mainly focused on characterization of the levels of gene expression that is reflected by the abundance of specific mRNA transcripts in a biological sample [22]. Gene expression profiling was used to compare the expression profiles of groups of genes with and without exposure to environmental chemicals. Although gene expression profiling experiments, especially those involving microarray technology, revolutionized numerous aspects of biological research and enabled thousands of gene transcripts to be monitored simultaneously, transcriptional responses often do not accurately reflect important toxicologically relevant biological responses. This could be due to the fact that there are important changes in proteins and metabolites within cells that are not detectable by just studying the levels of mRNAs. Therefore, proteomics and metabolomics are often used to complement genomics for toxicological and health studies [16, 3].

Proteomics is the large-scale study of protein expression and related biological functions. Proteins are ultimately functional molecules involved in most cellular processes. Toxic responses are driven by interactions between chemi-

cals and biomolecular targets, many of which are proteins. Most toxicological endpoints are preceded by changes of protein expression. Therefore, proteomic-based toxicity studies and biomarkers are highly-relevant to biological functions, adverse health outcomes, and health risk assessment.

Metabolomics, the study of metabolic profiles consisting of small metabolites, is the latest addition to the OMICS family. Metabolites can be collected from urine, saliva, and plasma. The formation of metabolites is probably the final manifestation of gene expression alterations. Therefore, metabolomics is a potentially useful tool for characterizing phenotypes under normal physiological and pathological conditions. Metabolic profiling supports and confirms the mechanisms derived from genomics and proteomics [23].

Although genomics, proteomics, and metabolomics target different molecules that may regulate and control biochemical pathways, as well as biological activities and events at different levels, the informative OMICS data of mRNA, proteins, and metabolites need to be integrated to achieve an effective and comprehensive understanding of modes of action and mechanisms of chemical-induced toxic responses and disease processes.

Implementation of OMICS technologies into environmental toxicology and environmental health research has been catalyzed by the report of the U.S. National Research Council (NRC) on Toxicity Testing in the 21st Century [24]. This report described how OMICS technologies could dramatically increase the efficiency and accuracy in evaluating both chemical toxicity and adverse human health outcomes through the development of novel OMICS toxicological endpoints, toxicity pathways, and biomarkers for chemical toxicity testing. Researchers working on environmental toxicology and health are increasingly turning to the application of OMICS technologies to answer fundamental questions in environmental sciences [15, 25].

The major objectives of environmental OMICS include elucidation of molecular mechanisms of toxicity, xenobiotic interactions with biological systems, and identification of mRNA, protein, and/or metabolite signatures or biomarkers for chemical toxicity testing, environmental monitoring, and human health risk assessment. Human health risk assessment evaluates long term effects of exposure to contaminants, frequently giving special attention to individual sensitivity such as preexisting disorders. For ecotoxicology, the focus is on monitoring of exposure and effect, which depends on the strengths of biomarkers and signatures for the stability of local and regional ecosystems. Additionally, environmental OMICS technologies have been widely used to address a variety of environmental and human health issues, such as effects of climate changes on different species [26], plant and animal responses to complex environmental stressors in soil and air [23, 27], effects of marine stressors such as acidification and hypoxia on sentinel organisms [28], chronic or acute exposures to metals in aquatic organisms and humans [29, 30], effects of exposure to novel pollutants such as nanomaterials [5, 31] and environmental microorganisms [32]. OMICS can often detect molecular changes before the appearance of visible morphological or physiological changes and thus can predict toxicity and reduce the time needed for more traditional toxicity testing.

Current applications of OMICS to environmental toxicology mainly focus on identifying and relating changes at RNA, protein, and metabolite expression levels, and how these factors reflect changes in networks or pathways in cells or tissues after exposure to toxicants with known adverse health outcomes. The ability of OMICS to efficiently identify the molecular changes within biological samples provides valuable information for understanding toxic processes, pathways, and mechanisms, which are the major focuses of traditional environmental toxicology.

OMICS approaches can also help to address many of the challenges that cannot be easily tested using traditional environmental toxicology approaches, e.g. toxicity data extrapolation to determine whether chemicals will affect human health. To date, many chemical toxicity testing approaches are primarily based on endpoints derived from animal toxicity studies. OMICS could be the solution for toxicity data extrapolation across species and doses. Among various species and different types of cells and tissues, it is remarkable that a common subset of genes involved in a set of conserved signaling pathways are conserved through evolution, Therefore, an extrapolation of potential toxic responses and toxicity outcomes across species and doses could be done at the molecular level through the comparison of OMICS data sets that reflect mRNA, protein, and metabolite levels. Combination of OMICS data from mRNA, proteins, and metabolites, cells, tissues, individuals, and ultimately populations will help to develop a much-improved predictive capacity for toxicity data extrapolation and health risk assessment.

2.2 *Environmental OMICS of single chemicals vs. chemical mixtures*

2.2.1 Single environmental chemicals

Much of current environmental OMICS research focuses on the toxicity of single chemicals. One reason for this is that the experimental methodologies, approaches, and technologies for studying the toxicity of single chemicals and their risks to human health are well-established. Arsenic is probably one of the most common and well-studied single environmental chemicals [33]. It is found ubiquitously in our environment including drinking water, foods, soil, and airborne particles, and there is a generally well-accepted dose-response relationship between arsenic ingestion and cancer incidence [34, 35]. Microarray-based expression profiling of the livers of zebrafish exposed to arsenic revealed global transcriptional changes and suggested that DNA and protein damages due to arsenic metabolism and the arsenic-induced oxidative stress are the major causes of cellular injuries observed in the liver [36]. Many genes encoding proteins in-

volved in DNA damage/repair, antioxidant activity, hypoxia induction, iron homeostasis, arsenic metabolism, heat shock proteins, and ubiquitin-dependent protein degradation were found to be differentially expressed [36]. cDNA microarray and enzyme-linked immunosorbent assays were also used to identify genes involved in arsenic-associated atherosclerosis [37]. Arsenic was also found to activate stress gene expression [38, 39] and to induce cell proliferation and apoptosis pathways [40, 3]. Using 2-D gel electrophoresis and MALDI-TOF-based proteomic approaches, several up-regulated proteins, including α-enolase, HSP90β, pyruvate kinase, aldolase reductase, GAPDH, phosphoglyceratemutase B, Cu-Zn SOD, and thioredoxin were identified in LEC transformed cells. Several proteins including intermediate filament proteins such as peripherin, cytokeratin 14, and cytokeratin 8 were down-regulated [41, 42]. Recently, RNA sequencing was used to acquire global transcriptome alterations and miRNA regulation in rice under As (III) treatment at different times and dosages to investigate the metabolic and regulatory network and their interactions in the plant [43]. Additionally, proteomics in conjunction with morphological, physiological, and biochemical analyses have been employed to unravel for the first time survival strategies of the diazotrophic cyanobacterium anabaena sp. PCC7120 under arsenic stress [44]. These studies together with other published studies on arsenic toxicity typically reflect many common interests, approaches, and potential problems that were found in the OMICS analysis of single environmental chemicals.

The focuses on single chemical OMICS has been on the identification of toxic pathways, mechanisms, and biomarkers of the chemicals. Previous studies on the changes of various genes, proteins, metabolites and their underlying toxicity pathways and toxic mechanisms in various tissues, organs, or in whole organisms in response to exposure to arsenic or other single chemicals have demonstrated the importance of incorporating OMICS data into the regulation frameworks for environmental chemicals. However, with few exceptions, to date most of these OMICS studies on arsenic and other single chemicals have been limited to a qualitative description of alteration in gene and protein levels from *in vitro* cells and animal tissues exposed to the chemicals at doses that are much higher than environmental exposure doses, with minimal reporting regarding the biological outcomes and with little correlation to toxicity or the contribution to human health risk assessment. To solve these problems, innovative approaches on how to differentiate between specific and unspecific changes in the genome, proteome, and metabolome, and how to decide the relevance of such changes for further toxicological research and biomarker identification, need to be developed. The appropriate use of controls including negative and positive controls, time points, and treatment dosages of single chemicals for OMICS studies is important for the evaluation and interpretation of the research results. Typically, toxicity is a persistent and easily identified endpoint. However, genomics, proteomic, and metabolomic responses are dynamic and only capture expression of mRNAs, proteins, and metabolites at a certain time point. Therefore, comparison of OMICS data requires sampling at multiple time points. The OMICS profiles collected at different time points also can help with the identification of true toxicity-specific changes. Analyzing samples treated with different dosages of toxicants is also very important in identifying toxicity-specific changes. Higher doses might provide additional sensitivity that could help in the initial identification of significant effects, while low dosage sample analysis could help establish thresholds that need to be exceeded prior to the initiation of the cascade of molecular responses leading to an adverse or toxic effect. Changes that are consistently presented in the OMICS expression profiles obtained from samples treated with different dosages of toxicants would indicate a toxicity-specific change. In addition, computational toxicology, bioinformatics, and system biology tools are also needed to integrate and model the complex OMICS data sets in order to understand the biological activities and toxic processes of other chemicals that lead to environmental toxicity and risks to human health.

2.2.2 Chemical mixtures

Humans are exposed to multiple chemicals. This may occur in the form of mixtures of chemicals, where multiple chemicals occur in a given environmental medium, or as a cumulative exposure, where multiple chemicals are encountered from multiple environmental media via multiple exposure routes [45]. For example, chemical disinfectants react with naturally occurring organic and inorganic matter in water to produce a wide variety of disinfection byproducts (DBPs), and more than 500 DBPs have been identified. Beyond environmental exposure scenarios, tobacco smoking is one of the most prominent drivers for damage to human health. Tobacco smoke contains more than 4000 chemicals, and at least 200 of them are toxic to humans and over 50 are recognized as known or probable human carcinogens [46]. Tobacco smoking remains a major public health problem, threatening the lives of over one billion people during this century, and tobacco use is estimated to kill more than 5 million people each year worldwide [47]. Some intrinsically complex mixtures such as diesel exhausts, welding fumes, coke oven emissions, and metal working fluids are routinely encountered in occupational settings. Chemical mixture effects are therefore a major issue in the environmental and health risk assessment of chemicals [48] and studying the environmental toxicity of chemical mixture and risks of chemical mixtures is more important than of individual chemicals.

However, to date very few biological systems and technologies have been available and suitable to address the toxic mechanisms and toxicity pathways of exposures to chemical mixtures. Such biological systems and technologies are needed to decipher all of the interactions among complex mix-

tures at molecular, cellular, and organ levels, which are critical to the toxicity characterization and risk assessment of chemical mixtures. The dilemma of the lack of biological systems, technologies, and scientific information versus the perception of the high risk from exposures to chemical mixtures poses an enormous challenge for the public health and risk assessment community. Therefore, there is an important need to develop novel and biology-based methodologies and approaches for both efficient analysis of environment toxicity pathways, biomarkers, and toxic mechanisms associated with exposure to chemical mixtures and accurate distinction of the chemicals in the mixture that present little or no concern from those with the greatest likelihood of causing an adverse effect in the target species. With support of OMICS-based technologies, this information gap can be filled.

Although approaches for determining the mechanisms by which a single chemical induces toxicity or carcinogenicity have been relatively well established, these approaches cannot be readily extended to the study of chemical mixtures. High-throughput and high-content OMICS technologies and methods applied to predictive toxicology provide opportunities to address these challenges. First of all, OMICS tools have the potential to improve our understanding and predictability of combined effects. The interplays between environmental stresses and the dynamic responses of organisms at the levels of genes, RNAs, proteins, and metabolites can be efficiently determined using OMICS technologies. Second, OMICS enables high throughput analysis and can identify multiple molecular targets, pathways, and environmental responses to exposed organisms simultaneously, which is critical to understanding the modes of actions of chemicals and determining the components that actually cause toxicity in a mixture. Third, OMICS offers great potential to identify novel molecular toxicity endpoints for identification, categorization, and prioritization of chemical mixtures [25, 49]. In a study of gene expression profiles of rainbow trout exposed to a simple mixture of chromium, 2,2,4,4-tetrabromodiphenyl ether, and 17b-ethynylestradiol, no single compound dominated gene expression profiles, and the toxicity of the mixture was not simply the sum of the toxicities of the individual chemicals [50]. However, combined effects of polyfluorinated and perfluorinated compounds on primary cultured hepatocytes from rain minnows were observed in a genomic study [51]. Thus, the relationships among the expression patterns, chemical interactions, and ultimate mixture toxicity are very complicated. Exposure to mixtures can result in common response [52], synergistic and antagonistic interactions depending on the genes [52, 53], and/or unpredicted combined effects with unique transcriptional signatures [54].

Proteomics and metabolomics have also been used to characterize chemical interactions and mixture toxicity [55-57]. These studies focused on the identification of protein or metabolite signatures associated with the toxicity of both individual chemicals and mixtures. This approach attempted to distinguish the exposure signatures of individual chemicals in the mixtures so as to identify those chemicals in the mixture that present major concern and the greatest likelihood of causing an adverse effect. The information generated from such studies both improves the certainty about the assumptions in predictive models used to quantify the environmental toxicity of chemical mixtures and helps to refine our current exposure monitoring and assessment of chemical mixtures. In addition, the integration of genomic, proteomic, and metabolomic approaches is important for improving our understanding and predictive capability of the combined effects of chemical mixtures. Presently there is an urgent need to develop well-accepted conceptual frameworks and standards, including experimental design, data interpretation, and modeling for the use of OMICS tools to study chemical mixture toxicity and associated risks to our environment and human health.

2.3 *Molecular modifications of genes, RNAs, proteins, and metabolites*

Genes, mRNA, proteins, and metabolites are frequently modified in response to environmental stresses. It has been widely recognized that knowledge of a gene, mRNA, or protein and its sequence is just a prelude to understanding the role of that molecule and its product within the cell and is only a starting point for a full description of its function. While the modified form of the molecule is essential to the understanding molecular functions and mechanisms, it is expected that more and detailed studies on the modified forms of the genes and functional molecules under a variety of environmental conditions will be carried out.

Oxidative stress and protein oxidation are examples that stress the importance, challenges, and directions of OMICS applications to the identification of gene and protein modifications. Environmental stressors frequently result in formation of ROS, and proteins absorb about 70% of the ROS [58]. Therefore, using proteomic technologies to detect protein changes and modifications in response to environmental pollutant-mediated oxidative stresses has been an important research topic. Redox proteomics is probably one of the most extensively studied areas for characterization of protein modifications under environmental stresses. Sheehan et al. [59] developed a "toolkit" of redox proteomic approaches that can detect quite low levels of redox lesions in two dimensional electrophoresis separations. An example of this approach is the use of activated thiol sepharose to select for protein thiols or redox variants, such as disulphides, using both 2-D gel electrophoresis and gel-free shotgun proteomics [60]. Oxidation of cysteine residues may not always result in an effect on protein function, and sometimes is a reversible process analogous to protein phosphorylation [61]. Protein thiol modification is an important signal transduction mechanism regulating oxidative and antioxidative processes and/or events [62]. In addition to protein thiol modifications, protein carbonylation in response to environmental exposures and stresses has also been studied exten-

sively [63, 64].

Carbonyls are relatively difficult to induce as compared to cysteinyl derivatives and may reflect more severe oxidative stress [59]. Protein carbonylation is generally associated with permanent loss of protein function and has been used as a marker for assessing oxidative damage [65]. Although technologies for the detection of oxidized proteins have been advanced recently and offer some advantages for identifying oxidative stress biomarkers, studies on the identification of oxidized protein biomarkers to determine chemically-induced oxidative stress and injury are still severely lacking. As proteomics studies progress, the goal will not only be to identify proteins in mixtures, but also to derive more information about protein modifications from the samples analyzed. The ability to identify oxidized proteins will yield in-depth information on protein structure and function, which, in turn, can facilitate the pace of studies to understand toxic processes, pathways, and mechanisms of environmental chemicals. In addition to protein modifications, RNA and metabolite modifications could also function as mechanistic linkages between environmental exposures and outcomes and serve as important biomarkers of environmental exposure, toxicity, and effects. However, little is known about the RNA and metabolite modifications.

2.4 *Environmental monitoring and health risk assessment*

Pollution is one of the most important environmental challenges we are currently facing. Some environmental chemicals, like hormones or drugs, can act at very low doses to disrupt ecological balance and threaten growth and development of precious species with knock-in effects on biodiversity and ecological services [66]. Environmental chemicals can also damage human health directly or indirectly through food chains. Monitoring of chemical toxicity and risk in ecological systems is important to natural resource protection, environmental sustainability, and human health. Assessments of ecosystem health play essential roles in the development of effective strategies for not only protecting the environment but human health as well. In aquatic ecosystems from oceans to river basins, water assessment has focused on providing specific information regarding the dynamics of pollutants and their effects on the health of different species.

Classical biomonitoring programs directly quantify the bioaccumulation of pollutants in tissues of exposed organisms or analyze different effects or responses to the xenobiotic in those exposure organisms. Bivalves have been extensively utilized as sentinel organisms due to their sessile nature, filter feeder habits, and high capacity to accumulate contaminants, providing therefore temporally and spatially integrated levels of contamination [67]. In the last decade, proteomics in bivalves has greatly contributed to the identification of more specific and sensitive markers of pollution, thereby providing an accurate estimation of ecosystem health [68, 69]. Historically, mussels have been utilized to evaluate a broad range of hazardous compounds such as metals [70], polycyclic aromatic hydrocarbons [71], anthropogenic pollution [71], and the biological consequences of oil spills in marine environments, such as after the accidents of the Exxon Valdez in Alaska [72] and Prestige in NW Spain [73]. The main reason that hinders the application of bivalve proteomics for large biomonitoring programs is the absence of an assembled and annotated genome sequence from bivalves. MS–proteomics approaches require a fairly complete genome annotation in publicly available databases, and unfortunately sequences from genes and proteins from bivalves are still scarce. Greater information could be obtained from mtDNA [74, 75]. Alternately, the development of assays for multiple reaction monitoring (MRM) or selected ion monitoring (SIM) could be explored. These approaches could improve environmental assessment and biomonitoring. In a more restricted way, this strategy has been used to analyze the toxin profile of *M. galloprovincialis* collected from an area with remarkable concentration of *Alexandrium ostenfeldii* cells in seawater [75].

Although the analysis of the effects of environmental stressors utilizing bivalves as sentinel organisms could be considered a pioneer area of research in environmental proteomics, it lags behind the achievement of proteomics in biomedicine. Important aspects of proteomics that were discussed in the First International Conference on Environmental OMICS included strategies to join efforts for the inter-laboratory validation of protein expression patterns (PES), developing proteomics methodologies for the verification or validation of selected target candidates, evaluating pollutant mixtures for possible synergies and modes of action, and systematically exploring protein translation modifications. Finally, the next generation of sequencing technologies has started to develop for environmental issues and applications.

The advantages of bivalves should not be forgotten for marine pollution assessment. The research in bivalves has been historically one of strongest research areas in biomonitoring, because the sessile nature of bivalves could provide a better correlation between analyzed stressors and topographic information [76]. The integration of genomics and proteomics data and the application of EST data for protein identification could hopefully improve biomonitoring programs and define the mechanisms underlying pollutant toxicity.

Environmental monitoring and health risk assessment require different and specific biomarkers, because environmental contaminants induce multiple responses in organisms that are not necessarily correlated. OMICS technologies offer the promise of fast, cost-effective, and broad-scale data gathering capacity, all of which are necessary to be able to identify environmental biomarkers and establish critical links between exposure, response, and disease. By comparing OMICS profiles of environmentally-exposed cells, tissues, and organisms to a database containing profiles induced by known toxicants, environmental biomarkers of exposure and toxicity were identified [77]. OMICS technologies are very

efficient and powerful in identifying environmental toxicity biomarkers. This approach is a popular application in ecotoxicology research. Although several candidate biomarkers for environmental monitoring have been reported [78], they must be validated for their potential usage in risk assessment, since these biomarkers could be the result of artifacts of specimen collection, bioinformatics bias, and/or experimental variations rather than truly toxicological responses.

In addition to the generation of large and information-rich datasets on the changes at expression levels of genes, proteins, and metabolites for the purpose of identifying environmental biomarkers of exposure, toxicity and effects of environmental chemicals, establishment of suitable experimental model systems for monitoring of environmental and health risks under different pollution situations or scenarios is also very important. The zebrafish is a well established experimental model for environmental monitoring and risk assessment, because it offers a combination of the advantages of *in vitro* and *in vivo* systems and many well-developed genetic tools including transgenesis. In recent years, the zebrafish has also been used to model human diseases, because the zebrafish is in the vertebrate family and shares many fundamental similarities in body plan, organ systems, physiology, pathology, and diseases with other vertebrates [79].

Two general approaches have been used to develop zebrafish-based environmental monitoring tools. The first approach is to employ selective, inducible promoters to generate transgenic fish that respond to environmental pollutants by expressing visible fluorescent colors [80]. The second approach is to employ zebrafish DNA chips or next generation sequencing to identify biomarkers from fish exposed to different environmental chemicals. In this way, these fish biomarkers can be used to identify environmental pollutants and to infer associated toxicity effects and risks [81], In two independent genomic studies [82, 83], zebrafish were exposed to polycyclic aromatic hydrocarbons, organic nitrogen compounds, organochlorine pesticides, endocrine disrupters, and metallo-compounds, and gene expression changes in exposed zebrafish were characterized. Both studies reported that chemically treated samples could be correctly grouped base

d on the hierarchical clustering of transcriptomic data, and subsets of genes could be linked to specific exposures [83, 82]. A recent proteomic study of zebrafish revealed that chronic toxicity of microcyetin is different from acute toxicity, and the reactive oxygen species pathway is the main toxic pathway [84]. Moreover, it has been determined that microcystin causes neurotoxicity in zebrafish at the proteomic level.

In addition to zebrafish, application of genomics to several other organisms such as fathead minnow (*Pimephales promelas*), medaka (*Oryzias latipes*) and several invertebrates such as water fleas (*Daphnia magna*) and soil nematode (*Caenorhabditis elegans*), have also been employed as tools for environmental monitoring [85, 86]. With the application of next generation sequencing technology in ecologically relevant organisms, ecotoxicogenomics and environmental risk assessment have been shown to be more feasible and productive [87].

Although significant accomplishments have been achieved in the OMICS-based environmental monitoring, there are still many important questions to be addressed and challenges to be overcome in the field of environmental monitoring. These include how to deal with the lack of sequencing data in environmental model organisms, how to validate the environmental monitoring biomarkers, and what validation criteria should be used.

2.5 *Meta-omics: a high throughput tool to study microbial community in ecosystems*

To solve environmental challenges and to make our natural environment sustainable, there is an urgent need to develop reliable and practical experimental technologies, systems, and methods for evaluating the large variety of unknown microbes in nature. Application of OMICS to systematic analysis of environmental microbial communities has been used to identify novel catalysts that can degrade environmental pollutants, to purify contaminated natural resources, to produce novel bioproducts, and to identify novel biomarkers for environmental risk monitoring. This has resulted in the emergence of new research fields: metagenomics, metaproteomics and metametabolomics. Since the publications in metagenomics have been previously reviewed [88, 89] and very few studies on metabolomics have been published, here we provide a brief review on some recent publications in metaproteomics and provide some recommendations for future meta-omics research.

Tringe et al. [90] investigated the collective environmental signature obtained from different microbial ecosystems using a gene-based bioinformatic approach and found that the predicted protein complement of a community is influenced by its environment. Since then, there have been a few studies describing metaproteomics, including the examination of protein expression profiles from activated sludge [91, 92], freshwater samples following exposure to heavy metals [93], contaminated soil and groundwater [94], endosymbiont [95], lake water [96], and extracellular proteins in activated sludge [97]. Notably, an intensive proteomic study of acid mine biofilms has been performed, in which approximately 2200 proteins were identified with one novel protein as a key component of energy conservation in that environment [98].

Despite the limited number of investigations, the metaproteomic approach has already highlighted its potential for providing functional insight into overall microbial ecosystem function. Kan et al. [99] introduced the metaproteomics approach to study a microbial community collected from the Chesapeake Bay. Power et al. [100] investigated dissolved proteins in seawater. Following these studies, metaproteomic studies on various marine environmental samples were conducted, and many metabolic and physiological activities including nutrient utilization and environmental adaptation,

were revealed [32]. Metaproteomics were also applied to the examination of protein expression in complex marine environmental samples, thereby opening a new window for marine microbial oceanography and microbial biogeochemistry [101]. Recently, it has been reported that new species of microbes feeding on the Deepwater Horizon oil spill may exist based on the oil spill degradation rate [102]. However, the analysis of microbial community responses to the oil spill has suffered from the lack of a reliable and comprehensive microbial database and analytical techniques and approaches. Meta-omics could potentially fill this scientific gap and generate the scientific data and basis for assessing ecological risks associated with oil spills and other environmental disasters. It could provide novel insights into the molecular bases and mechanisms through which microbes respond to environmental perturbations and produce microbial enzymes for the biodegradations of the spilled oils in oceans and other aquatic environments.

Our natural environment contains a large community of microbes that are already adapted to the background supply of environmental pollutants. Eventually, the microbes will "take care" of the pollution problem by consuming the pollutants that are biodegradable. Environmental OMICS provides a high-throughput and high-resolution tool to glean a more complete picture of microbial community composition, and to identify novel catalysts and microbes that can biodegrade environmental pollutants. Therefore, meta-omics research offers the potential to provide a solid scientific foundation for the understanding and management of environmental pollution and for the utilization of natural microbial resources to maintain environmental sustainability.

3. Conclusions

Environmental OMICS approaches have already had a big impact in helping to identify toxicity pathways, toxicity mechanisms, and environmental biomarkers for health risk assessment. Additionally, OMICS approaches have impacted environmental monitoring and sustainability through high throughput and simultaneous measurement of multiple genomic, proteomic, and metabolomic profiles and parameters in a given system under defined environmental conditions. Environmental OMICS is still in the early stage of OMICS data collection, and the field is transitioning from "profiling OMICS" to "functional OMICS."

Beyond providing a list of molecular changes, OMICS approaches emphasize the biological significances of the identified molecular changes, which is key to the success of environmental OMICS development. In the future, environmental functional OMICS approaches may help to solve environmental challenges, such as the elucidation of chemical mixture toxicity, protein modifications, environmental and health monitoring, and meta-omics.

The development of research standards, guidelines, and frameworks for sample collection and OMICS data generation, collection, validation, interpretation, and presentation will need to be made in the near future. Other needs identified for environmental OMICS are the combined analysis of multiple OMICS data sets, including genomics, proteomics, and metabolomics, and the integration of OMICS data with toxicity data to define the links between OMICS data and particular toxic processes or environmental diseases under investigation. Integration of OMICS data with classical toxicology endpoints and clinical observation will allow more sensitive and earlier detection of adverse health effects, precise identification of toxicity signatures and biomarkers, and development of predictive environmental toxicology for more effective environmental biomonitoring and human health risk assessment.

To accomplish these goals, international collaborations among environmental OMICS scientists worldwide is needed. These collaborations should also include partnerships between governmental agencies and nongovernmental research groups in both academia and industry.

Acknowledgments

We want to thank all of the 2011 ICEO conference participants for their presentations and discussions, which inspired this review article. The aim of the conference was to serve as an interdisciplinary scientific forum to present the most recent advances in environmental OMICS and OMICS technologies, to promote international communication and collaboration on environmental OMICS research, and to foster integration of the latest scientific developments into practical applications for the improvement of human health and the environment.

During the conference, it was proposed to establish an International Society for Environment OMICS (ISEO) to consolidate different environmental OMICS research laboratories, groups, and organizations worldwide, to engage in scientific and educational activities that promote environmental OMICS research and technologies, and to assist in coordinating shared public environmental OMICS initiatives.

The authors would like to thank Drs. Jeffrey Ross, Sheau-Fung Thai, Witold Winnik, and Charlene McQueen at EPA and Dr. Richard Woychik at the National Institute of Environmental Health Sciences (NIEHS) for their very helpful comments on this manuscript.

This document has been reviewed in accordance with U. S. Environmental Protection Agency and Food and Drug Administration policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

[1] B. Kuska, B. Bethesda. J Natl Cancer Inst. 21 (1998) 90 - 93.

[2] J. Lederberg, A. T. McCray. Scientist. 15 (2001).

[3] A. T. Lau, J. F. Chiu. Proteomics. 6 (2006) 1619 - 1630.

[4] T. Tsuchiya, T. Tanaka-Kagawa, H. Jinno, H. Tokunaga, K. Sakimoto, M. Ando, M. Umeda. Toxicol Sci. 84 (2005) 344 - 351.

[5] Y. Ge, B. Maribel, K. Wallace, W. Winnik, R. Y. Prasad. Proteomics. 11 (2011) 2406 - 2422.

[6] P. P. Li, X. J. Liu, H. Li, X. X. Peng. J Proteomics. 75 (2012) 2638 - 2648.

[7] N. Mei, L. Guo, R. Liu, J. C. Fuscoe, T. Chen. BMC. Bioinformatics 8 Suppl 7 (2007).

[8] P. A. Ortiz, M. E. Bruno, T. Moore, S. Nesnow, W. Winnik, Y. Ge. J Proteome Res. 9 (2010) 1268 - 1278.

[9] J. Wang, Y. Y. Wang, L. Lin, H. S. Hong, D. Z. Wang. J Proteomics. 75 (2012) 2038 - 2052.

[10] K. Matheis, D. Laurie, C. Andriaman- droso, N. Arber, L. Badimon, X. Benain. Today. 16 (2011) 600 - 608.

[11] P. Vineis, A. E. Khan, J. Vlaanderen, R. Vermeulen. Environmental Health. 8 (2009) 54 - 64.

[12] D. Yasokawa, H. Iwahashi. J Biosci Bioeng. 110 (2010) 511 - 522.

[13] D. L. Eaton, E. P. Gallagher, M. L. Hooper, D. Schlenk, P. Schmeider, C. Thompson. In Chapter 3 of the book "Genomic Approaches for Cross-Species Extrapolation in Toxicology" (Eds. Di Giulio RT and Benson WH) CRC press. (2006).

[14] K. J. Ralston-Hooper, B. C. Sanchez, J. Adamec. Environmental Toxicology and Chemistry. 30 (2011) 1197 - 1203.

[15] Y. Ge, M. Bruno, H. Foth . John Wiley & Sons Ltd. 1 (2011) 197 - 215.

[16] K. Kishi, E. Kitagawa, N. Onikura, A. Nakamura, H. Iwahashi. Genomics. 88 (2006) 241 - 251.

[17] M. Bruno, T. Moore, S. Nesnow, Y. Ge. J Proteome Res. 8 (2009) 2070 - 2078.

[18] K. Mezhoud, D. Praseuth, J. C. Francois, C. Bernard, M. Edery. Adv Exp Med Biol. 617 (2008) 419 - 426.

[19] C. Yi, T. Pan. Chem Res. 44 (2011) 1380 - 1388.

[20] E. Enan, F. Fumio Matsumura. J Biochemical Toxicology. 9 (1994) 159 - 170.

[21] M. Van den Berg, L. Birnbaum, A. T. C Bosveld, B. Brunstrom, P. Cook, M. Feeley. Environmental Health Perspectives. 106 (1998) 775 - 792.

[22] Y. Iwahashi, H. Hosoda, J. Park, J. Lee, Y. Suzuki, E. Kitagawa, S. Murata, N. Jwa, M. Gu, H. Iwahashi. J Agri Food Chem. 54 (2006) 1936 - 1942.

[23] K. Cho, J. Shibato, G. K. Agrawal, Y. H. Jung, A. Kubo, N. S. Jwa, S. Tamogami, K. Satoh, S. Kikuchi, T. Higashi, S. Kimura, H. Saji, Y. Tanaka, H. Iwahashi, Y. Masuo, R. Rakwal. J Proteome Res. 7 (2008) 2980 - 2998.

[24] National Research Council. http://www.nap.edu/openbook.php?record_id=11970. (2007).

[25] H. Iwahashi, E. Ishidou, E. Kitagawa, Y. Momose. Environ Sci Technol. 41 (2007) 7941 - 7946.

[26] L. Tomanek, M. J. Zuzow. Dev Dyn. 239 (2010) 3182 - 3191.

[27] V. J. Nesatyy, M. J. Suter. Environ Sci Technol. 41 (2007) 6891 - 900.

[28] W. R. Howard, J. Havenhand, L. Parker, D. Raftos, P. Ross, J. Williamson, R. Matear. In a marine climate change impacts and adaptation report card for Australia (Eds. E.S. Poloczanska, A.J. Hobday and A.J. Richardson), NCCARF Publication 05/09, ISBN 978-1-921609-03-9. (2009).

[29] D. Braconi, G. Bernardini, A. Santucci. J Proteomics. 74 (2011) 2324 - 2337.

[30] D. Z. Wang, H. P. Dong, Z. X. Xie, M. H. Dai, H. S. Hong. Limnol Oceanogr. 56 (2011) 1641 - 1652.

[31] M. Horie, H. Kato, K. Fujita, S. Endoh, H. Iwahashi. Chemical Research in Toxicology. 25 (2012) 605 - 619.

[32] R. M. Morris, B. L. Nunn, C. Frazar, D. R. Goodlett, Y. S. Ting, G. Rocap. The ISME J. 4 (2010) 673 - 685.

[33] G. Marshall, C. Ferreccio, Y. Yuan, M. N. Bates, C. Steinmaus, S. Selvin, J. Liaw, AH Smith. J Natl Cancer Inst. 99 (2007) 920 - 928.

[34] R. Baastrup, M. Srensen, T. Balstrm, K. Frederiksen, C. L. Larsen, A. Tjnneland, K. Overvad, O. Raschou-Nielsen. Environ Health Perspect. 116 (2007) 231 - 237.

[35] WHO. Environmental health criteria 224, 2nd Ed. World Heath Organization, Geneva, Switzerland. (2001).

[36] S. H. Lam, C. L. Winata, Y. Tong, S. Korzh, W. S. Lim, V. Korzh, J. Spitsbergen, S. Mathavan, L. D. Miller, E. T. Liu, Z. Gong. Physiol Genomics. 27 (2006) 351 - 361.

[37] J. C. States, S. Srivastava, Y. Chen, A. Barchowsky. Toxicol Sci 107 (2009) 312 - 323.

[38] M. Li, J. F. Cai, J. F. Chiu. J Cell Biochem. 87 (2002) 29 - 38.

[39] C. Martin-Chouly, C. Morzadec, M. Bonvalet, M. D. Galibert, O. Fardel, L. Vernhet. Mol Immunol. 48 (2011) 956 - 965.

[40] W. M. Cheung, P. W. K. Chu, Y. L. Kwong. Cancer Lett 246 (2007) 122 - 128.

[41] G. Li, L. S. Lee, M. Li, S.W. Tsao, J. F. Chiu. J Cell Physiol. 226 (2011) 3225 -3232.

[42] Y. Sun, J. Pi, X. Wang, E. J. Tokar, J. Liu, M. P. Waalkes. Toxicology. 262 (2009) 162 - 170.

[43] L. J. Yu, Y. F. Luo, B. Liao, L. J. Xie, L. Chen ,S. Xiao, J. T. Li, S. N. Hu, W. S. Shu. New Phytol. 195 (2012) 97 - 112.

[44] S. Pandey, R. Rai, L. C. Rai. J Proteomics. 75 (2012) 921 - 937.

[45] M. C. Celander. Acquatic Toxicology 105 (2011) 72 - 77.

[46] K. Husgafvel-Pursiainen. Mutat Res 567 (2004) 447 - 474.

[47] WHO. World Heath Organization, http://whqlibdoc.who.int/publications/2009/9789241563918_eng_full.pdf. (2009).

[48] A. Kortenkamp, T. Backhaus, F. Faust. Report to the European Commission. http://ec.europa.eu/environment/chemicals/pdf/report_Mixture%20toxicity.pdf. (2009).

[49] S. E. Hook, A. D. Skillman, B. Gopalan, J. A. Small, I. R. Schultz. Toxicol Sci. 102 (2008) 42 - 60.

[50] D. J. Spurgeon, O. A. Jones, J. L. Dorne, C. Svendsen, S. Swain, S. R. Stürzenbaum. Sci Total Environ. 408 (2010) 3725 - 3734.

[51] Y. Wei, X. Shi, X. Shi, H. Zhang, J. Wang, B. Zhou, J. Dai. Aquat Toxicol. 95 (2009) 27 - 36.

[52] F. Dondero, M. Banni, A. Negri, L. Boatti, A. Dagnino, A. Viarengo. BMC Genomics. 12 (2011) 195 - 205.

[53] C. M. Hutchins, D. F. Simon, W. Zerges, K. J. Wilkinson. Aquat Toxicol. 100 (2010) 120 - 127.

[54] F. A. Tilton, S. C. Tilto, K. Bammler, R. P. Beyer, P. L. Stapleton, N. L. Scholz, E. P. Gallagher . Toxicol. 102 (2011) 205 - 215.

[55] M. Merhi, C. Demur, C. Racaud-Sultan, J. Bertrand, C. Canlet, Y. Blas, F. Estada, L. Gamet-Payrastre. Toxicology. 267 (2010) 80 - 90.

[56] G. Pelletier, S. Masson, M. J. Wade, J. Nakai, R. Alwis, S. Mohottalage, P. Kumarathasan, P. Black, W. J. Bowers, I. Chu, R. Vincent. Toxicol Lett. 184 (2009) 176 - 185.

[57] H. Wu, W. X. Wang. Toxicol. 100 (2010) 339 - 345.

[58] D. Sheehan, B. McDonagh. Invertebrate Survival Journal. 5 (2008) 110 - 123.

[59] D. Sheehan. Biochemical and Biophysical Research Communications 349 (2006) 455 - 462.

[60] W. Hu, S. Tedesco, B. McDonagh, J. A. Barcena, K. Keane, D. Sheehan. Analytical Biochemistry 398 (2010) 245 - 253.

[61] J. M. Held, B. W. Gibson. Mol Cell Proteomics. 11 (2012) R111.013037

[62] C. C. Winterbourn, M. B. Hampton. Free Radic Biol Med. 45 (2008) 549 - 561.

[63] S. Chora, B. McDonagh, D. Sheehan, M. Starita-Geribaldi, M. Roméo, M. J. Bebianno. Chemosphere 81 (2010) 1212 - 1217.

[64] S. Tedesco, T. Nuri, A. Coeho, D. Sheehan. Journal of Integrated OMICS DOI: 10.5584/jiomics.v2012i2012.77 (2012).

[65] E. S. Cannizzo, C. C. Clement, R. Sahu, C. Follo, L. Santambrogio. J Proteomics. 74 (2011) 2313 - 2323.

[66] A. Campos, S. Tedesco, V. Vasconcelos, S. Cristobal. J proteomics. 75 (2012) 4346 - 4359.

[67] D. W. Kolpin, E. T. Furlong, M. T. Meyer, E. M. Thurman, S. D. Zaugg, L. B. Barber, H. T. Buxton. Environmental Science &Technology. 36 (2002) 1202-1211.

[68] I. Apraiz, J. Mi, S. Cristobal. Proteomics 7 (2006) 2997 - 3009.

[70] E. L. Thompson, D. A. Taylor, S. V. Nair, G. Birch, P. A. Haynes, D. A. Raftos. Aquat Toxico. 103(2011) 241 - 249.

[71] J. K. H. Fang, D. W. T Au, R. S. S. Wu, G. J. Zheng, A. K. Y. Chan, P. K. S. Lam, P. K. S. Shin. Marine Pollution Bulletin. 58 (2009) 615 - 620.

[72] R. E. Thomas, C. Brodersen, M. G. Carls, M. Babcock, S. D. Rice. Comparative biochemistry and physiology Part C, Pharmacology, toxicology & endocrinology.122 (1999) 153 - 163.

[73] I. Apraiz, M. P. Cajaraville, S. Cristobal. Mar Pollut Bull 58 (2009) 1815 - 1826.

[74] R. J. Hoffmann, J. L. Boore, W. M. Brown. Genetics. 13 (1992) 397 - 412.

[75] H. Doucet-Beaupre, S. Breton, E. G. Chapman, P. U Blier, A. E Bogan, D. T Stewart, W. R. Hoeh. BMC Evol Biol. 10 (2010) 50 - 61.

[76] I. Apraiz, D. Lindenstrand, J-O Persson, S. Cristobal. Journal of Proteomics and Bioinformatics. 2 (2009) 255 - 261.

[77] T. D. Williams, N. Turan, A. M. Diab, H. F. Wu, C. Mackenzie. PLOS Computional Biology 7 (2011) e1002126.

[78] M. L. Martín-Díaz, T. A. DelValls, I. Riba, J. Blasco. Cell Biol Toxicol. 2 (2008) 513 - 526.

[79] S. H. Lam, Z. Gong. Springer Verlag, Heidelberg. (2010) 827 - 843.

[80] Z. Zeng, T. Shan, Y. Tong, S. H. Lam, Z. Gong. Env Sci Tech. 39 (2005) 9001 - 9008.

[81] S. H. Lam, M. M. Hlaing, X. Zhang, C. Yan, Z. Duan, L. Zhu, C. Y. Ung, S. Mathavan, C. N. Ong, Z. Gong. PLoS One. 6 (2011) e28273.

[82] S. H. Lam, S. Mathavan, Y. Tong, H. Li, R. K. M. Karuturi, Y. L. Wu, V. B. Vega, E. T. Liu , Z. Gong. PLoS Genetics. 4 (2008) e1000121.

[83] L. Yang, J. R. Kemadjou, C. Zinsmeister, M. Bauer, J. Legradi, F. Müller, M. Pankratz, J. Jäkel, U. Strähle. Genome Biol. 8 (2007) 227 - 237.

[84] M. H. Wang, L. L. Chan, M. Z. Si, H. S. Hong, D. Z. Wang. Toxicol Sci.113 (2010) 60 - 69.

[85] T. Iguchi, H. Watanabe, Y. Katsu, M. H. Wang, L. L. Chan, D. Z. Wang.Toxicological Sciences. 131 (2010) 60 - 69.

[86] D. L. Villeneuve, N. Garcia-Reyero, B. L. Escalon, K. M. Jensen, J. E. Cavallin, E. A. Makynen, E. J. Durhan, M. D. Kahl, L. M. Thomas, E. J. Perkins, G. T. Ankley. Environ Sci Technol. 46 (2012) 51 - 59.

[87] A. C. Mehinto, C. J. Martyniuk, D. J. Spade, N. D. Denslow. Front Genet. 3 (2012) 62 - 72.

[88] J. A. Gilbert, M. Hughes. Methods Mol Biol. 733 (2011)195 - 205.

[89] W. R. Streit, R. A. Schmitz. Current Opinion in Microbiology. 7 (2004) 492 - 498.

[90] S. G. Tringe, C. von Mering, A. Kobayashi, A. Salamov, K. Chen. Science 308 (2005) 554 - 557.

[91] P. Wilmes, P. L. Bond. Water Sci Technol. 54 (2006) 217 - 226.

[92] P. Wilmes, A. F. Andersson, M. G. Lefsrud, M.

Wexler, M. Shah, B. Zhang, R. L. Hettich, P. L. Bond, N. C. VerBerkmoes, J. F. Banfield. ISME J. 2 (2008) 853 - 864.

[93] C. M. R. Lacerda, L. H. Choe, K. F. Reardon. J Proteome Res. 6 (2007) 1145 - 1152.

[94] D. Benndorf, G. U. Balcke, H. Harms, M. von Bergen. ISME J. 1 (2007) 224 - 234.

[95] S. Markert, C. Arndt, H. Felbeck, D. Becher, S. M. Sievert, M. Hugler, D. Albrecht, J. Robidart, S. Bench, R. A. Feldman, M. Hecker, T. Schweder. Science. 315 (2007) 247 - 250.

[96] M. Pierre-Alain, M. Christophe, S. Severine, A. Houria, L. Philippe, R. Lionel. Microb Ecol. 53(2007) 426 - 434.

[97] C. Park, J. T. Novak, R. F. Helm, Y. O. Ahn, A. Esen. Water Res. 42 (2008) 3879 -3889.

[98] R. J. Ram, N. C. VerBerkmoes, M. P. Thelen, G. W. Tyson, B. J. Baker, R. C. Blake, M. Shah, R. L. Hettich, B. J. Banfield. Science 308 (2005) 1915 - 1920.

[99] M. J. Powell, J. N. Sutton, C. E. Del Castillo, A. I. Timperman. Mar Chem. 95 (2005) 183 - 198.

[100] J. Kan, J. M. Hanson, K. Ginter. Saline Syst. 19 (2005) 7 - 15.

[101] M. H. Wang, Y. Y. Wang, J. Wang, L. Lin, H. S. Hong, D. Z. Wang. Aquatic Toxicology. 103 (2011) 129 - 139.

[102] M. Beazley, J. Robert, R. J. Martinez, S. Rajan, J. Powell, Y. M. Piceno, L. M .Tom, G. L. Andersen, T. C. Hazen, J. D. V. Nostrand, J. Zhou, B. Mortazavi1, P. A Sobecky. PLoS ONE. 7 (2012) e41305.

JOURNAL OF INTEGRATED OMICS
A METHODOLOGICAL JOURNAL
HTTP://WWW.JIOMICS.COM

ORIGINAL ARTICLE | DOI: 10.5584/jiomics.v3i2.128

Inorganic mass spectrometry-based metallomics for environmental monitoring of terrestrial ecosystems affected by metal pollution using *Mus spretus* as bioindicator

M.A. García-Sevillano[1,2,3], T. García-Barrera*[1,2,3], J.L. Gómez-Ariza*[1,2,3]

[1]Department of Chemistry and Materials Science, Faculty of Experimental Sciences, University of Huelva, Campus de El Carmen, 21007-Huelva, Spain. [2]Research Center on Health and Environment (CYSMA). University of Huelva. Spain. [3]International Campus of Excellence on Agrofood (ceiA3). University of Huelva. Spain.

Received: 01 February 2013 **Accepted:** 30 May 2013 **Available Online**: 18 July 2013

ABSTRACT

A metallomic approach based on the use of size-exclusion chromatography coupled to inductively coupled plasma-mass spectrometry (SEC-ICP-MS) has been used to characterize the biological response in liver, brain, kidneys, lungs and plasma of the free-living mouse *Mus spretus* in polluted areas located in Doñana National Park (southwest Spain) and the surroundings, mainly affected by agriculture, mining and industry activities, which are responsible for the presence of metallic contaminants. It is remarkable the high presence of Cu, Zn, Cd, As, Pb and Ni in the cytosolic extracts of different organs and plasma, especially in contaminated areas. In liver extracts, high intensity peaks traced by Cu, Zn, Pb and Cd at 7 kDa (matching with metallothionein I standard) are triggered by the presence of contaminants. In kidney, similar Cu and Cd-peaks at 7 kDa were observed but the equivalent Zn-peak was depleted by the competitive interactions of Cu-Cd-Zn for the active sites of these molecules. In addition, peaks traced by Cu and Zn at about 32 kDa in liver extract match with superoxide dismutase standard (Cu,Zn-SOD), which increase in accordance to contamination. An analogous behavior was observed for a Zn,Cu-peak at about 67 kDa that can be related with the bovine serum albumin standard (Cu,Zn-BSA) or other carrier protein such as transferrin (Cu-Tf) present in liver and plasma. Finally, low molecular mass arsenic metabolites were detected in mice captured in MAT site affected by mine waste.

Keywords: *Mus spretus*; SEC-ICP-MS; metal pollution; Doñana National Park; Bioindicators; metallomics.

Abbreviations

AsB: Arsenobetaine; **AsC**: Arsenocholine; **BSA**: Bovine serum albumin; **DMA^V**: Dimethylarsinate; **DNP**: Doñana National Park; **GSH**: Reduced glutathione; **HPLC**: High-performance liquid chromatography; **ICP-MS**: Inductively coupled plasma-mass spectrometry; **ICP-AES**: Inductively coupled plasma- atomic emission spectrometry; **LC**: Liquid chromatography; **MMA^V**: Methylarsonate; **MS**: Mass spectrometry; **MT**: Metallothionein; **PEEK**: Polyether ether ketone; **PMSF**: Phenylmethanesulfonyl fluoride; **PTFE**: Polytetrafluoroethylene; **SEC**: Size exclusion chromatography; **SOD**: Superoxide Dismutase; **TCEP**: Tris-(2-carboxyethyl)phosphine hydrochloride; **Tf**: Transferrin.

1. Introduction

The use of free-living organisms as bioindicators of terrestrial ecosystems has been proposed in numerous papers [1-4]. The interest in monitoring the exposure and associated response to heavy metals on living organisms has increased in the last years, since they can reflect the effect of pollutants on cellular metabolism, trafficking and global homeostasis [5]. In addition, the metabolism of trace elements can not be considered in isolation since different elements and their species act jointly in the cells, tissues or organs, and consequently, it is important to consider their possible synergistic or antagonist interactions [6]. Studies of small mammals, mainly rodents, have been used as bioindicators in numerous environmental studies because they can provide useful information to assess the risk of metals to humans. To this

***Corresponding author:** J.L. Gómez Ariza (ariza@dqcm.uhu.es); T. García-Barrera (tamara@dqcm.uhu.es). José Luis Gómez Ariza, Tamara García Barrera. Department of Chemistry and Materials Sciences. Faculty of Experimental Sciences. University of Huelva. Campus de El Carmen. E21071 Huelva. Phone: +34 959 219962. Fax: +34 959 219942.

end, the aboriginal species *Mus spretus* has been frequently used as bioindicator in the southwest Spain, especially in Doñana National Park (DNP). This important ecological area has a biodiversity unique in Europe, Doñana contains a great variety of ecosystems and shelters wildlife including thousands of European and African migratory bird, fallow deer, Spanish red deer, wild boar, European badger, Egyptian mongoose, and endangered species such as the Spanish Imperial Eagle and Iberian Lynx. However, the Park is affected by agricultural, mining and industrial activities [7,8] which makes necessary the regular monitoring of environment quality based in the analysis of pollutants and the biological responses to them. The application of proteomics for this purpose using the mouse *Mus spretus* as bioindicator has represented a reliable option [9], especially by the demonstrated genetic homology of this mouse with the classical inbred laboratory mouse *Mus musculus* that has been already sequenced. This fact allows the use of databases from *Mus musculus* for the identification of proteins and metalloproteins of *Mus spretus* [5,9] as well as for transcriptomics studies [10], avoiding the cumbersome work associated to *de novo* sequencing. On the other hand, the use of massive information methods, the *-omics*, is a promising alternative to the assessment of metal pollution by using biomarkers in environmental contamination [11,12]. Metallomics is one of the most recent *-omic* [13] which uses metals or metalloids, that are present in one third of biomolecules in cells, as heteroatomic markers or tags to track these molecules in complex biological matrices [14].

Metalloproteins have many different functions in cells, such as enzymatic activity, transport, storage, and signal transduction. Essential elements such as Cu, Zn, Fe, Mn and Se are integral constituents of numerous proteins where they have structural or catalytic roles, as is the case of transferrin [15], serum albumin [16], chaperones [16] and metallothioneins (MTs) [17], among many others. Deficiency or excess of these elements in mammalian organisms can result in metabolic dyshomeostasis which in turn can eventually lead to oxidative stress and toxicity that can ultimately cause disease and even death [18,19]. Carrier or transport proteins are involved in the movement of ion, small molecule, or macromolecules, such as other proteins, across biological membranes [20]. Storage proteins are biological reserves of metal ion and amino acids, used by organisms, as is the case of ferriti for iron, and finally, MTs are an example of signal transduction proteins [21]. For this reason, metallomics provides a good alternative to deep insight into the fate of elements in exposed organisms to metal pollution, and gives information about metals trafficking, interactions and homeostasis.

Metal ions and element species present in living systems are usually present in biological fluids and tissues at low picograms or nanograms per ml or mg concentrations. For this reason, the use of small mammals as bioindicators to evaluate terrestrial ecosystems requires analytical techniques that combine a high resolution separation technique with a sensitive elemental or molecular specific detector [22]. In this sense, the great potential of inductively coupled plasma-mass spectrometry (ICP-MS) for trace metals analysis has been highlighted in recent reviews since it provides great sensitivity, selectivity and precision, and allows the simultaneous detection of multiple heteroatoms in metalloproteins [23,24]. In addition, the use of ICP-MS requires small amount of sample, a very important advantage when working with micro-mammals. On the other hand, the drawback of polyatomic and/or isobaric interferences in the plasma as well as matrix effects can be minimized by the use of high resolution double focusing spectrometers [24] or collision and dynamic reaction cells (ICP-(ORC)-MS) in quadrupole analyzers [25].

In the present work, a metallomic analytical approach that consisted of high performance liquid chromatography (HPLC) based in size exclusion chromatography (SEC) coupled to ICP-MS has been used to characterize metal containing species (< 70 kDa) in liver, kidney, brain, lungs and plasma of *Mus spretus* captured in Doñana National Park (southwest Spain) and surroundings. For this purpose, Terrestrial ecosystems of this Park affected by differential contamination from mine, industrial and agricultural activities have been considered. The different size exclusion chromatographic profiles traced by the metals have been used to assess environmental contamination and the interactive relationships between metabolically active metals is discussed to explain the mice response against polluted sceneries.

2. Material and Methods

2.1. *Standard solutions and reagents*

For total element determination tissues and plasma, nitric acid (65 mass %), hydrochloric acid (30 mass %), hydrofluoric acid (40 mass %) and hydrogen peroxide (30 mass %) of Suprapur® grade (Merck, Darmstadt, Germany) were used for mineralization of the samples.

All reagents used for sample preparation of cytosolic extracts from the different organs were of the highest available purity. Phenylmethanesulfonyl fluoride (PMSF) and tris(2-carboxyethyl)phosphine hydrochloride (TCEP) (BioUltra grade, >98%) were obtained from Sigma Aldrich (Steinheim, Germany). Helium used as collision gas in a ICP-ORC-MS system, was of high-purity grade (>99.999%).

Standards used for mass calibration of the analytical SEC column of mass range 70-3 kDa were as follows: ferritin (440 kDa) (purity 95%), bovine serum albumin (67 kDa) (purity 96%), superoxide dismutase containing Cu and Zn (32 kDa) (purity > 70%), metallothionein I containing Cd, Cu and Zn (7 kDa) (purity > 95%) and arsenobetaine (179 Da) (purity > 98%). All these reagents were purchased from Sigma-Aldrich (Steinheim, Germany). On the other hand, the standards used for mass calibration of the analytical SEC column of mass range <10kDa were: bovine serum albumin (67 kDa) (purity 96 %), metallothionein I containing Cd, Cu and Zn

(7 kDa) (purity > 95 %), vitamin B_{12} (1.35 kDa) (purity > 96 %) and reduced glutathione (307 Da) (purity 98–100 %). All these reagents were purchased from Sigma–Aldrich (Steinheim, Germany). Standard stock solutions with a concentration of 10 mg mL^{-1} were prepared by dissolving the respective compound in 20 mM of ammonium acetate at pH 7.4 purchased from Merck (Darmstadt, Germany). The mobile phase solution used in SEC was 20 mM of ammonium acetate (Suprapure grade), which was prepared daily with ultrapure water (18 MΩcm) from a Milli-Q system (Millipore, Watford, UK) and the pH adjusted to pH 7.4 with ammonia solution, this later prepared by dilution of 20% (w/v) ammonia solution (Suprapur, Merck) with ultrapure water.

2.2. *Apparatus*

A cryogenic homogenizer SPEX SamplePrep (Freezer/Mills 6770) was used to prepare the homogenates. Homogenized tissues were subsequently disrupted with a glass/teflon homogenizer. The extraction was followed by ultracentrifugation with an ultracentrifuge Beckman model L9-90 K (rotor 70 Ti). Polycarbonate bottles of 10 ml with cap assembly (Beckman Coulter) were used for this purpose. A microwave oven (CEM Matthews, NC, USA, model MARS) was used for the mineralization of extracts.

Trace metals and metal-linked biomolecules were analyzed with an inductively coupled plasma mass spectrometer Agilent 7500ce (Agilent Technologies, Tokyo, Japan) equipped with an octopole collision/reaction cell. Chromatographic separations were performed using a Model 1100 HPLC pump with detector UV (Agilent, Wilmington, DE, USA) as the delivery system. ICP-MS measurement conditions (Table 1) for collision (He) mode were optimized using a 2% (v/v) HNO_3 aqueous solution of ^{59}Co, ^{7}Li, ^{89}Y and ^{205}Tl (1 mg L^{-1}). The flow of collision gas was fixed at 3.7 mL min^{-1} for He in order to avoid or reduce the polyatomic interferences. Before the SEC-ICP-MS with Superdex-Peptide column the extracts were filtered through Iso-Disc poly(vinylidene difluoride) filters (25-mm diameter, 0.2-μm pore size) to avoid column overloading or clogging and ultrafiltered and preconcentrated (10 times) with AMICON 30K (Millipore) by centrifugation at 10.000 g for 30 min to 4ºC.

2.3. *Sampling area and animals*

Free-living mice (*Mus spretus*) were collected during the autumn 2009 in three sampling areas from Doñana National Park (DNP), southwest Spain. Three areas were considered concerning their differential contamination (Fig. 1): (i) the control was the sampling point located at "Lucio del Palacio" (LDP- green spot) that is a non-contaminated area in the core of the Park; (ii) "La Rocina Stream" (ROC- red spot) with strawberry, citrus fruit and grape fields in the surroundings, that is in addition affected by diffused pollution from the petrochemical and chemical activities from the industrial

Table 1. Operating conditions for ICP-MS and SEC

SEC conditions		
Column	Superdex™-75 (10x300 x13μm)	Superdex™-Peptide (10x300x13μm)
Resolution range	3-70 kDa	<10 kDa
Mobile phase	Ammonium acetate 20 mmol L^{-1} (pH 7.4)	
Flow rate	0.7 mL min^{-1}	
Injection volumen	20 μL	
UV. Visible wavelength	254 nm	
ICP-MS conditions		
Foward power	1500 W	
Plasma gas flow rate	15.0 L min^{-1}	
Auxiliary gas flow rate	1.0 L min^{-1}	
Carrier gas flow rate	0.9 L min^{-1}	
Sampling depth	8 mm	
Sampling and skimmer cones	Ni	
He flow	3.7 mL min^{-1}	
Q_{oct}	-18 V	
Q_p	-16 V	
Dwell time	0.3 per isotope	
Isotopes monitored	^{63}Cu, ^{65}Cu, ^{64}Zn, ^{66}Zn, ^{75}As, ^{57}Fe, ^{103}Rh, ^{114}Cd, ^{60}Ni, ^{207}Pb, ^{208}Pb.	

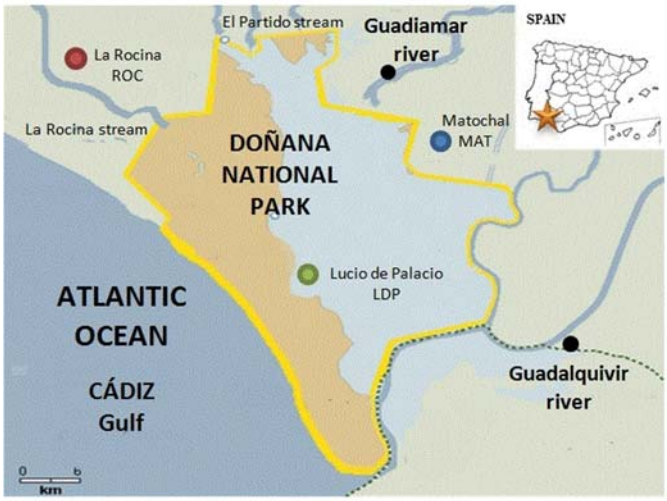


Figure 1. Sampling area in Doñana National Park and surroundings (SW Spain). Localization of sampling point: (i) "Lucio del Palacio" (LDP, green spot) non-contaminated site (control); (ii) "La Rocina" stream (ROC, red spot) and (iii) "Matochal" (MAT, blue spot) contaminated sites.

belt of Huelva, as well as by acid waters and metals from north-west mining metallurgical activities of Riotinto village (Huelva); and (iii) "el Matochal" (MAT- blue spot) site next to Guadiamar river that is affected by rice growing fields and suffered the input of metals transported by the Guadiamar river during the rupture of Aznalcollar mine tailing pond in 1998 [26]. A total of 68 *Mus Spretus* mice were caught with Sherman live-traps baited with a hazelnut cream over bread, which were mounted during the evenings and checked the next morning. Adult animals were taken alive to a laboratory at Doñana Biological Reserve, and site/date of capture, sex, weight and external measurements were recorded. In this study, 10 mice of each sampling area were caught. For each mouse, measurements of body size were recorded as body length to the nearest 0.5 mm (calculated by subtracting tail length from total length) and body weight measured to the nearest 0.1 g. On the basis of body weight and colour pattern all mice were considered as adults and they were used in this study. Then, mice were transported to the University of Huelva in other cleaned live-traps and they were sacrificed two hours later for plasma and organs extraction in a laboratory equipped for the manipulation of animals.

Mice were individually anesthetized by isoflurane inhalation and exsanguinated by cardiac puncture, dissected using a ceramic scalpel and finally, the organs were transferred rapidly to dry ice. Individual organs were excised, weighed in Eppendorf vials, cleaned with 0.9% NaCl solution, frozen in liquid nitrogen and stored at -80 ºC, until they were used for sample preparation. Plasma collection was carried out by centrifugation (4000rpm, 30 min, 4ºC) after addition of heparin (ANTICLOT) as an anticoagulant. Mice were handled according to the norms stipulated by the European Community. The investigation was performed after approval by the Ethical Committee of the University of Huelva (Spain).

2.4. *Analytical procedures*

2.4.1. Sample preparation

Livers, brain and plasma from 10 different male mice were treated following a procedure described elsewhere [5]. Briefly, individual organs were disrupted by cryogenic homogenization in a 6770 freezer/mill apparatus (2 min at rate 15) from Spex SamplePrep (Metuchen, NJ, USA). For SEC analysis, three pools of three individual homogenized organs were exactly weighed (0.500 g). After that the metal-biomolecules were extracted with a solution (3 ml of extractant solution per gram of cryhomogenized organs) containing 20 mM of ammonium acetate buffer solution at pH 7.4, 1 mM of tris(2-carboxyethyl)phosphine (TCEP) and 1 mM of phenylmethanesulfonylfluoride (PMSF) using a glass/teflon homogenizer in a cold chamber at a constant temperature of 4ºC. Then the extracts were centrifuged at 120,000 g for 1 h at 4 ºC. Extracts were stored under nitrogen atmosphere to avoid oxidation by air and at -80 ºC until analysis.

2.4.2. Determination of total metals concentration in the tissues and plasma of Mus spretus mice

For this purpose, five pools of two individual homogenized organs each were exactly weighed (0.100 g) in 5-ml microwave vessels and 500 µL of a mixture containing nitric acid and hydrogen peroxide (4:1 v/v) was added. After 10 min, the PTFE vessels were closed and introduced into the microwave oven. The mineralization was carried out at 400 W from room temperature, ramped to 160ºC for 15 min and hold for 40 min at this temperature. Then the solutions were made up to 2 ml and metals analyzed by ICP-MS. Rhodium was added as internal standard (1µg mL^{-1}). In the case of plasma, 100 µL of each mouse was weighed in 5-ml microwave vessels and the mineralization was carried out following the same conditions previously described for tissues digestion. For multielemental determination two replicates from each pool of two organs were carried out. For plasma, two replicates from each mouse specimen were carried out.

2.4.3. Analysis of cytosolic extracts and plasma of Mus spretus mice using SEC-ICP-MS

The SEC-ICP-MS online coupling was performed by connecting the outlet of the chromatographic column to the Micromist nebulizer inlet (GlasExpansion, Switzerland) of the ICP-MS by means of a 30cm PEEK tubing (0.17 i.d. mm). The quality control of the SEC-ICP-MS system was performed as described elsewhere [12] to overcome problems related to contamination, loss and stability of species. The retention times of the mass calibration standards used for the Superdex-75 column are the following: ferritin 12.0 min, bovine serum albumin (BSA) 13.5 min, superoxide dismutase containing Cu and Zn (Cu,Zn-SOD) 15.6 min, metallothionein I containing Cd, Cu and Zn (Cd,Cu,Zn-

MT1) 19.1 min, and reduced glutathione (GSH) 23.6 min. On the other hand, for the Superdex-Peptide column standards and retention times were: bovine serum albumin (BSA) 11.2 min, metallothionein I containing Cd, Cu and Zn (Cd,Cu,Zn-MT1) 14.6 min, vitamin B_{12} 19.3 min and reduced glutathione (GSH) 24.4 min.

3. Results and Discussion

The three sites considered in this study (Fig. 1) were selected by the differential presence of toxic metals. In the stream "La Rocina" (ROC) the presence of remarkable concentrations of nitrate, probably from agricultural sources, has been previously reported [27], and in this area metal levels in pore waters are also high, which suggests a transport of contaminants from the Iberian Pyrite Belt [28]. In addition, in 1998, a part of a 360 ha tailings dam of Aznalcóllar pyrite mine, located 60 km North of DNP, were released to Guadiamar stream, a tributary of Guadalquivir River, that were estimated at four cubic hectometers of acidic water and two cubic hectometers of mud. The high toxic metals content of this mud –35% Fe, 0.8% Zn, 0.8% Pb, 0.5% As, 0.2% Cu, 0.05% Sb, 0.006% Co, 0.005% Tl, 0.005% Bi, 0.0025% Cd, 0.0025% Ag, 0.0015% Hg, 0.001% Se—threatened DNP and the Guadalquivir Estuary [26]. This fact explains the high concentrations of non-essential elements (such as Cd and Hg) in tissues of shrews that can be associated to the pollution from mine activities [29], and high levels of Cu, Zn, Mn, Cd and As in soils and sediments [5].

Biological response of free-living mice *Mus spretus* to contamination at the above described site was studied in organs with high metabolic activity such as liver, brain and kidneys. In addition, plasma has been evaluated using the same procedure. The changes in the levels of metal–biomolecules caused by pollutants were traced by SEC-ICP-MS. The most interesting results were observed in relation to Cu, Zn, Fe and toxic elements such as Cd, As, Ni and Pb.

3.1. *Total metals concentration in the tissues and plasma of Mus spretus mice*

The presence of metals in the different organs and plasma of Mus spretus from the three sampling sites previously described (Fig 1) was evaluated. Recovery experiments were performed by spiking the extracts with 1, 5, 10, or 50 ìg L−1 of metals depending on the relative concentration of each one in the extracts; the results are also shown in Table 2 that confirm quantitative recoveries in all the cases. ICP-MS detection limits are also given in this table.

It is remarkable the higher presence of iron, copper and zinc in all organs studied and plasma. In contrast, the levels of arsenic, cadmium, lead and nickel were considerably lower. The lowest concentration of toxic metals was found in brain, lungs and plasma. The highest concentrations of metals such as iron, copper, zinc, arsenic, cadmium and lead have been found in mice captured in MAT, a contaminated area affected by metal pollution in soils and sediments [5], as a consequence of the breakdown of the dams from Aznalcóllar mines and their release from the rice growing fields. ROC presents an intermediate metal contamination between LDP and MAT (table 2). This fact can be explained by metallic contaminants from the Iberian Pyrite Belt and agricultural activities. In addition, the highest concentration of iron in plasma from mice captured in ROC could be related with the high content of this element in the Iberian Pyrite Belt [28], which is a site affected by the acid mine drainage. Numerous transport protein of iron can be found in plasma of mammals, such as transferrin, myoglobin and hemoglobin. On the other hand, the higher copper concentration in lungs from mice captured in LDP and MAT could be related with several agricultural activities developed around, which use pesticides and herbicides containing this element (table 2). The major concentration of arsenic in kidneys from mice sampled in MAT and ROC in comparison with mice captured in LDP could be explained by the proximity of the former to the Guadiamar river, affected by rice growing fields (pesticides and herbicides) that also suffered the input of metals transported by the river during the rupture of Aznalcollar mine tailing pond in 1998 [26],. The concentrations of arsenic in mice kidneys from ROC are also high, probably due to air pollution, since this site is nearest to the industry. In contrast, in ROC lower concentrations of this element in soils and sediments were found [5]. In addition, similar concentration of arsenic was found in lungs from mice captured in both areas (table 2). Finally, it is remarkable the higher concentration of cadmium in kidneys from mice sampled in LDP in comparison with ROC, which are in good agreement with a work previously published work [5].

3.2. *Size characterization of Cu, Zn, Cd, As, Ni and Pb-biomolecules in liver extracts from Mus spretus captured in contaminated and non-contaminated areas*

The study of accumulation and effects of heavy metals in living organisms is very important in connection with global environmental pollution. The induction of Cd and Zn-metallothioneins in *Mus musculus* mice exposed to industrial particles with high content of metals has been reported, and strong interactions between Cu–Zn–Cd-Pb have been confirmed [30]. Antagonistic interactions have been established between several heavy metals, particularly between Cd and Zn. In this sense, Cd replaces Zn from several proteins [31] and high amounts of Cd in the diet leads to deficiency of zinc in living organism [32]. In our study, we can observe potential interactions between these metals in liver of free-living mice from Doñana National Park and surroundings, especially in mice captured in MAT, specially Cu and Cd in the fraction of 7 kDa associated with MTs (Fig. 2).

It is remarkable the high intensities of the peaks traced by Cu, Zn and Cd at 7 kDa in the liver cytosolic extract, that match with the retention time of metallothionein I standard (Fig. 2). The intensity of a MT-Cu-peak from the contami-

Table 2. Concentration of elements (μg g-1) in the different organs and plasma of the mice

		Fe ($\mu g.g^{-1}$)	Cu ($\mu g.g^{-1}$)	Zn ($\mu g.g^{-1}$)	As ($\mu g.g^{-1}$)	Cd ($\mu g.g^{-1}$)	Ni ($\mu g.g^{-1}$)	Pb ($\mu g.g^{-1}$)
LIVER	*Mus Spretus* LDP	8.57 ±0.201	0.369 ±0.0371	4.80 ±0.103	0.00161 ±0.000221	0.0921 ±0.000247	0.0242 ±0.00281	<LOD
	Mus Spretus MAT	14.9 ±0.181	1.328 ±0.0674	10.658 ±0.199	0.00512 ±0.00138	0.00321 ±0.000522	0.0751 ±0.00493	0.00714 ±0.00152
	Mus Spretus ROC	10.3 ±0.220	1.45 ±0.0552	10.3 ±0.208	0.00421 ±0.00181	0.00143 ±0.000321	0.0315 ±0.00373	0.00601 ±0.00211
BRAIN	*Mus Spretus* LDP	1.94 ±0.0912	0.487 ±0.0211	2.94 ±0.0663	0.00102 ±0.000221	0.00112 ±0.000441	<LOD	<LOD
	Mus Spretus MAT	2.96 ±0.111	0.578 ±0.0382	2.96 ±0.0881	0.00622 ±0.00171	0.00944 ±0.00152	<LOD	0.00162 ±0.000511
	Mus Spretus ROC	2.01 ±0.0792	0.492 ±0.0873	2.93 ±0.102	0.00221 ±0.000822	0.00243 ±0.000741	<LOD	<LOD
KIDNEYS	*Mus Spretus* LDP	5.51 ±0.207	0.268 ±0.0232	3.38 ±0.111	0.00192 ±0.000311	0.0277 ±0.00342	0.00284 ±0.000742	<LOD
	Mus Spretus MAT	10.4 ±0.214	0.406 ±0.0214	5.15 ±0.104	0.00332 ±0.00123	0.0324 ±0.0111	0.0122 ±0.00212	<LOD
	Mus Spretus ROC	5.37 ±0.0812	0.467 ±0.0161	3.46 ±0.0972	0.00484 ±0.000942	0.0151 ±0.00922	0.00312 ±0.00121	<LOD
LUNGS	*Mus Spretus* LDP	10.7 ±0.524	0.158 ±0.0213	0.296 ±0.0252	<LOD	<LOD	<LOD	<LOD
	Mus Spretus MAT	13.6 ±0.705	0.201 ±0.0584	0.554 ±0.00612	0.0221 ±0.00182	0.00232 ±0.000914	<LOD	<LOD
	Mus Spretus ROC	9.048 ±0.458	0.147 ±0.0351	0.321 ±0.00344	0.0121 ±0.00122	0.00124 ±0.000522	<LOD	<LOD
PLASMA	*Mus Spretus* LDP	4.69 ±0.251	0.223 ±0.0380	1.641 ±0.0512	<LOD	<LOD	<LOD	<LOD
	Mus Spretus MAT	5.25 ±0.321	1.14 ±0.0842	2.18 ±0.0634	0.00241 ±0.000724	0.00154 ±0.000422	<LOD	0.00122 ±0.000341
	Mus Spretus ROC	10.5 ±0.503	0.553 ±0.0341	1.89 ±0.0607	0.0013 ±0.000411	<LOD	<LOD	<LOD
LOD ($\mu g.L^{-1}$)		**0.0471**	**0.0384**	**0.494**	**0.0272**	**0.00911**	**0.0244**	**0.0322**
Recovery (%)		**117**	**105**	**121**	**99**	**102**	**113**	**97**

LOD: Limit of Detection

nated area MAT is clearly higher than that from other sites, such as ROC and LDP. In addition, the presence of significant peaks of MT-Cd (Fig. 2) and MT-Pb (Fig. 3) in the same sampling site are also remarkable. In contrast, the intensity of this peak traced by Zn for mice captured in MAT presents lower intensity. Metallothioneins (MTs) are synthesized by mammals in response to heavy metal stress [33,34]. Detoxification of heavy metals and scavenging of free radicals are also reported as important actions of MT which are present in mammalian tissues in two major isoforms, MT-1 and MT -2. [35-37]. Due to the relative concentrations of Cd/Cu and Zn in MAT discussed before, is possible to suppose that cadmium and copper can replace the zinc binding to MTs, because these elements have greater affinity for sulfhydryl groups of cysteine present in these proteins [30]. For this reason, up-regulation of MTs triggered in liver by the presence of toxic elements such as Cd and high levels of Cu can be considered as biomarkers of free-living organisms exposure to environmental pollution, in particular to toxic metals.

Another peak traced by Cu, which matches with the retention time of superoxide dismutase (SOD) at about 32 kDa, shows higher intensity for samples from MAT and LDP (Fig. 2). This fraction has been purified and identified by mass

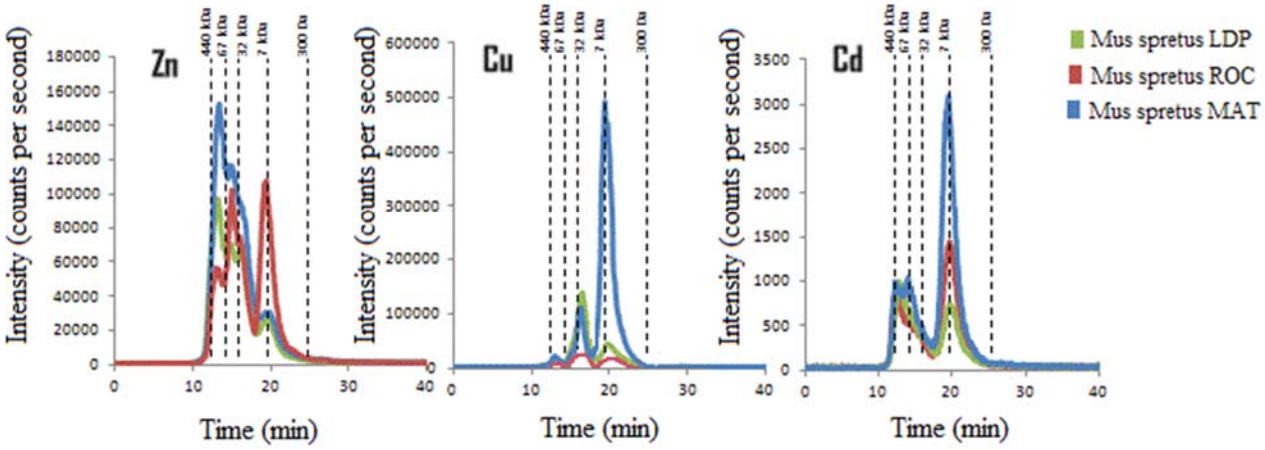


Figure 2. Zn, Cu and Cd-biomolecules complexes in liver of *Mus spretus* from different environmental areas assessed by molecular mass distribution with SEC-ICP-MS. Chromatographic conditions: column, Superdex™-75 (10x300x1.3 µm); mobile phase, ammonium acetate 20 mmol L^{-1} (pH 7.4); flow rate 0.7 ml min^{-1}; injection volume, 20 µL

spectrometry in a previous work [5] demonstrating that peak intensity of the copper can be attributed unequivocally to the enzyme superoxide dismutase (Cu/Zn-SOD), a biomarker used in numerous studies to assess environmental stress [1,8]. In Fig. 2 shows two signals traced by Zn that matches with standards of bovine serum albumin (Zn-BSA) and superoxide dismutase (Cu,Zn-SOD) at about 67 kDa and 32 kDa, respectively. The presence of pollutants in sites such as MAT and ROC enhances the intensity of these peaks, but in the non-polluted LDP the intensity decreases. It is remarkable the results obtained in samples from ROC where despite the significant presence of contaminants, the intensity of putative peaks matching with Zn-BSA and Zn-SOD decreases, which can be explained by zinc homeostasis and Cu, Zn and Cd interactions [30,31].

Finally, it is interesting to observe the presence of arsenic and nickel in liver extracts (Fig. 3). SEC-ICP-MS shows the presence of low molecular mass biomolecules of As, that are less abundant in MAT, and also some small Ni containing compounds, highly abundant in the same site. The high concentration of arsenic in soils and sediments analyzed in previous works [5] led us to conclude that arsenic species to can be bonded to thiol groups of proteins [38] in mice captured in MAT. In addition, the concentration of arsenic found in lung in this area could be related with volatile arsenic species inhaled (table 2). Arsenic and its compounds, especially the trioxide, are used in the production of pesticide, treated wood products, herbicide, and insecticide that maybe are

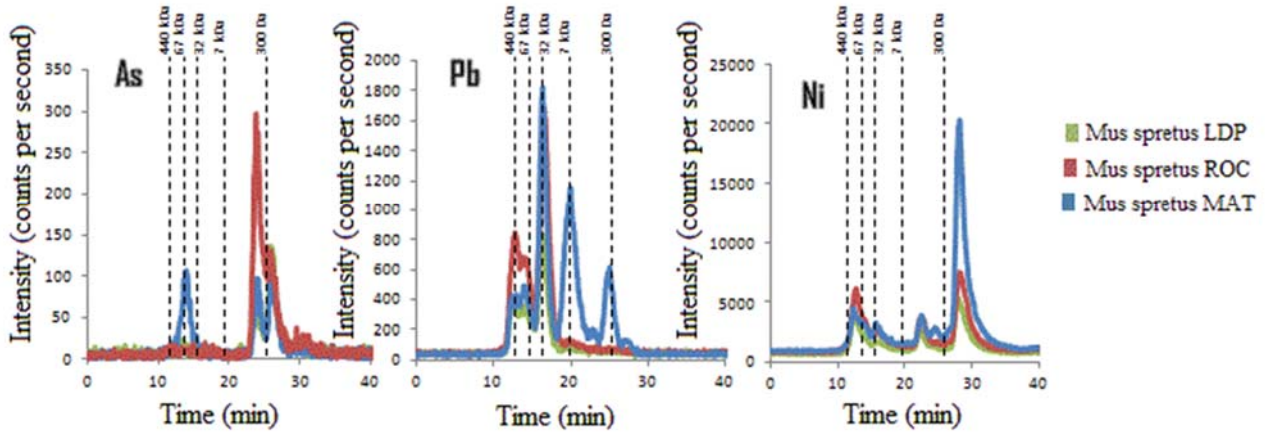


Figure 3. As, Pb and Ni-biomolecules complexes in liver of *Mus spretus* from different environmental areas assessed by molecular mass distribution with SEC-ICP-MS. Chromatographic conditions: column, Superdex™-75 (10x300x1.3 µm); mobile phase, ammonium acetate 20 mmol L^{-1} (pH 7.4); flow rate 0.7 ml min^{-1}; injection volume, 20 µL.

used in the rice and grape fields of MAT and ROC, respectively. When the biomethylation of inorganic arsenic is saturated in liver, the trivalent inorganic arsenic (the most abundant arsenic species in the terrestrial ecosystems) is chemically more reactive than pentavalent species and binds many carrier proteins, such as albumin and transferrin (Tf) (67kDa) [39]. On the other hand, the presence of low molecular mass nickel metabolites and nickel containing metalloproteins is remarkable in liver cytosolic extracts from mice captured in MAT. This element is rarely found in the terrestrial ecosystems; therefore this fact can be explained by the presence of this metal in the leachates from the breakdown of the ponds of Aznalcollar mines in 1998.

3.3. *Size characterization of Cu, Zn, Cd and As -biomolecules in brain extracts from Mus spretus captured in contaminated and non-contaminated areas*

In the case of brain the coupling SEC-ICP-MS was again used to obtain the Cu, Zn, Cd and As-traced peaks in the cytosolic fractions of M. spretus (Fig. 4). These chromatograms show a remarkable intensity of the peaks traced by Cu, Zn, and Cd at 7 kDa as in the case of liver extracts discussed before, which match well with the retention time of metallothionein, although the most abundant MT isoform in brain is MT-3 [35-37]. In MAT the intensities of the 7 kDa peak traced by Cd, Cu and Zn presents higher intensity than ROC and LDP, respectively. In the case of Cu tracing in MAT is not visible in Figure 4, unless it parallels exactly the ROC tracing. On the other hand, the presence of small arsenic-containing metabolites is significantly higher in brain of mice captured in the polluted areas (Fig. 4) that follows the range MAT>ROC>LDP. The presence of this toxic element in the brain can be explained because experiments performed with rats as model organisms have revealed that As can cross the blood brain barrier (BBB) producing an increasing presence of reactive oxygen species (ROS) as well as oxidative stress [40].

3.4. *Size characterization of Cu, Zn, Cd and As -biomolecules in plasma from Mus spretus captured in contaminated and non-contaminated areas*

The most interesting results obtained by the analysis of plasma samples by SEC-ICP-MS are related with Cu, Zn, Cd and As-biomolecules (Fig. 5). In the chromatogram traced by Cu we can observe a peak of about 67 kDa of high intensity. This peak is can be related with the transport proteins Cu/Zn-BSA and Cu-Tf of 67 kDa and 79kDa, respectively. The same 67 KDa-matched peak is also traced by Zn which agrees with this hypothesis (Fig. 5). Higher intensities of both metals were detected in plasma from mice captured in MAT which agrees with the increased presence of these metals in the sampling area both in soils and sediments [5]. The profile traced by Cd presents differences between mice captured in MAT with those captured in LDP and ROC (Fig. 5),

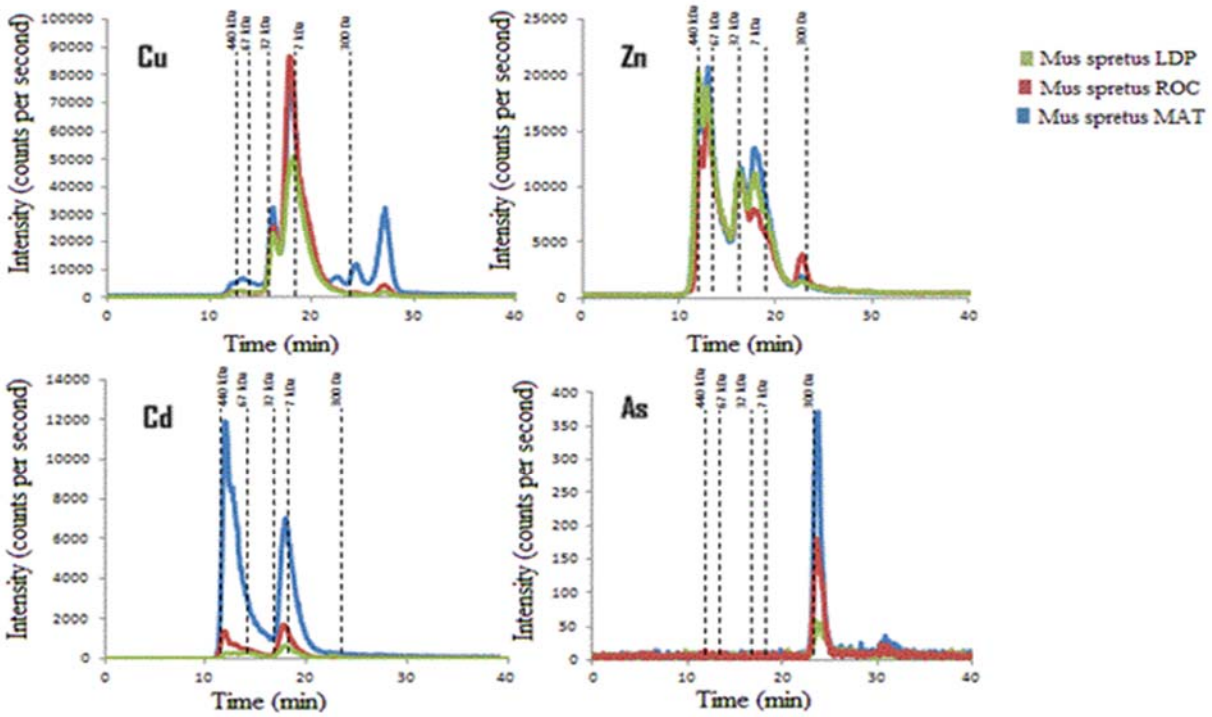


Figure 4. Cu, Zn, Cd and As-biomolecules complexes in brain of *Mus spretus* from different environmental areas assessed by molecular mass distribution with SEC-ICP-MS. Chromatographic conditions: column, Superdex™-75 (10x300x1.3 µm); mobile phase, ammonium acetate 20 mmol L^{-1} (pH 7.4); flow rate 0.7 ml min^{-1}; injection volume, 50 µL.

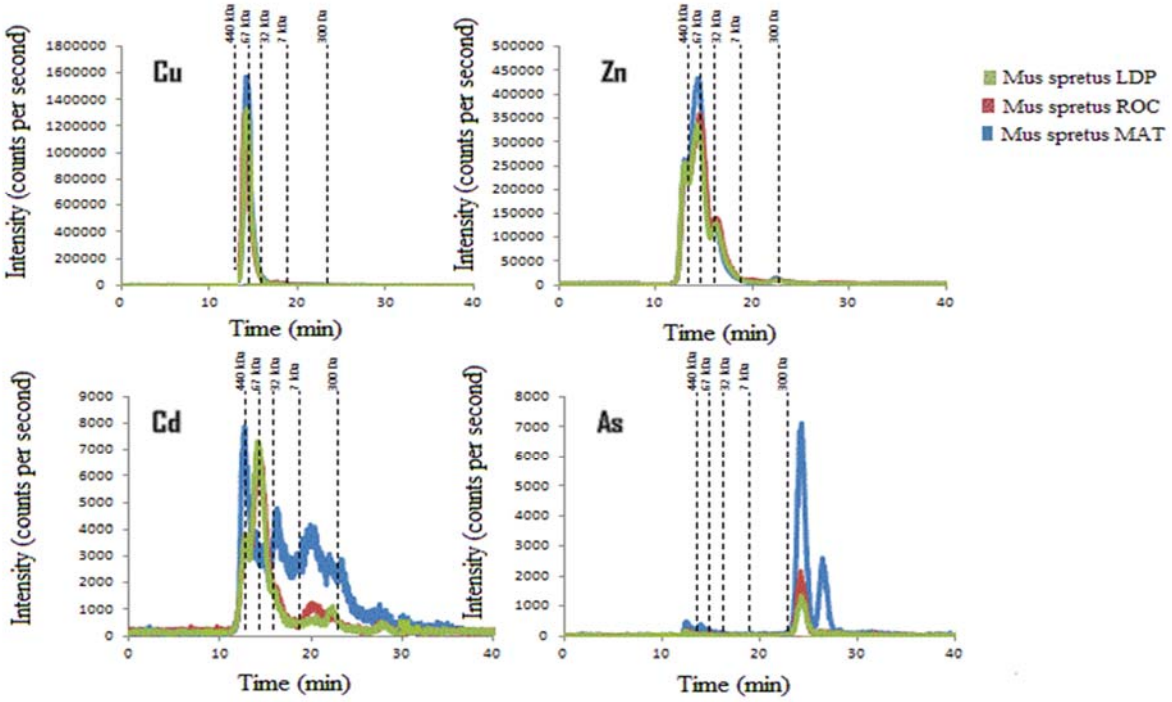


Figure 5. Cu, Zn, Cd and As-biomolecules complexes in plasma of *Mus spretus* from different environmental areas assessed by molecular mass distribution with SEC-ICP-MS. Chromatographic conditions: column, Superdex™-75 (10x300x1.3 μm); mobile phase, ammonium acetate 20 mmol L^{-1} (pH 7.4); flow rate 0.7 ml min^{-1}; injection volume, 50 μL.

and the chromatograms show several peaks that can be explained by the high affinity of this element for thiol groups of proteins as stated by other authors [41]. Finally, in the chromatogram obtained for As-biomolecules (Fig. 5) is also observed the high intensity obtained in plasma samples from mice captured in MAT that confirm the assertion of Suzuki et al [41] about the potential use of arsenic presence in blood and plasma as biomarker of exposure to this element.

3.5. *Size characterization of Cu, Cd and As-biomolecules in kidneys extracts from Mus spretus captured in contaminated and non-contaminated areas*

In order to evaluate potential interactions between Cu and Cd, the cytosolic extracts of kidneys were ultrafiltered using microcentrifugal filters with a cut-off mass of 30kDa and directly analyzed by SEC-ICP-MS using a Superdex-Peptide column (with mass range <10kDa). The results obtained shown different expression profiles for Cu, Cd and As-biomolecules in kidneys ultrafiltered extracts from mice captured in the different areas of study (Fig. 6). As previously discussed in liver (Fig. 2), potential interactions between Cu and Cd are observed in kidney from mice captured in MAT and ROC in the fraction of 7 kDa related with MTs. In MAT, the higher presence of cadmium in the ecosystem and the greater affinity of this element for MTs explain the high intensity of the cadmium peak about 7 kDa against the lower intensity of copper in the same peak (Fig. 6).

On the other hand, the presence of low-molecular-mass As species in kidney cytosolic extracts analyzed by SEC-ICP-MS (column Superdex-Peptide) can be seen in Fig. 6. The higher intensity of the signal was obtained from ROC, in which total As concentration in kidney organs is also the highest (Table 2). High concentrations of arsenic in kidneys have been explained by other authors by the diet [42], while in contrast, the major concentration of arsenic in plasma, liver and brain could be related with the inhalation of this element [43] arsenic. In addition, inorganic arsenic exposure in humans, by the inhalation route, has been shown to be strongly associated with lung cancer [44], while ingestion of inorganic arsenic in humans has been linked to a form of skin cancer and also to bladder and kidney cancer [45].

4. Concluding Remarks

The importance of monitoring the exposure and studying the effects of heavy metals in living organisms has increased in the last decades. Studies of small mammals, mainly free-living mice (Mus spretus), have been used as bioindicators in numerous environmental studies because they can provide useful information for assessment of risk of metals to humans. The use of SEC-ICP-MS coupling is a good choice to

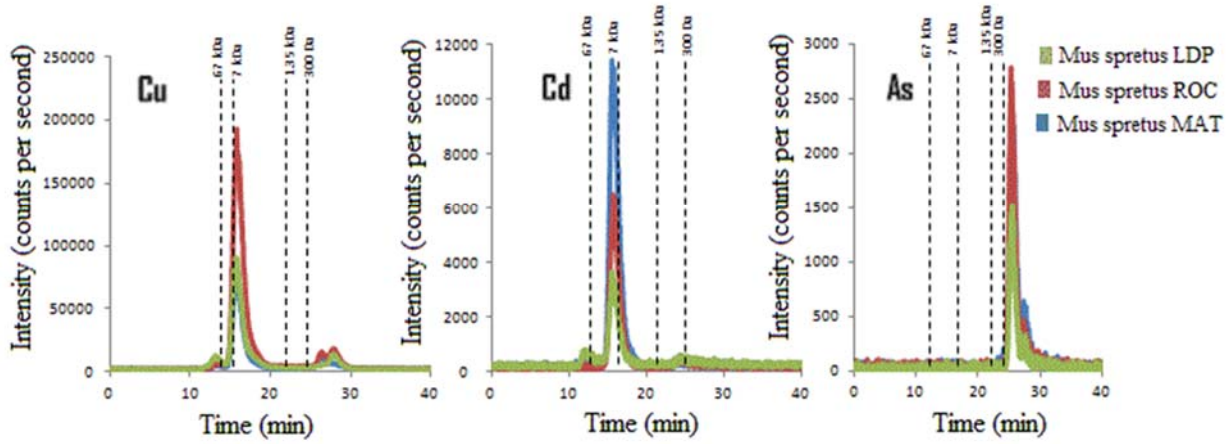


Figure 6. Cu, Cd and As-biomolecules complexes in kidneys of *Mus spretus* from different environmental areas assessed by molecular mass distribution with SEC-ICP-MS. Chromatographic conditions: column, Superdex™-Peptide (10x300x1.3 μm); mobile phase, ammonium acetate 20 mmol L^{-1} (pH 7.4); flow rate 0.7 ml min^{-1}; injection volume, 20 μL.

assess changes in the profiles of metal-binding biomolecules in environmental bioindicators (e.g. Mus spretus) caused by metal pollution. The study of liver, brain, kidneys, lungs and plasma from mice captured in contaminated and non-contaminated areas in Doñana National Park (southwest Spain) and the surroundings reveals differences in the expression of Cu, Zn, Cd, As, Pb and Ni-biomolecules that can be related with contamination episodes. These results, confirm the potential metallomics for environmental issues assessment and the characterization of metal interactions in organisms exposed to contamination. Future works are necessary to identify by organic mass spectrometry the metal-biomolecules traced by SEC-ICP-MS as well as the identification and quantification of the arsenic containing peaks by HPLC-MS. In addition, the use of micro-mammals as bioindicators to evaluate terrestrial ecosystems requires analytical techniques with sensitive elemental or molecular specific detectors. In this sense, the potential of inductively coupled plasma-mass spectrometry (ICP-MS) joining great sensitivity, selectivity, precision and multiple heteroatoms detection of metalloproteins makes essential its use in this type of environmental studies. On the other hand, the intensities of SEC-ICP-MS signals from metal-containing biomolecules show a complex metal-profile that cannot be well resolved, and for this reason, multidimensional chromatographic approaches will be used to overcome this problem in further studies, such as ionic exchange chromatography and, finally, the altered metalloproteins will be identified using organic mass spectrometry.

Acknowledgements

This work was supported by the project CTM2009-12858-C02-01 from the Ministerio de Ciencia e Innovación, and the projects P08-FQM-3554 and P09-FQM-4659 from Consejería de Innovación Ciencia y Empresa (Junta de Andalucía). M. Angel García Sevillano thanks the Ministerio de Educación for a predoctoral scholarship.

References

[1] D. Bonilla-Valverde, J. Ruiz-Laguna, A. Muñoz, J. Ballesteros, F. Lorenzo, J.L. Gómez-Ariza, J. López-Barea, Toxicology 197 (2004) 123–138. DOI: 10.1016/j.tox.2003.12.010

[2] B. Sure, Aquatic Ecology 35(2) (2001) 245-255. DOI: 10.1023/A:1011422310314

[3] R. Montes Nieto, T. García-Barrera, J.L. Gómez-Ariza, J. López-Barea, Environmental Pollution 158 (2010) 401-408. DOI: 10.1016/j.envpol.2009.09.005

[4] A. Romero-Ruiz, J. Alhama, J. Blasco, J.L. Gómez-Ariza, J. López-Barea, Environmental Pollution 156 (2008) 1340–1347. DOI:10.1016/j.envpol.2008.02.02

[5] M. A. García-Sevillano, M. González-Fernández, R. Jara-Biedma, T. García-Barrera, J. López-Barea, C. Pueyo and J. L. Gómez-Ariza, Analytical and Bioanalytical Chemistry 404 (2012) 1967-1981. DOI: 10.1007/s00216-012-6274-2

[6] T. García-Barrera, J.L. Gómez-Ariza, M. González-Fernández, F. Moreno, M.A. García-Sevillano, V. Gómez-Jacinto, Analytical and Bioanalytical Chemistry 403(8) (2012) 2237-2253. DOI:10.1007/s00216-012-5776-2

[7] V. Funes, J. Alhama, J. López-Barea, J. Peinado, Journal of Environmental Pollution 139 (2006) 214-223. DOI: 10.1016/j.envpol.2005.05.016

[8] A. Vioque-Fernández, E. Alves de Almeida, J. López-Barea, Science of the Total Environment 407 (2009) 1784-1797. DOI: 10.1016/j.scitotenv.2008.11.051

[9] R. Montes-Nieto, C.A. Fuentes-Almagro, D. Bonilla-Valverde, M.J. Prieto-Álamo, J. Jurado, M. Carrascal, J.L. Gómez-Ariza, J. López-Barea, C. Pueyo, Proteomics 7 (2007) 4376–4387. DOI: 10.1002/pmic.200700409

[10] J. Ruiz-Laguna, N. Abril, T. Garcia-Barrera, J.L. Gómez-Ariza, J. López-Barea, C. Pueyo, Environmental Sciences & Technologies 40 (2006) 3646-3652. DOI: 10.1021/es060056e

[11] J. López-Barea, J.L. Gómez-Ariza, Proteomics 6 (2006) 51-62. DOI: 10.1002/pmic.200500374

[12] M. González-Fernandez, T. García-Barrera, J. Jurado, M.J. Prieto-Álamo, J. López-Barea, C. Pueyo, J.L. Gómez-Ariza, Pure Applied Chemistry 80 (2008) 2609-2626. DOI: 10.1351/pac200880122609

[13] S. Mounicou, J. Szpunar, R. Lobinski, Chemical Society Reviews 38 (2009) 1119–1138. DOI: 10.1039/b713633c

[14] A. Sanz-Medel, Analytical and Bioanalytical Chemistry 390 (2008) 1–2. DOI: 10.1007/s00216-007-1705-1

[15] R.R. Crichton, M. Charloteaux-Wauters, European Journal of Biochemistry 164(3) (1987) 485-506. DOI: 10.1111/j.1432-1033

[16] J.R. Prohaska, American Journal Clinical Nutrition 88(3) (2008) 826–829.

[17] A. Sigel, H. Sigel, R.K.O. Sigel,Editors, ,Metal Ions in Life Sciences, Volume 5, Metallothioneins and Related Chelators (2009) The Royal Society of Chemistry (RSC),Cambridge 536 Hard cover, 266 (1-3), 55-56.

[18] D.P. Giedroc, A.I. Arunkumar, Dalton Transactions 2007, 3107–3120. DOI:10.1039/B706769K

[19] W. Zheng, M. Aschner, J. F. Ghersi-Egea, Toxicology Applied Pharmacology 192 (2003) 1–11. DOI: 10.1016/S0041-008X (03)00251-5

[20] D. Sadava et al. Life, the Science of Biology, 9th Edition. Macmillan Publishers, 2009.

[21] T. Kimura, N. Itoh, Journal of Health Science, 54 (3) (2008) 251-260.

[22] J. Bettmer, M. Montes Bayón, J. Ruiz Encinar, M.L. Fernández Sánchez, M.R. Fernández de la Campa, A. Sanz Medel, Journal of Proteomics 72 (2009) 989-1005. DOI: 10.1016/j.jprot.2009.05.003.

[23] 23. S. Mounico, J. Szpuna, R. Lobinsk, European Journal of Mass Spectrometry 16(3) (2010 243-253. DOI: 10.1255/ejems.1059.

[24] J.M. Marchante-Gayón, C. Sariego Muñíz, J.I. García Alonso, A. Sanz-Medel, Analytical Chimica Acta 400 (1999) 307–320. DOI:

[25] S.D. Tanner, V.I. Baranov, D.R. Bandura, Spectrochimica Acta Part B: Atomic Spectroscopy, 57(9) (2002) 1361-1452. DOI: S0584-8547(02)00069-1

[26] J.O. Grimalt, M. Ferrer, E. Macpherson, Science of the Total Environment 242 (1999) 3-11.

[27] G. Tortosa, D. Correa, A.J. Sánchez-Raya, A. Delgado, M.A. Sánchez-Monedero, E.J. Bedmar, Ecological Engineer 37 (2001) 539–548.

[28] A. Tovar-Sanchez, M.A. Huerta-Diaz, J.J. Negro, M.A. Bravo, S.A. Sañudo-Wilhelmy, Journal Environmental Manag 78 (2006) 286-293.

[29] A. Sánchez-Chardi, C.A. Oliveira Ribeiro, J. Nadal, Chemosphere 76 (2009) 387–394.

[30] S. Teodorova, R. Metcheva, M. Topashka-Ancheva, Environmental Research 91 (2003) 85–94.

[31] H.M. Perry, E.F. Perry, J.E. Putifoy, J.N. Erlanger, Trace Subst. Environ. Health 7 (1973) 281–288.

[32] A.W. Voors, M.S. Shuman, P.M. Galagher, Trace Subst. Environ. Health 6 (1972) 215–222.

[33] M. J. Stillman, C. F. Shaw and K. T. Suzuki, VCH, New York, 1992.

[34] J.F. Riodan, B.L. Valee, Methods in Enzymology, vol. 205, Academic Press, New York, 1991.

[35] M. Sato, K.T. Suzuki, Biomedical Research on Trace Element, 6 (1995) 13.

[36] Y. Uchida, Neurosignals 3 (1994) 211-215. DOI: 10.1159/000109547

[37] C.J. Quaife, S.D. Findley, J.C. Erickson, G.J. Froelick, E.J. Kelly, B.P. Zambrowicz, R.D. Palmiter, Biochemistry 33 (1994) 7250.

[38] Aposhian, H.V., Aposhian, M.M., Chemical Research in Toxicology, 19 (2006) 1-15. DOI: 10.1021/tx050106d

[39] M.G. Bogdana, A. Samgayo-Reyest, H.V. Aposhian, Toxicology 93 (1994) 175-193. DOI: 10.1016/0300-483X(94) 90077-

[40] F. Piao, N. Ma, Y. Hiraku, M. Murata, S. Oikawa, F. Cheng, L. Zhong, T. Yamauchi, S. Kawanishi, K. Yokoyama, Journal of Occupational Health 47 (2005) 445–449. DOI:10.1539/joh.47.445

[41] A.Y. Aoki, K.T. Suzuki, Journal of Biochemical Toxicology 2 (1) (2006) 67–71, DOI: 10.1002/jbt.2570020107

[42] 42. E.M. Kenyo, L.M. Del Raz, Toxicological Sciences 85 (1) (2005) 468-475. DOI: 10.1093/toxsci/kfi107

[43] S.W. Burchiel, L.A. Mitchell, F.T. Laver, X. Sun, J.D. McDonald, L.G. Hudson, K.J. Liu, Toxicology & Applied Pharmacology, 241 (2009) 253-259. DOI: 10.1016/j.taap.2009.09.019.

[44] H.J. Lubi, L.E. Moor, J.F. Fraumeni, K.P. Canto, Health Perspectives 116(12) (2008) 1661–1665. DOI: 10.1289/ehp.1151

[45] C.J. Chen, C.W. Chen, M.M. Wu, T.L. Kuo, British Journal of Cancer 66 (1992) 888-892.

JOURNAL OF INTEGRATED OMICS
A METHODOLOGICAL JOURNAL
HTTP://WWW.JIOMICS.COM

ORIGINAL ARTICLE | DOI: 10.5584/jiomics.v3i2.136

Global proteomic profiling of the membrane compartment of bovine testis cell populations

Michelle L. Colgrave[1,2†*], Sally Stockwell[1,2†], Aimee Grace[3,4], Mary McMillan[3,4], Rhonda Davey[3], Sigrid Lehnert[1,2], Sabine Schmoelzl[2,3,4]

[1]CSIRO Animal, Food and Health Sciences, 306 Carmody Rd, St Lucia, QLD 4067; [2]CSIRO Food Futures National Research Flagship, 5 Julius Avenue, North Ryde, NSW 2113, Australia; [3]CSIRO Animal, Food and Health Sciences, F. D. McMaster Laboratory, Armidale, NSW 2530; [4]School of Environmental and Rural Science, University of New England, Armidale, NSW 2351; [†]*These authors have contributed equally to the work.*

Received: 24 April 2013 **Accepted:** 03 June 2013 **Available Online**: 16 July 2013

ABSTRACT

Spermatogonial stem cells hold enormous potential in mammalian reproductive medicine through the preservation of gametes, the restoration of fertility, enhancement of germ-lineage genetic manipulation and the improvement in our understanding of stem cell biology. Here we describe the protein profiles of the membrane compartment of bovine testicular cell isolates which were enriched for germ cells using differential plating. The isolated cells were characterised with antibodies to UCHL1 (previously known as PGP9.5) for type A spermatogonia; DDX4 (previously known as VASA) for germ cells and vimentin for Sertoli cells. Ultracentrifugation techniques were used to specifically isolate cell membranes, with membrane protein identifications significantly increased when compared to whole cell lysates. We utilised the filter-aided sample preparation protocol for improved digestion efficiency of membrane proteins. Using ESI-LC-MS/MS, we compared the proteins present in two cell populations. A total of 1,387 proteins were identified in bovine testis cell isolates, of which 39% were membrane-associated. A total of 64 proteins were differentially expressed in the non-adhered fraction (enriched for undifferentiated germ cells) compared to the adhered fraction, of which 16 were unique to this cell population and the remaining 48 showed a two-fold change (increase when compared to the adhered cell population) as judged by spectral counting. This analysis revealed a number of candidate germ cell markers including the known markers, DDX4 and UCHL1. The proteomic profiles generated in this study support and complement transcription data on gene expression and histological levels, and reinforce the potential of proteomics in identifying and characterising the protein effectors of self-renewal and/or differentiation in stem cells.

Keywords: Spermatogonial stem cells; germ cells; bull testis; cell surface; proteomics; membrane.

Abbreviations

BSA: bovine serum albumin; **DBA**: *Dolichos biflorus* agglutinin; **DMEM/F12**: Dulbecco's Modified Eagle Medium: Nutrient Mixture F12; **FASP**: filter-aided sample preparation; **PBS**: phosphate buffered saline; **PGP9.5/UCHL1**: ubiquitin carboxy-terminal hydrolase L1; **qRT-PCR**: Quantitative Reverse Transcribed Polymerase Chain Reaction; **RT**: room temperature; **SSC**: spermatogonial stem cell; **TBS**: Tris-buffered saline; **VASA/DDX4:** ATP-dependent RNA helicase DDX4.

1. Introduction

The mammalian spermatogonial stem cell (SSC) population is set aside early during embryonic development and differentiation of SSC in adults results in the continual production of sperm in the testis.[1]. Spermatogonial stem cells are a subset of the undifferentiated spermatogonial cell population, and are defined by the ability to colonise testis tissue after transplantation [2]. At present there are no definitive markers that can be used to distinguish SSC from other cell types in the testis of any species, limiting the ability to identify and purify SSC populations [3-5]. A gene expression array -based whole genome approach to identify biomarkers employed a rare condition of defective spermatogenesis to sepa-

***Corresponding author:** Michelle L. Colgrave, CSIRO Animal, Food and Health Sciences, 306 Carmody Rd, St Lucia, QLD 4067 Australia; Tel: +61 (0)7 3214 2697; Fax: +61 (0)7 3214 2900. Email address: michelle.colgrave@csiro.au

rate spermatogonial markers, but the markers identified by this approach could not be confirmed as spermatogonial stem cell markers [5]. Pluripotency markers such as SSEA4 have been found to be expressed in human repopulating SSC [6], but such markers have not been confirmed for bovine SSC.

Spermatogonial stem cells reside on the basal membrane of seminiferous tubules in close association to Sertoli cells. While no specific markers are known so far for SSCs, there are several established markers for spermatogonia and Sertoli cells in different species. Probable ATP-dependent RNA helicase (DDX4; VASA) is a highly conserved molecular marker for the germ cell lineage across species, expressed in primordial germ cells and undifferentiated spermatogonia as well as throughout the germ cell lineage [7, 8]. Established markers for bovine undifferentiated spermatogonia are ubiquitin carboxyl-terminal hydrolase isozyme L1 (UCHL1) [9], and binding of *Dolichos biflorus* agglutinin (DBA), a plant lectin binding to sugar residues on the surface of bovine undifferentiated spermatogonia [10-12]. Sertoli cells, which are the only other cell population present in seminiferous tubules besides germ cells, can be identified by their expression of GATA4 [13] and vimentin [14, 15].

Proteomics provides a feasible route to profiling testis cell populations, complementing the genomics and molecular biology efforts. Improvements in sensitivity, accuracy and speed of mass spectrometric analysis when combined with automated methods for data acquisition and processing have enabled protein profiling experiments in a high throughput manner. Over recent years, proteomic studies have been applied to the study of SSC from rats [16], mice [17-19], humans [20], chickens [21] and dogfish [22]. Proteomics can reveal information about protein expression (presence and quantity) and post-translational modification which facilitates our understanding of protein interactions and function in complex systems. For example, proteomics has played a critical role in understanding complex biological processes, such as that of spermatogenesis [23].

Methods for the isolation and culture of SSC from the testes of mice have been available for decades [24, 25]. Since then, human SSC have been successfully cultured [26]. A major impediment to the development of systems to culture cattle SSC is the lack of specific markers for isolation and comparison of stem cell populations. While genomic and transcriptomic studies in mice and humans have provided candidate markers, no such list or "molecular signature" is available for studies of bovine SSC. The identification of suitable markers is further confounded by the requirement that proteins with utility for sorting of cell populations must be surface-exposed. Plasma membrane proteins are typically of low abundance and/or of low solubility in cell lysates, hampering their identification and necessitating membrane fractionation prior to analysis.

In the current investigation, a global proteomic profiling experiment was applied to the analysis of membrane-enriched fractions from bovine testis cell populations generated by differential plating. The presence or absence of proteins in each population was assessed and the relative abundance was examined. The relative abundance of each candidate germ cell marker was assessed first by protein score and secondly by spectral counting. A short list of candidate germ cell markers was investigated further for differential protein or gene expression by immunohistochemistry and qRT-PCR.

2. Materials and Methods

2.1. *Cell isolation*

All animal experimentation was approved by the Armidale Animal Ethics Committee (Animal Research Authority 11-09). Testes of 15-30 g weight were harvested from prepubertal bull calves at slaughter and transferred to the laboratory in cold phosphate buffered saline (PBS; no calcium or magnesium; Invitrogen, Carlsbad, CA, USA) + penicillin (100 U/mL) and streptomycin (100 µg/mL) (Pen/Strep, Gibco, Grand Island, NY, USA). Single cell isolations were prepared according to Herrid *et al* [12] with some modification. The tunica vaginalis and the epididymis were removed and each testis weighed and washed in Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F12; Invitrogen) + Pen/Strep. The seminiferous tissue was dissected free, minced finely in fresh DMEM/F12 + Pen/Strep and ground gently in a tea strainer using a 5 mL syringe plunger to remove interstitial cells. The remaining tubules were then washed several times in a 50 mL Falcon tube with the DMEM/F12 solution and digested in collagenase (Type IV; Sigma-Aldrich, St Louis, MO, USA; 2 mg/mL; 20 mL/10 g of tubules) in a shaking 37°C water bath until individual tubules were observed by microscopy. The tubule suspension was washed repeatedly in PBS + Pen/Strep and treated with 0.025 % trypsin/ethylenediaminetetraacetic acid (Gibco) at 37°C for 5-10 min to digest the tubules. During incubation, 1 mL DNase1 (7 mg/mL; Sigma-Aldrich) was added. When dissociation of tubules was observed, the reaction was stopped with 2 mL of foetal bovine serum (Gibco). The solution was filtered through 70 µM cell strainers (BD Biosciences, North Ryde, NSW, Australia) and the resultant cell suspension pelleted by centrifugation. Red blood cells were removed by incubation at 37°C in 5 mL of 0.83% NH_4Cl for 7 min. The NH_4Cl was neutralised by the addition of PBS, the cells pelleted again and resuspended in DMEM/F12 + 10% foetal bovine serum for culture. Cell counts were performed using a haemocytometer and cell viability assessed using Trypan Blue.

2.2. *Cell culture and differential plating*

The isolated testis cells were cultured overnight at a concentration of 2 x 10^6 cells/mL in T175 flasks at 37°C and 5% CO_2. The following day, the supernatant containing non-adhered cells including the undifferentiated germ cells

was collected and the cells washed once with PBS. The non-enriched, adhered fraction was washed once and then gently resuspended in PBS using a cell scraper. Both cell fractions were collected by centrifugation at 800 x *g* for 5 min. Cell viability was again determined using Trypan Blue. Smears of both cell populations were air-dried on glass slides (Superfrost; Menzel Glaser, Braunschweig, Germany) and stored at -80°C until analysed. The remaining cells were used to prepare membrane extracts.

2.3. *Characterisation of cell smears*

Cell smears were equilibrated to room temperature prior to immunocytochemical staining. The cells were fixed in Modified Davidson's Fixative for 2 min, rinsed in Tris-buffered saline (TBS) (2 x 5 min) and blocked in 0.3% hydrogen peroxide for 15 min at RT. After further washes in TBS + 0.1% Tween 20 (TBS/Tween 20: 2 x 5 min) the cells were incubated with the primary antibody in TBST, 0.5% BSA for 30-60 min at RT. Germ cells were characterised using DDX4 (Abcam, Cambridge, UK; 1:2,000) and UCHL1 (Dako Australia, Campellfield, VIC, Australia; 1:700) and Sertoli cells were highlighted with Vimentin (Zymed, USA; 1:100). The cell smears were washed again in TBS/Tween 20 (2 x 5 min) and incubated for 30 min at RT with EnVision+ Dual Link System Peroxidase secondary antibody (Dako). Following further washes, the chromogen diaminobenzidine (Dako) was used to visualise the reaction. Cells were counterstained in haematoxylin, dehydrated through alcohol and xylene and mounted in SHUR/Mount (ProSciTech, Thuringowa, QLD, Australia). Primary antibodies were replaced with buffer for negative controls. Cell counts were performed manually from two fields of view and at least 500 cells.

2.4. *Immunohistochemistry*

Paraffin sections (5 µm) were dehydrated through xylene and graded alcohol dilutions to water. Antigen retrieval was performed by immersing the slides in heated 10 mM Tris-base, 1 mM EDTA, 0.05% Tween 20 and heated in an 800 W microwave on 30% power for 15 minutes and allowed to cool. The slides were rinsed in tap water. Binding of non-specific proteins was blocked by incubating sections in a detergent solution of tris-buffered saline 0.05% Tween 20 (TBST) for 30 minutes. For immunofluorescence, DDX6 (Santa Cruz Biotechnology, Santa Cruz, TX, USA) was used at 2 µg/mL, NAP1L4 (Abcam, UK) at 0.2 µg/mL and TKTL1 (Santa Cruz Biotechnology) at 0.5 µg/mL, GATA-4 (Santa Cruz Biotechnology) at 1 µg/mL, DDX4 (previously known as VASA) (Abcam) at 0.5 µg/mL, Vimentin (Zymed, USA) 1/100, and DBA-biotin (B1035, Vector Laboratories, Burlingame, CA, USA) at 10 µg/mL. Parallel sections were incubated with a mixture of rabbit and mouse immunoglobulin at 2 µg/mL as a control (Sigma-Aldrich). Cocktails of primary antibodies containing DBA-biotin, a mouse and a rabbit primary antibody, or control immunoglobulin, were incubated for 45 minutes at room temperature. Slides were washed in TBST, and then the secondary antibodies, streptavidin-alexafluor 350 (10 µg/mL), (Invitrogen), goat anti-mouse Alex 488 (A11001) 1:250, chicken anti-rabbit Alex 594 (Invitrogen) applied, incubated at room temperature for 30 minutes, washed in TBS, cover slipped in Prolong Gold (Invitrogen) and kept in the dark until photographed using a AxioImager and Axiovision software (Carl Zeiss, Oberkochen, Germany). For the brightfield images, antibody against DDX6 was used at 2 µg/mL, NAP1L4 at 0.2 µg/mL, TKTL1 at 0.5 µg/mL, and staining was visualised using the Expose Mouse and Rabbit Specific HRP/DAB detection kit (Abcam), counterstained with haematoxylin, dehydrated through alcohol and xylene and then mounted with DPX (Merck, Darmstadt, Germany).

2.5. *Membrane Fraction Preparation*

Differentially plated cells were harvested as described above from a single testis of two different animals. Protein samples were kept separate as biological replicates throughout the analysis. All membrane preparation steps were performed at 4°C where possible. The cell isolates were suspended in PBS and mechanically lysed by repetitive passes through a 21 G needle. The solution was further disrupted by probe sonication on ice for 5 min. Cell debris was removed by centrifugation at 1,000 x *g* for 10 min. The supernatant was clarified at 15,000 x *g* for 30 min removing larger membrane particles. Cell membranes were collected by centrifugation at 240,000 x *g* for 1 h. Membrane pellets were washed twice with PBS (20 min; 240,000 x *g*) and stored at -80°C if required.

2.6. *Protein Concentration Determination*

The isolated membrane pellet was resuspended in PBS and sonicated briefly at 4°C in a sonic water bath. Protein concentration was determined from an aliquot of the cell suspension using a commercial bicinchoninic acid assay kit (Pierce, Thermo Fisher Scientific, Scoresby, VIC, Australia).

2.7. *Filter-Aided Sample Preparation (FASP)*

Reduction, alkylation and digestion was performed using the filter-aided sample preparation protocol developed by Wisniewski *et al* [27]. In brief, up to 30 µL of the membrane pellet suspension containing up to 300 µg of protein was mixed with 200 µL of 8 M urea in 0.1 M Tris/HCl (UA solution; pH 8.5) in a Microcon YM10 filter unit (Millipore, Merck Kilsyth, VIC, Australia) and centrifuged at 14,000 x *g* for 15 min. The unit was washed with another 200 µL of UA (14,000 x *g* for 15 min). The flow-through from the collection tube was discarded and 100 µL of iodoacetamide solution (0.05 M in UA) was added. The filter unit was mixed at 600 rpm at 20°C for 1 min then incubated for a further 20 min at 20°C. Excess iodoacetamide was removed by centrifu-

gation at 14,000 x *g* for 10 min and the filter washed three times with 100 µL of UA (14,000 x *g*; 15 min). Ammonium bicarbonate (0.05 M in H_2O; 100 µL) was added to the column and washed through at 14,000 x *g* for 10 min. This step was repeated twice. Digests were performed on the filters with 40 µL of trypsin (Promega, Madison, WI, USA) in ammonium bicarbonate at an enzyme to protein ration of 1:60. The filter units were mixed at 600 rpm for 1 min and incubated overnight at 37°C in a humid chamber. The filter units were transferred to fresh collection tubes, centrifuged at 14,000 x *g* for 10 min to collect the tryptic peptides and washed with a further 40 µL ammonium bicarbonate. The eluate was dried down in a vacuum concentrator and resuspended as required for proteomic analysis.

2.8. *Chromatography*

Tryptic peptides were chromatographically resolved using a Shimadzu Prominence LC20 HPLC system with a C18 Vydac column (75 µm x 15 cm, 300 Å, 5 µm). Protein digests were reconstituted in 0.1% formic acid and 1 µg was injected on-column. A linear gradient at a flowrate of 800 nL/min from 1-40% solvent B over 80 min was utilised where solvent A was 0.1% formic acid and solvent B was 0.1% formic acid in 90% acetonitrile.

2.9. *Mass Spectrometry*

The eluate from the HPLC system was directly coupled to the nanoelectrospray ionisation source of the TripleTOF™ 5600 system (AB/Sciex, Foster City, CA, USA). Data were acquired in information dependent acquisition (IDA) mode. The IDA method consisted of a high resolution TOF-MS survey scan followed by 20 MS/MS in a second with a maximum accumulation time of 50 ms. First stage MS analysis was performed in positive ion mode over the mass range *m/z* 300-2000 with a 0.5 s accumulation time. The ionspray voltage was set to 2600 V, the curtain gas was set to 25, the nebuliser gas to 20 and the heated interface was set to 150°C. Tandem mass spectra were acquired over the mass range *m/z* 100-2000 using rolling collision energy (CE) for optimum peptide fragmentation.

2.10. *Database Searching and False Discovery Rate Analysis*

All data were processed using ProteinPilot v4.0 (AB/Sciex) with integrated false discovery rate analysis [28]. The spectral sets (either individually or combined) were searched against all bovine proteins present in Uniprot database (version 20110718; 70,452 proteins). Search parameters were defined as cysteine alkylation with iodoacetamide, trypsin as the digestion enzyme and no restrictions were placed on taxonomy. Modifications were set to the "generic workup" and "biological" modification sets provided with this software package, which consisted of all modifications listed in Unimod, for example, acetylation, methylation and phosphorylation. The generic workup modifications set contains 59 potential modifications that may occur as a result of sample handling, for example, oxidation, dehydration and deamidation. The identification of proteins was recorded in the Results section if the protein was identified at a 1% global false discovery rate (FDR).

2.11. *Quantitative Reverse Transcribed Polymerase Chain Reaction (qRT-PCR)*

Bovine specific primers were designed with Primer3 (http://frodo.wi.mit.edu) using bovine sequences in the NCBI GenBank databases. The primers used in this study are listed in Supplemental Table 1. Appropriate reference genes, *ATP5G2* (ATP synthase) and *RPS26*, were determined using the sheep GeNorm kit (PrimerDesign, Southampton, UK). Total RNA was extracted from frozen cell samples using an RNeasy Midi kit (Qiagen, Hilden, Germany). cDNA was synthesised using a Superscript III first strand synthesis kit (Invitrogen). qRT-PCR reactions were carried out in triplicate in an iQ5 real time thermal cycler (Bio-Rad, Hercules, CA, USA). Each reaction contained 1× IQ SYBR Green Supermix (Bio-Rad), 0.5 µM of each forward and reverse primer and cDNA transcribed from 10 ng RNA. Interplate controls and negative controls were included in each assay.

2.12. *Statistical analysis*

Cell counts from immunocytochemical characterisation were expressed as percentages of positive cells and analysed using analysis of variance (ANOVA). Differences with p values < 0.05 were considered significant.

C_t values from qRT-PCR were converted into expression data using the Excel add-in Genex (Bio-Rad). Statistical analyses of gene expression data were performed using GraphPad Prism software (GraphPad Software, San Diego, CA, USA). A ranked Mann Whitney test was applied to determine statistically significant differences between treatment groups. Differences of $p < 0.05$ were considered to be significant. Results are presented as dot plots with means ±SEM shown.

2.13. *Bioinformatics*

Identified proteins were classified according to known gene ontology using the GO analysis toolkit for the agricultural community (AgriGO) [29]. The GO annotation terms were further reduced using REViGO [30] to remove redundant terms and to visualise the data. Enrichment was determined against the bovine genome. The molecular function and biological processes of the identified proteins were obtained from Protein ANalysis THrough Evolutionary Relationships (PANTHER) [31], the Database for Annotation, Visualisation and Integrated Discovery (DAVID) [32, 33] or the European Bioinformatics Institute's web based tool for gene ontology searching (QuickGO) [34]. Where annota-

tions for bovine proteins were unavailable, molecular function and process were determined from sequence homology and classification in other species. Predictions of transmembrane domains, signal peptides and non-classical secretion were derived from the Centre for Biological Sequence Analysis (CBS) prediction servers TMHMM [35], SignalP 4.0 [36] and SecretomeP 2.0 [37] respectively.

3. Results

Differential plating

Testis tissue from two pre-pubertal bull calves was subjected to enzymatic digestion and subsequent cell isolation. At this stage of development, the germ cell population is more likely to contain only gonocytes and undifferentiated spermatogonia [17]. After overnight culture, adhered and non-adhered cell populations were harvested for cell characterisation and membrane fractionation. Cell viability at harvest was greater than 90%. Cell smears were treated with antibodies to the known germ cell markers UCHL1 and DDX4, and the Sertoli cell marker Vimentin. These markers were present in both cell populations, however, cell counts demonstrate significant germ cell enrichment in the non-adhered fraction (UCHL1 = 9%, representing a 5x increase, DDX4 = 6% representing a 7x increase, $p < 0.01$) and slightly higher Sertoli cell numbers in the adhered population (Vimentin = 24% representing a 1.3x increase) (Figure 1).

Global Protein Profiling

Mass spectrometry analysis, followed by database searches identified a total of 1,387 proteins at a 1% false discovery rate (FDR) when all data were combined and searched together (Supplemental Table 2). In the non-adhered populations 1,150 proteins were identified, while in the adhered population 988 proteins were identified (Supplemental Tables 3,4). A total of 767 proteins (55%) were found to be expressed in both sub-populations.

The goal of this study was to help identify proteins that are expressed by undifferentiated spermatogonia, but not present on other cells in bovine testis tissue. To this end, we performed a comparative analysis to identify proteins that were either: (1) present in the non-adhered population, but not in the adhered population; or (2) present in greater abundance in the non-adhered fraction.

The preliminary abundance levels were assessed based on the total protein score, the number of peptides mapping to a given protein and the percentage of sequence coverage (Supplemental Table 2). In practice, proteins that were identified by at least two unique peptides (resulting in a score $\geq$ 4.0) in both non-adhered populations (NA1 and NA2), but not detected in either of the adhered cell populations (A1 and A2) were considered unique. Proteins with at least two unique peptides in both NA1 and NA2 and that were detected in A1 and/or A2 were considered common. Common proteins with a protein score ratio $\geq$ 2 were considered to be present in greater abundance. Proteins passing either of these criteria are listed in Supplemental Table 5 (proteins more abundant in non-adhered) and Supplemental Table 6 (proteins more abundant in adhered) and were subjected to further scrutiny.

Based on the above criteria, a total of 89 proteins were differentially expressed in the non-adhered fraction, of which 21 were only detected in the non-adhered cell population and the remaining 68 were present with protein scores that differed by two-fold. A total of 105 proteins appeared to be differentially expressed in the adhered cell population, with 20 proteins only detected in the adhered population and 85 proteins showing protein scores that differed by two-fold. As the protein score and number of peptides identified are dependent on the protein size and other factors (ionisation efficiency, peptide size), we also examined the relative abundance by spectral counting, that is the total number of MS/MS spectra that identify a given protein was determined for each analysis. Only peptide spectrum matches (PSM) identified with $\geq$ 95% confidence were considered.

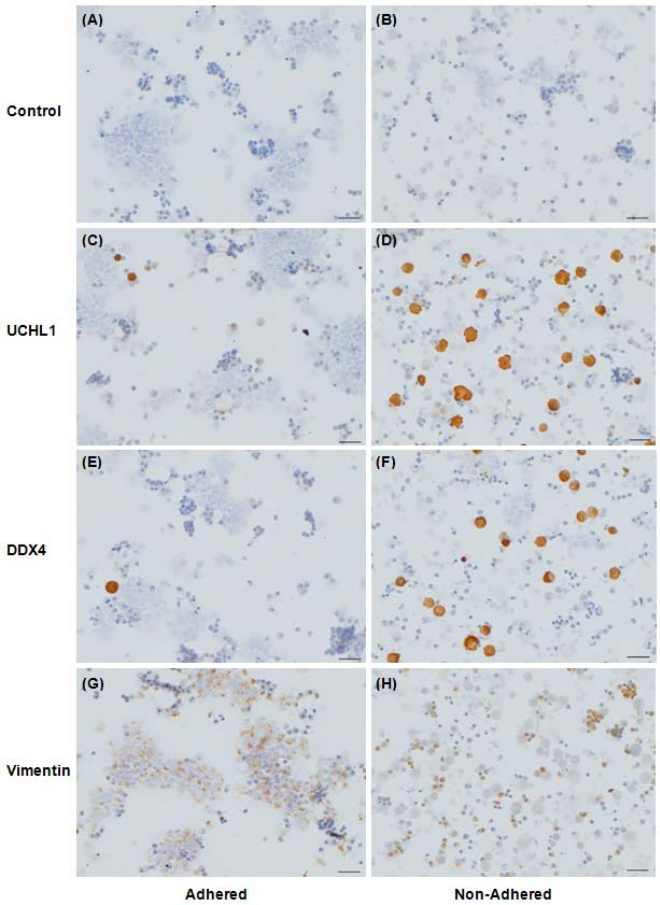


Figure 1. Immunocytochemistry of adhered (A,C,E,G) and non-adhered (B,D,F,H) bovine testis cell isolates. A,B: no primary antibody (control); C,D: UCHL1 antibody (germ cells); E,F: DDX4 antibody (germ cells); G,H: Vimentin antibody (Sertoli cells). The scale bar represents 50 μm.

Only proteins with spectral counts in two replicates greater than 4 were considered. Proteins that showed variation in the spectral count between the biological replicates were excluded (Supplemental Tables 7 and 8).

With the added stringency of spectral counting, a total of 64 proteins passed the criteria for differential expression in the non-adhered fraction (Table 1), of which 16 were only detected in the non-adhered cell population and the remaining 48 were present with protein scores that differed by two-fold. Likewise, a total of 78 proteins were differentially expressed in the adhered cell population, with 19 proteins only detected in the adhered population and 59 proteins showing protein scores that differed by two-fold.

Of the proteins identified in the literature as mammalian germ cell markers, DDX4 and UCHL1 were detected in the non-adhered population, however, these two proteins did not pass the strict criteria applied to be considered differentially expressed. DDX4 was detected only in the non-adhered population, but was identified by only a single peptide in both replicates. Likewise, UCHL1 was detected in both non-adhered populations, but by a single peptide in one replicate. For Sertoli cell markers, vimentin was detected in both, adhered and non-adhered fractions with only sightly higher representation in the adhered cell population (ratio 0.9) (Supplemental Table 8).

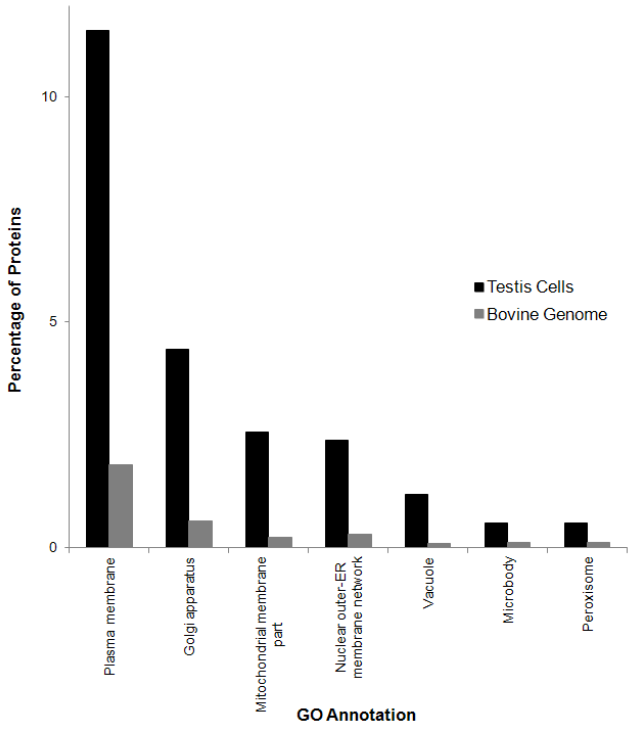


Figure 2. Subcellular localization annotations of membrane proteins in bovine testicular cell isolates. Localisation annotations determined from Gene Ontology using AgriGO (http://bioinfo.cau.edu.cn/agriGO/)

Functional annotation

AgriGO analysis of the cellular component of the total list of 1,387 proteins demonstrated a clear enrichment for membrane associated proteins. Of the total number, 536 (38.65%) were membrane proteins compared to 28.45% present in the bovine genome. Cell organelle localisation shows that the greatest proportion of these membrane proteins are associated with the plasma membrane (11.45%) (Figure 2), however, all organelle membrane proteins are enriched in our samples when compared to genomic levels.

The web-based PANTHER classification system was used to group the differentially expressed adhered and non-adhered proteins according to their molecular function and the biological processes in which they are involved. Figure 3 shows the distribution of the molecular functions in both groups. More proteins in the adhered group are classified as having a binding function (44.8%) when compared to the

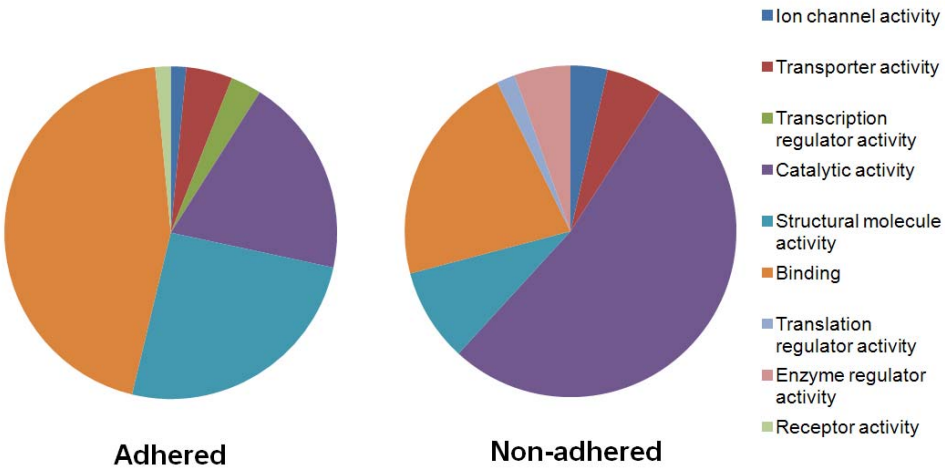


Figure 3. Distribution of molecular function annotations of membrane proteins in adhered and non-adhered subpopulations of bovine testicular cell isolates. Annotations determined from Gene Ontology using PANTHER (http://www.pantherdb.org/)

non-adhered (21.8%). Similarly, more proteins with a structural function (e.g. keratin) were identified in the adhered cells (25.4%) than the non-adhered population (9.1%). The non-adhered fraction contains a higher proportion of proteins associated with catalytic activity (52.7%) than that of the adhered cells (19.4%).

The broader GO annotation terms displayed in Figure 3 encompass a number of "child" terms and these are listed for the non-adhered cells only in Table 1 along with their biological processes. The prediction of transmembrane helices (TM) in the identified proteins are also listed in Table 1 along with the presence of a signal peptide cleavage site (SP) or the involvement of the protein in non-classical secretion (NCS), that is, secretion not triggered by cleavage of a signal peptide. Of the 64 proteins found to be differentially expressed in non-adhered cells, only two, malectin and hypoxia up-regulated protein 1, contain a transmembrane domain, however, both proteins are associated with the ER membrane. The same two proteins also have a signal peptide cleavage site along with 4 others, and 30 candidates demonstrate non-classical secretion.

Gene expression

To identify candidates for gene expression markers for testicular germ cells, a number of candidate molecules were selected from the proteins found exclusively or preferentially in the non-adhered cells (Table 1 and Supplemental Tables 2 and 5). The expression patterns of genes corresponding to these proteins were examined in human testis tissue images at www.proteinatlas.org. Genes which showed an expression pattern similar to early spermatogonia (situated in low numbers on the basement membrane in seminiferous tubules) were included as candidates for investigation of RNA expression by qRT-PCR. Ten candidate genes (*ASB9, ATIC, DDX6, FSCN1, IQGAP1, NAP1L4, PFN1, PHGDH, TKTL1* and *TLN1*) and four established testis cell marker genes (*DDX4, GATA4, UCHL1* and *VIM*) were assessed for transcription levels relative to the reference genes *ATP5G2* (ATP synthase) and *RPS26* in samples from four non-adhered and four adhered cell fractions (Supplemental Table 1). Of the ten investigated candidates, three showed higher ($p < 0.05$) expression levels in non-adhered cells (*DDX6, NAP1L4, TKTL1*) (Figure 4). *ASB9* showed a trend towards higher gene expression in non-adhered cells, but the difference was not statistically significant.

Immunohistochemistry

For three candidate germ cell markers which showed significantly higher levels of RNA expression in non-adhered cells (*DDX6, NAP1L4 and TKTL1)*, protein expression was investigated by antibody staining of peri-pubertal testis tissue sections (Figure 5). None of the three candidate markers showed any co-staining with vimentin, indicating no expression by Sertoli cells. DDX6 did show co-staining with DBA-biotin, consistent with expression by early spermatogonia and also stained cells which did not show binding of DBA-biotin or vimentin-antibody, consistent with expression by more advanced germ cells (Figure 5 B, E-H). Although NAP1L4 showed high background when visualized for brightfield micrography, in fluorescence micrography the same antibody exhibited a tight overlap with DBA-biotin staining, consistent with expression by early spermatogonia (Figure 5 C, I-L). TKTL1 showed staining by cells towards the centre of seminiferous tubules, consistent with expression in advanced spermatogonia (Figure 5 D, M-P).

4. Discussion

In this study, membrane fractionation combined with proteomics analyses was employed to examine enriched undifferentiated spermatogonial cell populations with the aim of uncovering novel surface markers.

The application of proteomics to mammalian spermatogenesis has been recently reviewed [23] Proteome reference maps have been generated for whole testes from pigs (yielding 337 identifications) [38], mice (504 identifications) [39] and humans (1430 identifications) [40]. 2D-gels have been the proteomic tool of choice for studies of germline stem cells in chickens [21], in germline and embryonic SCs in mice [18] and of rat spermatogonia [41] revealing 56, 166 and 102 proteins respectively. More recently, the spermatogonial stem cell niche of dogfish has been examined yielding 16 protein identifications [22]. In this study, we report the comprehensive proteomic profiling of bovine testis cell populations which enabled the identification of 1,387 proteins, of which 1,150 were identified in the population enriched for undifferentiated spermatogonia. Using optimised membrane fractionation and enzymatic digestion protocols, we show an enrichment for membrane proteins (39% compared to 28% present in the total bovine proteome) and we demonstrated a 6-fold enrichment for proteins associated with the plasma membrane (Fig. 3).

In contrast to global proteomic profiling studies employing whole cell lysates, this study focussed specifically on the membrane-fraction of testis cell populations. Methods for proteomic studies of cell membrane preparations include "cell shaving" and FASP technology [27]. Since the testis cell isolation procedure included the use of trypsin, an enzyme-based method for retrieval of cell surface-exposed proteins was not considered a promising approach. Application of the FASP technology allows an increased rate of membrane protein identification as these rather insoluble proteins may be solubilised and treated in the presence of detergents and chaotropes not typically used in gel-free proteomic studies because of their interference in down-stream analyses. The resulting protein preparation was enriched in membrane proteins and can therefore be expected to yield an increased identification rate for low abundance proteins that may be useful as markers of undifferentiated spermatogonia.

The testicular cell isolates were differentially plated to yield

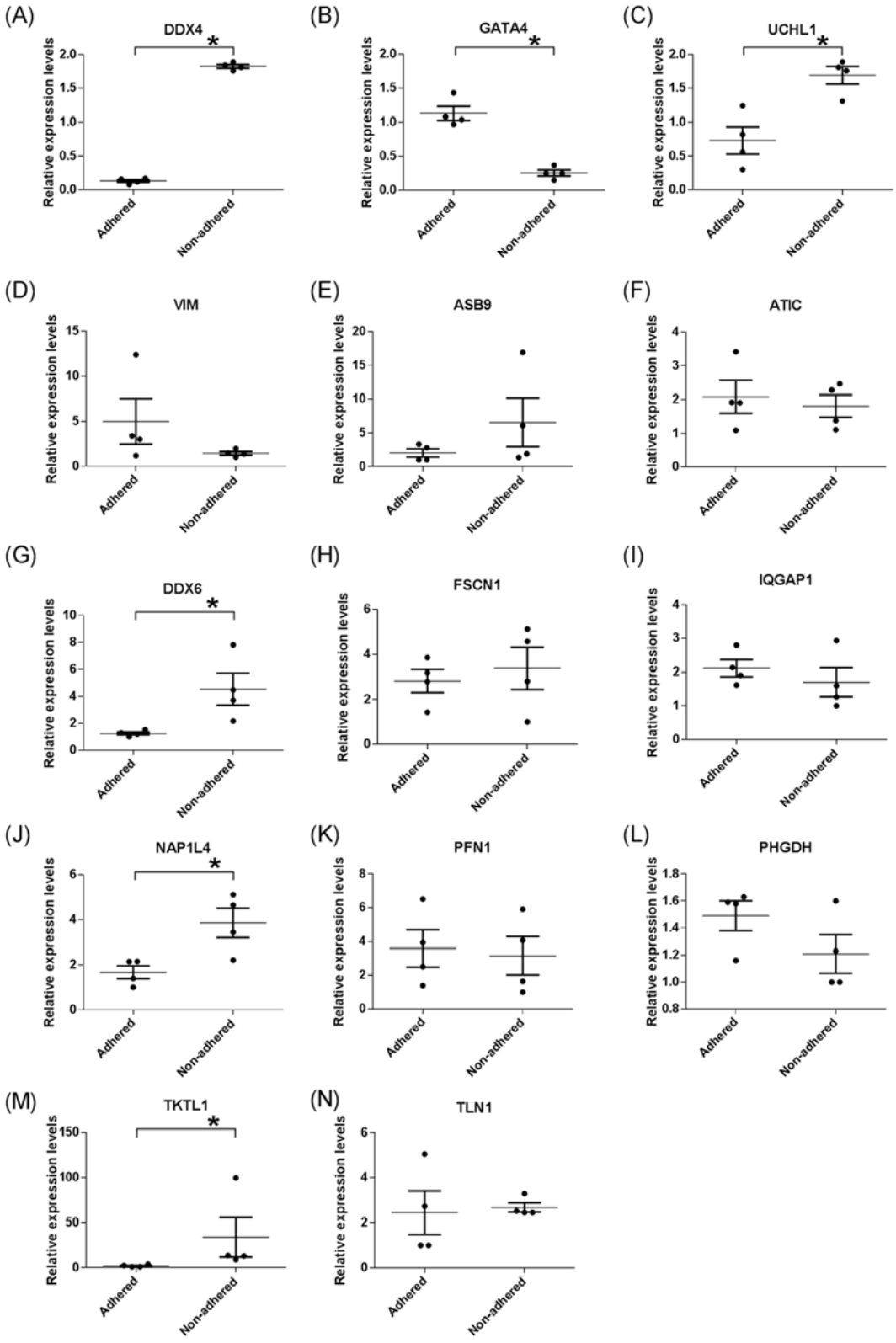


Figure 4. Comparison of candidate gene expression in adhered and non-adhered testis cell fractions. (A-D) Established testis cell markers. (E-N) Candidate markers. Gene expression levels are graphed as a ratio relative to reference gene expression levels (y-axes). Each data point represents the average of three technical replicates. Horizontal bars represent the mean of four biological replicates, with error bars representing the standard error of the mean. * denotes a statistically significant difference between cell fractions in a Mann-Whitney U test ($p < 0.05$).

two cell populations: adhered and non-adhered. The non-adhered population, enriched for undifferentiated spermatogonia, was the target of our investigation as it is well established that the majority of undifferentiated spermatogonia remain in suspension during the initial culture period, whereas their support cells, the Sertoli cells are expected to be present in both populations with a slight enrichment in the adhered population [12]. This was confirmed by immunocytochemical examination of cell smears that showed significant enrichment for the known germ cell markers UCHL1 and DDX4 in the non-adhered population. Label-free relative quantification of the identified proteins following LC-MS/MS analysis was undertaken and the proteins were classed as differentially expressed if they passed multi-

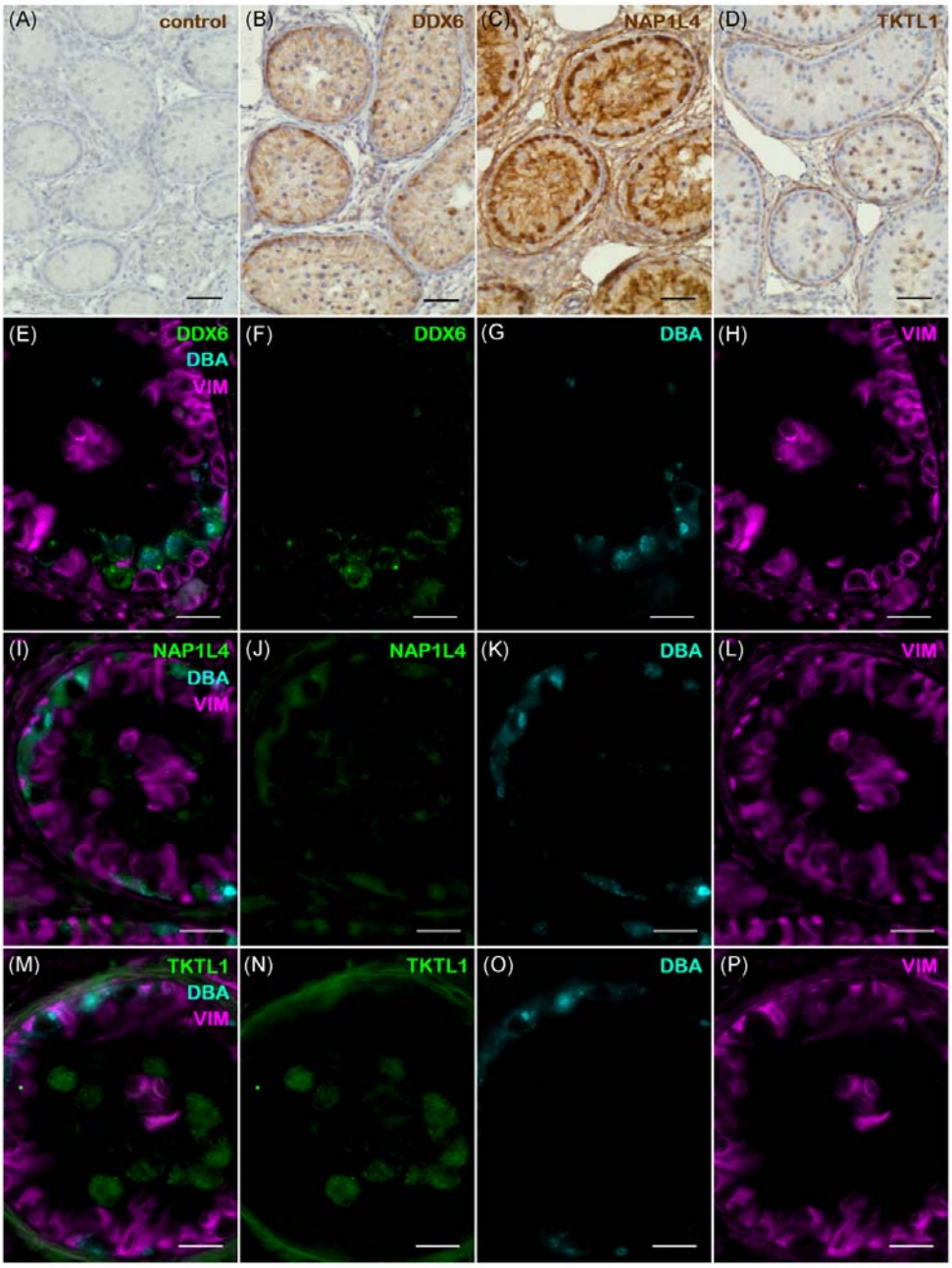


Figure 5. Expression of candidate markers in peripubertal bovine testis tissue. (A-D) Overview brightfield micrographs, scale bars are 50 µm. (E-P) Detailed fluorescence micrographs with (E, I, M) merged images, scale bars are 20 µm. (A) Control, section parallel stained with secondary antibody cocktail. (B, F) Tissue stained with DDX6-antibody. (C, J) Tissue stained with NAP1L4-antibody. (D, N) Tissue stained with TKTL1-antibody. (E-H) Tissue co-stained with DDX6-antibody (E, F), DBA-biotin (E, G) and Vimentin-antibody (E, H); (I-L) tissue co -stained with NAP1L4-antibody (I, J), DBA-biotin (I, K) and Vimentin-antibody (I, L); (M-P) tissue co-stained with TKTL1-antibody (M, N), DBA-biotin (M, O) and Vimentin-antibody (M, P).

ple criteria based on the number of peptides identified, the total protein score and the spectral count. One of the constraints in our experimental system was that the differential plating did not result in a homogenous population, it only served to enrich the fraction for undifferentiated spermatogonia. However, it is well understood that other cell types would be present, for example red blood cells. In fact, both hemoglobin subunit alpha and beta were present in the differentially expressed lists for the non-adhered population. An additional confounding factor lies in the fact that the protein profile would be expected to change as a result of the attachment of the cells to the plates. This was observed, with structural proteins such as the keratins, collagens and spectrin present in higher abundance in the adhered fraction. With these constraints in mind, the protein populations were classified according to their molecular function. As expected, the adhered population was dominated by proteins associated with binding and structural molecule activity. In contrast, the non-adhered population was dominated by proteins associated with catalytic activity and to a lesser extent enzyme and translation regulator activities.

Of the 64 proteins that appear to be more abundant in the non-adhered cell population, 15 show localisation or close association with the plasma membrane (NSF, TGM2, ATIC, PSMC6, MDH2, ACLY, PHGDH, ENO1, IQGAP1, EHD2, ATP6V1A, ATP6V1B1, TLN1, FSCN1 and CAP1) according to the Human Protein Atlas (www.proteinatlas.org) or literature reference. For 6 of those proteins (TGM2, ATIC, MDH2, EDH2, TLN1, and CAP1), no previous reports exist that would indicate a specific role in mammalian testis or germ cells.

Vesicle-fusing ATPase (or *N*-ethylmaleimide-sensitive fusion protein, NSF) is primarily localised to microvesicles and is involved with vesicle-to-plasma membrane fusion during exocytosis. It plays a critical role in the sperm acrosome reaction, a calcium-dependent exocytosis event, which is required for fertilization [42]. In this study, NSF was uniquely detected in the non-adhered cell fraction (5 and 7 peptides in the two biological replicates).

Several of the proteins identified that appear to be more abundant in the non-adhered population share similar functions including cell surface remodelling, cell adhesion and migration. TGM2 was detected with higher protein score and spectral count ratios in the non-adhered fraction. Likewise, IQGAP1, TLN1, EHD2 and FSCN1 were also detected in the non-adhered fraction with higher protein score ratios along with higher spectral count ratios. The protein glutamine gamma-glutamyltransferase 2 (TGM2) has been demonstrated to play a role in conferring stem cell-like properties to mammary epithelial cells [43]. TGM2 additionally plays a role in remodelling extracellular matrices and promotes cell adhesion with the highest concentrations observed at cell-cell and cell-substratum contact points [44]. Although initally thought to be an intracellular enzyme, TGM2 has been demonstrated to bind to the pericellular fibronectin coat [45]. Ras GTPase-activating-like protein IQGAP1 is localised to the plasma membrane and cell junctions and also plays a role in cell surface remodelling (through organisation of the actin cytoskeleton) and cell adhesion. IQGAP1 through its interaction with CDC42 co-localises to the periphery of both Sertoli and germ cells regulating their adhesion [46]. Talin-1 (TLN1) is a plasma membrane protein that shows higher concentrations in areas of cell-cell contact. TLN1 binds to integrin-β and plays a critical role in cell adhesion and morphogenesis [47]. A recent report by Sharma et al [48], demonstrated that talin-1 enhances the survival , migration and differentiation of cardiac stem cells. EH-domain containing 2 protein (EHD2) is a cell membrane protein also involved with membrane reorganisation and membrane trafficking between the plasma membrane and endosomes [49]. A recent genome-wide profiling study found that EHD2 was up-regulated in primordial germ cells and embryonic germ cells compared to embryonic stem cells [50]. Fascin (FSCN1) plays a role in cell motility and migration through its involvement in actin bundling and formation of cell protrusions [51]. Fascin-3 is a testis-specific isoform [52] that has been shown to increase during sperm capacitation [53].

Two of the proteins identified as more abundant in the non-adhered cell population were enzymes initially assumed to be cytosolic in nature. ATP citrate lyase (ACLY) was detected with a higher protein score ratio (5.28) and spectral count ratio (4.75) in the non-adhered fraction, whereas α-enolase was uniquely detected in the non-adhered cell fraction (6 peptides detected in each of the two biological replicates). ACLY is the primary enzyme responsible for the synthesis of acetyl-CoA in many tissues and is primarily localised to the cytoplasm but also the plasma membrane. Lipid synthesis is required not only for membrane production, but also for lipid-based post-translational modification of proteins. Rat testicular germ cells were observed to possess ACLY and post-meiotic increases in the activity of anabolic enzymes were noted [54]. Alpha-enolase (ENO1) is a multifunctional cytoplasmic enzyme involved in glycolysis and cell growth control that is also expressed on the cell surface where it binds to plasminogen [55]. Alpha-enolase is known to be present in the tail of mature sperm and is associated with post-translational modification during sperm maturation. It was recently identified as a potential biomarker of bull fertility [56].

Lastly, proteins representing the CCT complex (TCP1, CCT2, CCT4, CCT5, CCT7, CCT8) and PMSC6 of the 26S proteasome, all of which were expected to be cytosolic in location were identified as more abundant in the non-adhered cell populations. Interestingly, these proteins were also identified in a recent study in our laboratory examining the plasma membrane fraction of bovine sperm, implying that they are indeed associated with the plasma membrane [57]. The CCT complex is present on the surface of capacitated spermatozoa playing a role in binding to the zona pellucida during fertilization [58], and the membrane proteasome of mammalian sperm is required for the acrosome

reaction and fertilization [59].

While the proteomic analysis was limited to two biological replicates, we were able to compare and confirm some of the results with gene expression data utilising four biological replicates, and with immunohistochemistry data of independent testis tissue sections. We identified a group of ten candidate genes from the non-adhered cell fraction which showed a cell localisation pattern akin to undifferentiated spermatogonia in human testis tissue images at the Human Protein Atlas (www.proteinatlas.org). Higher gene expression levels observed in the non-adhered cell populations for *DDX6*, *NAP1L4* and *TKTL1* correlated with the results of the proteomic analysis where DDX6 and NAP1L4 were identified only in the non-adhered cell fraction, and TKTL1 showed higher abundance in the non-adhered fraction. These three genes are known to be associated with spermatogonial cell function: *DDX6* codes for an ATP-dependent RNA helicase that has been found highly expressed in mammalian spermatogonia [60], *NAP1L4* encodes a nucleosome assembly chaperone protein that has been shown to be required for the incorporation of the testis-specific H3t histone variant into nucleosomes [61] and the *TKTL1* gene product was identified as a germ cell biomarker in a proteomics study of human seminal plasma [62].

Protein expression analysis of the differentially expressed genes DDX6, NAP1L4 and TKTL1 confirmed expression in spermatogonia. DDX6 and NAP1L4 were found to be expressed in early spermatogonia identified by DBA-lectin binding, indicating expression in an undifferentiated subset of spermatogonia. In comparison, probable ATP-dependent RNA helicase (DDX4) is expressed throughout the germ lineage [7, 8]. UCHL1 is an established marker for bovine undifferentiated spermatogonia [9], which includes but is not limited to DBA-lectin binding spermatogonia [12]. NAP1L4 in particular showed an expression profile more similar to DBA-lecting binding, and hence might be closer to a putative SSC marker than other established spermatogonial markers in the bovine. Further analysis will confirm expression patterns of DDX6 and NAP1L4 in bovine spermatogonia, and the utility of the new candidates when compared to established markers. From our data, further candidates for characterisation of the non-adhered cell population, which show enrichment for undifferentiated spermatogonia, are NSF, TGM2, ATIC, PSMC6, MDH2, ACLY, PHGDH, ENO1, IQGAP1, EHD2, ATP6V1A, ATP6V1B1, TLN1, FSCN1 and CAP1 as well as proteins of the CCT complex.

5. Concluding Remarks

This study represents the most comprehensive proteomic profile of bovine testicular cells with 1,387 proteins identified at a 1% false discovery rate. Furthermore, 39% of the proteins identified are associated with cellular membranes, primarily the plasma membrane. We have identified a number of potential surface markers that warrant further investigation. Fourteen proteins were assessed by qRT-PCR including the four known markers (DDX4, GATA4, UCHL1 and VIM). Of the ten candidate markers, three (DDX6, NAP1L4 and TKTL1) are also differentially expressed at the RNA level, and for those proteins expression in spermatogonia was confirmed by immunohistochemistry. Proteome maps such as this one provide the foundation for future studies of spermatogenesis and germ cell biology.

6. Supplementary material

Supplementary data and information is available at: http://www.jiomics.com/index.php/jio/rt/suppFiles/136/0

Supplemental Table 1 – Primers used for qRT-PCR.
Supplemental Table 2 – List of proteins identified in bovine testis isolates using nanoflow HPLC-MS/MS.
Supplemental Table 3 – List of proteins identified in non-adhered cell population of bovine testis isolates.
Supplemental Table 4 – List of proteins identified in adhered cell population of bovine testis isolates.
Supplemental Table 5 – List of proteins that appeared to be more abundant (based on protein score) in non-adhered cell population of bovine testis isolates. Proteins that were excluded from the final list (based on spectral count) are indicated in red italics.
Supplemental Table 6 – List of proteins that appeared to be more abundant (based on protein score) in adhered cell population of bovine testis isolates. Proteins that were excluded from the final list (based on spectral count) are indicated in red italics.
Supplemental Table 7 – List of proteins that appeared to be more abundant (based on protein score and spectral count) in non-adhered cell population of bovine testis isolates.
Supplemental Table 8 – List of proteins that appeared to be more abundant (based on protein score and spectral count) in adhered cell population of bovine testis isolates.

Acknowledgements

The authors wish to thank Miss Keren Byrne for assistance with bioinformatic analysis. Dr Shelly Hope and Dr Gene Wijffels provided helpful feedback and advice on the manuscript. Mass spectrometry equipment was provided by the Molecular & Cellular Proteomics Mass Spectrometry Facility at The University of Queensland. We wish to thank the staff at Highchester Abattoirs, Gleneagle, Queensland for their cooperation in obtaining bovine testes at slaughter.

References

[1] C. Dores, W. Alpaugh, I. Dobrinski, Cell Tissue Res, 349 (2012) 691-702. doi: 10.1007/s00441-012-1457-x

[2] R.L. Brinster, J.W. Zimmermann, Proc Natl Acad Sci U S A, 91 (1994) 11298-11302.

[3] P.M. Aponte, M.P. van Bragt, D.G. de Rooij, A.M. van Pelt,

APMIS, 113 (2005) 727-742. doi: 10.1111/j.1600-0463.2005.apm_302.x

[4] J.M. Oatley, R.L. Brinster, Methods Enzymol, 419 (2006) 259-282. doi: 10.1016/S0076-6879(06)19011-4

[5] K. von Kopylow, C. Kirchhoff, D. Jezek, W. Schulze, C. Feig, M. Primig, V. Steinkraus, A.N. Spiess, Hum Reprod, 25 (2010) 1104-1112. doi: 10.1093/humrep/deq053

[6] F. Izadyar, J. Wong, C. Maki, J. Pacchiarotti, T. Ramos, K. Howerton, C. Yuen, S. Greilach, H.H. Zhao, M. Chow, Y.C. Chow, J. Rao, J. Barritt, N. Bar-Chama, A. Copperman, Hum Reprod, 26 (2011) 1296-1306. doi: 10.1093/humrep/der026

[7] T. Noce, S. Okamoto-Ito, N. Tsunekawa, Cell Struct Funct, 26 (2001) 131-136.

[8] E. Raz, Genome Biol, 1 (2000) REVIEWS1017. doi: 10.1186/gb-2000-1-3-reviews1017

[9] R. Rathi, A. Honaramooz, W. Zeng, S. Schlatt, I. Dobrinski, Reproduction, 130 (2005) 923-929. doi: 10.1530/rep.1.00912

[10] F. Izadyar, G.T. Spierenberg, L.B. Creemers, K. den Ouden, D.G. de Rooij, Reproduction, 124 (2002) 85-94.

[11] C. Ertl, K.H. Wrobel, Histochemistry, 97 (1992) 161-171.

[12] M. Herrid, R.J. Davey, J.R. Hill, Cell Tissue Res, 330 (2007) 321 -329. doi: 10.1007/s00441-007-0445-z

[13] T. Imai, Y. Kawai, Y. Tadokoro, M. Yamamoto, Y. Nishimune, K. Yomogida, Mol Cell Endocrinol, 214 (2004) 107-115. doi: 10.1016/j.mce.2003.10.065

[14] K. Steger, M. Schimmel, K.H. Wrobel, Arch Histol Cytol, 57 (1994) 17-28.

[15] K.H. Wrobel, D. Bickel, R. Kujat, Acta Anat (Basel), 153 (1995) 263-272.

[16] E. Guillaume, B. Evrard, E. Com, E. Moertz, B. Jegou, C. Pineau, Mol Reprod Dev, 60 (2001) 439-445. doi: 10.1002/mrd.1108

[17] X. Guo, W. Ying, J. Wan, Z. Hu, X. Qian, H. Zhang, F. He, Electrophoresis, 22 (2001) 3067-3075. doi: 10.1002/1522-2683(200108)22:14<3067::AID-ELPS3067>3.0.CO;2-V

[18] H. Dihazi, G.H. Dihazi, J. Nolte, S. Meyer, O. Jahn, G.A. Muller, W. Engel, J Proteome Res, 8 (2009) 5497-5510. doi: 10.1021/pr900565b

[19] H. Dihazi, G.H. Dihazi, O. Jahn, S. Meyer, J. Nolte, A.R. Asif, G.A. Mueller, W. Engel, J Proteome Res, 10 (2011) 1962-1973. doi: 10.1021/pr1012015

[20] L. Harkness, H. Christiansen, J. Nehlin, T. Barington, J.S. Andersen, M. Kassem, Stem Cell Res, 1 (2008) 219-227. doi: 10.1016/j.scr.2008.06.001

[21] B.K. Han, J.G. Jung, J. Nam, J.K. Moon, J.N. Kim, S.I. Lee, J.K. Kim, H. Kim, J.Y. Han, J Androl, 30 (2009) 690-702. doi: 10.2164/jandrol.108.007401

[22] G. Loppion, R. Lavigne, C. Pineau, P. Auvray, P. Sourdaine, Comp Biochem Physiol Part D Genomics Proteomics, 5 (2010) 157-164. doi: 10.1016/j.cbd.2010.03.004

[23] S. Chocu, P. Calvel, A.D. Rolland, C. Pineau, Syst Biol Reprod Med, 58 (2012) 179-190. doi: 10.3109/19396368.2012.691943

[24] A.R. Bellve, J.C. Cavicchia, C.F. Millette, D.A. O'Brien, Y.M. Bhatnagar, M. Dym, J Cell Biol, 74 (1977) 68-85.

[25] M. Nagano, M.R. Avarbock, E.B. Leonida, C.J. Brinster, R.L. Brinster, Tissue Cell, 30 (1998) 389-397. doi: S0040-8166(98)80053-0 [pii]

[26] S. Conrad, M. Renninger, J. Hennenlotter, T. Wiesner, L. Just, M. Bonin, W. Aicher, H.J. Buhring, U. Mattheus, A. Mack, H.J. Wagner, S. Minger, M. Matzkies, M. Reppel, J. Hescheler, K.D. Sievert, A. Stenzl, T. Skutella, Nature, 456 (2008) 344-349. doi: 10.1038/nature07404

[27] J.R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann, Nat Methods, 6 (2009) 359-362. doi: 10.1038/nmeth.1322

[28] W.H. Tang, I.V. Shilov, S.L. Seymour, J Proteome Res, 7 (2008) 3661-3667. doi: 10.1021/pr070492f

[29] Z. Du, X. Zhou, Y. Ling, Z. Zhang, Z. Su, Nucleic Acids Res, 38 (2010) W64-70. doi: 10.1093/nar/gkq310

[30] F. Supek, M. Bosnjak, N. Skunca, T. Smuc, PLoS One, 6 (2011) e21800. doi: 10.1371/journal.pone.0021800

[31] P.D. Thomas, A. Kejariwal, M.J. Campbell, H. Mi, K. Diemer, N. Guo, I. Ladunga, B. Ulitsky-Lazareva, A. Muruganujan, S. Rabkin, J.A. Vandergriff, O. Doremieux, Nucleic Acids Res, 31 (2003) 334-341.

[32] W. Huang da, B.T. Sherman, R.A. Lempicki, Nucleic Acids Res, 37 (2009) 1-13. doi: 10.1093/nar/gkn923

[33] W. Huang da, B.T. Sherman, R.A. Lempicki, Nat Protoc, 4 (2009) 44-57. doi: 10.1038/nprot.2008.211

[34] D. Binns, E. Dimmer, R. Huntley, D. Barrell, C. O'Donovan, R. Apweiler, Bioinformatics, 25 (2009) 3045-3046. doi: 10.1093/bioinformatics/btp536

[35] A. Krogh, B. Larsson, G. von Heijne, E.L. Sonnhammer, J Mol Biol, 305 (2001) 567-580. doi: 10.1006/jmbi.2000.4315

[36] T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, Nat Methods, 8 (2011) 785-786. doi: 10.1038/nmeth.1701

[37] J.D. Bendtsen, L.J. Jensen, N. Blom, G. Von Heijne, S. Brunak, Protein Eng Des Sel, 17 (2004) 349-356. doi: 10.1093/protein/gzh037

[38] S.Y. Huang, J.H. Lin, Y.H. Chen, C.K. Chuang, E.C. Lin, M.C. Huang, H.F. Sunny Sun, W.C. Lee, Proteomics, 5 (2005) 4205-4212. doi: 10.1002/pmic.200401284

[39] Y.F. Zhu, Y.G. Cui, X.J. Guo, L. Wang, Y. Bi, Y.Q. Hu, X. Zhao, Q. Liu, R. Huo, M. Lin, Z.M. Zhou, J.H. Sha, J Proteome Res, 5 (2006) 2217-2225. doi: 10.1021/pr0600733

[40] X. Guo, P. Zhang, R. Huo, Z. Zhou, J. Sha, Proteomics Clin Appl, 2 (2008) 1651-1657. doi: 10.1002/prca.200780120

[41] E. Com, B. Evrard, P. Roepstorff, F. Aubry, C. Pineau, Mol Cell Proteomics, 2 (2003) 248-261. doi: 10.1074/mcp.M300010-MCP200

[42] C.N. Tomes, G.A. De Blas, M.A. Michaut, E.V. Farre, O. Cherhitin, P.E. Visconti, L.S. Mayorga, Mol Hum Reprod, 11 (2005) 43-51. doi: 10.1093/molehr/gah126

[43] A. Kumar, H. Gao, J. Xu, J. Reuben, D. Yu, K. Mehta, PLoS One, 6 (2011) e20701. doi: 10.1371/journal.pone.0020701

[44] C.A. Gaudry, E. Verderio, R.A. Jones, C. Smith, M. Griffin, Exp Cell Res, 252 (1999) 104-113. doi: 10.1006/excr.1999.4633

[45] C.A. Gaudry, E. Verderio, D. Aeschlimann, A. Cox, C. Smith, M. Griffin, J Biol Chem, 274 (1999) 30707-30714.

[46] W.Y. Lui, D.D. Mruk, C.Y. Cheng, J Cell Physiol, 202 (2005) 49-66. doi: 10.1002/jcp.20098

[47] J. Liu, X. He, Y. Qi, X. Tian, S.J. Monkley, D.R. Critchley, S.A. Corbett, S.F. Lowry, A.M. Graham, S. Li, Mol Cell Biol, 31 (2011) 3366-3377. doi: 10.1128/MCB.01403-10

[48] U.C. Sharma, D. M Vu, Journal of Stem Cell Research & Therapy, 02 (2012) 123. doi: 10.4172/2157-7633.1000123

[49] S. Benjamin, H. Weidberg, D. Rapaport, O. Pekar, M. Nudelman, D. Segal, K. Hirschberg, S. Katzav, M. Ehrlich, M. Horowitz, Biochem J, 439 (2011) 433-442. doi: 10.1042/BJ20111010

[50] N. Pashai, H. Hao, A. All, S. Gupta, R. Chaerkady, A. De Los

Angeles, J.D. Gearhart, C.L. Kerr, PLoS One, 7 (2012) e39088. doi: 10.1371/journal.pone.0039088

[51] J.C. Adams, Curr Opin Cell Biol, 16 (2004) 590-596. doi: 10.1016/j.ceb.2004.07.009

[52] B. Tubb, D.J. Mulholland, W. Vogl, Z.J. Lan, C. Niederberger, A. Cooney, J. Bryan, Exp Cell Res, 275 (2002) 92-109. doi: 10.1006/excr.2002.5486

[53] M.A. Baker, G. Reeves, L. Hetherington, R.J. Aitken, Proteomics, 10 (2010) 482-495. doi: 10.1002/pmic.200900574

[54] M. Bajpai, G. Gupta, S.K. Jain, B.S. Setty, Andrologia, 30 (1998) 311-315.

[55] S. Moscato, F. Pratesi, A. Sabbatini, D. Chimenti, M. Scavuzzo, R. Passatino, S. Bombardieri, A. Giallongo, P. Migliorini, Eur J Immunol, 30 (2000) 3575-3584.

[56] Y.J. Park, W.S. Kwon, S.A. Oh, M.G. Pang, J Proteome Res, 11 (2012) 4162-4168. doi: 10.1021/pr300248s

[57] K. Byrne, T. Leahy, R. McCulloch, M.L. Colgrave, M.K. Holland, Proteomics, 12 (2012) 3559-3579. doi: 10.1002/pmic.201200133

[58] M.D. Dun, N.D. Smith, M.A. Baker, M. Lin, R.J. Aitken, B. Nixon, J Biol Chem, 286 (2011) 36875-36887. doi: 10.1074/jbc.M110.188888

[59] R. Sanchez, M. Deppe, M. Schulz, P. Bravo, J. Villegas, P. Morales, J. Risopatron, Andrologia, 43 (2011) 114-120. doi: 10.1111/j.1439-0272.2009.01031.x

[60] K. Matsumoto, O.Y. Kwon, H. Kim, Y. Akao, Developmental dynamics : an official publication of the American Association of Anatomists, 233 (2005) 1149-1156. doi: 10.1002/dvdy.20429

[61] H. Tachiwana, A. Osakabe, H. Kimura, H. Kurumizaka, Nucleic Acids Res, 36 (2008) 2208-2218. doi: 10.1093/nar/gkn060

[62] A.D. Rolland, R. Lavigne, C. Dauly, P. Calvel, C. Kervarrec, T. Freour, B. Evrard, N. Rioux-Leclercq, J. Auger, C. Pineau, Hum Reprod, 28 (2013) 199-209. doi: 10.1093/humrep/des360

JOURNAL OF INTEGRATED OMICS
A METHODOLOGICAL JOURNAL
HTTP://WWW.JIOMICS.COM

ORIGINAL ARTICLE | DOI: 10.5584/jiomics.v3i2.137

Determining the C-Terminal Amino Acid of a Peptide from MS/MS Data

Jens Allmer

Molecular Biology and Genetics, Izmir Institute of Technology, Urla, Izmir, Turkey

Received: 26 April 2013 **Accepted:** 16 June 2013 **Available Online**: 15 July 2013

ABSTRACT

Proteomics is currently chiefly based on mass spectrometry (MS) which is the tool of choice to investigate proteins. Two computational approaches to derive the tandem mass spectrum precursor's sequence are widely employed. Database search essentially retrieves the sequence by matching the spectrum to all entries in a database whereas de novo sequencing does not depend on a sequence database. Both approaches benefit from knowledge about the enzyme used to generate the peptides. Most algorithms default to trypsin for its abundant usage. Trypsin cuts after arginine and lysine and thus the c-terminal amino acid is not known precisely and usually either of the two. Furthermore, 90% of protein terminal peptides may not end with either arginine or lysine and may thus contain any of the other amino acids. Here an algorithm is presented which predicts the c-terminal amino acid to be arginine, lysine or any other.

Here an algorithm, named RKDecider, to sort the c-terminal amino acid into one of three groups (arginine, lysine, and other) is presented. Although around 90% accuracy was achieved during data mining spectra for rules that determine the c-terminal amino acid, the implementation's (RKDecider) accuracy is a little less and achieves about 80%. This is due to the fact that the decision trees were implemented as a rule-based system for speed considerations. The implementation is freely available at: http://bioinformatics.iyte.edu.tr/RKDecider.

Keywords: mass spectrometry; proteomics; trypsin cleavage; database search; *de novo* sequencing; fragmentation analysis, c-terminal amino acid.

Abbreviations

MS: Mass Spectrometry; **Da**: Dalton; **m/z**: Mass to charge ratio.

1. Introduction

Proteomics analyzes the sequence, localization, modifications, and other parameters of proteins. Currently mass spectrometry is the tool of choice in proteomics [1]. Since peptides are easier to bring into gas phase than large proteins, bottom-up proteomics is widely applied [2]. In short, proteins are digested by a protease and the resulting peptides are fed into a mass spectrometer where their mass to charge ratio and/or their fragmentation spectra are measured. Usually trypsin is used for the purpose of digesting proteins since it leads to peptides of desirable size for mass spectrometry and since it cleaves highly specific and efficient [3,4]. Despite the widespread use of trypsin its proteolytic process has not been analyzed in depth, but has recently been modeled from mass spectrometric data [5]. Since mass spectrometry can only measure mass to charge ratios, the measurements need to be translated back to peptide sequences and then integrated into proteins. To derive the sequence of a fragmentation spectrum (MS/MS spectrum), generally two different methods are employed. The first is database search, where the MS/MS spectrum is compared to all sequences in a database to determine the best fit [6]. The other method is *de novo* sequencing which needs no other information than the MS/MS spectrum to assign a protein sequence [7]. Many algorithms for database search and *de novo* sequencing require the knowledge of the enzyme used for digestion. In case trypsin was used, it is not clear whether the c-terminal amino acid is arginine or lysine. However, for many purposes it would be beneficial to know the c-terminal amino acid precisely (see below).

Peptide fragmentation in gas phase leads to the cleavage of peptides into smaller fragments, a process for which many

***Corresponding author:** Jens Allmer, tel.: 00902327507310, fax: 00902327507303. Email Address: jens@allmer.de; address: Assoc. Prof. Dr. Jens Allmer, Molecular Biology and Gentics, Izmir Institute of Technology, Gulbahce Campus, Urla, Izmir, Turkey.

methods are available [7]. The fragmentation mechanism, even for a time tested method like collision induced dissociation (CID) [8] is not completely understood and has not been completely integrated into algorithms that assign amino acid sequences to MS/MS spectra, although statistics have been derived from larger amounts of spectra [9]. These fragments are important for database search and *de novo* sequencing algorithms to derive the sequence of the peptide precursor that gave rise to the MS/MS spectra. Among the resulting peaks there may be some which could facilitate the distinction of which amino acid terminates a peptide. Previously, it has been shown that a peak, resulting from 17 Dalton or 42 Dalton eliminated from the precursor ion, may be diagnostic for arginine [10].

Olsen and colleagues showed that the largest amount of peptides derived from a tryptic digest have either arginine or lysine at the c-terminus and conclude that non tryptic peptide assignments, done by some algorithms, should not be trusted [4]. They acknowledge, however, that not all peptides that result from a tryptic digest have a c-terminal arginine or lysine [4]. Taking into account that there are 20 possible amino acids terminating a protein, 90% of them are not arginine or lysine. Some studies use non-specific cleavage to increase the number of identified spectra but this is controversial and likely only increases the number of false positive identifications [4]. Instead of globally turning a search engine to non-specific cleavage for all spectra, such costly and dangerous operations can be performed on the basis of the decision whether the c-terminus of the peptide, that gave rise to the mass spectrum, is not tryptic. Thus only spectra which are not tryptic trigger searches with non-specific enzymatic cleavage settings. Additionally, it is of help to know the c-terminus so that peptide candidates in database search can either be pre-filtered with the knowledge of the c-terminal amino acid or the results can be evaluated in respect to proper sequence selection based on the terminal amino acid.

In order to reap these benefits an approach and its implementation to decide whether the c-terminus of peptides, underlying CID spectra from LTQ instruments, is tryptic and which amino acid is at the c-terminus, is presented here. 12 potential diagnostic losses for arginine and lysine were defined and data mining on about 8500 LTQ spectra was performed. From these, rules were derived that in data mining practice can distinguish between arginine, lysine, or other amino acids with an accuracy of about 90%. The practical implementation reaches an accuracy of about 80% since other limitations like speed and unknown charge state had to be taken into account.

2. Methods

2.1. *Spectral Dataset*

44 synthetic peptides, mostly derived from cytochrome-c and bovine serum albumin, were designed and ordered for a different purpose than this study and will be published elsewhere. The dataset had to be prepared in this way to have a ground truth for the analysis [11]. These peptides were directly injected into a Thermo Finnigan LTQ mass spectrometer and their collision induced dissociation fragmentation spectra were recorded. This resulted in 8447 MS/MS spectra with an average of 192 spectra per peptide. 42 peptides were tryptic while two (QVYQGCGV and YKELGFQG) were not and ended in valine and glycine, respectively. The complete list of peptides and the number of measured MS/MS spectra are available as Supplementary File 1. The resulting spectra were predominantly from singly charged precursors (5353); and spectra from doubly charged precursors (2616) as well as spectra from triply charged precursors (478) were less. 63% of the spectra are derived from a precursor terminating with lysine, 34% from a precursor terminating with arginine, and 3% had a precursor that terminated with a different amino acid.

2.2. *Diagnostic Fragments and Data Mining*

In order to determine whether a spectrum derives from a peptide precursor ending in arginine, lysine, or another amino acid, 12 parameters were defined. These parameters are relative losses of the precursor ion. A loss of 156.1 Da of the precursor ion, for instance, signifies the loss of arginine. In the following -156.1 shall signify a loss of 156.1 Da from the precursor. The chosen diagnostic fragments are listed and briefly explained in Table 1. Not all of the selected fragments have to exist in practice. Their absence can also be learned by the machine learning algorithms employed here and they can hence still be used as diagnostic fragments. The chosen diagnostic fragments are the following: -16, -17, -32, -33, -34, -42, -43, -57, -128.1, -129, -156.1, -175.

Orange Canvas [12] was used for all data mining and if not otherwise stated in the text, the default settings were used. For example in all cases 10 fold cross validation was used for learning and testing instead of the default which is only 5 fold.

2.3. *Software Implementation*

The implementation of the RKDecider software was written in Java™ using the Netbeans integrated development environment version 7.1. The implementation is available for download at http://bioinformatics.iyte.edu.tr/RKDecider.

3. Results

The recorded LTQ spectra were analyzed and their charge state was determined using the recorded precursor mass and the theoretically calculated mass. Since the sequence was known and there was no mixture this procedure guaranteed accurate charge determination. A window of +/- fragment tolerance (0.3 Da) around the peak with the diagnostic loss was extracted from all spectra and in case multiple peaks were found in the window their abundances were summed.

Table 1. In order to learn whether a spectrum derives from a peptide ending in arginine or lysine, a number of peaks, described as losses from the precursor ion, have been defined. Here the loss from the precursor ion is given in Daltons and the fragment is briefly explained.

Loss (Da)	Reasoning
-16	As -17 but accommodating for one hydrogen difference (accounting for possible dependence on the acidity of the solution)
-17	Loss of NH3 which is possible for arginine and lysine
-32	As -34 but accommodating for two hydrogen difference (accounting for possible dependence on the acidity of the solution)
-33	As -34 but accommodating for one hydrogen difference (accounting for possible dependence on the acidity of the solution)
-34	Loss of 2 NH3 which is only possible for arginine.
-42	As -43 but accommodating for one hydrogen difference (accounting for possible dependence on the acidity of the solution)
-43	Partial elimination of the terminal part of the side chain from arginine (C1N2H4).
-57	Full elimination of the terminal part of the side chain from arginine (C1N3H6).
-128.1	Molecular weight of lysine without water; elimination of the amino acid.
-129	Elimination of the immonium ion of arginine.
-156.1	Molecular weight of arginine without water; elimination of the amino acid.
-175	Molecular weight of arginine including water and one hydrogen; full elimination, leading to a radical ion.

For all spectra and all diagnostic losses the relative abundances of the windows (normalized to the total ion current) were recorded and then submitted to data mining. Since the charge state was known, the diagnostic fragments for doubly charged precursors were divided by two and for triply charged precursors by three. Although the resulting mass changes, the nomenclature is not changed here, for convenience. Due to this data mining is also simplified because the addition of parameters only valid in certain subsets of the overall data set can be avoided. Consequently, all charges are treated with the same learned model. However, many classification algorithms require two classes and cannot work on data sets with more than two classes. Therefore the three class problem (arginine, lysine, and other) was split up into two data sets. One approach investigates whether the c-terminus is arginine or any other amino acid. The other whether the c-terminus is lysine or not.

3.1. *Determining C-Terminal Arginine*

In order to investigate the importance of the selected diagnostic losses, their information content was ranked using a number of algorithms available in Orange Canvas (Table 2).

Most algorithms agree in general, except for SVM Weight. It can be gathered from Table 2 that the first six parameters contain the majority of the information needed to distinguish between arginine and other amino acids at the c-terminus. Nonetheless, all parameters were included when training the learners since they may still contribute some of the distinguishing power.

When looking at the resulting decision tree (Figure 1) which was created with a different algorithm from the ones in Table 2, some of the parameters, that did not receive good scores in Table 2, are used early on in the decision trees which means they have great distinguishing power between classes.

The collected data was analyzed using the supervised learning algorithms available in Orange Canvas. As expected, not all algorithms performed equally well, with the best one being Random Forest (Table 3).The distance to k-Nearest Neighbor, CN2 Rules, and Classification Tree are not very significant, though. The following algorithm, Logistic Regression

Table 2. The information content of the attributes according to different measures is presented. Rows are sorted in respect to Random Forest since its classification was most accurate for predicting the arginine or other status of peptide c-termini.

Diagnostic Loss	Random Forest	ReliefF	Inf. Gain	Gain Ratio	Gini	SVM Weight
-156.1	6.57	0.06	0.07	0.04	0.02	36.71
-175	4.02	0.00	0.01	0.00	0.00	0.09
-42	3.20	-0.01	0.19	0.11	0.05	216.31
-43	2.67	-0.01	0.09	0.05	0.03	22.68
-17	2.34	0.04	0.12	0.06	0.03	303.88
-16	2.04	0.01	0.14	0.08	0.04	162.40
-34	1.65	0.00	0.15	0.08	0.04	277.53
-33	1.36	0.01	0.07	0.04	0.02	89.99
-129	1.28	0.00	0.05	0.04	0.02	132.19
-57	0.88	-0.02	0.01	0.01	0.00	14.61
-128.1	0.78	0.01	0.04	0.03	0.01	43.33
-32	0.62	0.00	0.00	0.00	0.00	8.18

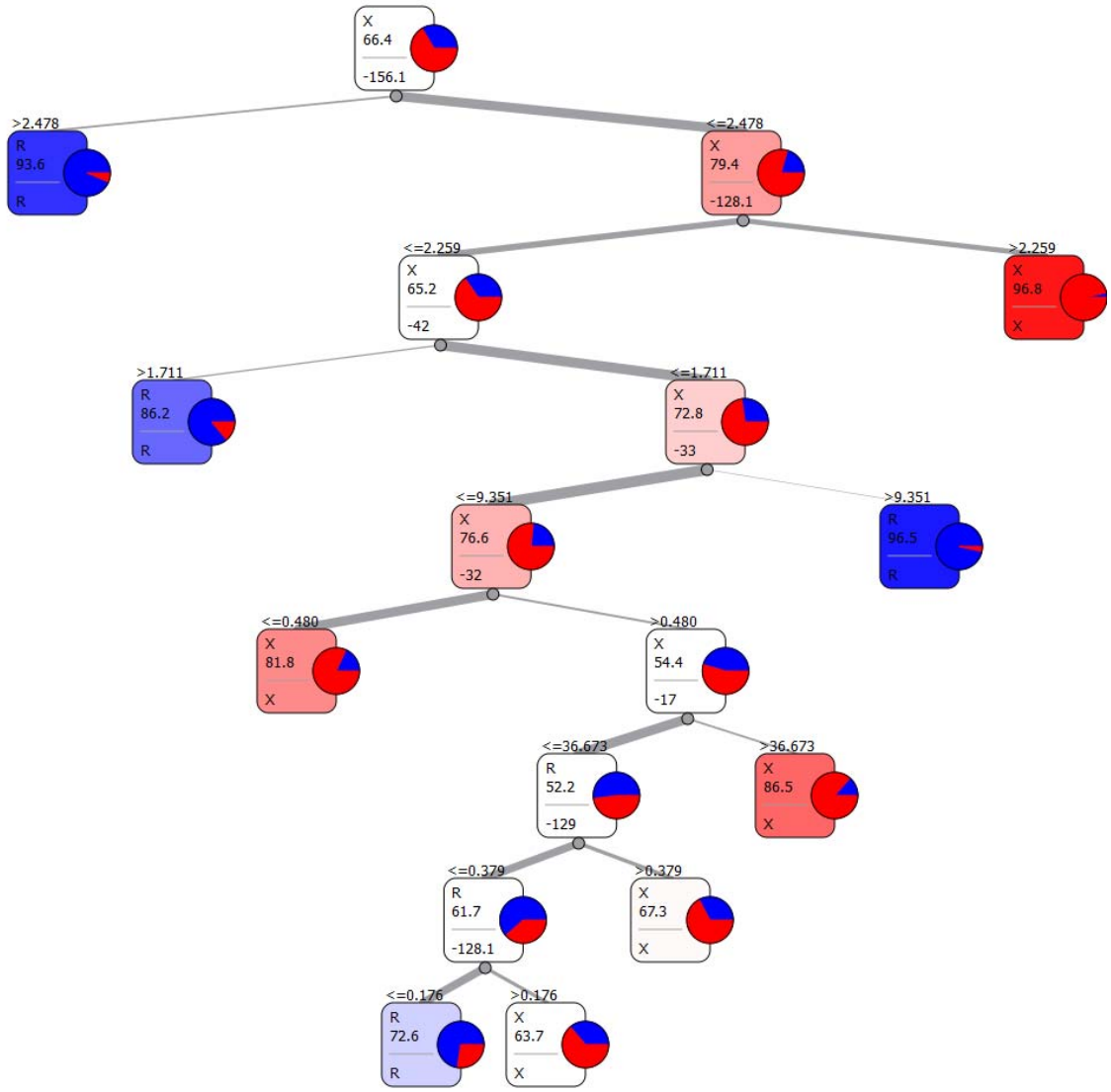


Figure 1. Decision tree for determining whether an MS/MS spectrum has a C-Terminal R. For best split exhaustive search was applied. When nodes reached more than 95% for majority class or less than 10 examples remained in a leaf no further splits were made. The nodes depict the majority class and its percentage in the upper left corner. Below the line in the nodes the next splitting parameter is shown. The pie chart graphically conveys the distribution of examples between classes. When classes become more pure the node is colored with a stronger shade of the associated color (red for any amino acid, and blue for arginine).

performs, however, much worse on this data.

Although the Random Forest algorithm achieves the best results on this data, it does not allow the construction of a decision tree within Orange Canvas, which could be implemented into software. Therefore, the Classification Tree algorithm was used to construct a decision tree (Figure 1) although it is slightly less accurate (Table 3).

Figure 1 shows that the most discriminating parameter is -156.1 which is used for the initial split of the data. This is followed by -128.1 and then by -42. This is slightly different from the data presented in Table 1 which is due to the fact that a different algorithm is employed. After the 6th split there is not much further benefit to prediction accuracy as the number of examples strongly decrease and the accuracy of the splits also decrease (Figure 1). This analysis is only for the decision whether a peptide precursor terminates with arginine or not. Additionally, the decision between lysine and other amino acids needs to be made.

3.2. *Determining C-Terminal Lysine*

The same process made for predicting arginine as the terminal amino acid was repeated for lysine. First the information content is ranked according to several algorithms. The assumption here is that different parameters should be important for the decision than the ones observed for arginine. Obviously, there is some overlap of important parameters such as -156.1 and -42 (Table 4). However, whether the

Table 3. Classification accuracy and other quality measured for all classification algorithms available in Orange Canvas for the prediction of terminal arginine or any other amino acid. The table is ordered by classification accuracy.

Classification Algorithm	Accuracy	Sensitivity	Specificity	AUC	F1	Precision	Recall
Random Forest	0.9022	0.8942	0.9157	0.9684	0.9200	0.9473	0.8942
kNN	0.9015	0.9252	0.8613	0.9599	0.9220	0.9188	0.9252
CN2 rules	0.8860	0.9390	0.7962	0.9341	0.9120	0.8865	0.9390
Classification Tree	0.8696	0.8988	0.8201	0.9378	0.8966	0.8944	0.8988
Logistic Regression	0.7758	0.7602	0.8022	0.8655	0.8101	0.8669	0.7602
Naive Bayes	0.7416	0.6618	0.8769	0.8408	0.7632	0.9011	0.6618
SVM	0.7281	0.6647	0.8356	0.8328	0.7546	0.8727	0.6647
Majority	0.6290	1.0000	0.0000	0.5000	0.7722	0.6290	1.0000

absence or presence of a parameter is important is not conveyed in this analysis. Thus the absence of a parameter could be support in one case whereas in the other case its presence could mean support. It can be observed that the most important parameter in Table 4 is one of the least important ones in Table 2.

So the assumption holds and different parameters or a different semantic of the same parameter are important for lysine prediction. This difference is further supported by the significantly different make of the decision tree (Figure 2) when compared to the arginine decision tree (Figure 1).

Just like for the classification of terminal arginine, several supervised learning algorithms were tried for learning from the data (Table 5). In this case the k-Nearest Neighbor (kNN) algorithm separates best among the two classes (lysine and other). In this analysis, the distance between the top scoring algorithms is even closer than for the prediction of arginine. Only Majority vote is significantly worse on this data than the other algorithms.

The classification accuracy for decision between lysine and other amino acids is slightly better than for the decision between arginine and other amino acids. Unfortunately, kNN does not produce rules or decision trees, either so that again the Classification Tree algorithm provided by Orange Canvas had to be employed for visualizing how decisions can be made between the classes (Figure 2).

Figure 2 shows that the decision for whether the c-terminus is lysine or any other amino acid is much more involved than for arginine (see Figure 1). Another aim for making the decision tree was to implement it into a rule based system for predicting the c-terminal amino acid of the precursor of an MS/MS spectrum.

Unfortunately, it turned out that this is only possible if the charge state of the precursor is known. Furthermore, there is accuracy dependence between charge and observed accuracy with higher charges leading to less accurate classification (Table 6). Therefore, software was designed to perform different decisions, if the charge is known and default to charge one if it is not provided.

3.3. *Software Implementation of the RKDecider*

The data mining results presented above show the theoretically possible accuracy based on the measured data set. One complication that would occur in practice is that the charge may not be annotated for recorded mass spectra. In order to develop software, which can be used to predict the status of the c-terminal amino acid (arginine, lysine, or other) the data set was split into six subsets based on the charge state and the terminal amino acid. For each resulting subset, a decision tree was built using Orange Canvas. These decision trees were subsequently implemented into software, named RKDecider.

This software, available as a console application, that implements the decision trees produced by Orange Canvas and combines them into one algorithm is available for download at: http://bioinformatics.iyte.edu.tr/RKDecider. The application first decides whether an incoming spectrum has a c-terminal arginine based on a decision tree (Supplementary File 2). The first check is for arginine as the prediction accuracy is higher than for lysine prediction (Table 6). The spectra that were not assigned arginine status are checked using the decision trees built for terminal lysine (Supplementary File 2). Finally, all unassigned spectra are labeled unknown. This either means that the spectral quality was not well enough, or did not contain the diagnostic peaks, to allow a proper identification or that the spectrum does not originate from a tryptic peptide like 90% of all protein terminal tryptic fragments which do not have a c-terminal lysine or arginine.

For this split data set, the accuracies that can be reached are limited by the algorithm which was used for decision tree building. The regular Classification Tree building algorithm in Orange Canvas was used to produce the decision trees (Supplementary File 2).

The implementation was tested on about 8500 LTQ spectra

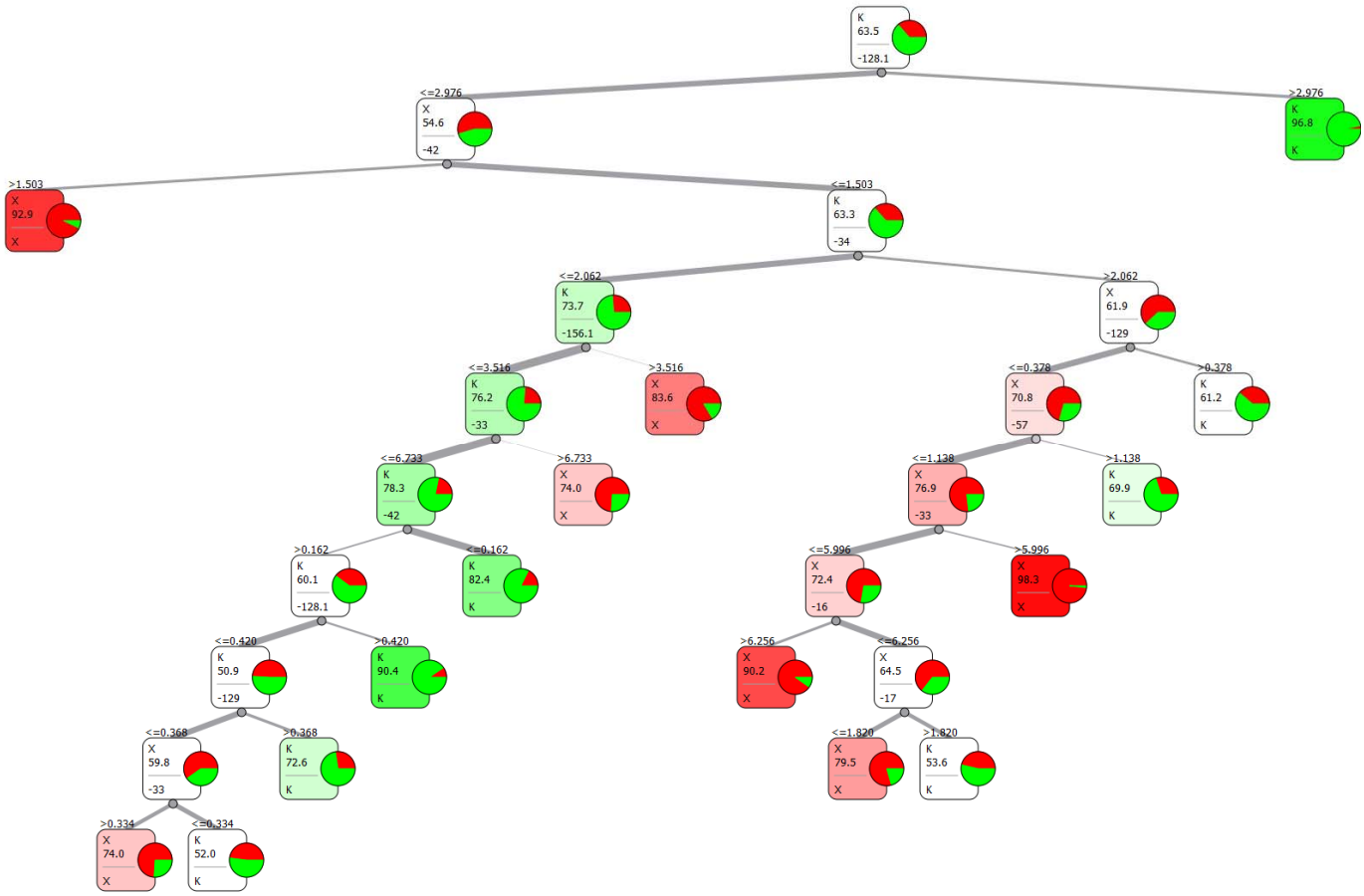


Figure 2. Decision tree for determining whether an MS/MS spectrum has a C-Terminal K. For best split exhaustive search was applied. When nodes reached more than 95% for majority class or less than 10 examples remained in a leaf no further splits were made. The nodes depict the majority class and its percentage in the upper left corner. Below the line in the nodes the next splitting parameter is shown. The pie chart graphically conveys the distribution of examples between classes. When classes become more pure the node is colored with a stronger shade of the associated color (red for any amino acid, and green for arginine).

Table 4. The information content of the attributes according to different measures is presented. Rows are sorted in respect to Random Forest since its classification was most accurate for predicting the lysine or other status of a peptide c-terminus.

Diagnostic Loss	Random Forest	ReliefF	Inf. Gain	Gain Ratio	Gini	SVM Weight
-128.1	9.30	0.03	0.22	0.12	0.06	131.42
-156.1	7.05	0.02	0.19	0.13	0.06	0.80
-34	6.18	0.01	0.20	0.10	0.06	34.51
-42	5.51	0.03	0.21	0.13	0.07	17.29
-17	3.53	0.02	0.05	0.02	0.02	1.64
-33	3.22	0.00	0.13	0.07	0.04	36.35
-16	2.32	0.00	0.11	0.06	0.04	4.72
-129	1.77	0.00	0.08	0.05	0.02	40.68
-43	1.68	0.02	0.07	0.04	0.02	0.23
-57	1.57	0.01	0.01	0.00	0.00	0.25
-175	0.89	0.00	0.05	0.03	0.02	2.26
-32	0.84	0.00	0.01	0.01	0.00	0.28

and achieved a combined accuracy of 80%. The implementation's accuracy of 80%, which in comparison to the accuracy of about 90%, achieved during data mining, seems low, but may be largely due to the fact that the spectra that were used for testing, had the annotation of their proper charge removed so that many larger peptides are likely predicted wrong. Since missing charge annotation is a general problem in mass spectrometry-based proteomics, charges were not given to the algorithm in this test.

4. Conclusion and Outlook

There is a need to know the identity of the c-terminal amino acid of a peptide. It can help to pre-filter database search candidates or to adjust database search settings so that the scope can easily be extended to include peptides with non tryptic termini, something which currently is controversial (Olsen et al., 2004). Furthermore, the confidence in the results can be elevated if the reported terminal amino acid is equal to the one predicted by RKDecider. It can be of benefit for *de novo* sequencing or sequence tag searches since it fixes one variable and thus leads to more precise results. This in-

Table 5. Classification accuracy and other quality measured for all classification algorithms available in Orange Canvas for the classification between terminal lysine and any other amino acid. The table is ordered by classification accuracy.

Classification	Accuracy	Sensitivity	Specificity	AUC	F1	Precision	Recall
Random Forest	0.9236	0.9390	0.8969	0.9686	0.9398	0.9406	0.9390
kNN	0.9148	0.9646	0.8282	0.9455	0.9349	0.9071	0.9646
CN2 rules	0.9081	0.9278	0.8739	0.9011	0.9277	0.9275	0.9278
Classification	0.8969	0.9894	0.7361	0.9792	0.9241	0.8670	0.9894
Logistic	0.8460	0.9536	0.6590	0.9259	0.8871	0.8294	0.9536
Naive Bayes	0.8280	0.9517	0.6130	0.8991	0.8754	0.8104	0.9517
SVM	0.8241	0.9441	0.6156	0.9033	0.8720	0.8102	0.9441
Majority	0.6348	1.0000	0.0000	0.5000	0.7766	0.6348	1.0000

Table 6. Classification accuracy at different charges and for separate prediction of terminal arginine or terminal lysine. Two algorithms from the Orange Canvas package were used, classification tree and random forest.

		Classification Accuracy	
Charge	C-Terminus	Classification Tree	Random Forest
1	R vs. other	98.45	99.51
2	R vs. other	82.68	88.42
3	R vs. other	82.41	86.39
1	K vs. other	97.66	98.51
2	K vs. other	82.68	88.42
3	K vs. other	82.41	86.39

formation can also be used as a simple pre-selection filter which would just reject all peptides that do not seem to have a proper tryptic c-terminus.

About 8500 LTQ CID spectra were recorded for 44 synthetic peptides and this data was mined for rules that could predict whether the c-terminus is arginine, lysine, or some other amino acid. The overall data mining accuracy peaked at around 90%. The rules, that were learned, have been implemented into software which achieves an accuracy of 80% and is freely available at http://bioinformatics.iyte.edu.tr/RKDecider.

In the future, it will be important to extend this analysis to other mass spectrometers and fragmentation methods, but for this to be successful, proper, accurate, and thus trustable benchmark data sets need to be created first [11]. The algorithm can also be improved by incorporating additional diagnostic losses which either support or contradict the existence of a specific amino acid at the c-terminus.

5. Supplementary material

Supplementary data and information is available at: http://www.jiomics.com/index.php/jio/rt/suppFiles/137/0

Acknowledgements

The author would like to thank Talat Yalcin and the Biological Mass Spectrometry and Proteomics facility at the Izmir Institute of Technology for their support and for vivid discussions about manuscript and project. This work was in part supported by an award for outstanding young scientists, received from the Turkish Academy of Sciences (TÜBA, GEBIP, www.tuba.gov.tr).

References

[1] M. Mann, R.C.C. Hendrickson, A. Pandey, Annual Review of Biochemistry 70 (2001) 437–73. DOI: 10.1146/annurev.biochem.70.1.437

[2] S.P. Mirza, M. Olivier, Physiological Genomics 33 (2008) 3–11. DOI: 10.1152/physiolgenomics.00292.2007

[3] H. Steen, M. Mann, Nat Rev Mol Cell Biol 5 (2004) 699–711.

[4] J. V Olsen, S.-E. Ong, M. Mann, Molecular & Cellular Proteomics: MCP 3 (2004) 608–14. DOI: 10.1074/mcp.T400003-MCP200

[5] S. Aiche, K. Reinert, C. Schütte, D. Hildebrand, H. Schlüter, T.O.F. Conrad, PloS One 7 (2012) e40656. DOI: 10.1371/

journal.pone.0040656

[6] E. Kapp, F. Schütz, Current Protocols in Protein Science / Editorial Board, John E. Coligan ... [et Al.] Chapter 25 (2007) Unit25.2. DOI: 10.1002/0471140864.ps2502s49

[7] J. Allmer, Expert Rev. Proteomics 8 (2011) 645–57. DOI: 10.1586/epr.11.54

[8] J.M. Wells, S.A. McLuckey, Methods in Enzymology 402 (2005) 148–185. DOI: 10.1016/S0076-6879(05)02005-7

[9] F. Schütz, E.A. a Kapp, R.J.J. Simpson, T.P.P. Speed, F. Schutz, Biochemical Society Transactions 31 (2003) 1479–83. DOI: 10.1042/

[10] M. Garzotti, M. Hamdan, Rapid Communications in Mass Spectrometry: RCM 12 (1998) 843–8. DOI: 10.1002/(SICI)1097-0231(19980715)12:13<843::AID-RCM250>3.0.CO;2-W

[11] J. Allmer, Journal of Integrated OMICS (2012). DOI: 10.5584/jiomics.v2012i2012.113

[12] T. Curk, J. Demsar, Q. Xu, G. Leban, U. Petrovic, I. Bratko, G. Shaulsky, B. Zupan, Bioinformatics 21 (2005) 396–8. DOI: 10.1093/bioinformatics/bth474

JOURNAL OF INTEGRATED OMICS
A METHODOLOGICAL JOURNAL
HTTP://WWW.JIOMICS.COM

ORIGINAL ARTICLE | DOI: 10.5584/jiomics.v3i2.139

Revisiting Protocols for the NMR Analysis of Bacterial Metabolomes

Steven Halouska[1,†], Bo Zhang[1,†], Rosmarie Gaupp[2], Shulei Lei[1], Emily Snell[1], Robert J. Fenton[2], Raul G. Barletta[2], Greg A. Somerville[2], Robert Powers[1,*]

[1]Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588-0304. [2]School of Veterinary Medicine and Biomedical Sciences, University of Nebraska-Lincoln, Lincoln, NE 68588-0905. [†]Equal contribution.

Received: 07 May 2013 **Accepted:** 09 July 2013 **Available Online**: 17 July 2013

ABSTRACT

Over the past decade, metabolomics has emerged as an important technique for systems biology. Measuring all the metabolites in a biological system provides an invaluable source of information to explore various cellular processes, and to investigate the impact of environmental factors and genetic modifications. Nuclear magnetic resonance (NMR) spectroscopy is an important method routinely employed in metabolomics. NMR provides comprehensive structural and quantitative information useful for metabolomics fingerprinting, chemometric analysis, metabolite identification and metabolic pathway construction. A successful metabolomics study relies on proper experimental protocols for the collection, handling, processing and analysis of metabolomics data. Critically, these protocols should eliminate or avoid biologically-irrelevant changes to the metabolome. We provide a comprehensive description of our NMR-based metabolomics procedures optimized for the analysis of bacterial metabolomes. The technical details described within this manuscript should provide a useful guide to reliably apply our NMR-based metabolomics methodology to systems biology studies.

Keywords: NMR; metabolomics; chemometrics; Mycobacterium; Staphylococcus; bacteria.

Abbreviations

NMR, Nuclear Magnetic Resonance; **MS**, Mass Spectrometry; **LC**, Liquid Chromatography; **GC**, Gas Chromatography; **CE**, Capillary Electrophoresis; **PCA**, Principal Component Analysis; **OPLS-DA**, Orthogonal Projection to Latent Structures Discriminant Analysis; **HSQC**, Heteronuclear Single Quantum Coherence; **TOCSY**, Total Correlated Spectroscopy; **TCA**, Tricarboxylic Acid; **TSB**, Tryptic Soy Broth; **MADC**, Middlebrook Albumin Dextrose Catalase; **CFU**, Colony Forming Unit; **MIC_{50}**, Minimum Inhibitory Concentration required to inhibit the growth of 50% of organisms; **SOGGY**, Solvent Optimized Gradient-Gradient Spectroscopy; **$HSQC_0$**, Time-Zero HSQC; **TMSP-d_4**, 3-(trimethylsilyl)propionic acid-2,2,3,3-d_4.

1. Introduction

Metabolomics is the study of small molecules in a biological system that participates in the metabolic reactions responsible for cell growth, survival, and other normal cellular functions [1-3]. Additionally, the metabolome responds to transcriptional and translational alterations associated with genotypical, epigenetic, or environmental perturbations [4-7]. Thus, metabolomics provides an assessment of global perturbations with respect to an altered genome, proteome, or environment [2, 8, 9]. The simultaneous integration of genomic, transcriptomic and proteomic data has enabled an in-depth analysis of the interplay, interaction, and regulation of DNA, RNA and proteins [10-12]. Along this line, monitoring the bacterial metabolome and integrating the results with other "omics" data has provided valuable insights into bacterial adaptability [13], biofilms [14], evolution [15], pathogenesis [16], and drug resistance [17].

Depending on the organism and growth state, the total number of metabolites within a cell varies between several

***Corresponding author:** Robert Powers, Department of Chemistry, 722 Hamilton Hall, University of Nebraska-Lincoln, Lincoln, NE 68588-0304, Tel: (402) 472-3039; Fax (402) 472-9402. Email Address: rpowers3@unl.edu

hundred to a few thousand, with a corresponding diversity in physical and chemical properties, such as size, stability, and concentration [18]. In addition to the challenge of the simultaneous study of all the metabolites within a given biological system [19], the selection of an analytical technique will influence which metabolites are observed. NMR and MS are commonly employed for metabolomics, where both instruments can be interfaced with LC, GC, and CE systems to select and emphasize specific components of the metabolome [20-24]. NMR has a number of advantages in analyzing the metabolome that includes minimal sample handling and that it is not reliant on chromatography to purify or separate metabolites. In addition, multiple resonances from a single molecule increase the accuracy of metabolite identification and quantitation. This accuracy can be further enhanced by the application of ^{13}C- and ^{15}N-isotope labeling to enhance specific regions of the metabolome [25, 26]. Importantly, the choice of ^{13}C- or ^{15}N-labeled metabolite determines the region of the metabolome observed by NMR, providing significant flexibility in experimental design. In contrast to MS, NMR is a relatively insensitive technique and only observes the most abundant ($\geq$ 1 to 5 μM) metabolites. In addition, MS has the advantage of detecting a wider-range of the metabolome. However, because of the relatively low molecular-weight range of the metabolome, MS methods generally require chromatography to separate metabolites before analysis [27]. Additionally, variations in ionization and the occurrence of ion suppression in a complex mixture add uncertainty in detecting specific metabolites by MS [28]. Finally, quantitation by MS is typically more challenging than NMR. Taken together, NMR and MS each have strengths and weaknesses but should be viewed as complementary techniques [29].

NMR-based metabolomics have been used to study a wide range of biological systems such as tissues [30], biofluids [31], mammalian cell cultures [32], plants [33] and bacteria [34-36]. The overall procedure for an NMR-based metabolomics study includes the following general steps: cell growth and harvesting, metabolite extraction, NMR data collection and analysis, multivariate statistical analysis, metabolite identification and quantification [37]. Typically, one-dimensional (1D) ^{1}H NMR spectra are used for a multivariate analysis such as principal component analysis (PCA) or orthogonal projection to latent structures discriminant analysis (OPLS-DA) [38, 39]. Both PCA and OPLS-DA provide global profiles of metabolome changes [40, 41]. Two-dimensional (2D) ^{1}H,^{13}C Heteronuclear Single Quantum Coherence (HSQC) or ^{1}H,^{1}H TOtal Correlated SpectroscopY (TOCSY) NMR experiments are used for the quantitative assessment of metabolite changes resulting from genetic modification or external stimuli [5, 14]. The ability to generate global profiles and quantitative differences coupled with the ease of applying NMR-based metabolomics has contributed to the rapid growth of the NMR metabolomics field. While NMR data acquisition and analysis methods are improving, care must be taken to ensure that the methods are appropriate to the task at hand and generate biologically relevant information. As an example, protocols to efficiently extract metabolites without inducing cellular changes are essential for success [32, 42]. In brief, the observed changes in the metabolome should reflect a change in the state of the system instead of how the cells are handled and processed. Similarly, variations in instrument performance, choice of procedures for data collection and processing, and invalidated models from multivariate analysis may induce unintended biases or incorrect interpretation of metabolomics data [43-46].

Since NMR-based metabolomics is a relatively new and still developing technology, improving and enhancing the experimental protocols is necessary to advance the field and ensure continued success. Toward this end, we describe our recently developed and optimized protocols for the application of NMR metabolomics to microbial samples. We present our current methodology and also discuss the challenges associated with each major step of the process: (i) sample preparation, (ii) NMR data collection and processing, (iii) multivariate statistical analysis, (vi) metabolite identification and network generation. Specifically, the overall methodology will be discussed in detail, where a number of key features will also be highlighted, such as automation, bioinformatics, experimental design, and harvesting the metabolome. The focus of our efforts has been to identify and minimize procedural steps that negatively influence the outcome of an NMR-based metabolomics experiment.

2. Experimental Design

A general protocol for the analysis of bacterial metabolomes using NMR is shown in Figure 1. The flow diagram illustrates procedures for both a global analysis of metabolome changes (metabolomics fingerprinting); and the identification and quantitation of specific metabolites correlated with the biological process (metabolomics profiling). The overall process consists of the following key steps: bacterial cultivation and harvesting, metabolite extraction, NMR data collection and analysis, multivariate statistical analysis, metabolite identification and quantification. Successful metabolomics sample preparation involves three steps. The first step is the simultaneous growth of all of the bacterial cultures or as many as is practical at a time. The bacteria are grown in a standard medium for fingerprint analysis, whereas the medium is supplemented with a ^{13}C-labeled metabolite for metabolomics profiling [47, 48]. After the bacteria are grown for a defined time or they have achieved a specified cell density, the second sample preparation step involves harvesting the bacteria, quenching to halt all enzymatic processes, and washing to remove the medium. The third sample preparation step involves lysing the cells and extracting the metabolome. A variety of solvents are routinely employed depending on the solubility of the targeted metabolites (cytosolic

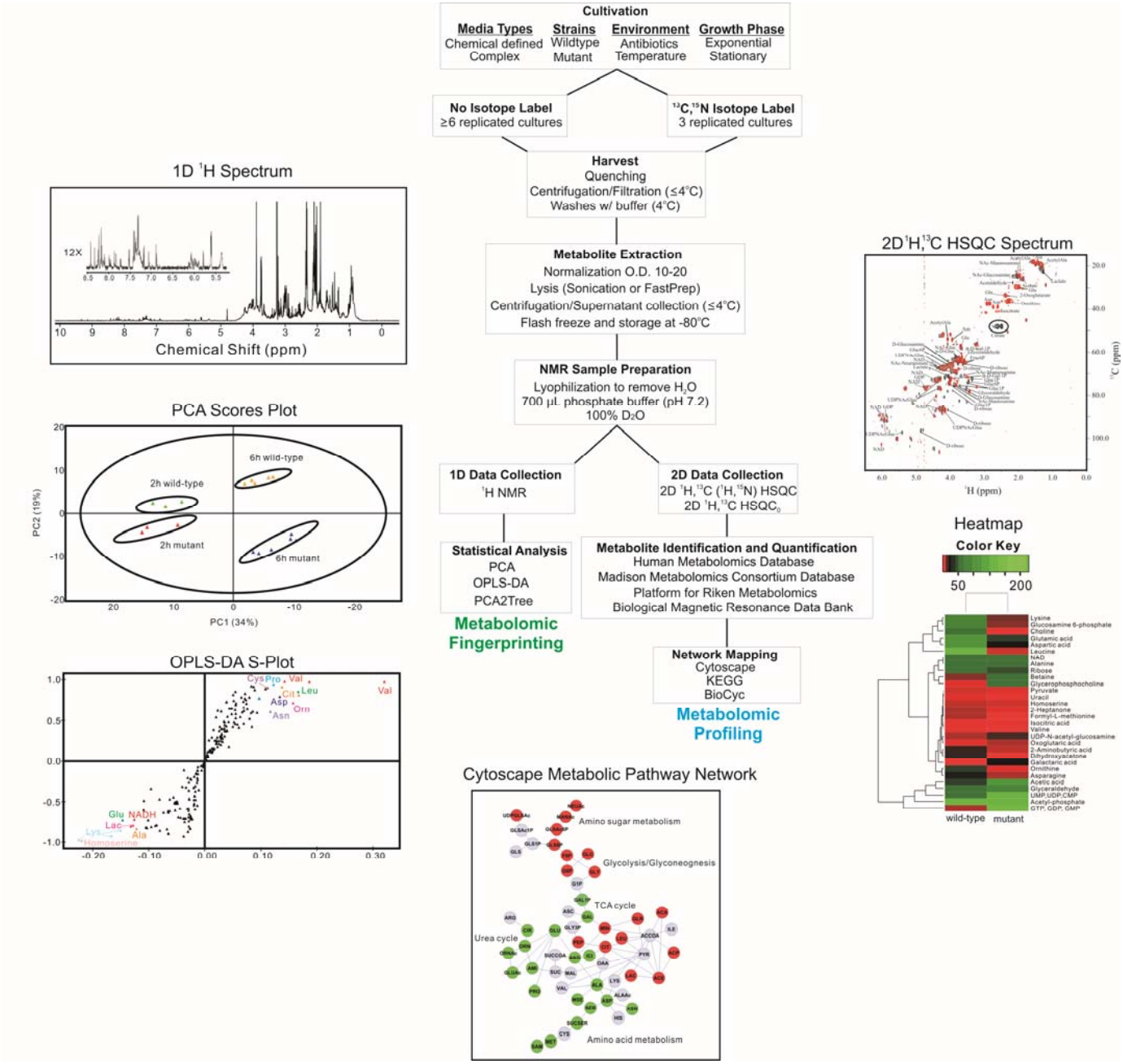


Figure 1. A flow chart of our protocol used for the NMR analysis of bacterial metabolomes.

metabolites, lipids, *etc*.). The metabolomics samples are then used to generate a series of NMR spectra, which are used for the multivariate statistical analysis, metabolite identification and quantification. The individual steps of the NMR-based metabolomics protocol will be discussed in detail highlighting challenges associated with each step.

2.1 *Identify the appropriate biological system for a metabolomics study*

NMR-based metabolomics is an important tool in systems biology research. The quantitative and qualitative measurement of metabolites from cytosolic extracts can be extremely valuable for investigating cellular processes, pathogenesis, and the effects of drugs or the environment on bacteria. Unfortunately, the bacterial metabolome is a complex mixture of metabolites and numerous interconnected metabolic and signaling pathways. This high interconnectivity may result in significant metabolite concentration changes far from the origin of the perturbation (inhibited, inactivated or down-regulated protein). Correspondingly, it is easier to observe changes to the metabolome than deduce the primary source of the perturbation after its impact has rippled throughout the metabolome. As an illustration, treating a bacterial cul-

ture with a particular drug would be expected to lead to a global change in the metabolome, but interpreting these changes to identify the therapeutic target is extremely challenging. To address this challenge, the *in vivo* mechanism of action of a potential drug lead may be determined by comparing these metabolome changes to other drugs with known biological targets [49] or to a mutant bacteria where a specific protein target is ablated or modified by genetic inactivation [50, 51]. This example illustrates that the comparative analysis between two or more metabolomes is an effective application of metabolomics. In order to obtain reliable insights into the physiology of bacteria or any other organism, it is essential to identify and establish at least two reference metabolomes (wild-type vs. mutant, drug-resistant vs. drug susceptible, nutrient-rich vs. nutrient-limited, *etc*.) for a comparative analysis. Once the reference conditions are established, bacteria can be exposed to any range of experimental variables such as a drug treatment, environmental stimuli (pH, temperature, nutrient change), or gene knockout (mutants, RNAi, inhibitor) to determine if similarities exist with the reference metabolome. The similarity between metabolomes infers an overlap in the underlying physiological processes or responses that gave rise to the metabolome changes. We have used this approach to demonstrate the similarity of *Staphylococcus epidermidis* metabolomes resulting from exposure to divergent environmental stressors that are known to facilitate biofilm formation [5, 14, 52]. Our results suggested that the tricarboxylic acid (TCA) cycle acts as a metabolic signaling network to transduce external stresses into internal metabolic signals. This conclusion was only possible because the experimental design was based on comparing the metabolomes of the *S. epidermidis* wild-type strain 1457 and an aconitase mutant strain 1457-*acnA::tetM* with and without the treatment of biofilm stressors. In summary, the successful outcome of a metabolomics study hinges on the experimental design and the proper choice of the cellular metabolomes to be compared.

2.2 *Minimize unintended bias and biologically irrelevant variations*

In addition to the proper choice of bacterial strains to compare in a metabolomics study, the experimental protocols must be optimized to reduce unwanted variation or bias in the collection of cell-free lysates. It is essential to ensure that any metabolome changes are limited to biologically relevant factors and are not caused by the handling or processing of the samples. Thus, the key to metabolomics is establishing an efficient methodology that closely captures the true state of the metabolome [53, 54]. Fundamental to a successful metabolomics experiment is maximizing the uniformity of the preparation, handling, processing, and analysis of each replicate sample [35, 45, 55-58]. In instances where cultivation and/or processing variation is unavoidable (*e.g.*, if multiple incubators are required to accommodate all the replicates), then the cultures should be randomly distributed between the incubators to minimize bias. Ideally, all of the metabolomics samples should be handled by the same person because subtle differences in individual techniques may influence the outcome. If multiple investigators are required to efficiently handle the samples, each researcher should be assigned a specific set of tasks that are consistently applied to each sample. For example, one investigator lyses all the bacterial cells while another performs the metabolome extraction procedure on every sample.

2.3 *Optimization of the number of replicates to maximize statistical significance*

As with sample cultivation and preparation methods, the NMR spectra generated from metabolomics samples need to accurately represent the state of the system. In other words, the NMR spectra must reflect the actual concentrations and identity of the metabolites present in the biological sample at the time of harvest. If the sample preparation and data acquisition represent the metabolic status at the time of harvest, then multivariate statistical techniques, such as PCA and OPLS-DA, will enhance the identification of similarities or differences in the NMR spectra, and, correspondingly, between the bacterial metabolomes [39]. These multivariate statistical techniques typically involve multiple replicates of 1D ^{1}H NMR or 2D ^{1}H,^{13}C HSQC spectra for each bacterial class or group (*e.g.,* wild-type, mutant, drug treated, *etc*.). The exact number of replicates is dependent on a number of factors: (i) the variance within a group, (ii) the variance between groups, (iii) the number of variables, and (iv) type of statistical analysis performed [59, 60].

In most metabolomics experiments, the number of biological samples is significantly smaller than the number of variables; in this case, the variables correspond to peaks in the NMR spectra or the detectable metabolites [60]. For this reason, a larger number of replicates ($\geq$ 6) per class are required to obtain a statistically significant PCA or OPLS-DA model. While greater numbers of replicates are desirable, there are practical considerations to increasing the number of replicates, including increased experimental time, availability of incubator space, and practical limits on the number of samples that can be simultaneously prepared and processed within a reasonable time frame. The increased time, larger number of samples, and added complexity may be detrimental to maintaining consistency between samples, where metabolite stability may become more of an issue [61]. So the potential benefit in improving the reliability of the PCA or OPLS-DA models may be negated by too large of a sample size if sample consistency is sacrificed. In general, 6 to 10 replicates per class can be routinely handled while providing a statistically significant PCA or OPLS-DA model. Lastly, to increase the sample consistency, the application of an automated sample changer or flow-probe can minimize variability by eliminating human involvement and providing a uniform and consistent protocol for NMR data collection.

3. Sample Preparation

3.1 *Techniques for consistent bacterial cultivation*

Consistency is critical to metabolomics, where variations in a bacterial metabolome may be introduced by cultivation protocols. To achieve the reproducible cultivation of bacteria requires consideration of three variables: bacterial strain, culture medium, and cultivation conditions. Strain selection is often driven by investigator preference, availability, or cultivability. The choice of culture medium will largely depend on which, if any, isotopically-labeled metabolite is being followed. For example, when using ^{15}N-arginine, it is impractical to add labeled arginine to a complex medium containing an unknown concentration of unlabeled arginine. In this example, to achieve maximal labeling of the bacteria, it would be best to use a chemically defined medium lacking arginine. Importantly, the culture medium has to be consistently employed throughout a specific metabolomics study of a defined set of bacterial strains. A different culture medium cannot be used for metabolomics fingerprinting and profiling, it cannot vary based on the requirements of the bacterial strain or to accommodate an experimental variable. For example, if a mutant bacterial strain requires the addition of a supplement for viability, then it is also necessary to add the same amount of the supplement to all other bacterial cultures in the study. Simply, any difference in the composition of the culture media will induce changes in the metabolome that will mask or complicate any analysis. Of course, the culture medium needs to be optimized for the specific requirements of each species and bacterial strain and, correspondingly, will likely vary between metabolomic studies. A metabolomics study employing *Staphylococcus epidermidis* will use a different culture medium than a study involving *Mycobacterium smegmatis*.

Bacterial cultures also need to be properly handled in order to avoid inducing biologically irrelevant changes. For example, pre-warming the culture medium prevents temperature shock and minimizes variation between biological replicates. Similarly, randomizing the samples from each group and class also minimizes bias that may occur if all the samples are processed in a predefined order. Importantly, different cell types may require special care or different handling protocols. Cultivation conditions will also vary depending upon the experiment; however, consideration must be given to each of the following: temperature, pH buffering (if used), % CO_2 (if used), the flask-to-medium ratio, the revolutions per minute of agitation (if used), the use of baffled or non-baffled flasks, and the inoculum dose. In effect, one protocol does not necessarily "fit all" and a general metabolomics protocol needs to be optimized for each experiment and cell type.

3.2 *Sample optimization to maximize NMR sensitivity*

Identifying an optimal culture size is an important next step in the design of a metabolomics study. The volume of the bacterial culture should be large enough to provide a sufficient number of cells to maximize the NMR signal-to-noise, but small enough to simplify the handling of numerous replicate samples. An appropriate cell density must be determined empirically for each species and bacterial strain, which will also limit the culture size. Similarly, the growth phase chosen for harvesting bacteria will also contribute to defining the optimal culture size since cell density changes drastically between the lag, exponential and stationary phases. In our experience with staphylococcal and mycobacteria cultures, media volumes between 15 to 50 mL are used to grow cells to an optical density at 600 nm ($O.D._{600nm}$) of 1-2 for bacterial cultures collected during the exponential phase. Conversely, media volumes of between 3 to 5 mL are used to grow cells to an $O.D._{600nm}$ of 3 to 7 for bacterial cultures collected during the stationary phase (e.g., 6 to 7 for *Staphylococcus epidermidis*, and 3 to 4 for *Mycobacterium smegmatis*). The overall goal is to have an $O.D._{600nm}$ of 10 to 20 after the bacterial cells have been concentrated to a final volume of 1 ml. This will ensure metabolite concentrations sufficient for detection by NMR. These culture volumes and $O.D._{600nm}$ values should be viewed as guidelines and targeted goals that may require further optimization for different bacterial strains or species.

Ideally, each bacterial culture should contain the same number of cells and be at the same growth phase when harvested. In reality, differences in cultivation conditions, media, and/or bacterial strains may substantially affect growth rates and/or growth yields. The two more common approaches to compensate for different bacterial growth rates are: collect the bacteria when they have reached the same cell density, but at different times to account for the different growth rates; and harvest the bacteria at the same time but harvest equivalent cell numbers. The number of bacterial cells in a given culture is estimated by measuring the culture turbidity using a standard optical density method. As examples, in staphylococci, the exponential and stationary growth phases were typically analyzed at the 2 h and 6 h time points, respectively [5]. For our mycobacterial experiments, a consistent growth phase was achieved by harvesting bacteria at a uniform $O.D._{600nm}$ of 1.2. In practice, it is extremely difficult to harvest every bacterial culture with an identical $O.D._{600nm}$ value. To correct for this variability, all the bacterial cultures are normalized to the same $O.D._{600nm}$ value. Simply, the cultures are suspended into a phosphate buffer until the $O.D._{600nm}$ values are equal. Alternatively, the bacterial cell cultures can be normalized based on colony-forming units (CFU), if OD-CFU calibration curves are available, or total protein concentration.

3.3 *Sample preparation protocols to maximize isotope labeling efficiency*

Metabolomics profiling requires ^{13}C- or ^{15}N-labeled metabolites and defines the choice of culture media. In our la-

boratories, we typically label staphylococci using ^{13}C-glucose in the complex medium tryptic soy broth (TSB) that is devoid of unlabeled glucose [6, 7]. This medium allows for maximal biomass generation, while assuring that nearly all (~99%; 1.1% is due to naturally occurring ^{13}C) of the ^{13}C-labeled metabolites in the metabolome were derived from glucose. Similarly, we have labeled mycobacteria using ^{13}C-glucose or ^{13}C-glycerol in Middlebrook 7H9 Albumin Dextrose Complex (MADC; Becton-Dickinson) media. We have also supplemented culture media with ^{13}C-alanine, ^{13}C-aspartate, ^{13}C-glutamate, ^{13}C-proline and ^{13}C-pyruvate as a more targeted approach to the analysis of the metabolome. These metabolites are associated with a limited number of metabolic pathways. The analysis of the metabolome can be further focused by using a targeted metabolite where only one or a few specific carbons in the metabolite are ^{13}C labeled. Only the metabolic pathways involving the specific ^{13}C -labeled carbon will be observable by NMR. The concentration of the ^{15}N-, or ^{13}C-labeled metabolite needs to be high enough (≥ 1 to 5 μM) to be detected by NMR. In our experience with staphylococcal and mycobacterial cultures, the volumes range from 25 mL to 100 mL and the culture media should be supplemented with approximately 2.5 to 4 g/L of $^{13}C_6$-glucose or ~10-15 mg/L of a targeted metabolite like ^{13}C -D-alanine in order to acquire a 2D ^{1}H-^{13}C HSQC spectrum with acceptable signal-to-noise.

3.4 *Protocols for determining an optimal drug dosage or administering other stress treatments*

To ensure consistency, the experimental variable such as a drug treatment, environmental stimuli, or gene knockout needs to be uniformly applied to the "treatment" class. An additional consideration for treatment of cultures, is that the impact on the metabolome should be strong enough to detect [49]. In other words, a particular drug dosage needs to be large enough to affect the cellular metabolome relative to untreated cells, but should not induce cell death. In our experience, a drug concentration that inhibits bacterial growth by 50% relative to untreated cultures is a desirable target [49, 50]. The availability from the literature of a minimal inhibitory concentration for the strain (MIC), or otherwise for the population isolates (MIC_{50}), provides a good starting point for optimizing a drug dosage, but the actual dosage must be determined empirically for each set of cultivation conditions. In our experience, literature MIC or MIC_{50} values tend to be too low for cultivation conditions used for metabolomics. We typically test drug concentration ranges at between 1 to 24 times the reported MIC or MIC_{50} values in order to identify an optimal drug dosage. Importantly, this also implies that drugs with a range of biological activity will require different drug concentrations in a metabolomics study; hence, the use of the 50% inhibition of growth is used as a metric as opposed to drug concentration. Typically, in our experiments the drug treatments are normally administered during the exponential phase and the bacteria are allowed to grow for at least one generation before harvesting. In our experience, this provides a sufficient amount of time for the drug to affect cell physiology and induce a perturbation in the metabolome. Administering a drug at an earlier time point can be problematic because of the inability to harvest enough bacteria.

3.5 *Quenching, washing and harvesting the bacterial cells*

Speed is critical to harvesting bacteria without inducing a change to the metabolome. Changes occur quickly because of different metabolite turnover-rates, varying stabilities, and the induction of stress responses, among other factors [61-63]. As bacteria are being harvested, the environment is changing dramatically: (i) the bacteria are either adhered to the surface of filter paper or at the bottom of a centrifuge tube under anaerobic conditions, (ii) the temperature is changed from 37°C to ~ 0°C, and (iii) the growth media is replaced with a phosphate buffer. To prevent perturbations to the metabolome caused by handling of the cell samples, the bacteria need to be rapidly quenched in order to stop all cellular processes from responding to these changes. Quenching efficiency has been widely discussed in the literature [42, 64-66]. Importantly, the quenching technique employed also defines the washing protocol and the order that quenching, washing and cell separation takes place. Our quenching techniques consist of filtered cells being quickly submerged into liquid nitrogen or the cells and media being directly added to -60°C cold ethanol or methanol solution while being vortexed. The media and ethanol/methanol volumes are at an equal 1:1 ratio. After centrifugation, the supernatant is decanted and disposed of, and the cell pellet is ready for washing. Unfortunately, there is a possibility of cell leakage and loss of metabolites when the cells are directly added to the cold ethanol or methanol solution.

Before intracellular metabolites can be analyzed, the bacteria need to be rapidly separated from the culture media. Filtration and centrifugation are both routinely used in our laboratory to separate bacterial cells from the media. Filtration has a definitive advantage because it is significantly faster than centrifugation, but challenges in removing and collecting intact cells from filter paper may lead to sample variability. Conversely, the variability between metabolome replicates is expected to be reduced with centrifugation because of the ease in handling the cells. Nevertheless, our experience with washing bacterial cells using either filtration or centrifugation has resulted in essentially identical metabolomics fingerprints (Figure 2a); thus, any undesirable variation within a group likely occurs during sample preparation.

The use of centrifugation or filtration also determines the quenching protocol [36]. Harvesting bacteria using centrifugation requires quenching the bacteria using the direct addition to -60°C cold ethanol or methanol. The bacteria, culture media, and quenching solution are in a properly sized conical centrifuge tube that is centrifuged for

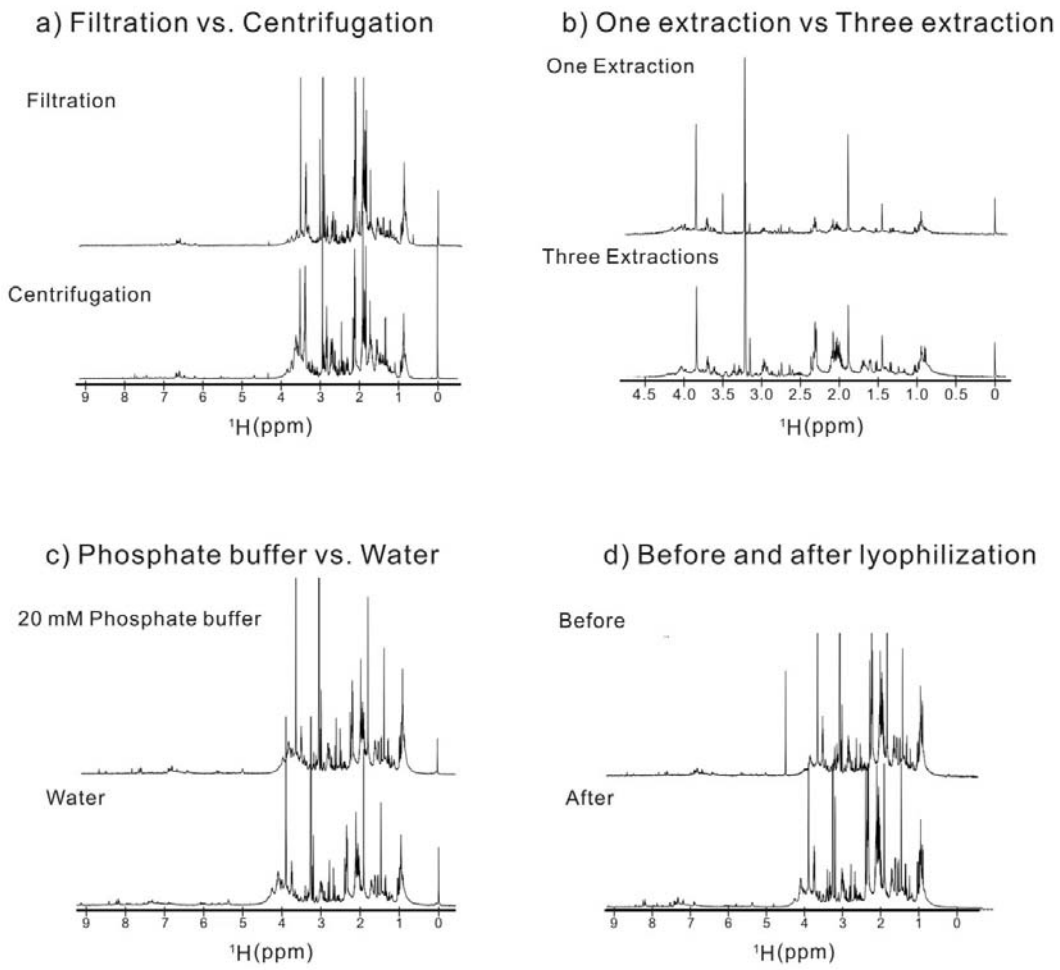


Figure 2. Illustrations of the impact of (a) filtration and centrifugation, (b) number of extraction steps, (c) type of wash buffer, and (d) lyophilization on the composition of the metabolome.

8 minutes at 4,284 *g* (bucket rotor) and ≤ 4°C. Following centrifugation, the culture media and quenching solution are decanted and the bacteria are suspended in 30 mL of an ice cold wash. We routinely wash bacteria with either a phosphate buffer (20 mM, pH 7.2), or phosphate buffered saline (PBS; 6 mM phosphate buffer, pH 7.4, 137 mM NaCl and 2.7 mM KCl) to remove residual media and avoid contamination of the metabolome. The buffered wash eliminates any impact on the cells from a pH or osmotic change that may lead to cell leakage and loss of metabolites. The bacteria are centrifuged again, the wash is decanted off and the process is repeated. After two washes, the cell pellet is suspended in 1 mL of the ice cold wash and transferred to a 2 mL vial for cell lysing. Additional washings provide an insignificant benefit in removing media contaminates, but results in an undesirable increase in time. The cells are kept on ice throughout this entire process.

Harvesting bacteria by vacuum filtration collects the bacteria on sterile filter paper (0.45 μm pore size; Millipore), while simultaneously removing the media. The number of bacteria that can be harvested onto a filter must be empirically determined to prevent a filter blockage. Under proper conditions, removing the media should take less than a minute, and should never exceed two minutes. If this cannot be achieved, then the bacteria need to be harvested using centrifugation. After filtration, the filter paper containing the cells is then quickly placed into a 50 mL conical centrifuge tube and submerged into liquid nitrogen to freeze and quench the cells. The conical vial is then warmed by placing it into a bucket of ice for ~1 to 2 minutes. This prevents freezing of the 1 mL of wash that is added to the conical vial. The cells are gently removed from the filter paper with the wash and then transferred to a 1.5 mL microcentrifuge tube. The cells are centrifuged and washed once (1 mL).

3.6 *Cell lysing and metabolite extraction*

The cells need to be lysed in order to extract the cellular metabolome. Cells can be lysed by chemical or physical

means, but the use of chemicals runs the added risk of contaminating the metabolome. Thus, the FAST-Prep bead beating method of lysing cells is our preferred approach. Each sample is placed into a 2 mL micro-centrifuge tube with small glass beads (Lysing Matrix B; MP Biomedical) and 1 mL of extraction buffer. The cells are crushed by bead beating for 40 to 60 seconds in the FAST-Prep instrument at a speed of 6.0 m/s. This process is repeated after keeping the crushed cells on ice for 5 minutes. The sample is then centrifuged for 2 minutes at 17,000 g to pellet the cell debris. The supernatant with the extracted metabolites is collected. The cell debris is washed 1 to 3 times with 1 mL of the extraction buffer to maximize the metabolome yield (Figure 2b). Also, double distilled water or a phosphate buffer are routinely used as the extraction buffer, since both approaches provide similar results (Figure 2c). All extracts per sample are combined for lyophilization, where the sample is then dissolved in 700 µL of a phosphate buffer in D_2O at pH 7.2 (uncorrected). Lyophilization may negatively impact some volatile metabolites, but, in general, no effect is observed (Figure 2d). A major concern during the extraction step is maximizing the overall yield while minimizing any perturbation to the metabolome. In our experience, the cell lysing and metabolite extraction process will require approximately 45 minutes for 30 cultures. The metabolomics samples can be stored in a -80°C freezer or directly lyophilized over-night.

4. NMR Spectroscopy

4.1 *One-dimensional ^{1}H NMR methodology*

One-dimensional (1D) ^{1}H (proton) NMR is an unbiased, nonselective, and nondestructive approach that requires no modification of the samples, where the data can be collected in a high-throughput manner. A 1D ^{1}H NMR spectrum contains numerous proton signals generated from a complex metabolomics mixture, where the chemical shift of each signal describes the structural characteristic of a specific metabolite. Moreover, the peak intensities or volumes are directly proportional to the concentration of each metabolite. Quantification of metabolites can be achieved by using an internal standard with a known concentration, where we routinely use 50 µM 3-(trimethylsilyl) propionic acid-2,2,3,3-d_4 (TMSP-d_4, Sigma). Thus, 1D ^{1}H NMR experiments combined with multivariate statistics are commonly used for the global analysis of the metabolome.

Collecting 1D ^{1}H NMR data for metabolomics is fast and simple, and provides highly reproducible and accurate results. Importantly, the NMR experimental parameters need to be identical for each metabolomics sample in order to collect reliable metabolomics data. Any per sample variation will erroneously bias the resulting clustering patterns in the PCA and OPLS-DA scores plot. To avoid this and maintain sample consistency, we use a BACS-120 sample changer, Bruker ICON-NMR, an automatic tuning and matching (ATM) unit, and autoshim to automate the NMR data collection. Nevertheless, instrument drift may still occur during the high-throughput metabolomics screen, so it is also important to randomize the samples during NMR data collection. If an NMR spectrum is collected first for all the control samples followed subsequently by each treatment class, there is a significant potential of inducing a biologically irrelevant bias into the analysis. The clustering pattern in the PCA and OPLS-DA scores plot may be dominated by the order of data collection instead of the expected biological differences.

In our laboratory, a typical 1D ^{1}H NMR spectrum is collected using 128 scans and 32k data points on a Bruker 500 MHz Avance DRX NMR spectrometer equipped with a triple-resonance, Z-axis gradient cryoprobe. The acquisition time is approximately 10 minutes per sample. The goal is to obtain optimal signal to noise while minimizing the total experimental time. We previously demonstrated that spectral noise is detrimental to the resulting PCA and OPLS-DA scores plot [55]. Random noise fluctuations results in large and irrelevant variations in the scores clustering. To avoid this problem, spectral noise needs to be removed prior to PCA and OPLS-DA. Correspondingly, the quality of the within class clustering in PCA and OPLS-DA scores plot is directly dependent on the spectral signal-to-noise (Figure 3). The within class variance decreases dramatically as the number of scans (signal-to-noise) is increased from right to left in the scores plot. Importantly, the spectral noise was still removed prior to PCA. Thus, the accuracy of identifying similarities or differences between multiple classes is dramatically improved by reducing within class variance, which is achieved by improving spectral sensitivity. Also, correctly identifying class differences improves with the number of replicates (Figure 4). The statistical significance of cluster separation as measured by *p*-value [67] is shown to decrease as both a function of group variance and the number of replicates. As a result, we prefer to use ten replicates per class and strive to achieve an average signal-to-noise ratio of > 100 to 200. This is achieved by simply increasing the number of scans or the number of cells, whichever is more practical. While signal-to-noise has a dramatic impact on scores clustering, PCA and OPLS-DA is indifferent to changes in spectral resolution unless the number of data points is ≤ 2K.

A D_2O phosphate buffer is the typical solvent of choice for aqueous metabolomics samples in order to efficiently remove residual water signals and avoid interference from buffer signals. Water and buffer signals are problematic since they can distort the NMR spectrum and may overlap and obscure important metabolite signals. Most NMR processing software can automatically remove the residual water peak, but extra data processing is required to correct for baseline distortions induced by the solvent. Unfortunately, simply applying a baseline correction changes the PCA and OPLS-DA clustering patterns [68]. Furthermore, different baseline correction protocols will induce variable changes into the scores plot. Also, removing the residual water peak may

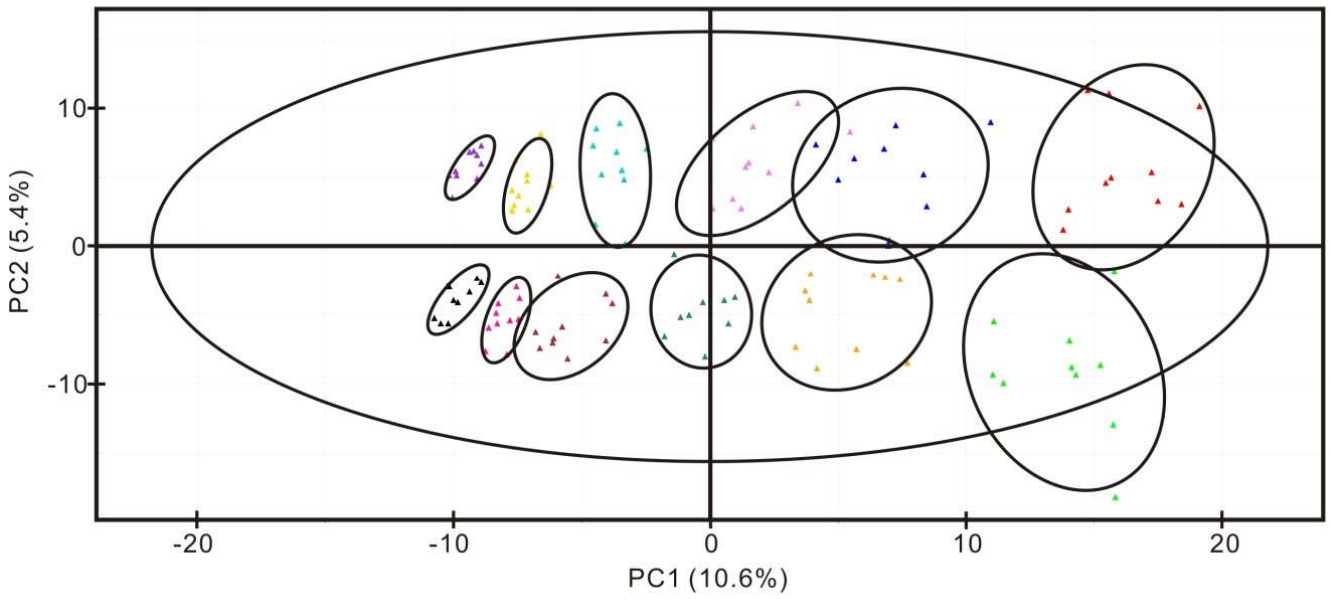


Figure 3. Illustration of the impact of the NMR signal-to-noise on within class variation in a PCA scores plot. From right to left, the 1D ^{1}H NMR spectra were collected with an increasing number of scans (1, 2, 4, 8, 16, and 32) resulting in a proportional increase in signal-to-noise. All other experimental parameters were kept constant.

result in a potential loss of information by also removing metabolite peaks near the water signal. Instead, a water suppression technique that experimentally removes the water peak without inducing baseline distortions is the preferred alternative.

There are a variety of NMR pulse sequences for water suppression that are available to the metabolomics community, such as WATERGATE, water pre-saturation, WET, and PURGE [69-73]. Our preferred choice for a water suppression pulse sequence is Solvent-Optimized Gradient-Gradient Spectroscopy (SOGGY). SOGGY does an outstanding job in eliminating the water signal without

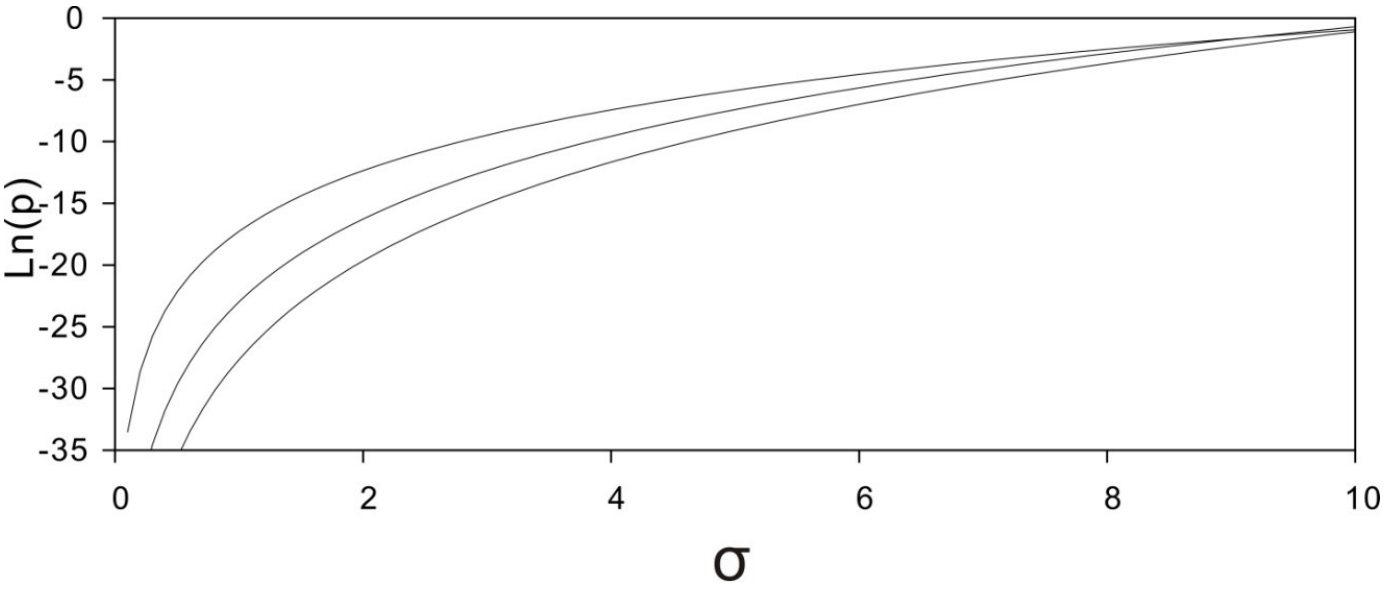


Figure 4. Illustration of the impact of within group variation and the number of replicates on the *p* values calculated between clusters in a simulated PCA scores plot. From top to bottom, *p* values from the simulated PCA scores plot were calculated with an increasing number of replicates (6, 8, 10) resulting in a proportional decrease in *p* values. Similarly, increasing the group variation by increasing the standard deviation (σ) per cluster resulted in a significant increase in *p* values.

inducing any base line distortions (Figure 2) [73]. SOGGY is a variant of excitation sculpting that employs a pulsed field gradient with a simple phase-alternating composite pulse. SOGGY offers the flexibility to optimize the 180 degree hard pulse to achieve optimal excitation of the water signal, and adjusting the 180 degree soft pulse to optimize the range of the water frequency to be suppressed [73, 74]. As a result, SOGGY efficiently suppresses the water signal while removing any phase cycle artifacts. A flat baseline is obtained while also maintaining metabolite signals near the water signal [73]. SOGGY completely eliminates the need to apply any baseline correction.

4.2 *Two-dimensional ^{1}H-^{13}C HSQC NMR methodology*

The severe overlap of signals in a 1D ^{1}H NMR spectrum is a challenge for metabolite identification. The difficulty arises because hundreds to thousands of peaks occupy a small chemical shift range (~10 ppm), where multiple metabolites share similar chemical shifts. Thus, we typically do not use 1D ^{1}H NMR spectra to assign metabolites. Instead, we routinely use 2D ^{1}H-^{13}C HSQC experiments for metabolite assignments. The 2D ^{1}H-^{13}C HSQC experiment is a more reliable approach for metabolite identification because of the significantly higher resolution and the correlation between ^{1}H and ^{13}C chemical shifts for each C-H pair in a molecule [75, 76]. Also, the 2D ^{1}H-^{13}C HSQC experiment simplifies the analysis of the metabolome because only compounds containing a ^{13}C-carbon derived from the ^{13}C-labeled metabolite added to the media will be detected.

In our laboratory, we use a standard Bruker 2D ^{1}H-^{13}C HSQC pulse sequence on our 500 MHz spectrometer, where an acceptable signal-to-noise is achievable using 64 scans. Similarly, a reasonable digital resolution is achieved by collecting 2K and 128 data points in the direct and indirect direction, respectively, with a corresponding spectral width of 10 ppm and 140 ppm along the ^{1}H and ^{13}C axis, respectively. Since some aromatic C-H pairs have a ^{13}C chemical shift greater than 140 ppm, the spectrum will contain folded peaks, but the folded peaks will not interfere with or overlap with other metabolite peaks due to their unique position along the ^{1}H axis (~ 7.0 ppm). This folding technique allows for an increase in the digital resolution without incurring an increase in acquisition time. In general, the 2D ^{1}H-^{13}C HSQC experiment requires approximately 4 hours per sample on our system.

A conventional 2D ^{1}H-^{13}C HSQC spectrum is useful for detecting metabolite changes by overlaying multiple spectra to identify missing peaks or peaks with significant intensity changes. Unlike 1D ^{1}H NMR spectra, obtaining metabolite concentrations is more difficult because peak intensities are dependent on J-couplings, dynamics and relaxation, in addition to metabolite concentrations [77, 78]. To quantify absolute metabolite concentrations, we use the Time-Zero HSQC ($HSQC_0$) experiment [77]. This approach requires collecting a series of three HSQCs spectra ($HSQC_1$, $HSQC_2$, $HSQC_3$) with an increased number of pulse sequence repetitions. A natural log plot of peak areas or intensities versus the increment number (1, 2, 3) allows for an extrapolation back to increment 0 or zero-time. The experimental parameters used in the $HSQC_0$ experiment is similar to the conventional method, but with some minor variations. The number of scans is increased to 128 due to the decrease in signal-to-noise in $HSQC_2$ and $HSQC_3$. To partially compensate for the increase in experimental time, the number of data points in the indirect dimension is reduced to 64. In general, the $HSQC_0$ set of experiment requires approximately 6 hours per sample on our system.

5. Data analysis

5.1 *Preprocessing of 1D ^{1}H NMR data*

The 1D ^{1}H NMR spectra are minimally processed (Fourier transformed and phase corrected) using ACD/1D NMR Manager (Advanced Chemistry Development). Each NMR metabolomics sample contains 50 μM of TMSP-d_4 as an internal standard, where each NMR spectrum is referenced to the TMSP-d_4 peak and uniformly aligned to 0.00 ppm. Also, all peak heights are normalized to the intensity of the TMSP-d_4 peak. Intelligent bucketing within the ACD/1D NMR Manager is then used to integrate each spectral region with a bin size of 0.025 ppm. The spectra are normalized; noise regions and residual solvent and buffer resonances are removed, and then the remaining bins are scaled prior to PCA and OPLS-DA using the commercial SIMCA12.0+ (UMETRICS) statistical package (http://www.umetrics.com/).

The need for data normalization and scaling prior to multivariate statistical analysis has been extensively discussed in the literature [79, 80]. Normalization adjusts for experimental variations between replicates, different number of cells, varying signal-to-noise, etc., and minimizes these contributions to the clustering patterns in PCA and OPLS-DA scores plot. We have encountered significant success in using a Z-score or center averaging the spectrum:

$$Z = \frac{X_i - \overline{X}}{\sigma} \quad (1)$$

where $\overline{X}$ is the average signal intensity in a given spectrum, σ is the standard deviation in the signal intensity, and X_i is the signal intensity within bin i (Figure 5a). After normalization, all the noise bins are uniformly removed. This was initially accomplished by manually identifying a "reference" noise region above 10 ppm and below 0 ppm; and calculating an average noise value. If a bin across all replicates had an integral value of less than twice the average noise, it was also identified as noise and removed (Figure 5b). The protocol for identifying noise regions has been recently improved upon and results in smaller within class variations (Figure 5c). This also results in an improved separation between truly distinct classes and removed erroneous separations. For

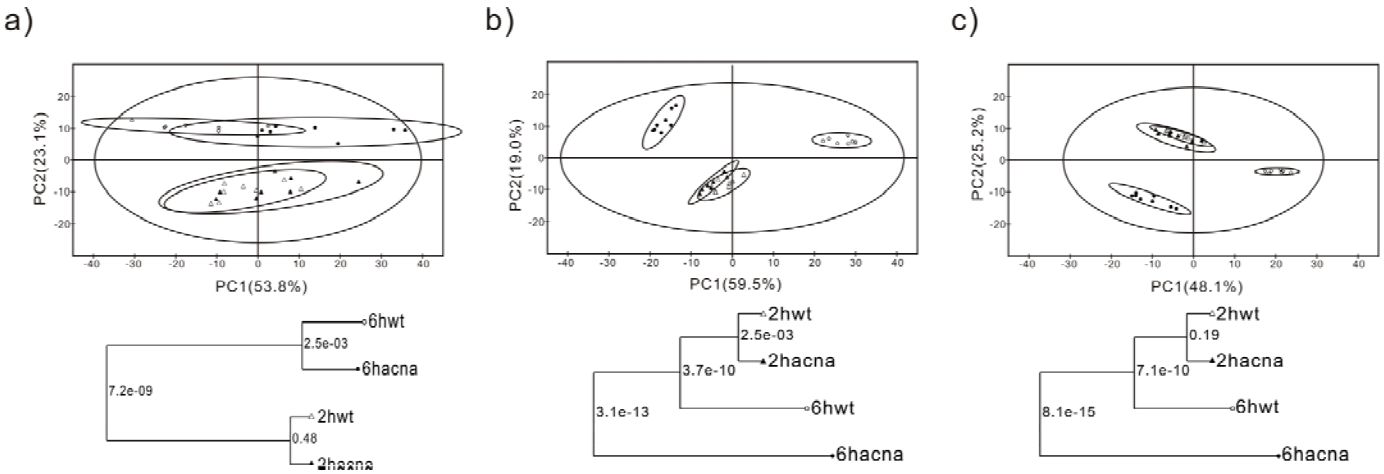


Figure 5. Illustration of the impact of NMR preprocessing on within and between class variations in a PCA scores plot. (a) The 1D ^{1}H NMR spectra was not properly preprocessed. The spectra were not normalized and the noise was not removed. The spectra were only Fourier transformed, phased corrected, and the residual H_2O resonance was removed. (b) The 1D ^{1}H NMR spectra were processed as in (a) with the addition of normalization using center averaging, but without noise removal. (c) The 1D ^{1}H NMR spectra were processed as in (b) with the addition of noise removal. Each spectrum was binned using intelligent bucketing with a bin size of 0.025 ppm. The ellipses correspond to the 95% confidence limits from a normal distribution for each cluster. The PCA scores plots compare the metabolomes of *S. aureus* wild-type (wt) strain SA564 with an aconitase mutant (acna) strain SA564-*acnA::tetM* at either two hours (2h) or six hours (6h) of cell growth. Below each PCA scores plot is a corresponding dendrogram generated from the scores using Mahalanobis distances, with *p* values for the null hypothesis reported at each branch.

example, the statistical significance between clusters 6hwt and 6hacna improved from a *p*-value of $3.1x10^{-13}$ to $8.1x10^{-15}$, while the small, but biologically irrelevant, separation between clusters 2hwt and 2hacna (*p*-value $2.5x10^{-3}$) was removed (Figure 5). Instead of using an average minimal signal intensity to define noise, we now define noise based on a relative standard deviation. This is based on the expectation that real NMR peaks from metabolites will have a higher intrinsic variability compared to the noise because of biological variations that naturally occur even between within class replicates. Conversely, the variability of the noise should be effectively constant for a given spectrometer operating within normal parameters. Simply, the standard deviation and average is calculated for each bin, where the standard deviation is normalized by the average peak intensity. This avoids eliminating weak peaks with a relatively small standard deviation. The same is done for the reference noise region, which is then used to define noise:

$$\textit{Noise}: \quad \sigma_i' \leq \sigma_0' \qquad (2)$$

$$\textit{Cutoff}: \quad \sigma_0' = \text{avg}(\sigma_n') + 2*\text{sd}(\sigma_n') \qquad (3)$$

where σ_i',σ_n' are the relative standard deviations (absolute standard deviation divided by the mean) for the *i*th bin in the spectral region and *n*th bin in the reference noise region, respectively, and avg(σ_n') and sd(σ_n') are the mean and standard deviation of σ_n' respectively. In effect, any peak that falls within the normal distribution of the reference noise region is defined as a noise bin. This approach is better at defining noise peaks in crowded and overlapping regions of the NMR spectra.

In addition to normalization, each bin or column in the data matrix also needs to be scaled to account for the large dynamic range in peak intensities. PCA and OPLS-DA emphasizes the absolute variation in bins between classes. Correspondingly, the relative variation of an intense peak may be insignificant compared to a weak peak, but the absolute changes in its intensity may completely mask biologically relevant changes in a small peak. Scaling increases the weight of the low intensity peaks so strong peaks do not dominate in PCA and OPLS-DA [79, 80]. In our experience, unit variance scaling, also known as autoscaling or a Z-score (see eqn. 1), has been shown to be effective in generating reliable clusters with the correct separation based on biologically relevant class distinctions. Also, within class variance is reduced using autoscaling, which is our default scaling method.

5.2 *Multivariate statistical analysis of 1D ^{1}H NMR data*

We routinely apply PCA, a non-supervised technique, to determine if the 1D ^{1}H NMR data can easily distinguish between the various test classes. PCA provides an unbiased view of group clustering in the resulting 2D scores plot. We only use a three-dimensional (3D) scores plot if class separation in a 2D scores plot is insufficient and the PC_3 contribution is significant (> 5 to 10%). OPLS-DA is only used if class separation is observed in the PCA scores plot. OPLS-DA is a supervised technique and assesses a relationship between the NMR data class designations. We limit OPLS-DA to only two class designations that

differentiate between the single control group (0) and the entire treatment group (1). As a supervised technique, OPLS -DA maximizes a separation between these two classified groups, while minimizing within class variations [39]. Thus, OPLS-DA identifies the important spectral features (metabolites) that primarily contribute to class separation. We routinely use an OPLS-DA S-plot or loading plot (Figure 1) to readily identify the key metabolites that contribute to class separation. Since OPLS-DA is a supervised technique and can generate a class separation even for random data [81], it is essential to verify the model [46]. But this is also an advantage of OPLS-DA over PCA; the statistical significance of the model is quantified. We cross-validate OPLS-DA models using a modified leave-one-out method [82, 83] and CV-ANOVA [84]. The modified leave-one-out method provides a quality assessment score (Q^2) and R^2 values, where CV-ANOVA provides a standard *p*-value. The theoretical maximum for Q^2 is 1, where a value of ≥ 0.4 is an empirically acceptable value for biological samples [85], but Q^2 does not have a critical value for inferring significance. It is still possible for an invalid model to produce a large Q^2 value. Similarly, the R^2 values only provide a measure of the fit of the data to the model. But large differences between Q^2 and R^2 ($R^2 >> Q^2$) does suggest an over-fit model. Conversely, a *p*-value << 0.05 from CV-ANOVA provides clear validation of the OPLS-DA model.

In addition to validating the OPLS-DA model, it is also extremely important to verify the statistical significance of the clustering patterns in the PCA and OPLS-DA scores plot. Is the between group difference larger than the within group variations? One key factor is the number of replicate samples. We have previously shown that increasing the number of replicates improves the statistical significance of cluster separation [86]. This finding is also supported by the increase in *p*-values seen with an increase in within class variations (Figure 4). Again, increasing the number of replicates improves the statistical significance of the class separation (lower *p*-value) even when within class variation increases. Correspondingly, we routinely use 10 replicates per group in our metabolomics study to improve the likelihood of observing statistical significant class separations.

It is also important to visually define each group or class within the PCA and OPLS-DA scores plot and to classify the statistical significance of the class separation. We developed a free PCA and OPLS-DA utilities software package (http://bionmr.unl.edu/pca-utils.php) [67] that draws ellipses or ellipsoids around each group cluster in a scores plot, where the ellipse corresponds to the 95% confidence limits from a normal distribution for each cluster. Visual separation of the ellipses infers a class separation. The same software package is also used to generate a metabolomics tree diagram based on the group clusters in the scores plot [67, 86]. Simply, a centroid from each cluster is used to calculate a Mahalanobis distance between clusters, where dendrograms are then generated from the resulting distance matrix. The significance of each node (cluster separation) is determined by using standard bootstrapping techniques and returning a bootstrap number [87], where a value above 50 infers a significant separation; or from Hotelling's T^2 and *F*-distributions that returns a *p*-value, where a number << 0.05 infers a statistically significant separation.

Observing a statistically significant difference in the global metabolome between two or more bacterial samples is typically the first objective of a metabolomics investigation. While this difference may infer some biological significance, the ultimate goal is to identify the underlying metabolites and associated pathways that are the primary contributors to the observed class separation in the PCA and OPLS-DA scores plot. One approach is to generate an S-plot (Figure 1) from the resulting OPLS-DA analysis. The S-plot identifies the key bins or ^{1}H chemical shifts that are correlated or anti-correlated with the separation between the two classes in an OPLS-DA scores plot. The ^{1}H chemical shifts can then be compared against a number of online NMR metabolomics databases [88-92] to assign the metabolites. Unfortunately, an unambiguous assignment is rarely possible because of the low chemical shift dispersion and the large number of potential metabolites. Instead, 2D NMR experiments combined with the biological knowledge of the system under investigation are required to improve the accuracy of metabolite identification.

5.3 *Metabolite Identification*

5.3.1 Automated peak picking of 2D NMR data

2D ^{1}H,^{13}C HSQC and ^{1}H,^{1}H TOCSY spectra are commonly used for metabolite identification because of the increase in chemical shift resolution achieved by spreading the information out into two-dimensions. Also, the correlation between ^{1}H chemical shifts for each J-coupled H pair; and the correlation between ^{1}H and ^{13}C chemical shifts for each C-H pair significantly reduces the assignment ambiguity. This occurs because both chemical shifts have to match a single metabolite in the database to make an assignment. Despite the advantages, peak picking and organizing a table of intensities from a 2D NMR experiment is a time consuming process, especially when multiple spectra are involved. Numerous software packages are available to automate the peak picking of 2D NMR spectra, however; it is extremely difficult, if not impossible, to align and match multiple sets of spectra with different peak patterns due to unique metabolomes.

For example, three different sets of cell cultures (different cell types, treatments or environmental conditions, etc.) will each exhibit a distinct set of peaks in the NMR spectrum due to the presence of unique metabolites. These unique peaks will be mixed with other peaks common to all three groups, but the relative peak intensities are likely to vary due to different metabolite concentrations. Thus, if the control group is designated as the reference spectrum for automated

peak picking, a peak list will be generated that only contains peaks observed in the control spectrum that are above the designated noise threshold. Correspondingly, peaks unique to the other two groups will be missed when this peak list is used to peak pick their spectrum. In addition, weak peaks may also be missed due to different noise levels between the spectra and a corresponding difference in the threshold setting for peak picking. Instead, a composite reference spectrum for automated peak picking needs to be generated that captures *all* the peaks present in the three separate groups. We accomplish this task by using the addNMR function in the free NMRpipe software package (http://spin.niddk.nih.gov/NMRPipe/) [93]. As the name implies, addNMR mathematically sums all spectra together from the three groups to make a single spectrum. This resulting "master spectrum" contains all the peaks observed throughout the set of 2D experiments and is used to generate a peak list for automated peak picking of each individual spectrum. Critically, the 2D NMR spectra need to be collected and processed using identical experimental parameters (spectral width, data points, zero-filling, etc.) and needs to be aligned to an internal reference (TMSP-d_4). In our experience, all the peaks from the complete set of NMR spectra are routinely matched to the reference list by using a chemical shift error-tolerance of 0.04 ppm and 0.25 ppm in the ^{1}H and ^{13}C dimensions, respectively. This approach has greatly simplified and increased the efficiency of a previously laborious procedure. The addNMR command can also be used to generate a difference spectrum that clearly highlights the major spectral changes between two classes (Figure 6).

5.3.2 Assignment of an NMR peak to a metabolite

Metabolite identification is an extremely important component of the metabolomics process because it enables the determination of the key metabolites perturbed by the treatment or the metabolites primarily contributing to class distinction. This includes the discovery of important biomarkers associated with drug efficacy or drug resistance. Also, metabolite identification is important to the drug

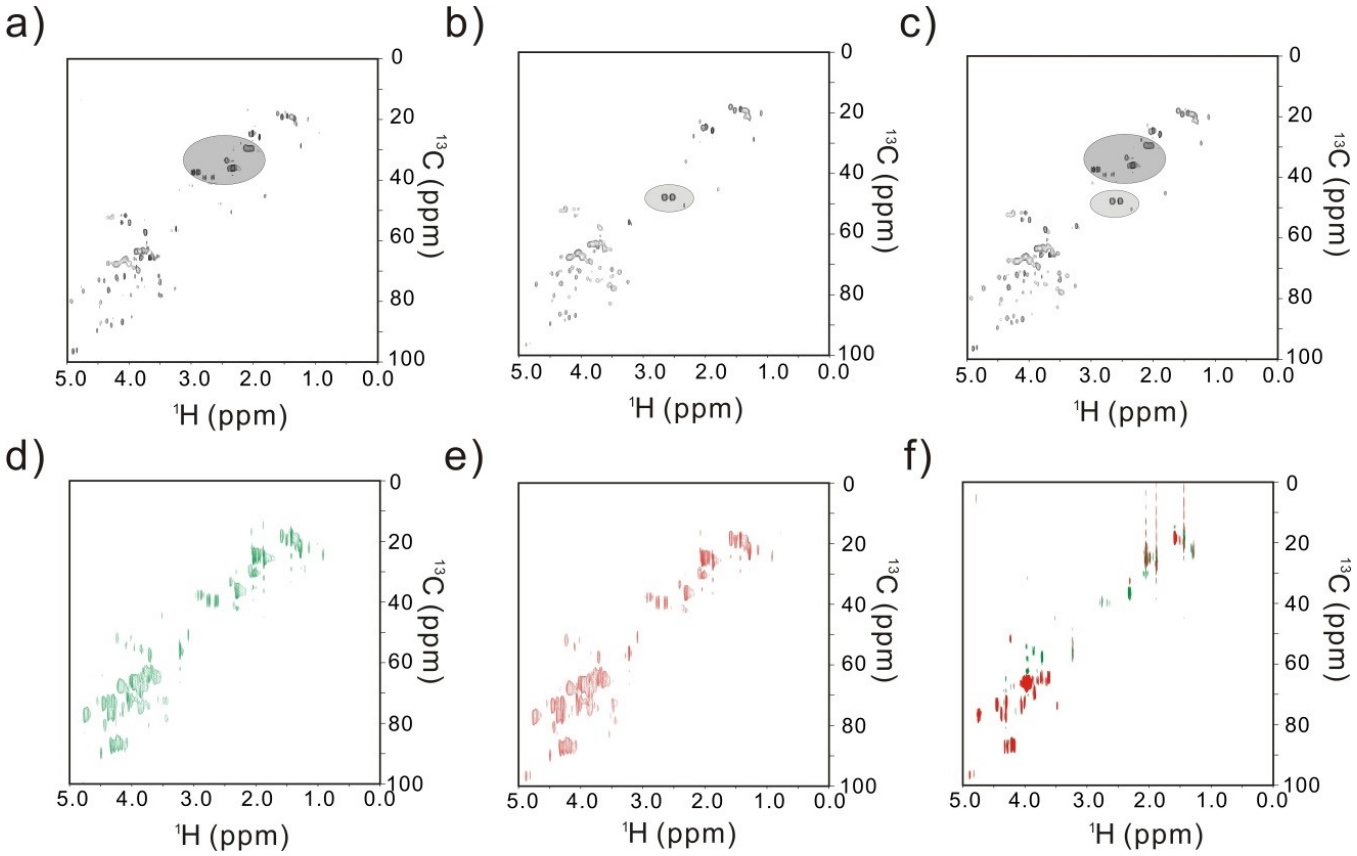


Figure 6. (a-c) Illustration of the procedure to generate a "master spectrum" and facilitate automated peak picking by creating a complete peak list. (a-b) Representative 2D ^{1}H-^{13}C HSQC spectra obtained from two distinct bacterial cultures, where some major spectral differences are highlighted. (c) The two 2D ^{1}H-^{13}C HSQC spectra from (a-b) were added to yield a master spectrum that contains all the observed NMR peaks. (d-f) Illustration of the procedure to generate a "difference spectrum" to facilitate metabolite identification by creating a signed (+, -, null) peak list. (d-e) Representative 2D ^{1}H-^{13}C HSQC spectra obtained from two distinct bacterial cultures. (f) The two 2D ^{1}H-^{13}C HSQC spectra from (d-e) were subtracted to yield a difference spectrum that identifies the NMR peaks, and correspondingly metabolites, that differ between the two bacterial cell cultures. Positive peaks, increased metabolite concentration, are colored green and negative peaks, decreased metabolite concentration, are colored red.

discovery process by either identifying metabolic pathways affected by a drug to evaluate efficacy or potential toxicity; or by identifying potentially new therapeutic targets. Nevertheless, accurate metabolite identification is very difficult and labor-intensive. The success of metabolite spectral assignment relies largely on the completeness of metabolomics databases. We routinely use a combination of the following freely-accessible databases: Human Metabolome Database (http://www.hmdb.ca/) [88], Madison Metabolomics Consortium Database (http://mmcd.nmrfam.wisc.edu/) [89], Platform for RIKEN Metabolomics (PRIME) (http://prime.psc.riken.jp/) [90], BioMagResBank (http://www.bmrb.wisc.edu/) [91], and Metabominer (http://wishart.biology.ualberta.ca/metabominer/) [92], which provide both redundant and complementary NMR spectral data. Importantly, the reference NMR spectra in the various databases were obtained under different buffer conditions and use different internal standards. This results in a range of potential chemical shifts for a given metabolite. Thus, the database with sample conditions that closely match our experimental conditions are used for chemical shift matching. The overall goal is to identify a complete set of metabolites as quickly and accurately as possible without any bias, by matching the experimental chemical shifts from the 2D NMR spectra with the values in the database.

For a 2D ^{1}H-^{13}C HSQC experiment, it is important to realize that metabolites may be heterogeneously labeled by the carbon-13 source present in the growth media. Correspondingly, all the peaks for a specific metabolite may not be detectable in the 2D ^{1}H-^{13}C HSQC experiment. Also, a reference spectrum for the metabolite may not be present in any of the available databases. The assignment of a particular peak might still be ambiguous because multiple metabolites may contain the same chemical shift or contain an identical substructure (*e.g.,* ATP, ADP, AMP or NAD, NADPH). Therefore, a few automated filters are applied to overcome some of these ambiguities during the peak assignment process.

The first filter is to verify that the bacteria can actually produce the proposed metabolite. This is routinely accomplished by searching the freely-accessible Biocyc (http://biocyc.org/) [94] and KEGG (http://www.genome.jp/kegg/) [95] database for metabolites known to exist in the bacteria under investigation. The second filter is based on a differential peak list. All the NMR peaks potentially assigned to a specific metabolite should have the same trend relative to the control. Obviously, the metabolite can only have one concentration and all the NMR peaks need to be consistent with this single concentration. Correspondingly, all the peaks have to be increased, decreased or the same relative to the same peaks in the control spectrum. This is easily and quickly visualized by subtracting the two sets of spectra and generating a signed (+, -, null) peak list. Peaks assigned to the same metabolite have to have the same sign. The third filter is based on a biological relationship with other metabolites. Simply, the likelihood of a correct assignment increases if other metabolites in a specific metabolic pathway have also been assigned. It is more likely to observe multiple metabolites from the same pathway than various metabolites from unrelated pathways. Similarly, if there is a direct metabolic path between two or more metabolites, then their assignments are more likely to be correct. The final filter is the application of our biological knowledge of the bacterial system under investigation. The pathways or metabolites that are expected to be perturbed by the treatment would be given precedent in the assignment process. As a simple example, a comparison between wild-type and mutant bacterial strains where aconitase has been inactivated would reasonably be expected to lead to changes in metabolites associate with the TCA cycle. Likewise, a comparison between untreated and drug-treated cells would be expected to lead to changes in metabolic pathways inhibited by the drug.

5.3.3 Statistical analysis of the 2D ^{1}H-^{13}C HSQC data.

After assigning the 2D ^{1}H-^{13}C HSQC spectra to a set of metabolites, the next goal is to determine metabolite concentration differences between the various bacterial culture conditions under investigation. Unfortunately, peak intensities in a standard 2D ^{1}H-^{13}C HSQC experiment are dependent on multiple parameters [77, 78], so only a relative percent change in a metabolite concentration can be determined [5]. Alternatively, an absolute concentration can be determined using $HSQC_0$, which requires a set of three HSQC experiments per sample. We routinely employ both approaches [77].

A relative difference in peak intensities is determined by using a triplicate set of a conventional 2D ^{1}H-^{13}C HSQC experiment for each bacterial culture condition. Prior to calculating a relative percent change in peak intensities, a detailed normalization process is required, which was previously described in detail [5]. First, the peak intensities within each spectrum are normalized by dividing each peak by the internal standard, the intensity of the TMSP-d_4 peak. Each peak pertaining to a specific chemical shift across each triplicate data set is then normalized by the most intense peak in the set of three peaks. Specifically, the maximal intensity for each peak across all data sets would be set to 100 and all other intensities are scaled relative to this peak intensity. Then all the normalized intensities for a given metabolite for each triplicate set are averaged together, and a relative percent error can be calculated between different cultures. A Student's t-test or ANOVA is then used to determine if the relative change in peak intensities is statistically significant at the 95% confidence limit. Calculating a relative difference in metabolite concentrations can be beneficial to understanding broader changes to the system, especially when a cluster of metabolites in a specific pathway exhibit a similar trend in concentration changes inferring an important role for the metabolic pathway.

Nevertheless, this approach is rather cumbersome and does not allow for a direct comparison between different metabolites.

Alternatively, we routinely use the $HSQC_0$ experiment to determine absolute metabolite concentrations. The overall protocol for the extrapolation of peak intensities to time-zero and the determination of the associated concentration has been previously described in detail [77, 78]. A distinct advantage of this method is that a single calibration curve can be made using multiple compounds with known concentrations to correlate the time-zero peak intensity with a concentration. Figure 7 illustrates such a calibration curve using 5 different mixtures, each consisting of 9 different ^{13}C-labeled metabolites ranging in concentrations from 5 to 300 μM. Also, the concentration for each metabolite was randomized in each mixture. For example the concentration of ^{13}C-D-alanine in the 5 mixtures is 300, 10, 25, 5, and 100 μM, respectively. The data was fitted using a weighted linear least squares calculation. Notably, the best-fit line (R^2 0.997) has a y-intercept close to zero as expected for a concentration of zero. Also, the correlation between peak volume and concentration is independent of the metabolite. Importantly, the accurate application of the calibration curve requires collecting and processing $HSQC_0$ spectra *identical* to the parameters used to obtain the original calibration curve. Critically, the receiver gain must be the same for all samples, because any change in the receiver gain influences the slope of the calibration curve. Also, the addition of 500 μM TMSP-d_4 as an internal standard is crucial, because both the calibration samples and experimental samples must both be normalized to the TMSP-d_4 peak. As an example, if the TMSP-d_4 peak volumes for the calibration mixtures are 1000, 500, and 250 for each $HSQC_i$ (i = 1, 2, 3) spectrum, respectively, then the experimental results for all *in vivo* metabolite extracts must be normalized so that the internal standard (TMSP-d_4) peak volumes are also 1000, 500, and 250. The concentrations are measured in triplicate, where a Student's t-test or ANOVA is used to determine if the concentration changes are statistically significant at the 95% confidence limit.

5.4 *Metabolomics Network Map*

Metabolites are highly interconnected through numerous metabolic pathways that form an extremely complex network [96]. Correspondingly, it is not uncommon to observe correlated changes between distantly connected metabolites. In effect, metabolomics depends on these complex interactions to understand the phenotype of a bacterial cell. Thus, a metabolomics network map provides an efficient approach to visualize and summarize the overall changes to the metabolome, to validate metabolite assignments based on clear connections to other metabolites, and the identification of key metabolic pathways.

We have routinely used Cytoscape (http://www.cytoscape.org/) to easily and quickly generate metabolomics network maps [97-100]. Cytoscape is a free, user-friendly software package with plug-ins related to metabolomics. Cytoscape simply requires a list of the metabolites and their associated concentration changes as input. The connections between nodes (metabolites) in the map are based on metabolic pathways from the freely-

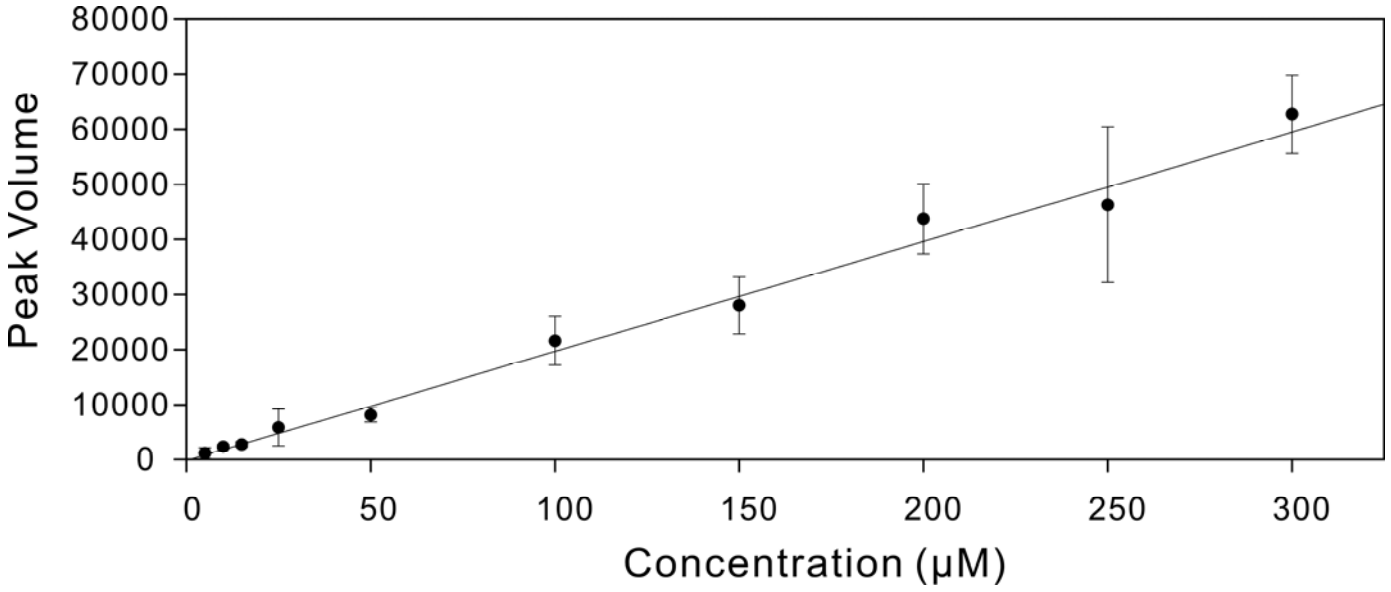


Figure 7. A strong correlation between NMR peak volumes and metabolite concentrations (R^2 0.997) is demonstrated by linear regression plot generated from $HSQC_0$ data. $HSQC_0$ NMR spectra were collected for five different metabolite mixtures containing nine ^{13}C-labeled compounds with concentrations ranging from 5 μM to 300 μM. The relationship between peak volume and metabolite concentration is independent of the metabolite.

accessible MetaCyc database (http://metacyc.org/) [101]. An example of a typical Cytoscape map summarizing the observed changes in the *S. epidermidis* metabolome caused by environmental stimuli associated with biofilm formation is shown in Figure 1. The metabolomics network map can be easily modified to highlight specific features of the metabolome. Edges can be broadened to highlight specific pathways; and the color and size of nodes can be adjusted to reflect the direction and magnitude of the concentration changes, respectively [102]. Cytoscape also provides a range of map design choices. Unfortunately, the resulting network maps (Figure 1) do not resemble standard metabolic pathways. Thus, Cytoscape maps are simply used as a template to manually draw more traditional looking metabolic pathways. Since Cytoscape maps are so easily generated, we also use the software to assist in metabolite assignments. Potential lists of metabolite assignments are input into Cytoscape to identify metabolites that are isolated nodes excluded from the main network map. These metabolites are likely misassigned and are reevaluated. In addition to Cytoscape, we also use the free R statistics package (http://www.r-project.org/) [103] to create heat maps from absolute metabolite concentrations or percent relative concentration changes.

6. Conclusion

NMR metabolomics is an invaluable tool for systems biology and its application is rapidly expanding. Global changes in the metabolic state of bacterial cells occur as a result of environmental stressors, genetic modifications, drug treatments, or numerous other factors. A detailed analysis of the differences in the NMR spectra is commonly used to identify the key metabolite changes that differentiate between these bacterial classes (*e.g.*, controls versus treated). In addition, metabolite identification by NMR allows for the subsequent identification of the important metabolic pathways that are affected by the treatment, providing further insight into the underlying biological process. The appeal of NMR metabolomics is its simplicity, but unfortunately it is also easy to obtain unreliable results. The observed changes in the metabolome should be biologically relevant, but because the metabolome is so sensitive to any environmental change; it is also easily perturbed by the experimental protocol. This is clearly an undesirable outcome. To address this issue, we described in detail our optimized protocols for the NMR analysis of bacterial metabolomes. We also highlighted common problems and potential sources of mistakes. We discussed the entire process that includes growing and harvesting bacterial cells, extracting the metabolome, NMR data collection, processing and analysis, statistical analysis, metabolite and network identification. The protocols described have been successfully applied to a number of systems biology projects [5, 49, 50, 52, 104-106].

Acknowledgements

This manuscript was supported in part by funds from the National Institute of Health (AI087668, P20 RR-17675, P30 GM103335), the American Heart Association (0860033Z), and the Nebraska Research Council. S. Halouska and R. J. Fenton were partially supported by O. Chacon's NIH grant (R21 AI087561) to standardize NMR techniques included in this publication. The research was performed in facilities renovated with support from the National Institutes of Health (RR015468-01).

References

[1] L.F. Shyur, N.S. Yang, Curr. Opin. Cell Biol., 12 (2008) 66-71. 10.1016/j.cbpa.2008.01.032.

[2] J.L. Spratlin, N.J. Serkova, S.G. Eckhardt, Clin Cancer Res, 15 (2009) 431-440. 10.1158/1078-0432.CCR-08-1059.

[3] S. Rochfort, J. Nat. Prod., 68 (2005) 1813-1820. 10.1021/np050255w.

[4] L.M. Raamsdonk, B. Teusink, D. Broadhurst, N. Zhang, A. Hayes, M.C. Walsh, J.A. Berden, K.M. Brindle, D.B. Kell, J.J. Rowland, H.V. Westerhoff, K. van Dam, S.G. Oliver, Nat. Biotechnol., 19 (2001) 45-50. 10.1038/83496.

[5] B. Zhang, S. Halouska, C.E. Schiaffo, M.R. Sadykov, G.A. Somerville, R. Powers, J Proteome Res, 10 (2011) 3743-3754. 10.1021/pr200360w.

[6] J.C. Ewald, T. Matt, N. Zamboni, Mol. Biosyst., 9 (2013) 440-446. 10.1039/c2mb25423a.

[7] O. Fiehn, Plant Mol. Biol., 48 (2002) 155-171.

[8] R. Kaddurah-Daouk, B.S. Kristal, R.M. Weinshilboum, Annu. Rev. Pharmacol. Toxicol., 48 (2008) 653-683. 10.1146/annurev.pharmtox.48.113006.094715.

[9] H.C. Keun, Pharmacol. Ther., 109 (2006) 92-106. 10.1016/j.pharmthera.2005.06.008.

[10] W. Zhang, F. Li, L. Nie, Microbiology, 156 (2010) 287-301. 10.1099/mic.0.034793-0.

[11] T. Ideker, V. Thorsson, J.A. Ranish, R. Christmas, J. Buhler, J.K. Eng, R. Bumgarner, D.R. Goodlett, R. Aebersold, L. Hood, Science, 292 (2001) 929-934. 10.1126/science.292.5518.929.

[12] P. Pir, B. Kirdar, A. Hayes, Z.Y. Onsan, K.O. Ulgen, S.G. Oliver, BMC Bioinf., 7 (2006) 203. 10.1186/1471-2105-7-203.

[13] A.C. Mosier, N.B. Justice, B.P. Bowen, R. Baran, B.C. Thomas, T.R. Northen, J.F. Banfield, mBio, 4 (2013). 10.1128/mBio.00484-12.

[14] M.R. Sadykov, B. Zhang, S. Halouska, J.L. Nelson, L.W. Kreimer, Y.F. Zhu, R. Powers, G.A. Somerville, J Biol Chem, 285 (2010) 36616-36624. DOI 10.1074/jbc.M110.152843.

[15] S.H. Yoon, M.J. Han, H. Jeong, C.H. Lee, X.X. Xia, D.H. Lee, J.H. Shim, S.Y. Lee, T.K. Oh, J.F. Kim, Genome Biol., 13 (2012) R37. 10.1186/gb-2012-13-5-r37.

[16] R. Yang, Z. Du, Y. Han, L. Zhou, Y. Song, D. Zhou, Y. Cui, Front. Cell Infect. Microbiol., 2 (2012) 157. 10.3389/fcimb.2012.00157.

[17] G.P. Bisson, C. Mehaffy, C. Broeckling, J. Prenni, D. Rifat, D.S. Lun, M. Burgos, D. Weissman, P.C. Karakousis, K. Dobos, J. Bacteriol., 194 (2012) 6441-6452. 10.1128/JB.01013-12.

[18] I. Nobeli, H. Ponstingl, E.B. Krissinel, J.M. Thornton, J. Mol. Biol., 334 (2003) 697-719. 10.1016/j.jmb.2003.10.008.

[19] W. Weckwerth, Bioanalysis, 2 (2010) 829-836. 10.4155/bio.09.192.

[20] R. Powers, Magn Reson Chem, 47 Suppl 1 (2009) S2-11. 10.1002/mrc.2461.

[21] S. Moco, R.J. Bino, R.C.H. De Vos, J. Vervoort, Trac-Trends in Anal. Chem., 26 (2007) 855-866. DOI 10.1016/j.trac.2007.08.003.

[22] S.G. Villas-Boas, S. Mas, M. Aakesson, J. Smedsgaard, J. Nielsen, Mass Spectrom. Rev., 24 (2005) 613-646. 10.1002/mas.20032.

[23] I.D. Wilson, R. Plumb, J. Granger, H. Major, R. Williams, E.M. Lenz, J. Chromatogr., B: Anal. Technol. Biomed. Life Sci., 817 (2005) 67-76. 10.1016/j.jchromb.2004.07.045.

[24] N.V. Reo, Drug Chem. Toxicol., 25 (2002) 375-382. 10.1081/dct-120014789.

[25] T.W.M. Fan, A.N. Lane, J. Biomol. NMR, 49 (2011) 267-280. 10.1007/s10858-011-9484-6.

[26] I.A. Lewis, R.H. Karsten, M.E. Norton, M. Tonelli, W.M. Westler, J.L. Markley, Anal. Chem., 82 (2010) 4558-4563. 10.1021/ac100565b.

[27] D.B. Kell, Curr. Opin. Microbiol., 7 (2004) 296-307. 10.1016/j.mib.2004.04.012.

[28] T.O. Metz, J.S. Page, E.S. Baker, K. Tang, J. Ding, Y. Shen, R.D. Smith, TrAC, Trends Anal. Chem., 27 (2008) 205-214. 10.1016/j.trac.2007.11.003.

[29] Z. Pan, D. Raftery, Anal. Bioanal. Chem., 387 (2007) 525-527. 10.1007/s00216-006-0687-8.

[30] M.R. Viant, E.S. Rosenblum, R.S. Tjeerdema, Environ. Sci. Technol., 37 (2003) 4982-4989. 10.1021/es034281x.

[31] O. Beckonert, H.C. Keun, T.M.D. Ebbels, J. Bundy, E. Holmes, J.C. Lindon, J.K. Nicholson, Nat. Protoc., 2 (2007) 2692-2703. 10.1038/nprot.2007.376.

[32] S. Dietmair, N.E. Timmins, P.P. Gray, L.K. Nielsen, J.O. Krömer, Anal. Biochem., 404 (2010) 155-164. 10.1016/j.ab.2010.04.031.

[33] H.K. Kim, Y.H. Choi, R. Verpoorte, Nat. Protoc., 5 (2010) 536-549. 10.1038/nprot.2009.237.

[34] R. Marcinowska, J. Trygg, H. Wolf-Watz, T. Mortiz, I. Surowiec, J Microbiol Methods, 87 (2011) 24-31. 10.1016/j.mimet.2011.07.001.

[35] M. Mashego, K. Rumbold, M. De Mey, E. Vandamme, W. Soetaert, J. Heijnen, Biotechnol. Lett., 29 (2007) 1-16. 10.1007/s10529-006-9218-0.

[36] X.H. Wu, H.L. Yu, Z.Y. Ba, J.Y. Chen, H.G. Sun, B.Z. Han, Biotechnol J, 5 (2010) 75-84. 10.1002/biot.200900038.

[37] M. Cuperlovic-Culf, D.A. Barnett, A.S. Culf, I. Chute, Drug Discov. Today, 15 (2010) 610-621. 10.1016/j.drudis.2010.06.012.

[38] Z. Ramadan, D. Jacobs, M. Grigorov, S. Kochhar, Talanta, 68 (2006) 1683-1691. 10.1016/j.talanta.2005.08.042.

[39] M. Bylesjo, M. Rantalainen, O. Cloarec, J.K. Nicholson, E. Holmes, J. Trygg, J. Chemom., 20 (2006) 341-351. Doi 10.1002/Cem.1006.

[40] A.R. Fernie, R.N. Trethewey, A.J. Krotzky, L. Willmitzer, Nat. Rev. Mol. Cell Biol., 5 (2004) 763-769. Doi 10.1038/Nrm1451.

[41] C.J. Clarke, J.N. Haselden, Toxicol. Pathol., 36 (2008) 140-147. 10.1177/0192623307310947.

[42] M. Faijes, A.E. Mars, E.J. Smid, Microb. Cell Fact., 6 (2007) 27. 10.1186/1475-2859-6-27.

[43] M.F. Van Batenburg, L. Coulier, F. van Eeuwijk, A.K. Smilde, J.A. Westerhuis, Anal. Chem., 83 (2011) 3267-3274. 10.1021/ac102374c.

[44] H. Kanani, P.K. Chrysanthopoulos, M.I. Klapa, J. Chromatogr. B, 871 (2008) 191-201. 10.1016/j.jchromb.2008.04.049.

[45] O. Teahan, S. Gamble, E. Holmes, J. Waxman, J.K. Nicholson, C. Bevan, H.C. Keun, Anal. Chem., 78 (2006) 4307-4318. 10.1021/ac051972y.

[46] B. Worley, R. Powers, Current Metabolomics, 1 (2013) 92-107. 10.2174/2213235X130108.

[47] C. Birkemeyer, A. Luedemann, C. Wagner, A. Erban, J. Kopka, Trends Biotechnol, 23 (2005) 28-33. 10.1016/j.tibtech.2004.12.001.

[48] T.M. Fan, L. Bandura, R. Higashi, A. Lane, Metabolomics, 1 (2005) 325-339. 10.1007/s11306-005-0012-0.

[49] S. Halouska, R.J. Fenton, R.G. Barletta, R. Powers, ACS Chem. Biol., 7 (2012) 166-171. 10.1021/Cb200348m.

[50] S. Halouska, O. Chacon, R.J. Fenton, D.K. Zinniel, R.G. Barletta, R. Powers, J. Proteome Res., 6 (2007) 4608-4614. 10.1021/Pr0704332.

[51] P. Forgue, S. Halouska, M. Werth, K.M. Xu, S. Harris, R. Powers, J. Proteome Res., 5 (2006) 1916-1923. 10.1021/Pr060114v.

[52] M.R. Sadykov, M.E. Olson, S. Halouska, Y. Zhu, P.D. Fey, R. Powers, G.A. Somerville, J. Bacteriol., 130 (2008) 7621-7632. 10.1128/JB.00806-08.

[53] K. Suhre, C. Gieger, Nat. Rev. Genet., 13 (2012) 759-769. 10.1038/Nrg3314.

[54] D.K. Zinniel, R.J. Fenton, S. Halouska, R. Powers, R.G. Barletta, J. Vis. Exp., 67 (2012) e3673.

[55] S. Halouska, R. Powers, J. Magn. Reson., 178 (2006) 88-95. 10.1016/j.jmr.2005.08.016.

[56] C.J. Bolten, P. Kiefer, F. Letisse, J.C. Portais, C. Wittmann, Anal. Chem., 79 (2007) 3843-3849. 10.1021/ac0623888.

[57] N.J.C. Bailey, M. Oven, E. Holmes, M.H. Zenk, J.K. Nicholson, Spectrosc. Int. J., 18 (2004) 279-287. 10.1155/2004/862164.

[58] M. Defernez, I.J. Colquhoun, Phytochemistry, 62 (2003) 1009-1017. 10.1016/S0031-9422(02)00704-5.

[59] D.I. Broadhurst, D.B. Kell, Metabolomics, 2 (2006) 171-196. 10.1007/s11306-006-0037-z.

[60] R.G. Brereton, Trac-Trends in Anal. Chem., 25 (2006) 1103-1111. 10.1016/j.trac.2006.10.005.

[61] A. Canelas, C. Ras, A. ten Pierick, J. van Dam, J. Heijnen, W. van Gulik, Metabolomics, 4 (2008) 226-239. 10.1007/s11306-008-0116-4.

[62] E.J. Saude, B.D. Sykes, Metabolomics, 3 (2007) 19-27. 10.1007/s11306-006-0042-2.

[63] A.M. Giuliodori, C.O. Gualerzi, S. Soto, J. Vila, M.M. Tavio, Ann. N. Y. Acad. Sci., 1113 (2007) 95-104. 10.1196/annals.1391.008.

[64] F. Schadel, F. David, E. Franco-Lara, Appl. Microbiol. Biotechnol., 92 (2011) 1261-1274. DOI 10.1007/s00253-011-3377-1.

[65] M. Wellerdiek, D. Winterhoff, W. Reule, J. Brandner, M. Oldiges, Bioprocess Biosyst. Eng., 32 (2009) 581-592. DOI 10.1007/s00449-008-0280-y.

[66] C. Wittmann, J.O. Kromer, P. Kiefer, T. Binz, E. Heinzle, Anal. Biochem., 327 (2004) 135-139. 10.1016/J.Ab.2004.01.002.

[67] B. Worley, S. Halouska, R. Powers, Anal. Biochem., 433 (2013)

102-104. 10.1016/j.ab.2012.10.011.

[68] B.C.M. Potts, A.J. Deese, G.J. Stevens, M.D. Reily, D.G. Robertson, J. Theiss, J. Pharm. Biomed. Anal., 26 (2001) 463-476. 10.1016/s0731-7085(01)00430-7.

[69] D.I. Hoult, J. Magn. Reson., 21 (1976) 337-347. 10.1016/0022-2364(76)90081-0.

[70] V. Sklenar, M. Piotto, R. Leppik, V. Saudek, J. Magn. Reson. Series A, 102 (1993) 241-245. 10.1006/jmra.1993.1098.

[71] M.L. Liu, X.A. Mao, C.H. Ye, H. Huang, J.K. Nicholson, J.C. Lindon, J. Magn. Reson., 132 (1998) 125-129. 10.1006/jmre.1998.1405.

[72] R.J. Ogg, P.B. Kingsley, J.S. Taylor, J. Magn. Reson. Series B, 104 (1994) 1-10. 10.1006/jmrb.1994.1048.

[73] A.J. Simpson, S.A. Brown, J. Magn. Reson., 175 (2005) 340-346. 10.1016/j.jmr.2005.05.008.

[74] T.L. Hwang, A.J. Shaka, J. Magn. Reson. Series A, 112 (1995) 275-279. 10.1006/jmra.1995.1047.

[75] J. McKenzie, A. Charlton, J. Donarski, A. MacNicoll, J. Wilson, Metabolomics, 6 (2010) 574-582. 10.1007/s11306-010-0226-7.

[76] Y. Xi, J.S. de Ropp, M.R. Viant, D.L. Woodruff, P. Yu, Anal. Chim. Acta, 614 (2008) 127-133. 10.1016/j.aca.2008.03.024.

[77] K.F. Hu, W.M. Westler, J.L. Markley, J. Amer. Chem. Soc., 133 (2011) 1662-1665. 10.1021/Ja1095304.

[78] K.F. Hu, J.J. Ellinger, R.A. Chylla, J.L. Markley, Anal. Chem., 83 (2011) 9352-9360. 10.1021/Ac201948f.

[79] R.A. van den Berg, H.C. Hoefsloot, J.A. Westerhuis, A.K. Smilde, M.J. van der Werf, BMC genomics, 7 (2006) 142. 10.1186/1471-2164-7-142.

[80] A. Craig, O. Cloareo, E. Holmes, J.K. Nicholson, J.C. Lindon, Anal. Chem., 78 (2006) 2262-2267. 10.1021/Ac0519312.

[81] K. Kjeldahl, R. Bro, J. Chemom., 24 (2010) 558-564. 10.1002/cem.1346.

[82] J. Shao, J. Am. Stat. Assoc., 88 (1993) 486-494. 10.1080/01621459.1993.10476299.

[83] A. Golbraikh, A. Tropsha, J. Mol. Graphics Modell., 20 (2002) 269-276. 10.1016/S1093-3263(01)00123-1.

[84] L. Eriksson, J. Trygg, S. Wold, J. Chemom., 22 (2008) 594-600. 10.1002/cem.1187.

[85] J. Westerhuis, H. Hoefsloot, S. Smit, D. Vis, A. Smilde, E. van Velzen, J. van Duijnhoven, F. van Dorsten, Metabolomics, 4 (2008) 81-89. 10.1007/s11306-007-0099-6.

[86] M.T. Werth, S. Halouska, M.D. Shortridge, B. Zhang, R. Powers, Anal. Biochem., 399 (2010) 58-63. 10.1016/j.ab.2009.12.022.

[87] A.R. Henderson, Clin. Chim. Acta, 359 (2005) 1-26. 10.1016/j.cccn.2005.04.002.

[88] D.S. Wishart, C. Knox, A.C. Guo, R. Eisner, N. Young, B. Gautam, D.D. Hau, N. Psychogios, E. Dong, S. Bouatra, R. Mandal, I. Sinelnikov, J.G. Xia, L. Jia, J.A. Cruz, E. Lim, C.A. Sobsey, S. Shrivastava, P. Huang, P. Liu, L. Fang, J. Peng, R. Fradette, D. Cheng, D. Tzur, M. Clements, A. Lewis, A. De Souza, A. Zuniga, M. Dawe, Y.P. Xiong, D. Clive, R. Greiner, A. Nazyrova, R. Shaykhutdinov, L. Li, H.J. Vogel, I. Forsythe, Nucleic Acids Res., 37 (2009) D603-D610. 10.1093/Nar/Gkn810.

[89] Q. Cui, I.A. Lewis, A.D. Hegeman, M.E. Anderson, J. Li, C.F. Schulte, W.M. Westler, H.R. Eghbalnia, M.R. Sussman, J.L. Markley, Nat. Biotechnol., 26 (2008) 162-164. 10.1038/Nbt0208-162.

[90] K. Akiyama, E. Chikayama, H. Yuasa, Y. Shimada, T. Tohge, K. Shinozaki, M.Y. Hirai, T. Sakurai, J. Kikuchi, K. Saito, In silico biology, 8 (2008) 339-345.

[91] E.L. Ulrich, H. Akutsu, J.F. Doreleijers, Y. Harano, Y.E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C.F. Schulte, D.E. Tolmie, R. Kent Wenger, H. Yao, J.L. Markley, Nucleic Acids Res., 36 (2008) D402-408. 10.1093/nar/gkm957.

[92] J. Xia, T.C. Bjorndahl, P. Tang, D.S. Wishart, BMC Bioinf., 9 (2008) 507. 10.1186/1471-2105-9-507.

[93] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, J. Biomol. NMR, 6 (1995) 277-293. 10.1007/BF00197809.

[94] R. Caspi, T. Altman, K. Dreher, C.A. Fulcher, P. Subhraveti, I.M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, L.A. Mueller, Q. Ong, S. Paley, A. Pujar, A.G. Shearer, M. Travers, D. Weerasinghe, P. Zhang, P.D. Karp, Nucleic Acids Res., 40 (2012) D742-753. 10.1093/nar/gkr1014.

[95] M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, Y. Yamanishi, Nucleic Acids Res., 36 (2008) D480-D484. 10.1093/nar/gkm882.

[96] M. Kohlstedt, J. Becker, C. Wittmann, Appl. Microbiol. Biotechnol., 88 (2010) 1065-1075. 10.1007/s00253-010-2854-2.

[97] J.J. Bot, M.J. Reinders, Bioinformatics, 27 (2011) 2451-2452. 10.1093/bioinformatics/btr388.

[98] J. Gao, V.G. Tarcea, A. Karnovsky, B.R. Mirel, T.E. Weymouth, C.W. Beecher, J.D. Cavalcoli, B.D. Athey, G.S. Omenn, C.F. Burant, H.V. Jagadish, Bioinformatics, 26 (2010) 971-973. 10.1093/bioinformatics/btq048.

[99] M. Kohl, S. Wiese, B. Warscheid, Methods Mol. Biol., 696 (2011) 291-303. 10.1007/978-1-60761-987-1_18.

[100] M.E. Smoot, K. Ono, J. Ruscheinski, P.L. Wang, T. Ideker, Bioinformatics, 27 (2011) 431-432. 10.1093/bioinformatics/btq675.

[101] R. Caspi, H. Foerster, C.A. Fulcher, P. Kaipa, M. Krummenacker, M. Latendresse, S. Paley, S.Y. Rhee, A.G. Shearer, C. Tissier, T.C. Walk, P. Zhang, P.D. Karp, Nucleic Acids Res., 36 (2008) D623-D631. 10.1093/nar/gkm900.

[102] N.V. Chaika, T. Gebregiworgis, M.E. Lewallen, V. Purohit, P. Radhakrishnan, X. Liu, B. Zhang, K. Mehla, R.B. Brown, T. Caffrey, F. Yu, K.R. Johnson, R. Powers, M.A. Hollingsworth, P.K. Singh, Proc Natl Acad Sci U S A, 109 (2012) 13787-13792. 10.1073/Proc. Natl. Acad. Ssci. U. S. A..1203339109.

[103] R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2012.

[104] A.S. Nuxoll, S. Halouska, M. Sadykov, G.A. Somerville, K.W. Bayles, T. Kielian, R. Powers, P.D. Fey, PLoS Pathogens, 8 (2012) e1003033. 10.1371/journal.ppat.1003033.

[105] M.R. Sadykov, V.C. Thomas, D.D. Marshall, C.J. Wenstrom, D.E. Moormeier, T.J. Widhelm, A.S. Nuxoll, R. Powers, K.W. Bayles, J Bacteriol, 195 (2013) 3035-3044. 10.1128/JB.00042-13.

[106] M.R. Sadykov, B. Zhang, S. Halouska, J.L. Nelson, L.W. Kreimer, Y. Zhu, R. Powers, G.A. Somerville, J. Biol. Chem., 285 (2010) 36616-36624. 10.1074/jbc.M110.152843.

JOURNAL OF INTEGRATED OMICS
A METHODOLOGICAL JOURNAL
HTTP://WWW.JIOMICS.COM

ORIGINAL ARTICLE | DOI: 10.5584/jiomics.v3i2.142

Iberian Lynx (*Lynx pardinus*) from the captive breeding program as reservoir of antimicrobial resistant enterococci and *Escherichia coli* isolates

Alexandre Gonçalves[1,2,3,4], Gilberto Igrejas[1,2], Hajer Radhouani[1,2,3,4], Tiago Santos[1,2,3,4], Ricardo Monteiro[1,2,3,4], Catarina Marinho[1,2,3,4], Maria José Pérez[5], Rocio Canales[6], José Luis Mendoza[7], Rodrigo Serra[8], Carmen Torres[9], Patrícia Poeta[3,4,*]

[1]Center of Genomics and Biotechnology/Institute for Biotechnology and Bioengineering, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal. [2]Department of Genetics and Biotechnology, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal. [3]Center for Animal Science and Veterinary, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal. [4]Veterinary Science Department, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal. [5]La Olivilla Captive Breeding Centre, Iberian Lynx Ex Situ Conservation Program, Santa Helena, Spain. [6]El Acebuche Captive Breeding Centre, Iberian Lynx Ex Situ Conservation Program, Huelva, Spain. [7]La Granadilla Breeding Centre, Iberian Lynx Ex Situ Conservation Program, Zarza de Granadilla, Spain. [8]National Centre for Captive Breeding of the Iberian Lynx, Silves, Portugal. [9]Biochemistry and Molecular Biology Area, University of Rioja, Logroño, Spain.

Received: 30 May 2013 **Accepted:** 13 July 2013 **Available Online**: 18 July 2013

Abstract

A total of 98 faecal samples from captive specimens of Iberian lynx were collected and analysed for enterococci (96 isolates) and *Escherichia coli* (90 isolates) recovery. These 186 isolates were tested for antimicrobial resistance, molecular mechanisms of resistance, and detection of virulence genes. Among the enterococci, *Enterococcus hirae* was the most prevalent species (35 isolates), followed by *Enterococcus faecalis* (30 isolates), *Enterococcus faecium* (27 isolates), and *Enterococcus durans* (4 isolates). High rates of resistance to tetracycline, erythromycin and high-level-kanamycin were detected among enterococcal isolates (41%, 26%, and 19%, respectively). The *tet*(M) and/or *tet*(L), *erm*(B), *aac*(6′)-Ie-*aph*(2″)-Ia, *ant*(6)-Ia, or *aph*(3′)-IIIa genes were detected among resistant enterococci. Likewise, high rates of resistance were detected in *E. coli* isolates to tetracycline, streptomycin, sulfamethoxazole-trimethoprim (SXT), nalidixic acid, ampicillin, and ciprofloxacin (34%, 28%, 26%, 21%, 17%, and 16%, respectively). Furthermore, the bla_{TEM} or bla_{SHV}, *tet*(A) and/or *tet*(B), *aadA* or *strA-strB*, *aac*(3)-II and/or *aac*(3)-IV, and different combinations of *sul* genes were detected among most resistant isolates. Fifteen isolates contained class 1 and/or class 2 integrons and 3 different gene cassette arrays were identified (*aadA1*, *dfrA1+aadA1*, and *estX+psp+aadA2*). The *E. coli* isolates were ascribed to phylo-groups A (12%); B1 (40%); B2 (37%), and D (11%), being *fimA* the most prevalent virulence gene (n=84), followed by *aer* (n=13), *cnf1* (n=13), *papC* (n=10) and *papG*-allele III (n=9). This study showed specimens of Iberian lynx acting as reservoirs of resistance genes, and in future (re)introductions they could spread resistant bacteria throughout the environment.

Keywords: Antimicrobial resistance; enterococci; *Escherichia coli*; Iberian lynx; virulence genes.

1. Introduction

The selection pressure exerted by the intensive use of antimicrobial agents, in human and veterinary medicine, has contributed to the selection and dissemination of resistant bacteria [1]. Moreover, resistant commensal bacteria constitute a reservoir of resistance genes that might be transferred to other commensal or pathogenic bacteria [2]. *Enterococcus* spp. and *Escherichia coli* inhabit the gastrointestinal tract of human and animals and are occasionally associated with both community- and hospital-acquired infections [3]. The levels of antimicrobial resistance in these microorganisms can be used as indicators that might help to track the evolution of antimicrobial resistance in different ecosystems [4]. Recent studies in wild animals confirm that they might act as reservoirs of antimicrobial resistant genetic elements that

***Corresponding author:** Patrícia Poeta. Department of Veterinary Science, University of Trás-os-Montes and Alto Douro, 5001-801 Vila Real. Portugal. Tel: +351-25935466. E-mail address: ppoeta@utad.pt

could be spread across the environment [4].

The Iberian lynx is a critically endangered species native of the Iberian Peninsula [5]. This species typically hunts the European wild rabbit (*Oryctolagus cuniculus*). The decline of prey population along with habitat loss caused a restriction in the Iberian lynx geographical distribution. Total wild population for this species is estimated to a maximum of 150 adults, surviving in two breeding population on Southern Spain (Doñana and Sierra Morena) [6]. Studying captive animals of Iberian lynx is of capital importance since the susceptibility of this endangered species to bacterial infection may be affected by the presence of virulence and resistance genes.

Previous studies performed by our group analysed the carriage of *E. coli* producing extended-spectrum beta-lactamases or vancomycin resistant enterococci in faecal samples of Iberian lynx (free range and captive), using for that purpose antibiotic-supplemented plates for resistant-bacteria recovery [7, 8]. In addition, the faecal E. *coli* and enterococci population of free-range lynx (obtained in non-antibiotic-supplemented plates) has been recently evaluated for antimicrobial resistance and virulence [9], but no data do exist, so far, about this population in captive Iberian lynx.

The purpose of our study was to investigate the antimicrobial resistance, the molecular mechanisms of resistance, and the detection of virulence genes in faecal *Enterococcus* spp. and *E. coli* isolates from captive specimens of Iberian lynx.

2. Material and Methods

2.1. *Faecal samples and bacterial isolates*

Antimicrobial resistance in enterococci isolates and *E. coli* was studied in 98 fresh faecal samples recovered from captive specimens of Iberian lynx (*Lynx pardinus*), between 2008 and 2010. Each faecal sample collected belonged to a different animal. Sample collection from captive animals was obtained in the breeding facilities of the Iberian Lynx Captive Breeding Program, Iberian Peninsula, and took place during the daily handling or during clinical procedures.

For enterococci recovery, samples were seeded in Slanetz-Bartley agar (Oxoid Limited, Basingstoke, UK) plates. One colony with typical enterococcal morphology was identified to the genus and species level by Gram-staining, catalase test, bile-aesculin reaction, and by biochemical tests using the API ID20 Strep system (BioMérieux, La Balme, Les Grottes, France). Species identification was confirmed by polymerase chain reaction (PCR) using specific primers and conditions for *E. faecalis*, *E. faecium*, *E. casseliflavus* [10, 11], *E. gallinarum* [12], *E. hirae* and *E. durans* [13].

For *E. coli* isolation, samples were seeded in Levine agar (Oxoid Limited, Basingstoke, UK) plates. One colony per sample with typical *E. coli* morphology was selected and identified by classical biochemical methods (Gram staining, catalase, oxidase, indol, Methyl-Red-Voges-Proskauer, citrate, and urease) and by the API 20E system (BioMérieux, La Balme Les Grottes, France).

2.2. *Antimicrobial susceptibility testing*

Antimicrobial susceptibility was performed by the disk diffusion method according to the criteria of the CLSI [14]. The enterococci susceptibility to 11 antimicrobial agents (vancomycin; teicoplanin; ampicillin; chloramphenicol; tetracycline; erythromycin; quinupristin–dalfopristin; ciprofloxacin; streptomycin; gentamicin; and kanamycin) was tested. *Enterococcus faecalis* ATCC 29212 and *Staphylococcus aureus* ATCC 25923 strains were used for quality control.

Similarly, susceptibility of the *E. coli* isolates was performed to 16 antimicrobial agents [ampicillin, amoxicillin plus clavulanic acid, cefoxitin, cefotaxime, ceftazidime, aztreonam, imipenem, gentamicin, amikacin, tobramycin, streptomycin, nalidixic acid, ciprofloxacin, sulfamethoxazole-trimethoprim (SXT), tetracycline, and chloramphenicol] by the disk diffusion method [14]. *E. coli* ATCC 25922 was used as a quality-control strain. Additionally, ESBL-phenotypic detection was carried out by double-disk diffusion test [14].

2.3. *Assay of gelatinase and beta-hemolytic activities in enterococci isolates*

Gelatinase is an extracellular metalloendopeptidase that can hydrolyse gelatin, collagen, and other bioactive peptides which suggests that it may take part in inflammatory processes. Moreover, haemolysin has been demonstrated to contribute to the severity of enterococcal disease (Archimbaud et al., 2002). Evaluation of gelatinase and hemolysin production in enterococci isolates was performed as previously reported [15].

2.4. *Characterization of antimicrobial resistance mechanisms and detection of virulence genes*

Among resistant-enterococci, the presence of genes encoding resistance to erythromycin [*erm*(A) and *erm*(B)], tetracycline [*tet*(M), *tet*(L), and *tet*(K)], kanamycin [*aph*(3′)-IIIa], streptomycin [*ant*(6)-Ia], gentamicin [*aac*(6')-Ie-*aph*(2'')-Ia], chloramphenicol [*cat*A], quinupristin–dalfopristin [*vat*D and *vat*E] and the presence of Tn*916*- and Tn*5397*-specific sequences were analysed by PCR using primers and conditions previously reported [16-19]. The presence of genes encoding different virulence factors (*gel*E, *agg*, *ace*, *cpd*, *fsr*, *esp*, *hyl* and *cyl*LLLSABM) was also studied by PCR [19, 20].

Likewise, in the resistant *E. coli* isolates the presence of resistance genes to ampicillin [bla_{TEM} and bla_{SHV}], tetracycline [*tet*(A) and *tet*(B)], streptomycin [*aadA* and *strA-strB*], gentamicin [*aac(3)-II, aac(3)-IV*], SXT [*sul1*, *sul2* and *sul3*] and chloramphenicol [*cmlA* and *floR*] were studied by PCR [21]. Additionally, the incidence of the *intl1* and *intI2* genes, encoding classes 1 and 2 integrases, respectively, and their variable regions were also analysed by PCR and sequencing

[21].

Lastly, the phylogenetic groups and virulence determinants often found in pathogenic *E. coli* (*stx1-stx2, fimA*, *papG* allele III, *cnf1*, *papC* and *aer*) were investigated [3, 22].

Positive and negative controls from the collection of strains of the University of Trás-os-Montes and Alto Douro (Portugal) were included in all PCR assays.

3. Results

Enterococci were obtained from 96 (98%) of the 98 samples analysed. *E. hirae* was the most prevalent detected species (35 isolates), followed by *E. faecalis* (30 isolates), *E. faecium* (27 isolates) and *E. durans* (4 isolates). *E. coli* isolates were detected in 90 (92%) of the 98 Iberian lynx faecal samples.

Fifty-five of the 96 enterococci isolates (57%) showed susceptibility to all the antimicrobial agents tested (24 *E. hirae*, 19 *E. faecium*, 8 *E. faecalis*, and 4 *E. durans*). Most of *E. faecalis* isolates showed a multiresistant phenotype.

No ampicillin-, vancomycin- or teicoplanin-resistant enterococcal isolates were found in this study and quinupristin –dalfopristin resistance, as expected, was found in *E. faecalis* isolates. On the other hand, higher levels of resistance to tetracycline (41%), erythromycin (26%), and kanamycin (19%) were detected among our isolates. Moderate level of resistance was observed to quinupristin–dalfopristin (10%), ciprofloxacin (9%), and streptomycin (9%). Additionally, low level of resistance was detected to gentamicin (high-level, 6%) and chloramphenicol (3%).

Table 1 presents the antimicrobial resistance genes detected among resistant-enterococci. Different associations of *tet* (M) and *tet*(L) genes with the genetic elements *Tn*916 and *Tn*5397 were found in tetracycline resistant isolates. All erythromycin-resistant *E. hirae* and *E. faecium* isolates harbored the *erm*(B) gene. On the other hand, only 12 of the 20 *E. faecalis* isolates that showed resistance to erythromycin presented the *erm*(B) gene. The *aac*(6')-Ie-*aph*(2'')-Ia gene was observed in 4 of the 6 high-level gentamicin-resistant *E. faecalis* isolates and the *ant*(6)-Ia gene in 1 of the 9 high-level streptomycin–resistant isolates. The *aph*(3')-IIIa gene, encoding an aminoglycoside-modifying enzyme, was detected in all kanamycin resistant isolates (Table 1).

No virulence genes were detected among enterococci isolates and none of these microorganisms showed beta-hemolytic or gelatinase activities.

Fifty of the 90 *E. coli* isolates (56%) showed susceptibility to all tested antimicrobial agents. No resistances were detected to amoxicillin plus clavulanic acid, cefoxitin, cefotaxime, ceftazidime, aztreonam, and imipinem. High percentages of resistance were observed to tetracycline (34%), streptomycin (28%), sulfamethoxazole-trimethoprim (26%), nalidixic acid (21%), ampicillin (17%), and ciprofloxacin (16%), and lower levels to gentamicin (13%), chloramphenicol (10%), tobramycin (2%), and amikacin (1%). Additionally, it is noteworthy that 27 isolates (30%) showed a multiresistant phenotype.

Table 1 presents the antimicrobial resistance genes detected among resistant *E. coli* isolates. From the 15 ampicillin-resistant *E. coli* isolates, 6 harbored the bla_{TEM} gene and 1 harbored the bla_{SHV} gene. The *tet*(A) and *tet*(B) genes were detected in 18 and 7 tetracycline-resistant *E. coli* isolates, respectively. The presence of the *aadA* and the *strA-strB* genes were shown in 16 and 6 streptomycin-resistant *E. coli* isolates, respectively. Among the gentamicin-resistant *E. coli* isolates, 9 harbored the *aac*(3)-II gene and 1 isolate the *aac* (3)-II plus *aac*(3)-IV genes. Different combinations of *sul1*, *sul2*, and *sul3* genes were identified in 22 of the 23 sulfamethoxazole-trimethoprim-resistant *E. coli* isolates. In addition, the *cmlA* gene was identified in 7 of the 11 chloramphenicol-resistant isolates (Table 1).

The presence of integrons was studied by PCR and sequencing, and the presence of class 1 integrons were detected in 14 isolates (all contained the *intI1*, but two isolates lacked the *qac*EΔ1/*sul*I 3'-conserved region) containing the following gene cassette arrays: *aadA1* (6 isolates), *dfrA1+aadA1* (6 isolates), and *estX+psp+aadA2* (2 isolates, both lacking the *qac*EΔ1/*sul*I 3'-conserved region). The presence of class 2 integrons was confirmed in 3 isolates (2 of them also harboring a class 1 integron), containing the gene cassette array *dfrA1+aadA1*.

Among the *E. coli* isolates, 84 of them contained at least one virulence gene. The *fimA* virulence gene was the most prevalent, detected in the 84 isolates. The *aer* and/or *cnf1* genes were found in 13 *E. coli* isolates, respectively. Moreover, the *papC* and *papG*-allele III virulence genes were shown in 10 and 9 isolates, respectively. The distribution of the phylogenetic groups was: 11 isolates ascribed to the phylogenetic group A, 36 isolates to B1, 33 isolates to B2 and 10 isolates into the phylogroup D (Table 2).

4. Discussion

E. faecium and *E. faecalis* are usually the most prevalent enterococcal species among isolates recovered from wild animals [4, 19]. In our study *E. hirae* was the predominant species followed by *E. faecalis* and *E. faecium*. Similar results were obtained in wild boars and Iberian wolf [23, 24].

No vancomycin resistant enterococci (VRE) were detected in this study in faecal samples of captive lynxes, and similar results were also demonstrated in free range lynxes [9]. Even when vancomycin-supplemented plates were used for VRE recovery, negative results were also demonstrated both in captive and free range lynxes, with the exception of enterococci with an intrinsic-mechanism of vancomycin-resistance [8].

The *tet*(M) and *tet*(L) genes, coding for tetracycline resistance, are frequently reported in tetracycline-resistant enterococcal isolates [4, 25]. In our study, most of the tetracycline-resistant isolates harbored simultaneously these two genes. The *erm*(B) gene, responsible for acquired erythromycin resistance, is frequently linked with the *tet*(M) gene on

Table 1 Resistance genes detected among antimicrobial resistant enterococci and *E. coli* isolates obtained from faecal samples of Iberian lynx.

Bacteria	Antimicrobial agent	No. of resistant isolates	Genes detected by PCR	
			Resistance genes and genetic elements	No. of isolates
Enterococci (n= 96)				
E. hirae (n= 35)	Chloramphenicol (30 µg)	1	*catA*	1
	Tetracycline (30 µg)	11	*tet*(M) + *tet*(L)	3
			tet(M) + *tn*916	2
			tet(M) + *tet*(L) + *tn*916	5
	Erythromycin (15 µg)	2	*erm*(B)	2
	Kanamycin (120 µg)	1	*aph*(3')-IIIa	1
E. faecalis (n= 30)	Chloramphenicol (30 µg)	2	*catA*	0
	Tetracycline (30 µg)	22	*tet*(L)	8
			tet(M) + *tet*(L)	9
			tet(M) + *tet*(L) + *tn*916	3
			tet(M) + *tet*(L) + *tn*5397	2
	Erythromycin (15 µg)	20	*erm*(B)	12
	Quinupristin-Dalfopristin (15 µg)	10	*vatD* or *vatE*	0
	Ciprofloxacin (5 µg)	8	-	-
	Streptomycin (300 µg)	9	*ant*(6)-Ia	1
	Gentamicin (120 µg)	6	*aac*(6')-Ie-*aph*(2'')-Ia	4
	Kanamycin (120 µg)	16	*aph*(3')-IIIa	16
E. faecium (n= 27)	Tetracycline (30 µg)	6	*tet*(L)	2
			tet(M) + *tet*(L)	4
	Erythromycin (15 µg)	3	*erm*(B)	3
	Ciprofloxacin (5 µg)	1	-	-
	Kanamycin (120 µg)	1	*aph*(3')-IIIa	1
E. coli (n= 90)	Ampicilin (10 µg)	15	bla_{TEM}	6
			bla_{SHV}	1
	Tetracycline (30 µg)	31	*tet*(A)	18
			tet(B)	7
	Gentamicin (10 µg)	12	*aac*(3)-II	9
			aac(3)-II + *aac*(3)-IV	1
	Amikacin (30 µg)	1	-	-
	Tobramycin (10 µg)	2	-	-
	Streptomycin (10 µg)	25	*aadA*	16
			strA-strB	6
	Nalidixic acid (30 µg)	19	-	-
	Ciprofloxacin (5 µg)	14	-	-
	Sulfamethoxazol.Trimethropim (25 µg)	23	*sul1*	9
			sul2	3
			sul3	3
			sul1 + *sul2*	4
			sul1 + *sul3*	1
			sul2 + *sul3*	2
	Chloramphenicol (30 µg)	11	*cmlA*	7

Table 2 Virulence genes detected among the *E. coli* isolates collected from faecal samples of Iberian lynx.

Bacteria	Virulence genes profiles detected by PCR	No. of isolates	Phylogenetic group			
			A	B1	B2	D
E. coli (n= 90)	*fimA*	56	5	24	19	8
	aer; *fimA*	11	1	9	1	0
	*cnf*1; *fimA*	6	0	0	5	1
	papC; *fimA*	1	0	0	1	0
	*papG*III; *fimA*	1	1	0	0	0
	aer; *papC*; *fimA*	1	0	1	0	0
	papC; *papG*III; *fimA*	1	0	0	1	0
	papC; *cnf*1; *papG*III; *fimA*	6	1	0	5	0
	aer; *papC*; *cnf*1; *papG*III; *fimA*	1	0	0	1	0
	Virulence genes not detected	6	3	2	0	1

the highly mobile conjugative transposon Tn*1545*, which predominates in clinically important Gram-positive bacteria [18]. In our study, it was shown that 50% of the *erm*(B) and *tet*(M)-positive isolates harbored the Tn*916*/Tn*1545*-like element. Nonetheless, the association of *tet*(M) and *erm*(B) genes with this transposon was not analysed. The *aac*(6′)-*aph*(2″), *ant*(6)-Ia and *aph*(3′)-IIIa genes detected in our HLR-aminoglycoside enterococci, has been frequently detected in previous reports from wild animals [4, 19, 23, 25].

None of our *E. coli* isolates produced extended-spectrum β-lactamases (ESBL). Nevertheless, in a previous study, ESBL-producing *E. coli* were recovered from 9 faecal samples of captive specimens of Iberian lynx when Levine-supplemented agar plates where used [7]. Thus, ESBL-producing *E. coli* could be present in the faecal microbiota of captive animals in lower rates when compared with ESBL-nonproducing *E. coli* isolates. In our study, 44% of *E. coli* isolates showed resistance to at least one antimicrobial agent tested and high levels of resistance to tetracycline, streptomycin, SXT, nalidixic acid, ampicillin, and ciprofloxacin were found. Similar results were detected in other studies performed with faecal *E. coli* isolates from wild animals [26, 27], and food-producing animals [28]. Still, higher levels of resistance have been previously detected in *E. coli* isolates from wild birds in Portugal [25].

The classical TEM enzymes are the predominant plasmid-mediated β-lactamases of gram-negative bacteria and have been previously found in ampicillin-resistant *E. coli* isolates from different origins [29]. Similarly, in our study 6 of the 10 ampicillin-resistant *E. coli* isolates harboured the bla_{TEM} gene and 1 isolate harboured the bla_{SHV} gene. Likewise, the *tet*(A) and *tet*(B) genes, detected among our isolates, are the most frequent ones reported among resistant isolates from different origins [21, 27]. The CmlA family is the largest family of chloramphenicol specific exporters and the *cmlA* gene was identified in 7 of our chloramphenicol-resistant isolates. The *cmlA* determinant is typically associated with mobile genetic elements that carry additional resistance genes. This fact explains the persistence of chloramphenicol resistance determinants despite the longstanding prohibition of chloramphenicol use in food-producing animals or wild animals [30]. The *aadA* gene and the linked *strA-strB* gene pair found in our study encode for streptomycin resistance and are commonly found among streptomycin-resistant *E. coli* isolates from animals, food and humans [31]. Additionally, the AAC(3)-II and AAC(3)-IV gentamicin acetyltransferases have been recurrently reported within *E. coli* isolates of human, food, animal and environmental origin [21]. In our study, the genes encoding these enzymes, *aac*(3)-II and/or *aac*(3)-IV, were also detected in 10 gentamicin-resistant *E. coli* isolates. The acquisition and prevalence of alternative dihydropteroate synthase (DHPS) genes (*sul* genes) by our SXT-resistant isolates are in accordance with the findings of previously studies performed within *E. coli* isolates from wild animals in Portugal [4, 25].

The presence of integrons among our commensal *E. coli* isolates is a cause for concern, as this genetic structure might enhance the dissemination of resistance genes to other bacteria, by mobile elements such as plasmids and transposons. Gene cassettes that confer resistance to streptomycin/spectinomycin (*aadA1* and *aadA2*), and to trimethoprim (*dfrA1* and *dfrA12*) were present within the detected integrons. Among the variable region of these structures, 4 different gene cassette arrays were identified (*aadA1*, *dfrA1+aadA1*, *dfrA12+aadA2*, and *estX+psp+aadA2*). Similar structures have been reported in *E. coli* isolates from different sources [27, 28, 32].

A high number of virulence genes were detected among

our *E. coli* isolates. As showed, 95% of the isolates harbored at least one virulence gene. The predominance of *fimA* (type 1 fimbriae) is in accordance with previous works [25]. Still, this high value is unusual for wild animals and is higher than those previously found within *E. coli* isolates from wild animals and meat products [25, 33].

E. coli isolates can be ascribed into four main phylogenetic groups (A, B1, B2, and D) and whereas most commensal strains belong to groups A and B1, virulent extra-intestinal strains belong mainly to groups B2 and D [22]. The presence of virulence-encoding genes in our isolates is a cause of great concern especially since a vast number of them were ascribed to phylogenetic groups B2 and D. A similar ratio has been previously reported among intestinal *E. coli* clones from wild boars in Germany [34]. Nonetheless, this phylogenetic distribution was unexpected, as former studies performed within faecal *E. coli* isolates from wild animals detected a higher prevalence of strains belonging to groups A and B1 [25, 27].

There are several examples of resistant bacteria transfer between animals, and from animals to man via the food chain [2]. Therefore, Iberian lynx might be contaminated through the food chain, as the presence of resistant *E. coli* strains has been previously detected in their predominant prey [4]. Captive animals of Iberian lynx are fed with captive rabbits produced for human consumption. Previous works showed food-producing animals as reservoirs of antimicrobial resistant bacteria, which could explain the results obtained in this study [2, 35]. The use of antimicrobial agents in captive Iberian lynx is very restrict and infrequent. Still, their use during clinical procedures (e.g. clinical cases and surgical procedures) might create a selective pressure that could lead to a small increase in the levels of antimicrobial resistance.

The results obtained in this study with captive specimens are similar to those previously found in wild specimens of Iberian lynx [9]. The values of phenotypic resistance to the antimicrobial agents tested, the phylogenetic distribution, and the presence of virulence genes are closely related with the wild specimens of this species. Consequently, there is a possibility that the founder animals of the breeding program (caught in the wild) might have carried these resistant strains and resistance genes into the program.

5. Concluding Remarks

This study showed captive specimens of Iberian lynx acting as reservoirs of resistance genes. The same genes found in bacteria from environment and human origins were found in this study, indicating the possible circulation of bacteria and resistance genes between animal, environment and human ecosystems. Additionally, the high percentage of virulence determinants in *E. coli* isolates is a cause of concern since an increased risk of infection and therapeutic failure due to virulence/resistance genes can consequently represent a health problem for these animals.

Further research should be carried out in the future in order to better understand the dissemination and frequency of resistance genes and virulence determinants. This may be useful to verify the possible contamination of these animals through the food chain.

Acknowledgements

Alexandre Gonçalves (SFRH/BD/47833/2008) and Hajer Radhouani (SFRH/BD/60846/2009) have PhD fellowships granted by FCT-Fundação para a Ciência e a Tecnologia and ESF-European Social Fund. Part of this work was financed by project SAF2012-35474 from the Ministerio de Economía y Competitividad of Spain and FEDER. Collection of samples was possible through project LIFENAT/ES/00570.

References

[1] S. Schwarz, C. Kehrenberg, T.R. Walsh, Int J Antimicrob Agents, 17 (2001) 431-437. DOI:10.1016/S0924-8579(01)00297-7

[2] F.M. Aarestrup, H.C. Wegener, P. Collignon, Expert Rev Anti Infect Ther, 6 (2008) 733-750. DOI:10.1586/14787210.6.5.733.

[3] J. Ruiz, K. Simon, J.P. Horcajada, M. Velasco, M. Barranco, G. Roig, A. Moreno-Martinez, J.A. Martinez, T. Jimenez de Anta, J. Mensa, J. Vila, J Clin Microbiol, 40 (2002) 4445-4449. DOI: 10.1128/JCM.40.12.4445-4449.2002

[4] N. Silva, G. Igrejas, N. Figueiredo, A. Gonçalves, H. Radhouani, J. Rodrigues, P. Poeta, Sci Total Environ, 408 (2010) 4871-4876. DOI:10.1016/j.scitotenv.2010.06.046.

[5] IUCN, IUCN Red List of Threatened Species. Version 2012.2. http://www.iucnredlist.org Downloaded on 10 February 2013.

[6] P. Ferreras, A. Rodriguez, F. Palomares, M. Delibes, Biology and Conservation of wild Felids, Oxorfd University Press, 2010, pp. 507-520.

[7] A. Gonçalves, G. Igrejas, H. Radhouani, V. Estepa, E. Alcaide, I. Zorrilla, R. Serra, C. Torres, P. Poeta, Lett Appl Microbiol, 54 (2012) 73-77. DOI:10.1111/j.1472-765X.2011.03173.x.

[8] A. Gonçalves, G. Igrejas, H. Radhouani, M. Lopez, A. Guerra, F. Petrucci-Fonseca, E. Alcaide, I. Zorrilla, R. Serra, C. Torres, P. Poeta, Sci Total Environ, 410 (2011) 266-268. DOI:10.1016/j.scitotenv.2011.09.074.

[9] A. Goncalves, G. Igrejas, H. Radhouani, T. Santos, R. Monteiro, R. Pacheco, E. Alcaide, I. Zorrilla, R. Serra, C. Torres, P. Poeta, Sci Total Environ, 456-457 (2013) 115-119. DOI:10.1016/j.scitotenv.2013.03.073.

[10] S. Dutka-Malen, S. Evers, P. Courvalin, J Clin Microbiol, 33 (1995) 1434.

[11] S. Dutka-Malen, S. Evers, P. Courvalin, J Clin Microbiol, 33 (1995) 24-27.

[12] A. Miele, M. Bandera, B.P. Goldstein, Antimicrob Agents Chemother, 39 (1995) 1772-1778. DOI: 10.1128/AAC.39.8.1772

[13] C.A. Arias, B. Robredo, K.V. Singh, C. Torres, D. Panesso, B.E. Murray, J Clin Microbiol, 44 (2006) 1567-1570. DOI: 10.1128/JCM.44.4.1567-1570.2006.

[14] CLSI, Clinical and Laboratory Standards Institute, Wayne,

PA, USA., (2011).

[15] M. López, Y. Sáenz, B. Rojo-Bezares, S. Martinez, R. del Campo, F. Ruiz-Larrea, M. Zarazaga, C. Torres, Int J Food Microbiol, 133 (2009) 172-178. DOI: 10.1016/j.ijfoodmicro.2009.05.020.

[16] F.M. Aarestrup, Y. Agerso, P. Gerner-Smidt, M. Madsen, L.B. Jensen, Diagn Microbiol Infect Dis, 37 (2000) 127-137. DOI: 10.1016/S0732-8893(00)00130-9

[17] Y. Agerso, A.G. Pedersen, F.M. Aarestrup, J Antimicrob Chemother, 57 (2006) 832-839. DOI:10.1093/jac/dkl069.

[18] E. De Leener, A. Martel, A. Decostere, F. Haesebrouck, Microb Drug Resist, 10 (2004) 341-345. DOI:10.1089/mdr.2004.10.341.

[19] P. Poeta, D. Costa, Y. Sáenz, N. Klibi, F. Ruiz-Larrea, J. Rodrigues, C. Torres, J Vet Med B Infect Dis Vet Public Health, 52 (2005) 396-402. DOI: 10.1111/j.1439-0450.2005.00881.x.

[20] T.J. Eaton, M.J. Gasson, Appl Environ Microbiol, 67 (2001) 1628-1635. DOI: 10.1128/AEM.67.4.1628-1635.2001.

[21] Y. Sáenz, L. Briñas, E. Dominguez, J. Ruiz, M. Zarazaga, J. Vila, C. Torres, Antimicrob Agents Chemother, 48 (2004) 3996-4001. DOI: 10.1128/AAC.48.10.3996-4001.2004.

[22] O. Clermont, S. Bonacorsi, E. Bingen, Appl Environ Microbiol., 66 (2000) 4555-4558. DOI: 10.1128/AEM.66.10.4555-4558.2000

[23] P. Poeta, D. Costa, G. Igrejas, J. Rodrigues, C. Torres, Vet Microbiol, 125 (2007) 368-374. DOI: 10.1016/j.vetmic.2007.06.003.

[24] A. Goncalves, G. Igrejas, H. Radhouani, S. Correia, R. Pacheco, T. Santos, R. Monteiro, A. Guerra, F. Petrucci-Fonseca, F. Brito, C. Torres, P. Poeta, Lett Appl Microbiol, 56 (2013) 268-274. DOI: 10.1111/lam.12044.

[25] H. Radhouani, P. Poeta, A. Gonçalves, R. Pacheco, R. Sargo, G. Igrejas, J Med Microbiol, 61 (2012) 837-843. DOI:10.1099/jmm.0.038364-0.

[26] D. Costa, P. Poeta, Y. Sáenz, L. Vinué, A.C. Coelho, M. Matos, B. Rojo-Bezares, J. Rodrigues, C. Torres, Microb Drug Resist, 14 (2008) 71-77. DOI:10.1089/mdr.2008.0795.

[27] H. Radhouani, P. Poeta, G. Igrejas, A. Gonçalves, L. Vinué, C. Torres, Vet Rec, 165 (2009) 138-142. 165/5/138 [pii].

[28] B. Guerra, E. Junker, A. Schroeter, B. Malorny, S. Lehmann, R. Helmuth, J Antimicrob Chemother, 52 (2003) 489-492. DOI: 10.1093/jac/dkg362.

[29] L. Briñas, M. Zarazaga, Y. Sáenz, F. Ruiz-Larrea, C. Torres, Antimicrob Agents Chemother, 46 (2002) 3156-3163. DOI:10.1128/AAC.46. DOI:10.3156-3163.2002

[30] K. Poole, J Antimicrob Chemother, 56 (2005) 20-51. DOI:10.1093/jac/dki171.

[31] M. Sunde, M. Norstrom, J Antimicrob Chemother, 56 (2005) 87-90. DOI: 10.1093/jac/dki150.

[32] L. Vinué, Y. Sáenz, S. Martinez, S. Somalo, M.A. Moreno, C. Torres, M. Zarazaga, Clin Microbiol Infect, 15 (2009) 954-957. DOI:10.1111/j.1469-0691.2009.02803.x.

[33] L. Soufi, M.S. Abbassi, Y. Sáenz, L. Vinué, S. Somalo, M. Zarazaga, A. Abbas, R. Dbaya, L. Khanfir, A. Ben Hassen, S. Hammami, C. Torres, Foodborne Pathog Dis, 6 (2009) 1067-1073. DOI:10.1089/fpd.2009.0284.

[34] P. Schierack, A. Romer, J. Jores, H. Kaspar, S. Guenther, M. Filter, J. Eichberg, L.H. Wieler, App Environ Microbiol, 75 (2009) 695-702. DOI: 10.1128/AEM.01650-08.

[35] C. Novais, T.M. Coque, M.J. Costa, J.C. Sousa, F. Baquero, L.V. Peixe, J Antimicrob Chemother, 56 (2005) 1139-1143. DOI: 10.1093/jac/dki360.

JOURNAL OF INTEGRATED OMICS
A METHODOLOGICAL JOURNAL
HTTP://WWW.JIOMICS.COM

ORIGINAL ARTICLE | DOI: 10.5584/jiomics.v3i2.143

An integrated proteomic and physiological approach to understand the adhesion mechanism of the probiotic *Lactobacillus reuteri* Lb2 BM DSM 16143

Erika Mangiapane[1], Cristina Lamberti[2], Alessandro Pessione[1], Patrizia Ceruti[3], Francesco Novelli[3], Eugenio Galano[4], Ritva Virkola[5], Timo K. Korhonen[5], Enrica Pessione[1]*

[1]Department of Life Sciences and Systems Biology, Via Accademia Albertina 13, 10123, Torino University, Italy; [2]CNR ISPA, c/o Bioindustry Park S. Fumero, Via Ribes 5, 10010, Colleretto Giacosa, Italy; [3]Department of Medicine and Experimental Oncology, CeRMS, Via Cherasco 15, 10126, Torino University, Italy; [4]Department of Chemical Sciences, University of Naples "Federico II", Via Cinthia 4, 80126, Napoli, Italy; [5]Division of General Microbiology Department of Biosciences, University of Helsinki, FIN-00014 Helsinki, Finland.

Received: 13 June 2013 **Accepted:** 27 August 2013 **Available Online**: 27 September 2013

ABSTRACT

The adhesion ability of the probiotic *Lactobacillus reuteri* Lb2 BM DSM 16143 was tested to both enterocyte-like Caco-2 cells and to extracellular matrix proteins (laminin, fibronectin and collagen I and IV). The adhesiveness was lost after an alkaline treatment known to release moonlighting proteins from lactobacillar cell surface. To characterize the putative adhesive molecules, a 2-DE experiment in the p*I* range 4-7 was performed on the extracellular proteins. The expression of several moonlighting proteins involved in adhesion (*i.e.* GAPDH, EF-Tu, phosphoglycerate kinase) was demonstrated. Some of the identified adhesins were able to bind plasminogen (Plg), but did not convert it into plasmin (Plm), in absence of exogenous activators. This indicates that the moonlighting proteome of *L. reuteri* Lb2 BM DSM 16143 can contribute to adhesion processes.

Keywords: extracellular matrix protein; bacterial adhesion; moonlighting proteins; 2DE; plasminogen binding.

1. Introduction

Lactobacillus reuteri was demonstrated to possess several positive features supporting its employment as a probiotic. First it produces several antimicrobial compounds, useful to treat or prevent infectious diseases, such as: i) reuterin, biosynthesized during anaerobic catabolism of glycerol and active against a broad range of pathogenic microorganisms [1,2] ii) reutericin, a bacteriocin of 2,7 kDa displaying lytic activity [3] iii) reutericyclin, a highly hydrophobic tetramic acid derivative, with a molecular mass of 349 Da exhibiting a broad inhibitory spectrum [4]. Additional antimicrobial features have recently been described in this species like the ability to neutralize toxins produced by fungi [5], to modulate expression of toxins in different bacterial species [6] and to co-aggregate with toxinogenic *Staphylococcus aureus* [7]. Furthermore *L. reuteri* displays anti-inflammatory properties linked to a factor secreted by the biofilm-forming cells [8]. All these features highlight a strong attitude of this bacterial species not only to counteract infections but also to attenuate their severity.

More recently the use of a *Lactobacillus reuteri* Lb2 BM DSM 16143, the strain investigated in this study, has been suggested as nutraceutical supplement, for oxidative stress protection in the human host [9,10].

Besides the most common pre-requisites to define a strain

***Corresponding author:** Enrica Pessione, Via Accademia Albertina 13, 10123, Torino University, Italy; Tel: +390116704644; Fax: +390116704508; Email Address: enrica.pessione@unito.it

as a probiotic (pH, gastric enzymes and bile salt resistance) additional important features are adhesion and metabolic/ biochemical safety (GRAS status).

Adhesion of probiotics to the mucosal surface is critical for exerting beneficial effects to the host organism [11]. It is first driven by weak forces, like Coulomb and Van der Waals attractions, and mediated by several bacterial binding-effectors such as polysaccharides [12], teichoic and lipoteichoic acids [13] followed by the specific production of proteins named adhesins [14]. Several studies have demonstrated that in some *Lactobacillus* species, glycolytic enzymes (such as GAPDH, phosphoglycerate kinase and mutase), protein folding and stress responses-involved proteins (GroEL and DnaK), as well as transcription and translation proteins (elongation factor Tu, Ts and trigger factor) can exert an adhesive function when they are secreted and surface-exposed [15,16,17]. Some of these adhesins also play a role in plasminogen (Plg) binding. In pathogenic bacteria (*Staphylococcus aureus* and *Streptococcus pneumoniae*) this constitutes a problem since they also biosynthesize enzymes (staphylokinases and streptokinases) able to convert Plg to plasmin (Plm), its proteolytic active form, that can damage tissues opening the way for blood colonization and body invasion. To date, although Plg binding capability has also been described for probiotic bacteria [18], their intrinsic potential (kinases) for Plg activation to Plm has never been demonstrated.

The adhesins cited above are defined "moonlighting proteins", *i.e.* proteins displaying different functions according to their subcellular localization [19]. They lack any anchoring motif or surface retention domain [20] and can be easily released from bacteria, either due to the normal cell-wall turnover [21] or to disturbances in cell wall permeability resulting from pH-stress [18] or exposure to host antimicrobial peptides [22]. In LAB no classical signal peptide responsible for moonlighting protein export has been identified so far. Nevertheless, it has been demonstrated in *Bacillus subtilis* that these proteins contain one additional alpha-helix responsible for their secretion [23].

A proteomic approach applied on the cell wall and extracellular proteomes can confirm the external location of such proteins. This information could not be obtained by a classical genome-transcriptome-based characterization since these adhesins, belonging to the main metabolic pathways, are constitutively expressed.

In the present study the adhesive ability of *L. reuteri* Lb2 BM DSM 16143 has been tested both on Caco-2 cells and extracellular matrix (ECM) proteins. These experiments were combined in a wide integrated approach, with classic 2DE proteomic experiments on the extracellular proteins in order to verify the presence of moonlighting proteins involved in adhesion. In parallel, other experiments were performed to evaluate if the detachment of such proteins from the cell wall was pH-dependent and if it could result in a decrease in the adhesion ability of the strain. Finally to ensure safety we investigated the intrinsic ability of the strain to activate proteolytic cascades.

2. Material and Methods

2.1 *Bacterial strain and culture conditions*

Lactobacillus reuteri Lb2 BM DSM 16143 was isolated from a human faecal sample and it belongs to the collection of BioMan life science S.r.l. It was maintained in a modified MRS (de Man, Rogosa and Sharpe) medium (10 g/L Tryptone enzymatic digest from casein, 8 g/L Peptone from soybean, 10 g/L Yeast extract, 10 g/L Sucrose, 1 ml/L Tween80, 2 g/L Potassium phosphate dibasic, 5 g/L Sodium acetate, 2 g/L Ammonium citrate tribasic anhydrous, 0.2 g/L Magnesium sulfate, 0.05 g/L Manganese sulfate) at - 24°C in 0.5 mL aliquots with 0.5 mL of 40% glycerol.

The cultures were grown in closed screw cap bottles, at 37 ° C, without shaking. The pH of the medium was adjusted at 6.4 before the inoculum. The bacterial growth was monitored by 600 nm optical density (OD_{600}) measurement.

For all the cultures three biological replicates were performed.

2.2 *Adhesion assay on Caco-2 cells*

The adhesion ability of *Lactobacillus reuteri* Lb2 BM DSM 16143 to human intestinal cells was evaluated using enterocyte-like Caco-2 cells (ATCC HTB37). The cells were grown in six-well plates in a monolayer with DMEM (Dulbecco's modified Eagle's medium, Gibco TM life technologies) and incubated for 15 days at 37°C, in a 10% CO_2 atmosphere. The culture medium was replaced every 48 hours and the monolayers of Caco-2 cells were used at post-confluence when they were fully differentiated. Intestinal cells were used between passages 25 and 35. Before starting the adhesion trial, Caco-2 cells were washed twice with phosphate-buffered saline (PBS, pH 7.4) and then a bacterial pellet, harvested during the exponential growth phase and resuspended in 1 mL of PBS was added onto Caco-2 cells. The cells were incubated for 1 h at 37°C in 10% CO_2. Then the monolayers were washed five times with sterile PBS to remove all the non-adhered bacterial cells. The washed monolayers were treated with 1 mL of 0.05% Triton X-100 PBS solution for 10 min to detach the Caco-2 cells with the adhered bacteria from the plate. The recovered solution was immediately centrifuged (4000 *x g*, 15 min, 4°C) to pellet cells and bacteria in order to avoid potential bacterial lysis due to Triton X-100 action. The number of colony forming units/mL (CFU/mL) was determined by plating serial 10-fold dilutions onto MRS agar before and after adhesion and incubated for 48 h at 37°C. The adhesion capacity was described as the percentage of bacteria adhered to Caco-2 cells relative to the total number of bacteria added. For all the cultures three technical replicates and two biological replicates were performed. The same experiment was performed on the positive control *Lactobacillus rhamnosus* GG which is known for its adhesive proper-

ties [24].

2.3 *Adhesion assay on extracellular matrix proteins*

Adhesion ability of *L. reuteri* LB2 BM DSM 16143 to extracellular matrix proteins and pH-dependent variations were tested at pH 4 and pH 8. After growth in modified MRS broth bacteria were collected (4000 *x g*, 10 min), washed twice with either 50 mM Tris-HCl buffer at pH 4 or pH 8 and finally resuspended at a concentration of 1 x 10^9 bacteria/mL in the same Tris-HCl buffer. *Lactobacillus* adherence to surface coated laminin (Sigma), fibronectin (Collaborative Biomedical Products), collagen types I and IV (Sigma) were performed as already described using 2.5 pmol surface concentration [25,26]. Bovine Serum Albumin (BSA) (Sigma), a control protein, was coated on glass surface from a solution of 25 μg/mL. The bacteria were incubated with coated target proteins on diagnostic slides for 2 h at room temperature. After incubation the slides were washed with 50 mM Tris-HCl buffer either at pH 4 or pH 8. The adherent bacteria were stained with methylene blue and slides were analyzed with light microscopy using NIH image software (Research Services Branch, National Institute of Health) [27]. The number of bacteria in 20 microscopic fields of 1.6 x 10^4 μm^2 were counted.

2.4 *SDS-PAGE of surface associated proteins*

L. reuteri Lb2 BM DSM 16143 was grown in modified MRS medium until the exponential growth phase. Bacteria were centrifuged at 4000 *x g* for 20 minutes and supernatant was removed. Pellets were resuspended in 200 mL of 50 mM Tris -HCl pH 8, and incubated one hour at 37°C in shaking mode (GallenKamp Orbital Incubator). After a centrifugation at 4000 *x g* for 20 minutes (4°C, Thermo Scientific SL 16R), supernatants were filtered with Stericup filters (Millipore) and Trichloroacetic Acid 16% w/v was added to promote proteins precipitation under shaking over night at 4°C. The obtained suspensions were then ultracentrifuged (35000 *x g*, 90 min, 4°C). Pellets were dried, pulverized and resuspended in 50 mM Tris-HCl pH 7.3. 1 mL aliquots of the obtained sample were subjected to phenol extraction as described before [28]. Briefly, 1 mL phenol was added and the mixtures were incubated for 10 min at 70°C and for 5 min at 0°C and centrifuged (7000 *x g*, 10 min, room temperature). The upper phase was discarded and 1 mL of MilliQ water was added to the lower phase, which was then incubated for 10 min at 70° C and for 5 min at 0°C and centrifuged (7000 *x g*, 10 min, room temperature) again. The upper phase was discarded and 1 mL of ice cold acetone was added to the lower phase before incubating over night at -20°C. Precipitated proteins were recovered by centrifuging (15000 *x g*, 20 min, 4°C) and washed with ice cold acetone (15000 *x g*, 20 min, 4°C). Pellets were pulverized, resuspended in Laemmli loading dye [29] and loaded on a 9.5% resolving gel and run at 120V for 1 hour.

2.5 *Protein identification of SDS-PAGE bands*

Bands were excised from the dried gels and rehydrated with MilliQ water. They were washed twice with 50% v/v ACN in a 25 mM NH_4CO_3 and once in 100% v/v ACN and vacuum-dried. The proteins were in-gel digested with sequencing-grade, modified porcine trypsin (Promega, Madison, WI, USA) and added to a MALDI target plate as described by Hewitson *et al.* [30]. Positive-ion MALDI mass spectra were obtained using an Applied Biosystems 4700 Proteomics Analyzer (Applied Biosystems, Foster City, CA, USA) in reflectron mode. MS spectra were acquired over a mass range of m/z 800–4000 and monoisotopic masses were obtained from centroid of raw, unsmoothed data. Finally, the mass spectra were internally calibrated using the tryptic autoproteolysis products at m/z 842.509 and 2211.104. CID-MS/MS was performed on the 20 strongest peaks with an S/N greater than 40. A source 1 collision energy of 1 kV was used for CID-MS/MS, with air as the collision gas. The precursor mass window was set to a relative resolution of 50, and the metastable suppressor was enabled. Default calibration was used for the MS/MS spectra, which were baseline-subtracted (peak width 50) and smoothed (Savitsky-Golay with three points across a peak and a polynomial order of 4); the peak detection used a minimum S/N of 5, a local noise window of 50 m/z, and a minimum peak width of 2.9 bins. S/N 20 and 30 filters were used to generate peak lists from the MS and MS/MS spectra, respectively. The mass spectral data from the protein spots were submitted to a database search using a locally running copy of the MASCOT programme (Matrix Science, version 2.1). Batch-acquired MS/MS data were submitted to an MS/MS ion search through the Applied Biosystem GPS explorer software interface (version 3.6) with MASCOT. The search parameters allowed a maximum of one missed cleavage, the carbamidomethylation of cysteine, the possible oxidation of methionine, peptide tolerance of 100 ppm and an MS/MS tolerance of 0.1 Da. The spectra were searched against a recent version of the NCBI non-redundant protein database. The significance threshold was set at $p<0.05$, and identification required that each protein contain at least one peptide with an expect e-value <0.05.

2.6 *Preparation of extracellular protein extracts*

L. reuteri Lb2 BM DSM 16143 was grown in modified MRS medium in biological triplicate. The biomass of middle exponential phase was separated by centrifugation (4000 *x g*, 20 min, 4°C) and culture supernatants were filtered in stericup 0.22 μm filters (Millipore). The supernatant was treated as already described in the paragraph "SDS-PAGE of surface associated proteins". The obtained pellets were pulverized, resuspended in rehydration solution (6.5 M urea, 2.2 M thiourea, 4% w/v CHAPS, 5 mM Tris-HCl, pH 8.8, 0.5% IPG buffer (GE-Healthcare), 100 mM DTT) and stored at -20°C. Protein concentration was determined by Bradford assay (Bio Rad).

2.7 *2-DE*

Isoelectrofocusing (IEF) was performed as previously described [31]. Two hundred and seventy-five µg of proteins were separated in 13 cm IPG strips (GE Healthcare) with a linear gradient ranging from 4 to 7 using IPGphor (GE Healthcare) at 20°C, with 83000 Vh, after 10 h rehydration. After IEF, the strips were incubated at room temperature in 6 M urea, 30% v/v glycerol, 2% w/v SDS, 50 mM Tris-HCl, pH 8.6, enriched at first with 2% w/v DTT for 15 min and afterwards with 4.5% w/v iodoacetamide for 15 min. They were then sealed at the top of the 1.0 mm vertical second dimension gels. For each sample, SDS-PAGE was carried out on 11.5% T and 3.3% C acrylamide (Biorad Acrylamide) homogeneous gels. The running buffer was 25 mM Tris, 192 mM glycine, 0.1% SDS. The running conditions were 11°C, 600V constant voltage, 20 mA/gel, 60W for 15 min and 11° C, 600 V constant voltage, 40 mA/gel, 80W for about 2.5 h. The molecular weight markers were from the Low Mr Electrophoresis Calibration Kit (GE-Healthcare). The gels were automatically stained using Processor Plus (Amersham Biosciences) with freshly prepared Neuhoff stain (Colloidal Coomassie Blue) [32] and, after image acquisition, they were dried in a GD 2000 Vacuum Gel Drier System (GE Healthcare).

2.8 *Image analysis and statistical analysis*

2-DE gels were digitized with Personal Densitometer SI (Amersham Biosciences). Image analysis and spot detection were performed with Progenesis PG200 software (Non Linear Dynamics). Spot detection was automatically performed by using the algorithm named "2005 detection". After the establishment of some user seeds, matching was automatically performed and manually verified. Two analytical replicates of 2-DE maps of extracellular proteins obtained from each of the three biological replicates were performed. Only spots present in at least five out of six replicates were identified.

2.9 *Protein identification of 2DE gels*

Enzymatic digestion was carried out with 200 ng of trypsin in 50 µL of 10 mM NH_4CO_3 buffer, pH 7.8. Gel pieces were incubated at 37°C overnight. Peptides were then extracted by washing the gel particles with 10 mM NH_4CO_3 and 1% formic acid in 50% ACN at room temperature. The resulting peptide mixtures were filtrated using 0.22 PVDF filter from Millipore. The peptide mixtures were analysed by nanoLC-chip MS/MS**,** using a CHIP MS 6520 QTOF equipped with a capillary 1200 HPLC system and a chip cube (Agilent Technologies). After loading, the peptide mixture (8 µL in 0.1% formic acid) was first concentrated and washed at 4 µL/min in 40 nL enrichment column (Agilent Technologies chip), with 0.1% formic acid as eluent. The sample was then fractionated on a C18 reverse-phase capillary column (75 µm x 43 mm in the Agilent Technologies chip) at flow rate of 400 nL/min with a linear gradient of eluent B (0.1% formic acid in 95% ACN) in A (0.1% formic acid in 2% ACN) from 7 to 60 % in 50 min. Doubly and triply charged peptides were selected and analyzed using data-dependent acquisition of one MS scan (mass range from 300 to 2,000 m/z) followed by MS/MS scans of the three most abundant ions in each MS scan. Collision energy (CE) applied during peptide fragmentation is calculated by the sequent empirical equations: CE= 4V/100Da -2,4V. Raw data from nanoLC–MS/MS were analyzed and converted in common spectral file formats (.mgf mascot generic file), using Qualitative Analysis software (Agilent MassHunter Workstation Software, version B.02.00). MASCOT software (www.matrixscience.co) version: 2.4.0 was used for the protein identification against NCBInr database (NCBInr_20120920.fasta; 21582400 sequences; 7401135489 residues), with the taxonomy restriction to *Other Firmicutes* (2926062 sequences). The MASCOT search parameters were: "trypsin" as enzyme allowing up to 3 missed cleavages, carbamidomethyl on cysteine residues as fixed modification, oxidation of methionine and formation of pyroGlu N-term on glutamine were selected as variable modifications, 10 ppm MS/MS tolerance and 0.6 Da peptide tolerance. By data analysis, threshold provided to evaluate quality of matches for MS/MS data was found to be 41.

2.10 *Plasminogen activation assay*

tPA-specific plasminogen activation assay in the presence of *L. reuteri* Lb2 BM DSM 16143 cells or its surface associated proteins were measured as described by Lähteenmäki *et al.* [33]. Overnight grown bacteria were divided into two aliquots: bacteria of the first aliquot were collected and resuspended in PBS (pH 7.1) at a concentration of 1 x 10^{10} bacteria/mL. Bacteria of the second aliquot were incubated in 50 mM Tris-HCl pH 8 for 1 h at 37 °C to release surface associated proteins; after pelleting the bacteria, supernatant was filtrated through a 0.2 µm filter to remove the bacterial cells. Bacteria (20 µl) or surface associated proteins (50 µl) were incubated with 4 µg of human Glu-Plg (American Diagnostica), 2 ng of tPA (Biopool) and 0.45 mM chromogenic substrate of plasmin S-2251 (Kabivitrum) in a total volume of 200 µl. The plasmin formation was verified by evaluation of the increase in plasmin activity assessed by measuring the Optical Density at 405 nm every 30 minutes for a 4-hour long period. Three biological replicates were tested and for every culture two technical replicates were performed. Mean values were subtracted to the values measured at time 0.

3. Results and Discussion

The ability of a probiotic to adhere to the epithelial mucosa is crucial for intestinal colonization [34]**.** It is the first step required for competitive exclusion of entero-pathogens and for intestinal micro-ecology modification. Microbial coloni-

zation of gut epithelium requires that bacteria first bind to extracellular secreted constituents, and then adhere to cell-surface membranes or to extracellular matrix (ECM) proteins [35]. Therefore the adhesiveness of *L. reuteri* Lb2 BM DSM 16143 was tested in two biological models: the enterocyte-like Caco-2 cells and the purified proteins of the ECM.

3.1 *L. reuteri Lb2 BM DSM 16143 adhesion on enterocyte-like Caco-2 cells*

An *in vitro* binding assay was performed to evaluate the adhesion capacity of *L. reuteri* Lb2 BM DSM 16143 on enterocyte-like Caco-2 cells, a model for the intestinal epithelium that displays a good correlation to *in vivo* conditions [36]. The adhesion capacity (AC) is described as the percentage of bacteria adhered to Caco-2 cells relative to the total number of bacteria added. The obtained value for the strain in study is equal to 0,82 ± 0,17 % in comparison to the value 3,2 ± 0,41 % obtained for the positive control *Lactobacillus rhamnosus* GG. Therefore *L. reuteri* Lb2 BM proved to be a moderate adhesive strain and these results prompted us to go more in depth in the evaluation of its adhesion capability.

3.2 *L. reuteri Lb2 BM DSM 16143 adhesion assays on ECM proteins*

The first targets for bacteria attachment are the main extracellular matrix proteins: laminin, fibronectin and collagens I and IV. Therefore, the adhesiveness of *L. reuteri* Lb2 BM DSM 16143 was tested using the purified matrix proteins. The adhesive process is mainly mediated by surface-associated proteins expressed by microorganisms on their cell wall. In the literature this association has been proposed to be pH-dependent: surface proteins are positively charged at acidic pH values and therefore remain associated with the bacterial cell-wall by means of electrostatic interactions. Conversely moonlighting proteins of *L. crispatus* are released from cell surface in neutral or cationic buffers, which also reduces bacterial adhesiveness to ECM proteins [18]. For this reason the experiment was performed both at pH 4 and 8. As shown in Figure 1A, at pH 4 the strain can efficiently adhere to ECM proteins. Some differences in adhesiveness towards the four tested proteins can be observed: adhesion was stronger to collagen I, collagen IV and laminin than to fibronectin. As expected, a strong reduction of *L. reuteri* ability to adhere to ECM proteins was detected at pH 8 (Figure 1B), confirming two hypotheses: 1. the adhesion is mediated by the proteins associated to the cell wall; 2. this association is pH-dependent.

3.3 *SDS-PAGE of L. reuteri Lb2 BM DSM 16143 surface-associated proteins after detachment by alkaline treatment*

In order to further analyze the proteins released at high pH and involved in the adhesion process, bacteria were subjected to an alkaline treatment (1 hour, pH 8) to induce the release of surface-associated proteins. After SDS-PAGE and MALDI TOF-TOF analysis, fourteen different proteins were identified (Table 1). Among these, ten are reported to be surface associated in different bacterial species; furthermore ornithine carbamoyltransferase, elongation factor Tu, putative elongation factor Tu and phosphopyruvate hydratase (α-enolase) are directly involved in the adhesion process. Ornithine carbamoyltransferase is an arginine deiminase (ADI) pathway enzyme that was previously found on the surface of the opportunistic *Staphylococcus epidermidis* [37] and the pathogenic *Clostridium perfrigens* [38], where it acts as a fibronectin-binding adhesin. Elongation factor Tu (EF-Tu), a guanosine nucleotide-binding protein involved in protein synthesis, is responsible for adhesion when surface-exposed.

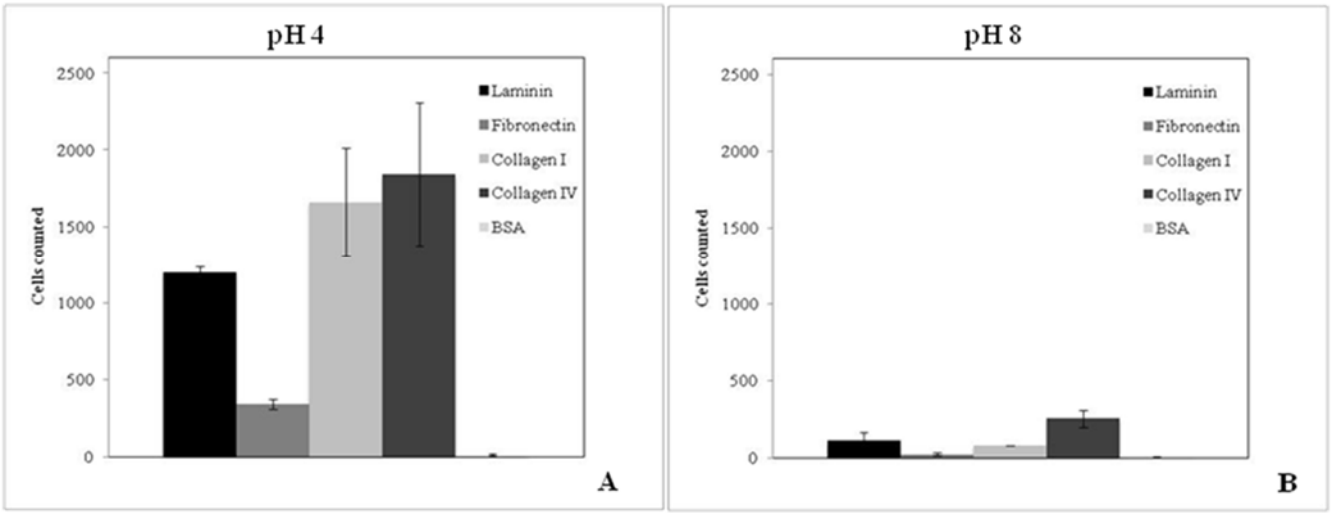


Figure 1. *Extracellular matrix adhesion assay.* Adhesion ability (reported as number of cells recovered after incubation with coated slides for 2 hours) of *L. reuteri* Lb2 BM DSM 16143 on ECM (extracellular matrix) proteins at pH 4 (A) and pH 8 (B). Values are reported as mean ± SEM.

Table 1. MALDI-TOF TOF identifications of weakly bound surface-associated proteins detached after alkaline (pH 8) treatment derived from SDS-PAGE bands. Bands 4, 6, 7 and 8 contained more than one protein, as demonstrated by the MS identification.

Bands	Score	Molecular weight	Identified protein	NCBI nr ID	Identified peptides
1	407	49920	Phosphopyruvate hydratase	gi|18415303	LGANAILAVSLAAAR GIHSFYNLSQQAR GNPTVEAEVYTEAGGVGR GIVPSGASTGEHEAVELR VDFQEFMIMPVGAPTVR
2	59	46790	NLP/P60 protein	gi|148544580	QSQWGDWYLFGNDGR
3	842	43405	Elongation factor Tu	gi|148543883	GISHDQIQR TLDLGEAGDNVGVLLR HYAHIDAPGHADYVK TKPHVNIGTIGHVDHGK GITINTAHVEYETEKR DLLSEYDFPGDDVPVVR TDLVDDDELVDLVEMEVR
4	226	37650	Ornithine carbamoyltransferase	gi|148543661	SFLTLADFNTR VLGGMFDGIEYR EMEVTDEVFESEHSVVFR
	225	36615	Alcohol dehydrogenase	gi|158544709	TVLDGIIVAGSLVGTR VGGVHAAVVTAVSASAFDQAVDSLRPDGK
	95	36284	Mannitol dehydrogenase	gi|45268465	EEIPADAYDIVVEAVGLPATQEQALAAAAR
5	70	31999	Elongation factor Ts	gi|148543917	DVAMHVAAINPEFMTR
6	148	29115	30S ribosomal protein S2	gi|14853916	FLGGIEDMPR
	216	28078	Putative elongation factor Tu	gi|22266054	TLDLGEAGDNVGVLLR
7	73	26162	Phosphoglyceromutase	gi|148543385	YGDEQVHIWR
	115	25849	Dehydratase, medium subunit	gi|148544952	IHYQAISAIMHIR
	57	24681	30S ribosomal protein S3	gi|148544690	IESYSDGTVPLHTLR
8	57	23947	Propanediol utilization protein	gi|148544946	SENFTLGIDAPIR
	156	22929	30S ribosomal protein S4	gi|148543735	QFSNLFVR EGTHGANFMALLER

Both the cell wall/extracellular localization and the involvement in the adhesion process of EF-Tu are well documented [39]: in *Mycoplasma pneumoniae* EF-Tu binds to fibronectin [40], and in *Lactobacillus johnsonii* it is able to bind mucin and thus intestinal epithelial cells, also displaying immune-modulatory properties [13]. Also α-enolase has widely been reported to be surface located in both commensal and pathogenic bacteria: in *L. plantarum* LM3 it is responsible for specifically binding human fibronectin [41], while in streptococci it mediates laminin binding [42]. Both EF-Tu and α-enolase, together with phosphoglycerate mutase, were also proved to be involved in plasminogen (Plg) binding: EF-Tu binds Plg in *Mycobacterium tuberculosis* [43] while in commensal bacteria, such as *L. crispatus* ST1

and *Bifidobacterium lactis* BI07, this function is mediated by α-enolase [18] and phosphoglycerate mutase [44] respectively.

3.4 *L. reuteri Lb2 BM DSM 16143 2DE extracellular proteome*

Adhesive proteins are mainly surface-located, but, as demonstrated in the previous paragraphs and in the literature, both pH [18] and physiological cell-wall turnover during the logarithmic growth phase [20,21], can give rise to a dynamic exchange of proteins between bacterial surface and extracellular space. The analysis of the extracellular proteome of a probiotic may shed light on its adhesion mechanisms and also confirm the safety of the strain in the tested conditions.

Figure 2 shows the extracellular proteome of *L. reuteri* Lb2 BM DSM 16143 in the 4-7 p*I* range. The 59 detected spots were all identified by nanoLC-chip MS/MS (Table 2, Table S1).

Considering p*I* isoforms, 21 proteins were identified. It is possible to hypothesize that spots 1, 2 and 15 are p*I* isoforms derived from post translational modifications, *i.e.* phosphorylations.

It is interesting to underline that seven of these proteins (phosphopyruvate hydratase, Nlp/P60 protein, elongation factor Tu, elongation factor Ts, alcohol dehydrogenase, mannitol dehydrogenase, phosphoglycerate mutase) were also identified as surface-proteins detached by the treatment at pH 8 described in the previous paragraph. This observation is a further evidence of the dynamic protein exchange between extracellular space and cell-wall.

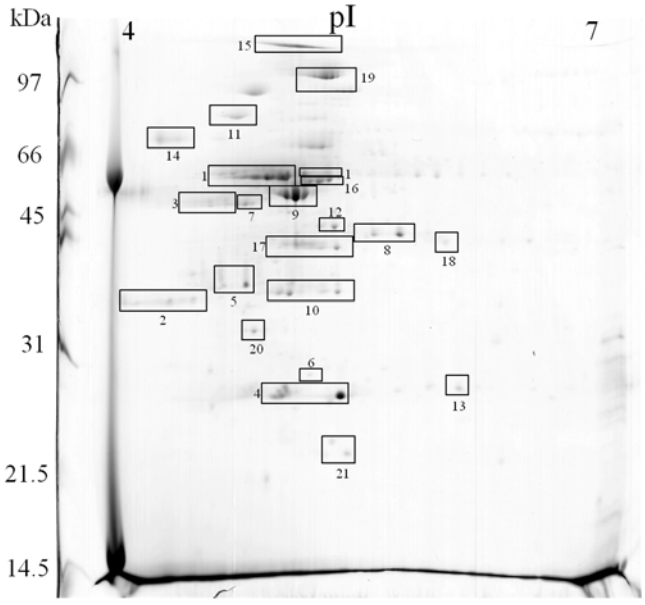


Figure 2. *2DE map of extracellular proteins*. Extracellular proteins maps in the acidic (4–7) p*I* range of *L. reuteri* Lb2 BM DSM 16143 grown in a modified MRS medium and collected at the middle exponential phase. Twenty-one proteins were identified by nanoLC-chip MS/MS from fifty-nine detected spots.

Identified proteins were divided into three main functional families and listed in Table 2: i) cell wall processing enzymes; ii) adhesion-involved proteins; iii) other proteins.

The first functional family is constituted by extracellular cell-wall processing enzymes. Bacterial cell wall is a very dynamic structure, especially during logarithmic growth and cell division [45]: all the proteins and protein domains detected are involved in the cell-wall re-arrangement and in the control of the cell shape during division. This finding is in agreement with the recovery of the bacterial cells during the middle exponential growth phase.

Mannosyl-glycoprotein endo-beta-N-acetylglucosamidase (spot 1) is a widespread enzyme in bacteria and catalyzes the hydrolysis of the glycosidic bond between *N*-acetyl-beta-D-glucosamine residues and the adjacent monosaccharides in peptidoglycan [46]. Nlp/P60 (spot 3) is an endopeptidase family with various roles in the dynamics of the bacterial cell wall such as the control of cell morphology and cell separation process [47]. Apf1-like protein (spot 6), is similar to Apf1 that was originally identified as one of the most abundant proteins in the supernatant of *L. gasseri* 4B2 [48]. It is directly related to the bacterial cell shape: an overproduction produces twisted cells, while a down-regulation causes no visible cell separation sites giving to the cells an elongated shape [49]. The peptidoglycan-binding LysM (spot 2, 4), is a repetitive domain consisting of 40 amino acids present in several peptidoglycan-binding enzymes [50]. It allows the non-covalent attachment of the majority of the extracellular proteins expressed by gram-positive bacteria to the cell-wall. Generally LysM domain is never present in Apf1 proteins. Bath and coworkers [51], nevertheless, demonstrated by bioinformatic analyses, that a LysM domain is present in *L. reuteri* ATCC 55730 Apf1-like protein. Interestingly in the same paper [51] it has been reported that the unknown extracellular protein lr1267 (spot 5), found also in this investigation, is an Apf-like protein.

It has to be taken into account that some bacterial cell-wall lysis enzymes are not only involved in autolysis and cell-wall rearrangement, but also in bacterial interspecies interactions, since they act as general cell-wall lysis factors, also degrading the cell-wall of bacterial species different from the producer one [52]. This is an appreciated feature for a probiotic strain, since it can be useful in killing competitive pathogenic bacteria, sharing the same ecological niche. Therefore their detection in the exoproteome of *L. reuteri* Lb2 BM can add consistence to the hypothesized antibacterial potential of this species [2,3,4].

The second functional group consists of adhesion-involved proteins. It is a well established matter that these proteins play important roles in both pathogenic and probiotic strains for their ability to interact with gut epithelial mucosa [15,20,21]. Most adhesins are cytosolic enzymes having well-proved moonlighting function when surface-bound or extracellularly secreted.

The evidences of the adhesive role of the surface-anchored

Table 2. The table lists the 21 identified proteins from extracellular 2DE maps of *L. reuteri* Lb2 BM 16143 divided in 3 different functional groups: cell wall processing enzymes, adhesion involved proteins and other proteins.

Functional role	Spot	Score	Molecular Weight	Identified protein	NCBI nr ID	N. of peptides	Sequence coverage (%)
Cell wall processing enzymes	1	1991	53593	mannosyl-glycoprotein endo-beta-N-	gi|148544583	41	87
		642	58147	N-acetylmuramoyl-L-alanine amidase	gi|148544581	19	47
	2	367	24886	peptidoglycan-binding LysM	gi|148544536	7	34
	3	796	46733	NLP/P60 protein	gi|148544580	24	55
	4	874	21648	peptidoglycan-binding LysM	gi|148543651	18	75
	5	435	26683	unknown extracellular protein lr1267	gi|68160846	9	34
	6	117	21600	Apf1-like protein	gi|33112857	4	16
Adhesion-involved proteins	7	1095	48010	phosphopyruvate hydratase (α-enolase)	gi|194468183	28	76
	8	1411	35971	glyceraldehyde 3-phosphate dehydrogenase	gi|184153036	38	94
	9	1518	43405	elongation factor Tu	gi|148543883	38	73
	10	1096	31999	elongation factor Ts	gi|148543917	36	60
	11	1529	67171	molecular chaperone DnaK	gi|148543938	57	77
	12	687	42934	phosphoglycerate kinase	gi|184153037	30	74
	13	234	26105	phosphoglyceromutase	gi|148543385	8	48
	14	1432	48717	trigger factor	gi|148543884	43	60
Other proteins	15	3558	167804	dextransucrase	gi|184153923	109	73
	16	523	55934	sucrose phosphorylase	gi|148544754	18	47
	17	500	35941	mannitol dehydrogenase	gi|45268465	12	49
	18	857	36102	alcohol dehydrogenase	gi|148544709	26	55
	19	1130	91346	phosphoketolase	gi|148544892	32	39
	20	233	27144	hypothetical protein Lreu_0552	gi|148543787	7	25
	21	383	20771	ribosome recycling factor	gi|148543919	17	85

elongation factor Tu has already been described in the previous paragraph. It is interesting to underline that EF-Tu (spot 9) has been detected also in the extracellular proteome of *Bifidobacterium animalis* subsp. *lactis* where it acts as moonlighting protein with adhesion roles contributing to the probiotic features of the strain [53]. EF-Ts (spot 10) has been classified as signal peptide-lacking exoprotein in both *Staphylococcus aureus* [54] and *Bacillus anthracis* [55] and as surface protein in *Lactococcus lactis* NZ9000 [56]. Also trigger factor (spot 14) was described to be exposed on the surface of *Lactobacillus plantarum* 299v [17]. It has been reported that surface adhesin P1 from *Streptococcus mutans* cannot efficiently work in the absence of both trigger factor and DnaK (spot 11) [57]. Furthermore, in *L. reuteri* NCIB 11951, a collagen I-binding protein that shares high sequence homology with *E. coli* trigger factor has been described [58].

The remaining identified proteins with adhesive function are moonlighting glycolytic enzymes. GAPDH (spot 8) is able to bind fibronectin, mucin and plasmin [59,60]; it has been found exposed on the surface of several *Lactobacillus* species (*L. gallinarum, L. gasseri, L. johnsonii, L. amylovorus, L. acidophilus* and *L. crispatus*) where it exerts the activity of fibronectin binding [61]. Phosphoglycerate kinase (PGK) (spot 12) has been referred to be extracellularly locat-

ed in *Lactobacillus rhamnosus* GG [62] and surface-associated in *Lactococcus lactis* IL1403 [63]. In eukaryotic cells it acts in the extracellular district as disulphide reductase and plasmin reductase [64]. Both activities are of interest in a probiotic strain since this enzyme (supposing the same biological function in prokaryotes) could contribute to the antioxidant effect of *L. reuteri* Lb2 BM and to the depletion from the extracellular environment of plasmin, a metabolite potentially able to activate proteolytic cascades [65].

Among the third protein group, ribosome recycling factor (spot 21) was previously described in streptococci as an extracellular located protein [66], while sucrose phosphorylase (spot 16) and dextransucrase (spot 15) are correlated to the probiotic potential of LAB. In *Lactobacillus acidophilus* NCFM sucrose phosphorylase expression is induced by sucrose. This enzyme is involved in the utilization and catabolism of human undigested sugars like fructooligosaccharides (FOS), considered prebiotic compounds, thus stimulating the growth of the producing strain and promoting competition of beneficial bacteria in the human gut [67]. Very recently [68] sucrose phosphorylase has been reported as a clue enzyme in generating fructose units, building blocks for the biosynthesis of the prebiotic functional sugars (*i.e*. FOS). Dextransucrase is an extracellular enzyme produced by several genera of lactic acid bacteria such as *Leuconostoc* and *Streptococcus* [69]. It is a glycosyltransferase catalyzing the cleavage of sucrose into glucose and fructose and the following polymerization of these glucosidic units into prebiotic molecules [70]. The finding of these proteins in the extracellular environment of *L. reuteri* Lb2 BM DSM 16143 underlines the probiotic potential of this strain for successfully colonizing the gut ecological niche.

The safety of the strain in the tested conditions is confirmed by the absence of potentially dangerous proteins such as extracellular serine protease and gelatinase recently described in the extracellular space of a strain of *Enterococcus faecalis*, a lactic acid bacterium whose use in food industry is still controversial [28].

3.5 *Potential of L. reuteri Lb2 BM DSM 16143 surface-associated proteins in plasminogen activation*

As reported in the previous paragraphs, some of the adhesive proteins identified in both the cell-wall and the extracellular space, also display the capability to bind plasminogen (Plg). Once localized to the bacterial surface, plasminogen can act as a cofactor in adhesion, or, following activation to plasmin (Plm), provide a source of potent proteolytic activity [71]. The activation to Plm may lead to damages to host tissues, opening the way for invasion by potential pathogenic bacteria, present in the same ecological niche. The conversion of Plg into Plm is a typical feature of several pathogenic bacteria, like staphylococci and streptococci, that possess staphylokinases and streptokinases which can act as specific activators [72].

The ability to convert Plg into Plm of *L. reuteri* Lb2 BM DSM 16143 was tested for both the whole bacterial cells and for surface-associated proteins recovered after cell treatment at pH 8. As shown in Figure 3 neither the whole cells nor the released proteins were able to induce Plm formation indicating the absence of the bacterial specific activators. The same assay was performed in presence of the mammalian tissue type Plg activator (tPA). As shown in Figure 3 in this condition both the whole bacterial cells and the surface-associated proteins were able, in presence of tPA, to produce Plm. This indicates that *L. reuteri* Lb2 BM DSM 16143 cells can immobilize Plg on its surface and that this property seems to depend on the surface-anchored moonlighting proteins.

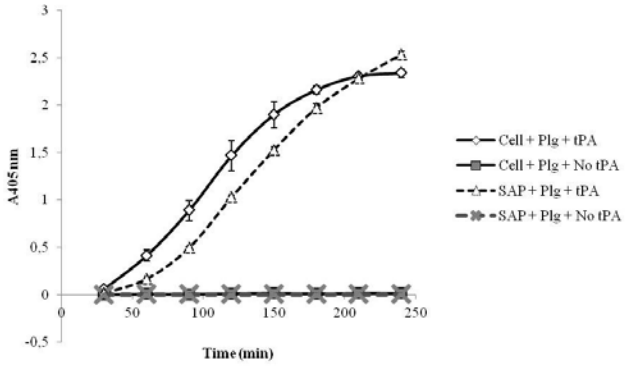


Figure 3. *Evaluation of Plg-binding ability.* Plasminogen activation assay of entire cells and surface-associated proteins (SAP) with and without mammalian tPA activator. Values were collected for 4 hours measuring absorbance at 405nm and are reported as mean ± SEM.

Even if the involvement of probiotic bacteria in the Plg/Plm system is yet to be elucidated, the proved ability of *Lactobacillus reuteri* to bind Plg on its surface could have a role in enhancing the colonization process at human epithelial surfaces, as already suggested by Candela and co-workers [44], or in localized dissolution of fibrin clots [22]. Furthermore the presence of Plg-binding proteins, either surface-bound or extracellularly released, could interfere in the interaction between Plg and gastrointestinal pathogens present in the same ecological niche, such as *Helicobacter pylori* and *Salmonella* sp., since probiotic bacteria could act as quenchers, as previously suggested by other authors [20].

4. Concluding Remarks

Integrating proteomic analyses with physiological studies proved to be a winning strategy to characterize a probiotic strain.

The ability of *L. reuteri* Lb2 BM DSM 16143 to adhere to both Caco-2 cells and extracellular matrix proteins, especially collagen and laminin, was proven by *in vitro* adhesion assays. The involvement of weakly cell-surface anchored proteins in this phenomenon was demonstrated by the drastic decrease in adhesion to ECM induced by their detachment obtained with incubation at pH 8. SDS-PAGE analysis

followed by MALDI-TOF TOF mass spectrometry revealed that the proteins involved in this process were mainly moonlighting proteins whose implication in adhesion has already been proved in both pathogenic and probiotic bacteria. The loss of adhesion ability induced by alkaline pH does not constitute a problem *in vivo* considering that the pH of colon, the main district of probiotic action, is around 6-7. However, *in vivo*, the strain adhesion potential could further be improved by the physiological activation of metabolic pathways leading to the expression of non-protein adhesive factors, such as the exopolysaccharides [73].

Extracellular proteomic profiles highlighted the presence of adhesive proteins also in this cell district. Among these are present glycolytic enzymes (GAPDH, PGM, PGK) as well as stress proteins (DnaK), trigger factor and protein synthesis enzymes (EF-Tu and Ts). The latter enzymes can also have an immune-modulating action. Since adhesion and immune system modulation are often related, the analysis of the extracellular proteomes proves to be a useful tool to obtain an overall picture of probiotic-host interaction, mainly in *L. reuteri* in which a strict link between adhesion and regulatory T-cell induction has been demonstrated [74]. Extracellular proteomics also revealed the presence of potential lytic factors, namely cell-wall hydrolases, useful in enhancing the antibacterial potential of the strain in the gut ecosystem and the total absence of virulence proteins. Furthermore proteins involved in the probiotic potential such as exopolysaccharide or fructooligosaccharide related enzyme have also been detected.

Some of the identified adhesins were reported to be able to link plasminogen; the strain was able to bind but not to activate plasminogen into plasmin without supplementation of exogenous activators like tPA. The expression of these Plg receptors is an useful feature because they can improve the adhesion ability of the tested strain and they can have a role in quenching Plg, subtracting it to pathogenic bacteria.

Although *in vivo* tests at different physiological pHs and in different individual enterotypes [75] are necessary to prove the adhesive and persistence capabilities of this strain, the present results enable us to assert that *L. reuteri* Lb2 BM can colonize the intestine, conferring to the host the benefits derived from its probiotic features.

5. Supplementary material

Supplementary data and information is available at: http://www.jiomics.com/index.php/jio/rt/suppFiles/143/0

Table S1 contains the sequences of peptides identified by nanoLC-chip MS/MS performed on the spots present in the extracellular 2DE maps.

Acknowledgements

This work has been supported by a grant of Bioman SRL who funded the post-doc work of Cristina Lamberti in the "Accordo Università-Regione: azioni A". We are indebted with Marco Cietto for Plasminogen activation experiments and with the Italian Proteomic Association (ItPA) for supporting the Marco Cietto's travel expenses to Finland.

References

[1] J.K. Spinler, M. Taweechotipatre, C.L. Rognerud, C.N. Oub, S. Tumwasorn, J. Versalovic, Anaerobe 14 (2008) 166-171. DOI: 10.1016/j.anaerobe.2008.02.001

[2] L.T. Axelsson, T.C. Chung, W.J. Dobrogosz, S.E. Lindgren, Microb Ecol Health Dis 2 (1989) 131-136. DOI: 10.3109/08910608909014021

[3] T. Kabuchi, T. Saito, Y. Kawai, J. Uemura, T. Itoh, Int J Food Microbiol 34 (1997) 145-156. DOI: 10.1016/S0168-1605(96)01180-4

[4] A. Holtzel, M.G. Ganzle, G.J. Nicholson, W.P. Hammes, G. Jung, Angewandte Chemie-International Edition 39 (2000) 2766-2768. DOI: 10.1002/1521-3773(20000804)39:15<2766::AID-ANIE2766>3.3.CO;2-7

[5] A. Hernandez-Mendoza, A.F. Gonzalez-Cordova, B. Vallejo-Cordoba, H.S. Garcia, J Basic Microbiol 51 (2011) 263-268. DOI: 10.1002/jobm.201000119

[6] J. Li, W. Wang, S.X. Xu, N.A. Magarvey, J.K. McCormick, PNAS 108 (2011) 3360-3365. DOI: 10.1073/pnas.1017431108

[7] J.A. Younes, H. C. van der Mei, E. van den Heuvel, H. J. Busscher, G. Reid, Plos One 7 (2012) e36917. DOI: 10.1371/journal.pone.0036917

[8] S. E. Jones, J. Versalovic, BMC Microbiol 9 (2009) 35. DOI: 10.1186/1471-2180-9-35

[9] C. Lamberti, E. Mangiapane, A. Pessione, R. Mazzoli, C. Giunta, E. Pessione, Proteomics 11 (2011) 2212-2221. DOI: 10.1002/pmic.201000747

[10] E. Galano, E.Mangiapane, J. Bianga, A. Palmese, E. Pessione, J. Szpunar, R. Lobinski, A. Amoresano, Mol Cell Proteomics (2013), in press

[11] J. Lukic, I. Strahinic, B. Jovcic, B. Filipic, L. Topisirovic, M. Kojic, J. Begovic, Appl Environ Microbiol 78 (2012) 7993-8000. DOI: 10.1128/AEM.02141-12

[12] G. Alp, B. Aslim, Z. Suludere, G. Akca, Microbiol Immunol 54 (2010) 658-665. DOI: 10.1111/j.1348-0421.2010.00227.x

[13] D. Granato, F. Perotti, I. Masserey, M. Rouvet, M. Golliard, A. Servin, D. Brassart, Appl Environ Microbiol 65 (1999) 1071-1077

[14] E. Pessione, Frontiers Cell Infect Microbiol 2 (2012) 86. DOI: 10.3389/fcimb.2012.00086

[15] D. Granato, G.E. Bergonzelli, R.D. Pridmore, L. Marvin, M. Rouvet, I.E. Corthesy-Theulaz, Infect Immun 72 (2004) 2160-2169. DOI: 10.1128/IAI.72.4.2160-2169.2004

[16] G.E. Bergonzelli, D. Granato, R.D. Pridmore, L.F. Marvin-Guy, D. Donnicola, I.E. Corthesy-Theulaz, Infect Immun 74 (2006) 425-434. DOI: 10.1128/IAI.74.1.425-434.2006

[17] H.C. Beck, S.M. Madsen, J. Glenting, J. Petersen, H. Israelsen, M.R. Nørrelykke, M. Antonsson, A.M. Hansen, FEMS Microbiol Lett 297 (2009) 61-66. DOI: 10.1111/j.1574-6968.2009.01662.x

[18] J. Antikainen, V. Kuparinen, K. Lahteenmaki, T. Korhonen T, J Bacteriol 189 (2007) 4539-4543. DOI: 10.1128/JB.00863-07

[19] J. C. Jeffery JC, Trends Genet 19 (2004) 415-417. DOI: 10.1016/S0168-9525(03)00167-7

[20] B. Sanchez, P. Bressolier, M.C. Urdaci, FEMS Immunol Med Microbiol 54 (2008) 1-17. DOI: 10.1111/j.1574-695X.2008.00454.x

[21] B. Sanchez, N. Saad, J.M. Schmitter, P. Bressollier, M.C. Urdaci, J Microbiol Biotechnol 20 (2010) 978-984. DOI: 10.4014/jmb.0911.11007

[22] V. Kainulainen, V. Loimaranta, A. Pekkala, S. Edelman, J. Antikainen, R. Kylväjä, M. Laaksonen, L. Laakkonen, J. Finne, T.K. Korhonen, J Bacteriol 194 (2012) 2509-2519. DOI: 10.1128/JB.06704-11

[23] C.K. Yang, H.E. Ewis, X. Zhang, C.D. Lu, P.J. Hu, Y. Pan, A.T. Abdelal, P.C. Tai, J. Bacteriol. 193 (2011) 5607-5615. DOI: 10.1128/JB.05897-11

[24] S. Leeber, I. Claes, H.L.P. Tytgat, T.L.A. Verhoeven, E. Marien, I. von Ossowski, J. Reunanen, A. Palva, W.M. de Vos, S.C.J. De Keersmaecker, J. Vanderleyden, Appl Environ Microbiol (2012) 185-193. DOI: 10.1128/AEM.06192-11

[25] B. Westerlund, P. Kuusela, J. Risteli, L. Risteli, T. Vartio, H. Rauvala, R. Virkola, T.K. Korhonen, Mol Microbiol 3 (1989) 329-337. DOI: 10.1111/j.1365-2958.1989.tb00178.x

[26] T. Toba, R. Virkola, B. Westerlund, Y. Björkman, J. Sillanpää, T. Vartio, N. Kalkkinen, T.K. Korhonen, Appl Environ Microbiol 61 (1995) 2467-2471

[27] R. Virkola, K. Lähteenmäki, T. Eberhard, P. Kuusela, L. van Alphen, M. Ullberg, T.K. Korhonen, J Infect Dis 173 (1996) 1137-1147

[28] A. Pessione, C. Lamberti, L. Cocolin, S. Campolongo, A. Grunau, S. Giubergia, L. Eberl, K. Riedel, E. Pessione, Proteomics 12 (2012) 431-447. DOI: 10.1002/pmic.201100468

[29] U.K. Laemmli, Nature 227 (1970) 680-685. DOI: 10.1038/227680a0

[30] J.P. Hewitson, Y.M. Harcus, R.S. Curwen, A.A. Dowle, A.K. Atmadja, P.D. Ashton, A. Wilson, R.M. Maizels, Mol Biochem Parasitol 160 (2008) 8-21. DOI: 10.1016/j.molbiopara.2008.02.007

[31] E. Pessione, R. Mazzoli, M.G. Giuffrida, C. Lamberti, E. Garcia -Moruno, C. Barello, A. Conti, C. Giunta, Proteomics 5 (2005) 687-698. DOI: 10.1002/pmic.200401116

[32] V. Neuhoff, N. Arold, D. Taube, W. Ehrhardt, Electrophoresis 9 (1988) 255-262

[33] K. Lähteenmäki, R. Virkola, R. Pouttu, P. Kuusela, M. Kukkonen, T.K. Korhonen, Infect Immun 63 (1995) 3659-3664

[34] M. Alander, R. Korpela, M. Saxelin, T. Vilpponen-Salmela, T. Mattila-Sandholm, A. von Wright, Lett Appl Microbiol 24 (1997) 361-364. DOI: 10.1046/j.1472-765X.1997.00140.x

[35] M.H. Coconnier, M.F. Bernet, S. Kerneis, G. Chauviere, J. Fourniat, A.L. Servin, FEMS Microbiol Lett 110 (1993) 299-305. DOI: 10.1016/0378-1097(93)90119-M

[36] C.N. Jacobsen, V. Rosenfeldt, A.E. Hayford, P.L. Moller, K.F. Michaelsen, A. Paerregaard, Appl Environ Microbiol 65 (1999) 4949-4956

[37] M. Hussain, G. Peters, G.S. Chhatwal, M. Herrmann, Infect Immun 67 (1999) 6688-6690

[38] S.I. Alam, S. Bansod, R.B. Kumar, N. Sengupta, L. Singh, BMC Microbiol 9 (2009) 162. DOI: 10.1186/1471-2180-9-162

[39] M.P. Velez, S.C.J. De Keersmaecker, J. Vanderleyden, FEMS Microbiol Lett 276 (2007) 140-148. DOI: 10.1111/j.1574-6968.2007.00908.x

[40] S. Balasubramanian, T.R. Kannan, J.B. Baseman, Infect Immun 76 (2008) 3116-3123. DOI: 10.1128/IAI.00173-08

[41] C. Castaldo, V. Vastano, R.A. Siciliano, M. Candela, M. Vici, L. Muscariello, R. Marasco, M. Sacco, Microb Cell Factories 8 (2009) 14. DOI: 10.1186/1475-2859-8-14

[42] S. Bergmann, M. Rohde, K.T. Preissner, S. Hammerchmidt, Thrombosis and Haemostasis 94 (2005) 304-311. DOI: 10.1160/TH05-05-0369

[43] W. Xolalpa, A.J. Vallecillo, M. Lara, G. Mendoza-Hernandez, M. Comini, R. Spalle, M. Singh, C. Espitia, Proteomics 7 (2007) 3332-3341. DOI: 10.1002/pmic.200600876

[44] M. Candela, S. Bergmann, M. Vici, B. Vitali, S. Turroni, B.J. Eikmanns, S. Hammerschmidt, P. Brigidi, J Bacteriol 189 (2007) 5929-5936. DOI: 10.1128/JB.00159-07

[45] J. Van Heijenoort, Glycobiology 11 (2001) 25R-36R. DOI: 10.1093/glycob/11.3.25R

[46] W. Vollmer, B. Joris, P. Charlier, S. Foster, FEMS Microbiol Rev 32 (2008) 259-286. S. 2008. DOI: 10.1111/j.1574-6976.2007.00099.x

[47] C. Bauerl, C. Pérez-Martinez, F. Yan, D.B. Polk, V. Monedero, J Mol Microbiol Biotechnol 19 (2010) 231-241. DOI: 10.1159/000322233

[48] R. Reniero, P. Cocconcelli, V. Bottazzi, L. Morelli, J General Microbiol 138 (1992) 763-768. DOI: 10.1099/00221287-138-4-763

[49] I. Jankovic, M. Ventura, V. Meylan, M. Rouvet, M. Elli, R. Zink, J Bacteriol 185 (2003) 3288-3296. DOI: 10.1128/JB.185.11.3288.3296.2003

[50] M.S. Turner, L.M. Hafner, T. Walsh, P.M. Giffard, Appl Environ Microbiol 70 (2004) 3673-3680. DOI: 10.1128/aem.70.6.3673-3680.2004

[51] K. Bath, S. Roos, W. Torun, H. Jonsson, FEMS Microbiol Lett 253 (2005) 75-82. DOI: 10.1016/j.femsle.2005.09.042

[52] O. Salazar, J.A. Asenjo, Biotechnol Lett 29 (2007) 985-994. DOI: 10.1007/s10529-007-9345-2

[53] O. Gilad, B. Svensson, A.H. Viborg, B. Stuer-Lauridsen, S. Jacobsen, Proteomics 11 (2011) 2503-2514. DOI: 10.1002/pmic.201000716

[54] M.J. Sibbald, A.K. Ziebandt, S. Engelmann, M. Hecker, A. de Jong, H.J.A. Harmsen, G.C. Raangs, I. Stokroos, J.P. Arends, J.Y.F. Dubois, J.A. van Dijl, Microbiol Mol Biol Rev 70 (2006) 755-788. DOI: 10.1128/MMBR.00008-06

[55] M. Gohar, N. Gilois, R. Graveline, C. Garreau, V. Sanchis,, D. Lereclus, Proteomics 5 (2005) 3696-3711. DOI: 10.1002/pmic.200401225

[56] A. Berlec, P. Zadravec, Z. Jevnikar, B. Strukelj, Appl Environ Microbiol 77 (2011) 1292-1300. DOI: 10.1128/AEM.02102-10

[57] P.J. Crowley, T.B. Seifert, R. Isoda,, M. van Tilburg, M.W. Oli, R.A. Robinette, W.P. McArthur, A.S. Bleiweis, L.J. Brady, Infect Immun 76 (2008) 2456-2468. DOI: 10.1128/IAI.01315-07

[58] P. Aleljung, W. Shen, B. Rozalska, U. Hellman, A. Ljungh, T. Wadstrom, Curr Microbiol 28 (1994) 231-236. DOI: 10.1007/BF01575966

[59] V. Pancholi, V.A. Fischetti, J Exp Medicine 176 (1992) 415-426. DOI: 10.1084/jem.176.2.415

[60] R.A. Alvarez, M.W. Blaylock, J.B. Baseman, Mol Microbiol 48 (2003) 1417-1425. DOI: 10.1046/j.1365-2958.2003.03518.x

[61] V. Hurmalainen, S. Edelman, J. Antikainen, M. Baumann, K. *Lahteenmaki K, T.K. Koh*ronen, Microbiol 153 (2007) 1112-1122. DOI: 10.1099/mic.0.2006/000901-0

[62] B. Sanchez, J.M. Schmitter, M.C. Urdaci, Lett Appl Microbiol 48 (2009) 618-622. DOI: 10.1111/j.1472-765X.2009.02579.x

[63] Y. Katakura, R. Sano, T. Hashimoto, K. Ninomiya, S. Shioya, Appl Microbiol Biotechnol 86 (2010) 319-326. DOI: 10.1007/s00253-009-2295-y

[64] A.J. Lay, X.M. Jiang, O. Kisker, E. Flynn, A. Underwood, R. Condron, P.J. Hogg, Nature 408 (2000) 869-873. DOI: 10.3410/f.1006146.76859

[65] A.J. Lay, X.M. Jiang, E. Daly, L. Sun, F.P. Hogg, J Biol Chem 277 (2002) 9062-9068. DOI: 10.1074/jbc.M111531200

[66] J.C. Wilkins, D. Beighton, K.A.Homer, Appl Environ Microbiol 69 (2003) 5290-5296. DOI: 10.1128/AEM.69.9.5290-5296.2003

[67] R. Barrangou, E. Altermann, R. Hutkins, R. Cano, T.R. Klaenhammer, PNAS 100 (2003) 8957-8962. DOI: 10.1073/pnas.1332765100

[68] J.S. Teixeira, V. McNeill, M.G.Gaenzle, Food Microbiol 31 (2012) 278-284. DOI: 10.1016/j.fm.2012.03.003

[69] R.K. Purama, A. Goyal, Biores Technol 99 (2008) 3635-3642. DOI: 10.1016/j.biortech.2007.07.044

[70] S. Patel, D. Kothari, A. Goyal Food Technology and Biotechnology 49 (2011) 297-303

[71] M.L. Sanderson-Smith, D.M.P. De Oliveira, M. Ranson, J.D. McArthur, J Biomed Biotechnol (2012) 272148. DOI: 10.1155/2012/272148

[72] K. Lahteenmaki, P. Kuusela, T.K. Kohronen, FEMS Microbiol Rev 25 (2001) 531-552. DOI: 10.1016/S0168-6445(01)00067-5

[73] P. Kanmani, R. Satish Kumar, N. Yuvaraj, K.A. Paari, V. Pattukumar, V. Arul, Critical Review in Food Sci Nutr 53 (2013) 641-658. DOI: 10.1080/10408398.2011.553752

[74] H.H. Smits, A. Engering, D. van der Kleij, E.C. de Jong, K. Schipper, T.M.M. van Capel, B.A.J. Zaat, M. Yazdanbakhsh, E.A. Wierenga, Y. van Kooyk, M.L. Kapsenberg, J Allergy Clin Immunol 115 (2005) 1260-1267. DOI: 10.1016/j.jaci.2005.03.036

[75] Q. Aziz, J. Doré, A. Emmanuel, F. Guarner, E.M. Quigley, Neurogastroenterol and Motility 25 (2013) 4-15. DOI: 10.1111/nmo.12046.

JOURNAL OF INTEGRATED OMICS
A METHODOLOGICAL JOURNAL
HTTP://WWW.JIOMICS.COM

ORIGINAL ARTICLE | DOI: 10.5584/jiomics.v3i2.140

Proteomic changes in an extended-spectrum beta-lactamase-producing *Escherichia coli* strain under cefotaxime selection

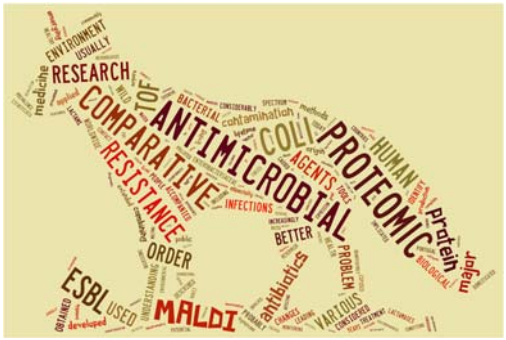
Hajer Radhouani[1-4], Júlio D. Nunes-Miranda[1,2], Ricardo J. Carreira[5], Hugo M. Santos[6], Luís Pinto[1-4], Ricardo Monteiro[1-4], Carlos Carvalho[1,2], Patrícia Poeta[3,4], Carlos Lodeiro[6], Jose-Luis Capelo-Martinez[6], Gilberto Igrejas[1,2]

[1]Institute for Biotechnology and Bioengeneering, Center of Genomics and Biotechnology, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal; [2]Department of Genetics and Biotechnology, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal; [3]Center of Studies of Animal and Veterinary Sciences, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal; [4]Veterinary Science Department, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal; [5]Biomolecular Mass Spectrometry Unit, Department of Parasitology, Leiden University Medical Center, Leiden, the Netherlands; [6]REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal.

Received: 15 May 2013 **Accepted:** 18 September 2013 **Available Online**: 09 October 2013

Abstract

Proteomics can be used to study the metabolic pathways and mechanisms involved in antimicrobial resistance. The aim of this comparative proteomic study was to establish the overall changes in the proteome of a naturally occurring ESBL-producing *E. coli* strain (C5478) stressed with its minimal inhibitory concentration (2 µg/mL) of cefotaxime, compared to the proteome of the same strain without antimicrobial stress, by using 2-D gel electrophoresis (2-DE) followed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). The comparative proteomic analysis revealed that the abundance of numerous protein species changed in the strain stressed by CTX compared to the non-stressed wild-type strain. A total of 188 spots were excised from the 2-DE gel of the wild-type strain, 112 of which were successfully identified by MALDI-TOF MS, representing 110 different proteins. Concerning the 2-DE gel of the CTX-stressed bacteria, 171 spots were excised and 156 were identified, representing 143 different proteins. The proteins identified in both strains were categorized according to their biological functions. These proteins were involved in metabolism, protein synthesis, cell division, stress responses, and antimicrobial resistance, among others. These findings will be helpful to further understand not only the antimicrobial resistance mechanisms, but also the role of wild animals as reservoirs in spreading antimicrobial-resistant bacteria into the environment.



Keywords: Proteomics; Cefotaxime; Antimicrobial resistance; *Escherichia coli*; MALDI-TOF-MS.

1. Introduction

Cefotaxime is one of the most widely used parenteral third generation cephalosporins in medicine [1]. However, shortly after the introduction of the first β-lactam antibiotics, some bacteria were found to be resistant, comprising the long-term usefulness of these drugs [2]. A common resistance mechanism is the expression of β-lactamase enzymes by bacteria, which inactivate the antibiotics by cleaving the β-lactam core. Resistant extended-spectrum β-lactamases (ESBL)-producing bacteria are thus a major clinical and public health concern in various countries, including Portugal, as they are commonly implicated in human infections [3].

***Corresponding author:** Gilberto Igrejas. Institute for Biotechnology and Bioengineering, Centre of Genetics and Biotechnology, Department of Genetics and Biotechnology, University of Trás-os-Montes and Alto Douro, 5001-801 Vila Real, Portugal. Tel: + 351 259 350 530; Fax: + 351 259 350 572. E-mail Address: gigrejas@utad.pt

The appearance of multiresistant bacterial strains in humans or animals in clinical situations can also lead to contamination of the environment [4]. The production of ESBLs by *Enterobacteriaceae*, and especially *Escherichia coli*, has been repeatedly described in apparently healthy animals in the last few years [5]. Wildlife is not usually exposed to antimicrobial agents in clinical use but may acquire antimicrobial-resistant bacteria through contact with humans, domesticated animals and the environment. Water polluted with faeces is probably the most significant route of contamination. Monitoring the prevalence of resistance in indicator bacteria such as *E. coli* in wild animals could show whether wildlife serves as an environmental reservoir or "melting pot" of bacterial resistance.

Proteomics techniques can be used to quantitatively compare the expression of protein in different conditions to obtain a better understanding of the cellular processes affecting their expression and/or in which they are involved. Mostly two-dimensional electrophoresis (2-DE) is used for extensive protein separation before identification. Mass spectrometry (MS) in combination with various protein resolution methods and bioinformatics tools is used routinely in proteomics research [6]. For example, 2-DE and MALDI-TOF MS have been used several times to identify antimicrobial-related proteins in *E. coli* strains [7-10].

The scientific community is developing several tools in order to address the problem of antimicrobial resistance [11, 12]. Proteomics research offers major opportunities to characterize bacterial pathogens, elucidate disease mechanisms and to identify new diagnostic markers and therapeutic targets [13]. Comparative proteomic analyses of different bacterial species have shown that a subproteome related to resistance mechanisms changes when some strains are challenged with antimicrobials, e.g., in *E. coli* isolates treated with piperacillin/tazobactam [14], nalidixic acid [15], tetracycline [16, 17], streptomycin [18], chloramphenicol [19] and ampicillin [17]; in *Salmonella* treated with fluoroquinolones [20]; in *Stenotrophomonas maltophilia* treated with imipenem [21]; and in *Acinetobacter baumannii* treated with tetracycline and imipenem [22].

To better understand the universal pathways that form barriers to antimicrobial agents, the aim of this comparative proteomic study was to establish the overall changes between a wild-type strain (ESBL-producing *E. coli* strain C5478) and the same strain stressed with cefotaxime.

2. Material and methods

2.1. *Isolation of bacteria from fox faeces*

The *E. coli* strain C5478, referred to as the wild-type strain, was characterized in a previous study [23]. Red fox faecal samples were collected from February-2008 to March-2009 in North of Portugal (where these animals live in the wild) during hunts of red foxes. This kind of hunting is organized all of the years during a short period of time having like main objective the ecological control of the animal population and is supervised by the Agriculture, Rural development and Fishery Ministry of Portugal under the Decree-Law no. 202/2004. The animals were obtained in collaboration with National Corporation of Forest Rangers. As far as we know, none of the animals had received antimicrobials. Faecal samples were seeded into Levine agar supplemented with CTX (2 μg/mL).

2.2. *Genetic characterization of E. coli C5478*

C5478 showed resistance to ampicillin, amoxicillin + clavulanic acid, CTX, ceftazidime, aztreonam, tetracycline, sulfamethoxazole/trimethoprim, streptomycin, nalidixic acid, ciprofloxacin and chloramphenicol. The β-lactamase genes detected in C5478 are $bla_{\text{SHV-12}} + bla_{\text{TEM-1b}}$. C5478 also harbours the genes *tet*(A) and *sul*2 and contained the non-classical class 1 integron (*intI*1-*dfr*A12-*orf*F-*aad*A2-*cml*A1-*aad*A1-*qacH*-IS*440*-*sul*3) with the PcH1 promoter. The presence of two amino acid changes in GyrA (S83L and D87G) and one in ParC (S80I) were responsible for the ciprofloxacin and nalidixic acid resistance phenotypes detected in C5478. This isolate was typed as sequence type ST1086, phylogroup A and carried the *fim*A virulence gene [23].

2.3. *Culture conditions and total protein extraction*

The "stressed strain", was obtained by culturing the "wild-type strain" (ESBL-producing *E. coli* strain C5478) into brain heart infusion (BHI) broth containing 2 μg/mL CTX (Oxoid, Cambridge, UK). For this study, *Escherichia coli* strain K-12 was used as the negative control. In the post-exponential phase (when the optical density of the culture OD_{540nm} = 6 corresponding to 2–3 × 10^9 cells/mL), the cells were pelleted at 10,000 rpm at 4ºC for 3 min. The pellet was resuspended in an equal volume of pre-warmed phosphate-buffered saline (PBS) pH 7.4 [24]. After a second centrifugation, the pellet was resuspended in 0.2 mL of SDS sample solubilization buffer. The sample was sonicated with an ultrasonic homogenizer (three 10-s bursts at 100 W and 4 °C). The disrupted cells were centrifuged in an Eppendorf microfuge at maximum speed (14,000 *g*) for 30 minutes at 4°C. For SDS-PAGE experiments the supernatant was collected and resuspended in an equal volume of buffer containing 0.5 M Tris HCl pH 8.0, glycerol, SDS and bromophenol blue.

2.4. *SDS-PAGE and staining*

SDS-PAGE was performed on vertical gels (12.5%T and 0.97%C) in a Hoefer™ SE 600 Ruby® (Amersham Biosciences) unit, according to the procedure described by Laemmli [25] with some modifications. Electrophoresis was carried out at constant amperage of 30 mA per gel until the dye-front reached the bottom of the gels that were then stained with Coomassie Brilliant Blue R250 and washed in water overnight. Gels were then fixed in 6% trichloroacetic acid for

4 h and in 5% glycerol for 2 h [26].

2.5. *Two-dimensional electrophoresis and proteomics analysis*

Immobiline™ pH Gradient (IPG) technology (Görg et al., 2007) was used for 2-DE [27]. Protein samples from the wild -type strain were compared to protein samples from the stressed strain by subculture at the minimal inhibitory concentration of CTX. For isoelectric focusing (IEF), precast 13-cm IPG strips with linear gradient of pH 3-10 (pH 3–10 NL, Amersham Biosciences, UK) were passively rehydrated overnight in a reswelling tray with rehydration buffer (8M urea, 1% CHAPS, 0.4% DTT, 0.5% carrier ampholyte IPG buffer pH 3-10) at room temperature. IPG strips were covered with DryStrip Cover Fluid (Plus One, Amersham Biosciences). Lysis buffer (9.5 M urea, 1% (w/v) DTT, 2% (w/v) CHAPS, 2% (v/v) carrier ampholytes (pH 3-10) and 10 mM Pefabloc® proteinase inhibitor) was added to the bacteria (1:1, v/v). Samples containing a total of 100 µg of protein were cup loaded on to the rehydrated IPG strips [28] and proteins were focused sequentially at 500 V for 1 h, 1000 V for 8 h, 8000 V for 3 h and finally 8000 V for 1 h 30 min incremented to 23135 V/h on an Ettan™ IPGPhor II™ (Amersham Biosciences, Uppsala, Sweden). Seven IEF replicate runs were performed according to Görg [27] and the GE Healthcare protocol for 13-cm IPG strips pH 3-10 to optimize the running conditions, resulting in a final 13 h 30 min run. Focused IPG strips were then stored in plastic bags at -80°C. Before running the second dimension, strips were incubated twice 15 min in equilibration buffer (6 M urea, 30% (w/v) glycerol, 2% (w/v) SDS in 0.05 M Tris-HCl buffer pH 8.8). For the first equilibration 1% DTT and bromophenol blue were added to the equilibration buffer and for the second 4% iodoacetamide and bromophenol blue were added. The equilibrated IPG strips were gently rinsed with SDS electrophoresis buffer, blotted to remove excessive buffer, and then applied to 12.52% polyacrylamide gels in a Hoefer™ SE 600 Ruby® (Amersham Biosciences) unit. IPG strips were inserted in the stacking gel for SDS-PAGE using a method modified to improve protein resolution [25, 26]. Gels were fixed in 40% methanol/10% acetic acid for 1 h then stained overnight in Coomassie Brilliant Blue G-250 [24]. Coomassie-stained gels were scanned on a flatbed scanner (Umax PowerLook 1100; Fremont, CA, USA) and the resulting digitized images were analyzed using Image Master 5.0 software (Amersham Biosciences; GE Healthcare).

2.6. *Protein digestion*

The excess of staining was removed from the gel, and the proteins were located. The spots from 2D-PAGE were excised manually with sterile tips under sterile dust-free environment and transferred to individual tubes for enzymatic digestion. For the protein digestion, the gel pieces were incubated with trypsin at 37°C overnight. Concerning the peptide extraction, the reaction media was acidified with Formic acid (FA) to stop the enzymatic digestion. Then, the supernatant containing the digested peptides was removed to another tube, and additional peptides were extracted from the gel with a 50 % acetonitrile/0.1 % TFA solution. After the extraction step, the supernatant and the extraction solutions were combined and analyzed by MS, or stored frozen for further analysis.

2.7. *Matrix formulation and sample deposition*

Matrix solution was prepared by dissolving 10 mg of α-CHCA in 1mL of H_2O/acetonitrile/TFA (49.95:49.95:0.1). The sample (10 µL) was mixed with 10 µL of the α–CHCA matrix solution with vortexing for 30 s. One microliter of the mixture was spotted on a MALDI-TOF MS target plate and allowed to dry for 5 min.

2.8. *MALDI-TOF MS analysis*

An Ultraflex II MALDI-TOF/TOF-MS instrument from Bruker Daltonics equipped with a 200 Hz Smart beam laser system, was used to obtain peptide mass fingerprints (PMF). MALDI mass spectra were acquired as recommended by the manufacturer. Measurements were taken in reflector positive ion mode in the mass range of 600-3500 Da. Close external calibration was performed with the monoisotopic peaks of the bradykinin (757.3992), angiotensin II (1046.5418), angiotensin I (1296.6848), substance P (1347.7345), bombesin (1619.8223), renin substrate (1758.9326), ACTH clip 1-17 (2093.0862), ACTH 18-39 (2465.1983), and somatostatin 28 (3147.4710). Mass spectrum analysis for each sample was based on the average of 500 laser shots. The peak lists were generated from the mass spectra using the peak detection algorithm SNAP from the FlexAnalysis 3.3 software (Bruker Daltonics). Peptide mass fingerprints were searched with the MASCOT search engines. MASCOT parameters were as follows: variable modifications, oxidation (M); fixed modifications, carbamidomethyl (C); Swiss-Prot database; one missed cleavage allowance, and a peptide tolerance of up to 100 ppm. The default significance threshold, $p < 0.05$, was used. A match was considered successful if the protein identification score was located outside the random region and the protein analysed scored in the first position. Figure 1 summarises the methodology used in the present study.

3. Results

In this study a comparative proteomic analysis was performed between strain C5478 without antimicrobial stress (wild-type) and C5478 stressed with cefotaxime.

A stress or shock involves a change from one environmental condition to another, so the natures of both the original condition and the shock are important in defining the appropriate response. Here, we reasoned that comparing the proteomes of a naturally occurring ESBL-producing *E. coli* strain (C5478 wild type strain) and the ESBL-producing *E.*

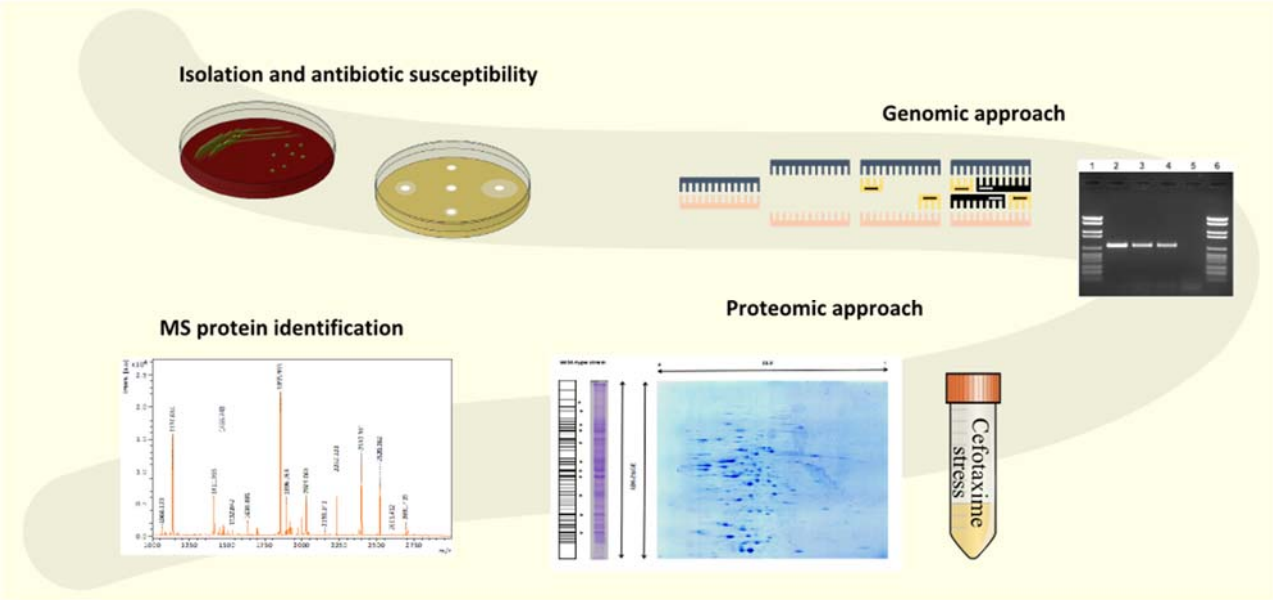


Figure 1. Typical genomic and proteomic workflow representing the classical components of genetical study and protein identification.

coli strain stressed by CTX (stressed strain) would reveal key differences that might be necessary for bacterial colonization in the presence of β-lactams. We used 2D-PAGE and MALDI-TOF MS to initially screen for significant differences in the distribution of proteins between wild-type and stressed strains (Supplementary material, Tables 1 and 2).

The comparative proteomic analysis revealed that the abundance of numerous protein species changed in the strain stressed by CTX compared to the non-stressed wild-type strain. A total of 188 spots were excised from the 2-DE gel of the wild-type strain, 112 of which were successfully identified by MS, representing 110 different proteins (Supplementary material. Table 1). Concerning the 2-DE gel of the CTX-stressed bacteria, 171 spots were excised and 156 were identified, representing 143 different proteins (Supplementary material, Table 2). The proteins identified in both strains were categorized according to their biological functions. These proteins were involved in metabolism, protein synthesis, cell division, stress responses, antimicrobial resistance, transport, and unknown functions among others (Figure 2).

Various proteins were found in more than one spot on the two gels. From the gel of proteins from the wild type ESBL-producing strain, Mdh (spots 74 and 117), Tuf1 (spots 110, 127 and 128), YajQ (spots 36, 56 and 157), GpmA (spots 46 and 50), PflB (spots 169 and 172), YeaD (spots 76 and 80), GadB and GadA (spots 114 and 126), DaaA (spot 32 and 34) were all present in more than one spot. In the gel of proteins from the CTX-stressed bacteria, Tsf (spots 243 and 337), Dps (spots 198 and 200), Pgk (spots 268, 339 and 344), Pnp (spots 303 and 316), rpoB (spots 252 and 315), Gnd (spots 280 and 345), PflB (spots 327 and 328), SlyA (spots 213 and 262), Prs (spots 285 and 360), Tpx (spots 205 and 281), TalA (spots 253 and 254), CysK (spots 253 and 254), MalE (spots 243, 266 and 338), Ssb (spots 207 and 208), AcnA (spots 322 and 323), TktA (329 and 330) and ClpB (spots 312 and 331) were present in more than one spot. Proteins that migrate at more than one distinctive position during electrophoresis have different charges and masses. Hence, these spots could correspond to proteins with different posttranslational covalent modifications [6, 14].

A total of 268 proteins were distinctively expressed in both samples. The proteins that decreased in percentage under CTX stress (Figure 3) were principally proteins involved in transport and ATP-binding (like ArgT, FliY and NmpC), glycolysis and gluconeogenesis (like GapA, GpmA, PykF and OckA), transcription and RNA processing (like DaaA, RopB, Fur and Pnp), translation (like RpsA and RpsB), proteolysis (PepQ and PrlC), fatty acid biosynthesis (like AccC and FabF), and cell shape (DdlB).

In contrast, the percentage of stress response proteins expressed was much greater in bacteria under CTX stress (Figure 4) such as those involved in mechanisms of acid resistance (GadA, GadB, WrbA), oxidative stress (like Tpx, SodB and OsmC), other general cellular stress responses involving cold and heat shock proteins (DnaK, ClpB, ClpP). Among the numerous changes noted in the stressed strain, the percentage of proteins related to DNA damage, cell division, cell redox, innate immune response (like FtsZ, FtsY, SsB and YbbN), starvation (Dps) and antimicrobial resistance (OmpX, GyrB, FabI) increased. In the stressed bacteria, there were also increased percentages of enzymes involved in protein biosynthesis (like AspS, FusA and GlyS), the tricarboxylic acid cycle (like MdH, SucB and SucC), amino-acid biosynthesis (like CysK and CarB), nucleotide and purine biosynthesis (PurA and Prs), the pentose shunt (like TalA, TalB and GnD) and transketolase activity (TktA) among others.

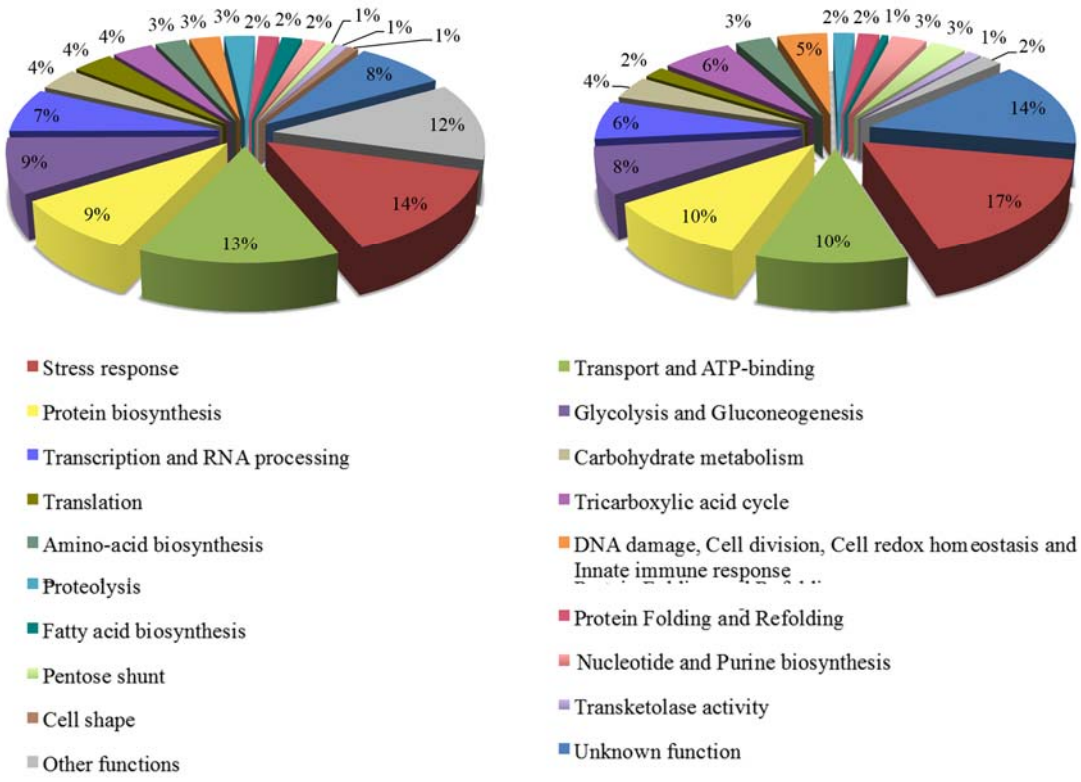


Figure 2. Distribution of biological functions of proteins expressed in wild type (left) and CTX stressed *E. coli* C5478 (right)

4. Discussion

The widespread use of antimicrobials to treat infections means that cases of bacterial resistance to these drugs are becoming more frequent, presenting a continuing challenge in medicine. Multidrug resistance is a main cause of clinical failure in treating bacterial infections. Increasing evidence suggests that bacteria can resist multiple antimicrobials through intrinsic mechanisms that rely on gene products such as efflux pumps that expel antimicrobials and specific membrane proteins that block the penetration of drug molecules.

The response of bacterial metabolism to environmental perturbations is characterized by a fast and appropriate adjusting of physiology on every level of the cellular and molecular network. The entire sequence of the *E. coli* genome has been available since 1997 greatly facilitating proteomic methodologies of 2-DE with MALDI-TOF MS analysis of proteins related to antimicrobial resistance of this bacterium.

In CTX-stressed bacteria, there were mainly decreases in percentages of proteins involved in transport/ATP-binding, glycolysis, transcription/RNA processing, translation, proteolysis, fatty acid biosynthesis and also some proteins related to oxidative and acid stress responses. Though enzyme analyses and other physiological data are still required, we can make some tentative conclusions about these overall changes in bacterial physiology. The consequences of decreased protection against oxidative stress and acid stress may have been compensated for by increased levels of proteins active in DNA damage repair, cell division, cell redox homeostasis, the innate immune response as well as proteins associated with antimicrobial resistance, tellurium resistance, general/heat stress and starvation. The stressed bacteria had acquired defence mechanisms normally related to the stress response, and an increased tolerance to antimicrobials was confirmed. Increases in the percentages of proteins involved in the stress response are often at the expense of proteins implicated in translation. Similarly we found that the increase in tolerance to cefotaxime in the stressed strain has deleterious effects on ribosomal activity.

These metabolic modifications could indicate the biological cost of antimicrobial resistance, though it is possible that other changes in these proteins that were not determined on these 2-DE gels could explain the results. Any such costs are likely to be mitigated by subsequent evolution as described by [29].

4.1. *Stress response*

Stress responses are usually a combination of specific responses aimed at minimizing deleterious effects (e.g. catalase production during oxidative stress) or repairing damage (e.g. chaperone expression under temperature stress) and general responses.

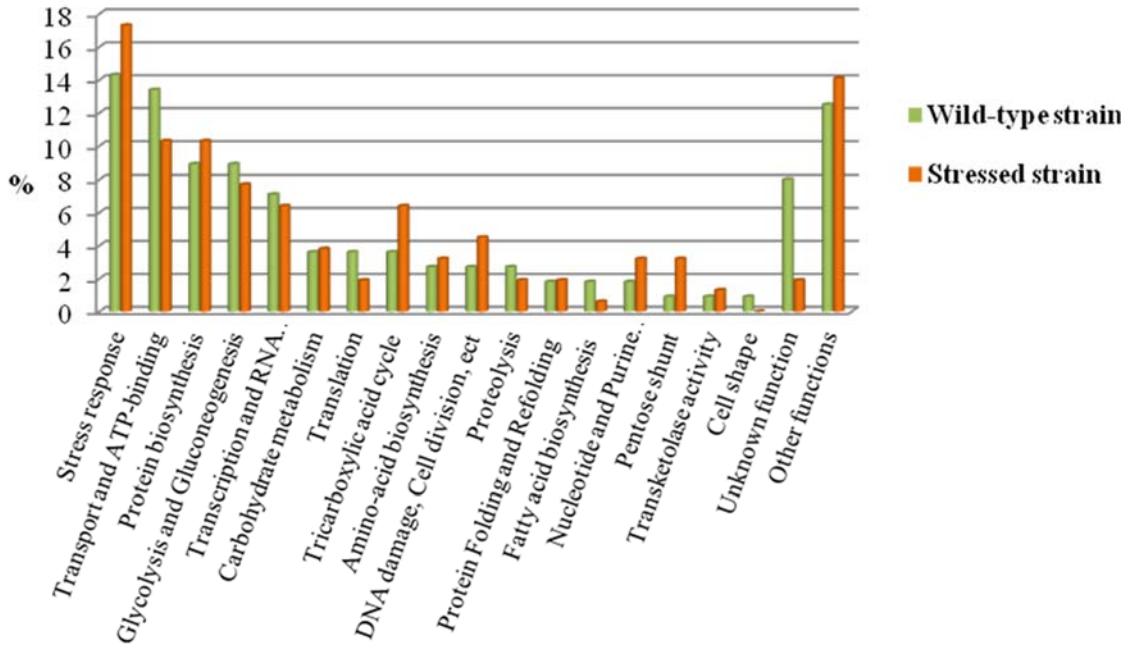


Figure 3. Percentage distribution by biological function of proteins identified in the wild type and CTX stressed *E.coli* C5478 strains

The results of this study revealed a significant increase in the percentage of stress response proteins, like GadA and GadB, present in the naturally occurring and stressed bacteria. GadA and GadB are the most efficient acid resistance system [30]. Defences against oxidative stress involve constitutive and tightly regulated adaptive mechanisms to prevent and scavenge oxidants as well as to repair damaged biomolecules. The widely accepted simplified paradigm of defence against oxidative stress is that superoxide is removed by superoxide dismutases such as SodB (spots 26 and 209) generating hydrogen peroxide, which is removed by catalases, like KatE (spots 174 and 324) and peroxidases like AhpC (spot 212, MW 20862 and PI 5.03). Several of these defences are controlled by regulators that respond to iron such as Fur (spot 2, MW 17012 and PI 5.68) or oxygen tension such as ArcA (spot 52) [31].

In the CTX-stressed bacteria, besides the ribosome-associated Trigger factor Tig (spot 284), the cytoplasm of the strain contains two of the three FK506-binding proteins (FKBPs) SlyD (spot 240) and FklB (spot 215). FKBPs are enzymes grouped into a superfamily of peptidyl-prolyl-cis/trans-isomerases (PPIases) that act as chaperones in folding proteins. Many proteins involved in disease processes require modification, thus PPIases can play an important role

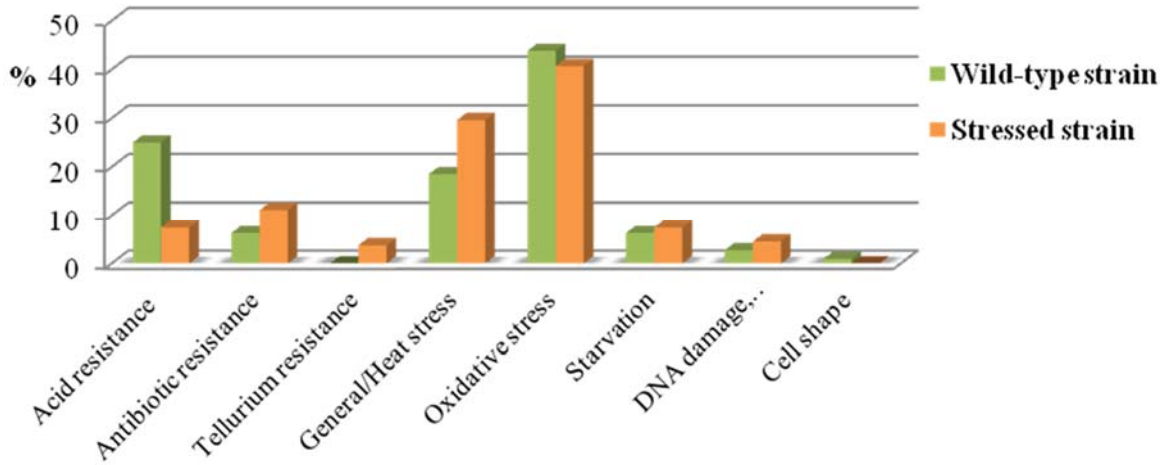


Figure 4. Percentage distribution of stress response proteins identified in wild type and CTX stressed *E. coli* C5478 strains

in pathogenesis [31].

Other proteins related to oxidative stress defence in the stressed strain were identified here such as CysQ (spot 219), YdhR (spot 323), YhfA (spot 191), MaeB (spot 310) and Tpx (spots 205 and 281). The osmotically inducible protein OsmC, which was identified in both strains (spots 7 and 190), is involved in hydroperoxide protection particularly in host–pathogen interactions [32].

Some heat shock proteins (HSPs) function as molecular chaperones or have functions linked with DNA replication, cell division, and maintenance of active protein conformations [33]. The number of proteins related to heat stress increased under CTX stress, for example ClpB (spots 312 and 331) and HtpG (spot 293), which participate in protein folding in stressed *E. coli* cells [34].

KlaB (TelA, MW 42130 and PI 6.05) was one of the stress response proteins identified in the stressed strain (spot 287). KlaB is responsible for tellurium resistance (Supplementary material, Tables 1 and 2). Tellurite is toxic to *E. coli* through an unknown mechanism that involves an increase in oxidative stress. However recently, biochemical, genetic, enzymatic and molecular approaches were used to demonstrate, for the first time, that tellurite toxicity in *E. coli* involves superoxide formation. Altogether, this study strongly indicate that the dramatic toxic effect of potassium tellurite in *E. coli* can be associated with the chemical activity of the tellurium oxyanion at various levels of bacterial metabolic pathways including, among others, inactivation of [Fe-S] center-containing dehydratases, cytoplasmic thiol oxidation, enzyme and protein carbonylation and membrane peroxidation [35].

4.2. *DNA damage, cell division, cell redox homeostasis and the innate immune response*

In the stressed strain derived from C5478, there were increased percentages of proteins related to DNA damage, cell division and cell redox homeostasis. Of the cell division proteins, FtsZ found in the stressed strain (spot 261) is of great interest as it is the earliest actor in the division pathway and its level dictates the frequency of division [36]. Furthermore, FtsY, also identified in the stressed strain (spot 314, MW 54480 and PI 4.46), is a signal recognition particle receptor in *E. coli* that mediates the targeting of integral membrane proteins to translocons by interacting with both signal recognition particle (SRP)-nascent polypeptide–ribosome complexes and the cytoplasmic membrane [37].

The SsB protein, one of the proteins in the DNA damage category, was found in the stressed strain in two spots (207 and 208). SsB protects and sequesters single-stranded DNA regions until the double helix can be reformed [38].

The relative amounts of cell wall proteins found in the stressed bacteria are quite high. Constant turnover of the cell wall provides ample decoy material for *E. coli* to engage host innate defences and the rate of cell wall production is directly proportional to the rate of colonization of the human tract [39]. Penicillin-binding proteins (PBPs), the target enzymes of β-lactam antimicrobials such as cefotaxime, play an essential role in bacterial cell elongation, cell division and cell wall biosynthesis. PBPs are membrane-bound enzymes that are widespread in *E. coli* [40].

The SOS response supports bacterial propagation by inhibiting cell division during repair of DNA damage and cell survival. The extent of induction of the SOS response can be determined by measuring β-galactosidase synthesis from a LacZ (spot 319, MW 117300 and PI 5.20) fusion with the SOS-regulated promoter of the *sfi*A gene, which prevents FtsZ (spot 261, MW 40299 and PI 4.63) polymerization and inhibits cell division when SOS is activated [37]. The immediate response of bacterial pathogens to antimicrobial therapy is of key interest when considering how to deal with infectious diseases.

The FtsZ protein, identified in the stressed bacteria, is crucial to the cell-division process. It assembles as a dynamic ring on the inner surface of the cytoplasmic membrane at the place where division will happen, with the formation of the ring being the signal for septation to begin [36].

It is important to also point out the presence of the FtsY protein in the stressed bacteria. The capability of transformants producing chimeric FtsY to process β-lactamase was evaluated in a recent study [37]. While β-lactamase is not dependent on signal recognition particle (SRP) for export, it does require SRP function for cleavage of the signal peptide, as leader peptidase is an SRP-dependent membrane protein in *E. coli* [41].

4.3. *Antimicrobial resistance*

More of the antimicrobial resistant proteins Ompx, GyrB and FabI were observed in the stressed strain. OmpX is present in the both strains (spot 14, MW 67648 and PI 5.30) and (spot 196, MW 18648 and PI 5.30). Because of the function of this integral outer membrane protein in membrane permeability, OmpX may be involved in controlling the penetration of antimicrobials such as β-lactams and fluoroquinolones through the enterobacterial outer membrane as is known for chloramphenicol, tetracycline, and kanamycin [42]. OmpX belongs to a family of highly conserved bacterial proteins that have been assigned key functions in promoting bacterial adhesion and entry into mammalian cells [43].

Usually, resistance to fluoroquinolones arises spontaneously because of point mutations that result in amino acid substitutions within the topoisomerase subunits GyrA, GyrB, ParC or ParE, a decreased expression of outer membrane porins, or an over expression of multidrug efflux pumps [44]. GyrB was identified (spot 325, MW 90179 and PI 5.72) in this study, which is consistent with the quinolone acid resistance observed in the strain. The wild-type strain was known to have two amino acid changes in GyrA (S83L+D87G) and one in ParC (S80I) responsible for ciprofloxacin and nalidixic acid resistance [23].

The presence of the protein FabI (enoyl reductase), an im-

portant enzyme in fatty acid biosynthesis, was observed in the stressed bacteria (spot 245, MW 28074 and PI 5.58). Triclosan, which has been described as a broad-spectrum antimicrobial and is found in many antibacterial consumer products, is now known to specifically target FabI in *E. coli* [45].

The likely loss of energy production in the stressed strain caused by the decreased percentage of proteins involved in the glycolytic pathway may be a cost that resistance imposes on the fitness of bacteria [14]. Besides glycolysis, other pathways such as the tricarboxylic acid (TCA) cycle are likely to be involved in regulating and responding to CTX bactericidal effects, which may explain the increase in the amount of TCA cycle related proteins. Further study of the relationships between iron, metabolism, and iron-sulfur cluster-containing proteins [46] is necessary to identify any post-transcriptional events not captured in gene expression studies that trigger this common cell death pathway [47].

4.4. *Transport proteins*

The wild-type strain is multidrug resistant. Both the wild-type strain and the stressed strain can resist antimicrobial agents by active efflux of the agents using translocation machinery [48]. This is borne out by our results showing that some of the most abundant proteins, more than 13% and 10% of those identified in the non-stressed and stressed conditions respectively, were linked to transport and ATP-binding functions (Figure 2). These included three proteins related to ATP-binding detected in the stressed strain, Acs (spot 329), GlnA (spot 295) and PpsA (spot 304). The ATP synthase subunit alpha (AtpA) and beta (AtpD) were present in the wild-type and stressed strain. In *E. coli* ATP synthase activity provides proton motive force for efflux pump activity [49]. Additionally, the SecD protein identified in the stressed strain (spot 202, MW 66648 and PI 8.62) is a transport protein implicated in bacterial pathogenesis and in the secretion of virulence proteins [39].

Antimicrobial resistance often entails a metabolic cost because the resistance mutations typically occur in genes of target molecules that have essential functions in the cell. In fact, mutation is a primary cause of bacterial resistance to antimicrobials. In *E. coli* active efflux systems are common mechanisms of reduced susceptibility to fluoroquinolones that may confer resistance, particularly when associated with mutation in GyrA, GyrB or ParC [49] like those in GyrA and ParC in the wild-type C5478 strain studied here.

As previously mentioned the enzyme FabI was identified in the stressed bacteria. Given the multi-drug efflux pumps in *E. coli*, over expression of one of the pumps could participate in resistance mediated by a FabI mutation, which might have the effect of spreading antimicrobials.

4.5. Amino acid and protein biosynthesis/metabolism

The number of proteins related to amino acids, protein biosynthesis and metabolism increased in the stressed bacteria. One example is the CysK protein (spots 253 and 254) that has elsewhere been shown to mediate tellurite resistance in *E. coli*. Pyridoxine 5'-phosphate synthase (PdxJ, spot 235) was present in the stressed bacteria. The genes *pdxJ* and *pdx*A encode proteins involved in vitamin B_6 biosynthesis. It was proposed that PdxJ enzymes could be ideal therapeutic targets against bacterial pathogens [50].

Transaldolase A (TalA), found in the stressed bacteria (spot 253 and 254), is involved in the nonoxidative branch of the pentose phosphate pathway. The gene coding for the isozyme TktA is also detected in the stressed strain (spots 229 and 230) and is known to be upregulated by stress conditions [51].

The Upp protein was found in the stressed bacteria (spot 214). It was shown that energy production genes such as SucB are involved in persister survival and tolerance of multiple antimicrobials and stresses in *E. coli* [52]. Upp, expressed naturally in *E. coli*, converts the chemotherapeutic drug 5-fluorouracil (5-FU) into 5-fluorouridine monophosphate, an extremely toxic compound. Adenoviral delivery of the Upp gene to human cancer cell lines increased their sensitivity to 5-FU [53]. The *E. coli* enzymes DeoD (spots 33, 217) have also been used in gene therapy studies [53].

The main HSPs DnaK, DnaJ, GroEL are molecular chaperones that assist in the correct folding and assembly of proteins and are implicated in various cellular processes including DNA replication, RNA transcription, flagella synthesis and UV mutagenesis [54]. HSPs participate in the immune response to bacterial infections and the development of autoimmune diseases [55]. Different classes of HSPs from different bacteria can directly induce cytokine expression and secretion in macrophages [56]. We found several chaperones related to the stress response were found in the both strains but they were more abundant in the stressed strain. It has been demonstrated previously that β-lactamase interacts with GroEL and GroES [57, 58] and that export of β-lactamase is defective in GroEL and GroES mutants [41].

Glycopeptides such as vancomycin and teicoplanin target the peptidoglycan cell wall of bacteria by selectively binding to the D-alanyl–D-alanine termini of peptidoglycan precursors preventing cross linking to adjacent strands. The presence of DdlB, an attractive target for developing novel antimicrobials [59], was detected in the wild-type strain.

5. Conclusion

In the present study the use of 2D-MALDI-TOF MS provided wide proteome coverage as demonstrated by the number of proteins identified (112 in the wild-type strain and 156 in the stressed strain). This made it possible to more comprehensively evaluate changes in protein expression in response to selection with CTX. The numerous changes in protein levels observed in the stressed strain suggest that resistant micro-organisms may adapt to adverse environmental conditions [60, 61] with implications for several aspects of bacterial metabolism, which may be mirrored in their virulence parameters.

The ability of resistant bacteria to survive in a population and in the community depends on numerous factors, such as biological fitness. In this case, antimicrobial resistance can be stabilized with fitness-restoring compensatory mutations and may permit completely resistant strains to compete successfully with susceptible strains in an antimicrobial-free environment [29]. In conclusion, we have demonstrated that a wide range of proteins change in abundance in a diverse secondary response to treatment with cefotaxime, a third-generation cephalosporin. For the proteins already associated with antimicrobial resistance, such as OmpX, FtsZ and FtsY, the modifications are likely due to the immediate effects of encountering CTX. These findings will be helpful for understanding more generally what constitutes a functional proteome and, in particular, how antimicrobial-resistant mechanisms work.

6. Supplementary material

Supplementary data and information is available at: http://www.jiomics.com/index.php/jio/rt/suppFiles/140/0

Table 1- Protein identification by MALDI-TOF/MS of 2-DE gel spots from wild-type strain ESBL-*E. coli* C5478.

Table 2 - Protein spots identification by MALDI-TOF/MS of 2-DE gel spots from stressed strain ESBL-E. coli C5478.

Acknowledgments

Hajer Radhouani, Júlio D. Nunes-Miranda, Hugo M. Santos and Luís Pinto were supported by grants SFRH/BD/60846/2009, SFRH/BD/80496/2011, SFRH/BPD/73997/2010 and SFRH/BD/81307/2011, respectively, and by FCT (Fundação para a Ciência e a Tecnologia) and POPH/FSE (Programa Operacional Potencial Humano/Fundo Social Europeu).

References

[1] D.M. Tracz, D.A. Boyd, L. Bryden, R. Hizon, S. Giercke, P. Van Caeseele, M.R. Mulvey, J Antimicrob Chemother, 55 (2005) 768-772. doi: 10.1093/jac/dki074

[2] B.T. Carter, H. Lin, S.D. Goldberg, E.A. Althoff, J. Raushel, V.W. Cornish, Chembiochem, 6 (2005) 2055-2067. doi: 10.1002/cbic.200500058

[3] D.L. Paterson, R.A. Bonomo, Clin Microbiol Rev, 18 (2005) 657-686. doi: 10.1128/CMR.18.4.657-686.2005

[4] M. Grobbel, A. Lubke-Becker, E. Alesik, S. Schwarz, J. Wallmann, C. Werckenthin, L.H. Wieler, Berl Munch Tierarztl Wochenschr, 120 (2007) 391-401.

[5] C. Ewers, M. Grobbel, A. Bethe, L.H. Wieler, S. Guenther, Berliner und Munchener tierarztliche Wochenschrift, 124 (2011) 94-101.

[6] P.R. Jungblut, H.G. Holzhutter, R. Apweiler, H. Schluter, Chem Cent J, 2 (2008) 16. doi: 10.1186/1752-153X-2-16

[7] L. Pinto, P. Poeta, H. Radhouani, A.C. Coelho, C. Carvalho, J. Rodrigues, C. Torres, R. Vitorino, P. Domingues, G. Igrejas, Journal of Integrated OMICS, 1 (2011) 42-48.

[8] L. Pinto, P. Poeta, H. Radhouani, A.C. Coelho, C. Carvalho, J. Rodrigues, C. Torres, R. Vitorino, P. Domingues, G. Igrejas, Journal of Integrated OMICS, 1 (2011). doi: 10.5584/jiomics.v1i1.19

[9] P. Roncada, F. Deriu, A. Gaviraghi, P.A. Martino, L. Bonizzi, Vet Res Commun, 33 Suppl 1 (2009) 157-160. doi: 10.1007/s11259-009-9260-1

[10] H. Radhouani, L. Pinto, P. Poeta, G. Igrejas, J Proteomics, (2012). doi: 10.1016/j.jprot.2011.12.035

[11] K. Reinhart, C.S. Hartog, Int J Antimicrob Agents, 36 Suppl 2 (2010) S17-21. doi: 10.1016/j.ijantimicag.2010.11.009

[12] G. Werner, T.M. Coque, A.M. Hammerum, R. Hope, W. Hryniewicz, A. Johnson, I. Klare, K.G. Kristinsson, R. Leclercq, C.H. Lester, M. Lillie, C. Novais, B. Olsson-Liljequist, L.V. Peixe, E. Sadowy, G.S. Simonsen, J. Top, J. Vuopio-Varkila, R.J. Willems, W. Witte, N. Woodford, Euro Surveill, 13 (2008).

[13] D. Josic, S. Kovac, Biotechnol J, 3 (2008) 496-509. doi: 10.1002/biot.200700234

[14] K.V. dos Santos, C.G. Diniz, C. Veloso Lde, H.M. de Andrade, S. Giusta Mda, F. Pires Sda, A.V. Santos, A.C. Apolonio, M.A. de Carvalho, M. Farias Lde, Res Microbiol, 161 (2010) 268-275. doi: 10.1016/j.resmic.2010.03.006

[15] X.M. Lin, H. Li, C. Wang, X.X. Peng, J Proteome Res, 7 (2008) 2399-2405. doi: 10.1021/pr800073c

[16] D.F. Zhang, B. Jiang, Z.M. Xiang, S.Y. Wang, Int J Antimicrob Agents, 32 (2008) 315-319. doi: 10.1016/j.ijantimicag.2008.04.015

[17] C. Xu, X. Lin, H. Ren, Y. Zhang, S. Wang, X. Peng, Proteomics, 6 (2006) 462-473. doi: 10.1002/pmic.200500219

[18] H. Li, B.C. Wang, W.J. Xu, X.M. Lin, X.X. Peng, J Proteome Res, 7 (2008) 4040-4049. doi: 10.1021/pr800310y

[19] H. Li, X.M. Lin, S.Y. Wang, X.X. Peng, J Proteome Res, 6 (2007) 3628-3636. doi: 10.1021/pr070307y

[20] N.G. Coldham, L.P. Randall, L.J. Piddock, M.J. Woodward, J Antimicrob Chemother, 58 (2006) 1145-1153. doi: 10.1093/jac/dkl413

[21] W. Liu, D. Zou, X. Wang, X. Li, L. Zhu, Z. Yin, Z. Yang, X. Wei, L. Han, Y. Wang, C. Shao, S. Wang, X. He, D. Liu, F. Liu, J. Wang, L. Huang, J. Yuan, J Proteome Res, 11 (2012) 4024-4033. doi: 10.1021/pr300062v

[22] S.H. Yun, C.W. Choi, S.O. Kwon, G.W. Park, K. Cho, K.H. Kwon, J.Y. Kim, J.S. Yoo, J.C. Lee, J.S. Choi, S. Kim, S.I. Kim, J Proteome Res, 10 (2011) 459-469. doi: 10.1021/pr101012s

[23] H. Radhouani, G. Igrejas, A. Gonçalves, V. Estepa, R. Sargo, C. Torres, P. Poeta, Archives of Microbiology, (2012). doi: 10.1007/s00203-012-0853-7

[24] A. Görg, W. Weiss, M.J. Dunn, Proteomics, 4 (2004) 3665-3685. doi: 10.1002/pmic.200401031

[25] U.K. Laemmli, Nature, 227 (1970) 680-685.

[26] G. Igrejas, in: Departement of Genetic and Biotechnology, Univesity of Trás-os-Montes and Alto Douro, Vila Real, 2000.

[27] A. Görg, A. Klaus, C. Lück, F. Weiland, W. Weiss, Technische Universität München, (2007).

[28] A. Görg, C. Obermaier, G. Boguth, A. Harder, B. Scheibe, R. Wildgruber, W. Weiss, Electrophoresis, 21 (2000) 1037-1053. doi: 10.1002/(SICI)1522-2683(20000401)21:6<1037::AID-ELPS1037>3.0.CO;2-V

[29] D.I. Andersson, Curr Opin Microbiol, 6 (2003) 452-456.

[30] M.P. Castanie-Cornet, T.A. Penfound, D. Smith, J.F. Elliott, J.W. Foster, J Bacteriol, 181 (1999) 3525-3535.

[31] K.L. Seib, H.J. Wu, S.P. Kidd, M.A. Apicella, M.P. Jennings, A.G. McEwan, Microbiol Mol Biol Rev, 70 (2006) 344-361. doi: 10.1128/MMBR.00044-05

[32] S. Mongkolsuk, J.D. Helmann, Mol Microbiol, 45 (2002) 9-15.

[33] C.A. Mason, J. Dunner, P. Indra, T. Colangelo, Appl Environ Microbiol, 65 (1999) 3433-3440.

[34] J.G. Thomas, F. Baneyx, Mol Microbiol, 36 (2000) 1360-1370.

[35] J.M. Perez, I.L. Calderon, F.A. Arenas, D.E. Fuentes, G.A. Pradenas, E.L. Fuentes, J.M. Sandoval, M.E. Castro, A.O. Elias, C.C. Vasquez, PLoS One, 2 (2007) e211. doi: 10.1371/journal.pone.0000211

[36] E.F. Bi, J. Lutkenhaus, Nature, 354 (1991) 161-164. doi: 10.1038/354161a0

[37] I. Maeda, A. Hirata, M. Shoji, S. Ueda, K. Yoshida, FEMS Microbiol Lett, 287 (2008) 85-90. doi: 10.1111/j.1574-6968.2008.01297.x

[38] E. Bore, M. Hebraud, I. Chafsey, C. Chambon, C. Skjaeret, B. Moen, T. Moretro, O. Langsrud, K. Rudi, S. Langsrud, Microbiology, 153 (2007) 935-946. doi: 10.1099/mic.0.29288-0

[39] G. Muthukrishnan, G.A. Quinn, R.P. Lamers, C. Diaz, A.L. Cole, S. Chen, A.M. Cole, J Proteome Res, 10 (2011) 2064-2078. doi: 10.1021/pr200029r

[40] G. Zhao, W.K. Yeh, R.H. Carnahan, J. Flokowitsch, T.I. Meier, W.E. Alborn, Jr., G.W. Becker, S.R. Jaskunas, J Bacteriol, 179 (1997) 4901-4908.

[41] D. Beha, S. Deitermann, M. Muller, H.G. Koch, J Biol Chem, 278 (2003) 22161-22167. doi: 10.1074/jbc.M300929200

[42] M. Dupont, C.E. James, J. Chevalier, J.M. Pages, Antimicrob Agents Chemother, 51 (2007) 3190-3198. doi: 10.1128/AAC.01481-06

[43] J. Mecsas, R. Welch, J.W. Erickson, C.A. Gross, J Bacteriol, 177 (1995) 799-804.

[44] A.M. Hammerum, O.E. Heuer, Clin Infect Dis, 48 (2009) 916-921. doi: 10.1086/597292

[45] S. Knobler, I.o.M.U.S.F.o.E. Infections, The resistance phenomenon in microbes and infectious disease vectors: implications for human health and strategies for containment : workshop summary, Knobler, S. ed., National Academies Press, 2003.

[46] K. Hantke, Curr Opin Microbiol, 4 (2001) 172-177.

[47] M.A. Kohanski, D.J. Dwyer, B. Hayete, C.A. Lawrence, J.J. Collins, Cell, 130 (2007) 797-810. doi: 10.1016/j.cell.2007.06.049

[48] K.A. Hassan, R.A. Skurray, M.H. Brown, J Mol Microbiol Biotechnol, 12 (2007) 180-196. doi: 10.1159/000099640

[49] N.G. Coldham, M.J. Woodward, J Proteome Res, 3 (2004) 595-603. doi: Doi 10.1021/Pr034129u

[50] A. Grubman, A. Phillips, M. Thibonnier, M. Kaparakis-Liaskos, C. Johnson, J.M. Thiberge, F.J. Radcliff, C. Ecobichon, A. Labigne, H. de Reuse, G.L. Mendz, R.L. Ferrero, MBio, 1 (2010). doi: 10.1128/mBio.00112-10

[51] F. Domain, X.R. Bina, S.B. Levy, Mol Microbiol, 66 (2007) 383-394. doi: 10.1111/j.1365-2958.2007.05928.x

[52] C. Ma, S. Sim, W. Shi, L. Du, D. Xing, Y. Zhang, FEMS Microbiol Lett, 303 (2010) 33-40. doi: 10.1111/j.1574-6968.2009.01857.x

[53] F. Kanai, T. Kawakami, H. Hamada, A. Sadata, Y. Yoshida, T. Tanaka, M. Ohashi, K. Tateishi, Y. Shiratori, M. Omata, Cancer Res, 58 (1998) 1946-1951.

[54] T. Yura, H. Nagai, H. Mori, Annu Rev Microbiol, 47 (1993) 321-350. doi: 10.1146/annurev.mi.47.100193.001541

[55] P.J. Murray, R.A. Young, J Bacteriol, 174 (1992) 4193-4196.

[56] C. Retzlaff, Y. Yamamoto, P.S. Hoffman, H. Friedman, T.W. Klein, Infect Immun, 62 (1994) 5689-5693.

[57] E.S. Bochkareva, N.M. Lissin, A.S. Girshovich, Nature, 336 (1988) 254-257. doi: 10.1038/336254a0

[58] A.A. Laminet, T. Ziegelhoffer, C. Georgopoulos, A. Pluckthun, Embo J, 9 (1990) 2315-2319.

[59] G. Triola, S. Wetzel, B. Ellinger, M.A. Koch, K. Hubel, D. Rauh, H. Waldmann, Bioorg Med Chem, 17 (2009) 1079-1087. doi: 10.1016/j.bmc.2008.02.046

[60] J.F. Linares-Rodriguez, J.L. Martinez-Menendez, Enferm Infecc Microbiol Clin, 23 (2005) 86-93.

[61] S.S. Justice, D.A. Hunstad, L. Cegelski, S.J. Hultgren, Nat Rev Microbiol, 6 (2008) 162-168. doi: 10.1038/nrmicro1820

JOURNAL OF INTEGRATED OMICS
A METHODOLOGICAL JOURNAL
HTTP://WWW.JIOMICS.COM

ORIGINAL ARTICLE | DOI: 10.5584/jiomics.v3i2.154

Proteomics Analysis of Morphogenic Changes of Human Umbilical Vein Endothelial Cells Induced by a Phorbol-Ester Mimicking Angiogenesis

Baudin Bruno*[1, 2], Ramani David[1], Bosselut Nelly[1], Marcelo Paulo[3], Lecornet Hélène[1], Margoline Gary[1], Vinh Joëlle[3], Vaubourdolle Michel[1]

[1]Biochimie A - Pôle Biologie Médicale et Pathologie - Hôpitaux Universitaires Est Parisien (APHP), Site Saint-Antoine - 184 Rue du Faubourg Saint-Antoine 75571 Paris cedex, France; [2] EA 4530 - UFR Pharmacie Université Paris-Sud, 5 rue Jean-Baptiste-Clément 92296 Châtenay-Malabry cedex, France; [3] CNRS-USR 3149, ESPCI Paris Tech, 10 rue Vauquelin, 75231 Paris cedex 05, France.

Received: 08 August 2013 **Accepted:** 01 October 2013 **Available Online**: 24 October 2013

ABSTRACT

Phorbol 12-myristate 13-acetate (PMA) can induce proliferation and migration of endothelial cells, mimicking vessel formation. We analysed by two-dimensional electrophoresis and MALDI-TOF/TOF the effects of PMA on cultured Human Umbilical Vein Endothelial Cells (HUVECs) to further investigate the complex mechanisms related to protein kinase C activation in this angiogenesis model. At 1 µg/ml for 24 hours, PMA induced transition of HUVECs from quiescent type into the proliferative-migrating phenotype. After 2D gel analysis, 15 differences were detected between PMA-treated samples and controls, including 8 increased proteins and 7 decreased proteins. The three main proteins identified by mass spectrometry and increased after PMA are directly involved in cell stress (α-glucosidase, heat-shock protein 70, and 150 kDa oxygen-regulated protein). Four other proteins varied in function of time, two increasing after PMA (heat shock protein 90β, protein-disulfide isomerase A3), and two other decreasing after this treatment (glucose-related protein 75, cathepsin B). These four proteins are involved in protein folding, apoptosis or tumour dissemination. Our data show that phorbol esters modify a number of proteins involved in multiple and intricate pathways for promoting a phenotype ensuring cell survival and cell migration for new vessels formation.

Keywords: Angiogenesis; endothelial cells; oxygen-regulated protein 150; phorbol esters; protein kinase C; proteomics.

Abbreviations

HUVECs, human umbilical vein endothelial cells; **PMA**, phorbol 12-myristate 13-acetate; **PKC**, protein kinase C; **HSP-70**, heat-shock protein 70; **HSP-90β**, heat-shock protein 90 beta; **GRP-75**, glucose-related protein 75; **ORP-150**, oxygen-regulated protein 150; **PDI-A3**, protein-disulfide isomerase A3; **VEGF**, Vascular endothelial growth factor.

1. Introduction

Angiogenesis is the formation of new capillaries from a pre-existing capillary network and is implicated in many physiological or pathological processes, particularly carcinogenesis and diabetic retinopathy. Both physiological and pathological angiogenesis are regulated by a tight balance between activators, such as "vascular endothelial growth factors" (VEGFs) and angiopoietins, and inhibitors, such as thrombospondin 1 and endostatin [1, 2]. Endothelial cells (ECs) play a key role in both mechanisms of angiogenesis, i.e. "sprouting" with EC migration, and "intussuceptive angiogenesis" with ramification of pre-existing blood vessels [3]. These two processes need pericytes recruitment and secretion of an extra-cellular matrix. Phorbol-esters, in particular phorbol 12-myristate 13-acetate (PMA), are protein kinase C (PKC) activators known to promote blood vessel formation [4, 5]. PMA is sufficient for the induction of angiogenesis and PKC is necessary for this process [6-8]. Activation of PKC by PMA has a variety of effects on cultured ECs, in-

***Corresponding author:** Bruno Baudin. Service de Biochimie A, Hôpital Saint-Antoine, 184 rue du Faubourg Saint-Antoine 75571 Paris cedex, France; Tel : 33 1 49 28 20 13; Fax : 33 1 49 28 20 77; E-mail Address: bruno.baudin@sat.aphp.fr

cluding induction of migration, proliferation and vessel formation [9-11], and essentially via induction of VEGF [12]. Since treatment of EC with PMA is able to mimic the *in vivo* angiogenic process, it has provided a valuable tool for the investigation of signalling pathways involved in angiogenesis [13]. In particular, in Human Umbilical Vein Endothelial Cells (HUVECs), PMA-activated PKC activates both Raf/MEK/ERK and MAPK signalling pathways [14-16]. ERK phosphorylates the protein complex binding the Egr-1 gene leading to the induction of gene expression. ERK is also able to activate *c-jun* forming the AP-1 complex. Both AP-1 and Egr-1 transcription factors bind their target sequence to promote gene expression. To date, at least 12 distinct members of the PKC superfamily have been identified in mammalian cells [17], but mainly PKC isoforms α and δ have been implicated in the regulation of angiogenesis [18, 19]. For example, down regulation of PKCα in HUVECs by antisense oligonucleotides inhibits vessel formation [20]. Other studies have shown that sphingosine kinase is activated by PKC leading to the formation of sphingosine 1-phosphate with angiogenic properties [21-23]. PMA is also effective in inducing morphogenic changes in HUVECs by activating both PI_3 kinase and Akt/protein kinase D (PKD) pathways. Moreover PKD is phosphorylated in response to PMA and was reported to directly stimulate ERK activation in ECs in response to VEGF [7, 24]. PMA and VEGF can induce CREB-mediated COX-2 expression through a PKCα dependent pathway in human endothelium [25]. These studies reveal the complexities of the pathways governing vessel formation. PMA was also shown to promote the adhesion of endothelial progenitor cells to vascular endothelium and extracellular matrix without particular toxicity and negligible impact on cell survival [26].

In our study, we analysed the effects of PMA on primary cultures of HUVECs to further investigate these complex mechanisms, using two-dimensional electrophoresis (2-DE) coupled with MALDI-TOF/TOF mass spectrometry for protein identification.

2. Material and Methods

2.1. *HUVEC culture*

This protocol was previously described in details [27]; only primary cultures were used for this study and they were obtained by mixing ECs from two umbilical cords. As observed by phase-contrast optic microscopy, all the cultures reached cell confluence 7 days after seeding in 35mm Petri dishes.

2.2. *PMA treatment and sample preparation*

At this stage, the cultures were washed in PBS (3 times) and half of them were incubated with PMA (Sigma, Saint-Quentin Fallavier, France) at a concentration between 0.5 and 2 µg/ml, the other half with culture medium (controls), for 14 to 48 hours at 37°C in 5% CO_2-95% air atmosphere saturated with water. For incubation with PMA, the foetal calf serum used for cell growth was replaced with Ultroser G (Biosepra, Cergy-Pontoise, France) at 1% in culture medium; PMA was diluted in this medium. After incubation, culture medium was eliminate; the remaining adherent cells were washed in PBS (3 times) and mechanically scrapped in 4% Chaps, 2.3% dithiothreitol (DTT), 5mM N-ethylmaleimide and 1mM 4-(2-aminoethyl)benzenesulfonyl fluoride (detergent, disulfide bond reducer and protease inhibitors, respectively). Protein concentration was determined after this stage, using 2D Quant Kit (General Electric Healthcare, Vélizy-Villacoublay, France) and bovine serum albumin (Sigma) as a standard. Finally, the samples with similar protein contents were treated with cold acetone (4/1 v/v) for protein precipitation (2 hours at - 20°C); after centrifugation (1500g, 15 min at + 4°C), the pellet was dissolved in buffer for 2-DE analysis performed as previously described [28].

2.3. *2-DE, statistical analysis and peptide preparation*

All these procedures were detailed in our previous works on proteomics of HUVECs [28, 29]. In this study, for the first dimension, we used IPG-Phor and pH 4-7 non linear IPG 18cm strips (GE Healthcare), and for the second dimension, home made 10% SDS-PAGE with Bio-Rad (Marnes-la-Coquette, France) materials. The gels were coloured with mass spectrometry compatible silver nitrate or colloidal Blue Coomassie as it applied, then scanned with ImageScanner and studied with ImageMaster 2D Platinium™ (GE Healthcare). For the first set of experiments (1 µg/ml PMA/control), the three 2D gels from PMA-treated cells were compared to the three gels from non-treated cells (control) after studying the reproducibility between the three gels of each class using bio-informatics set in the software. For the second set of experiments (kinetic study), six gels were performed, two for each time and condition (1 µg/ml PMA/Control). Statistical comparisons were performed using Wilcoxon tests ($p < 0.05$ was considered significant); they were based on normalized volumes of the spots.

For both sets of experiments, spots corresponding to statistically significant variations were manually excised and treated for peptide preparation as described in detail before, particularly using trypsin, DTT for disulfide bridge reduction, and iodoacetamide for thiol alkylation [30].

2.4. *Protein identification by MALDI-TOF/TOF*

MALDI-TOF/TOF-MS experiments were carried out using a tandem mass spectrometer 4800 MALDI TOF/TOF™ Analyser (Applera Applied Biosystems, Framingham, MA). MS analyses were performed in positive ion reflectron mode, with an accelerating voltage of 20kV; 0.3 µl of the samples were deposited with 0.6 µl of α-cyano-4-hydroxycinnamic acid matrix (4 mg/ml) dissolved in a 60% acetonitrile, 40% H_2O (v/v), 0.1% trifluoroacetic acid and 10mM ammonium citrate solution. For subsequent data processing, the GPS

Explorer™ Workstation (Applied Biosystems) was used before the identification with MASCOT software (http://ww.matrixscience.com). An external calibration was achieved with a standard peptide mix (Proteomix Peptide mix4, LaserBioLabs, Sophia-Antipolis, France) containing: bradykinin fragment 1-5 (573.3150 Da), human angiotensin II (1046.5424 Da), neurotensin (1672.9176 Da), ACTH fragment 18-39 (2464.1989) and oxidized insulin B chain (3494.6514 Da). The MASCOT program was used with the following parameters: human species, 2 missed cleavages by trypsin, monoisotopic peptide masses, mass deviation of 100ppm (MS) and 0.15 Da (MS/MS). Oxidation of methionine and carbamidomethylation of cysteine were taken into consideration. All the proteins identified should have a protein score (exponential score) greater than 60 (Confidence Interval with this software) and individual ion score greater than 40 was set as the threshold for acceptance (MASCOT linear score) with at least two different identified peptides.

3. Results

Used at 1 µg/ml PMA induced morphogenic effects in HUVECs with elongation of the cells but without new blood vessels promotion (Figure 1). At 0.5 µg/ml these effects were lower, whereas at much lower concentration (e.g. 0.1 µg/ml), PMA failed to induce this particular phenotype. When used at 2 µg/ml PMA induced toxic effects with cell disruption and detachment, but, conversely, in some areas tube-like structure could be observed (not shown). As at this concentration PMA exerted ambiguous effects, we chose the 1 µg/ml concentration for the following studies. Then, we performed a kinetic study with 1 µg/ml PMA for 14h, 24h or 48h of incubation: at 14h of incubation the elongation of the cells was already perceptible, but less intensive than at 24h or 48h. The most intense effects, and sometimes with tube-like structure formation, were shown after 48h, but some cells seemed to suffer at this longer time. Whatever the tested concentration or time, we could not observe fibroblast-like structures that may come from the differentiation of ECs into smooth muscle cells, since HUVECs retain this potential in vitro when deprived of fibroblast growth factor [31]. Indeed PMA mimics VEGF effects on HUVECs inducing transition from the quiescent into the proliferating-migrative phenotype.

In the first set of 2-DE experiments, three different cultures of HUVECs were performed in both conditions, 1 µg/ml PMA-treated cells versus non-treated cells (controls) for 24h of incubation. Two-D gels were compared with software and bio-informatics. The mean number of spots was 1176 ± 15 for the control group and 1247 ± 104 for the PMA group, a difference not statistically significant. The coefficient of correlation for matched spots was 0.825 between gels from control group, and 0.81 between gels from PMA group showing identical reproducibility for both groups. Moreover, between PMA-treated and control 2D gels the coefficient of correlation was 0.871 (the regression for matching spots was $y = 0.98 x + 0.001$) showing that most of proteins were not modified after PMA treatment. When the bio-informatics analysis was restricted to matched spots, 15 differences were detected, including 8 proteins with enhanced level and 7 proteins with a decrease level after PMA treatment. All the spots showing increased levels of proteins in PMA group were analysed by MALDI-TOF/TOF with peptide sequencing for protein identification: 6 proteins were identified, identically in one gel from each group when possible (Table 1). As keratins and cytokeratins could have contaminated the peptide preparation, they were not further taken into consideration. Albumin (in two forms) could have also con-

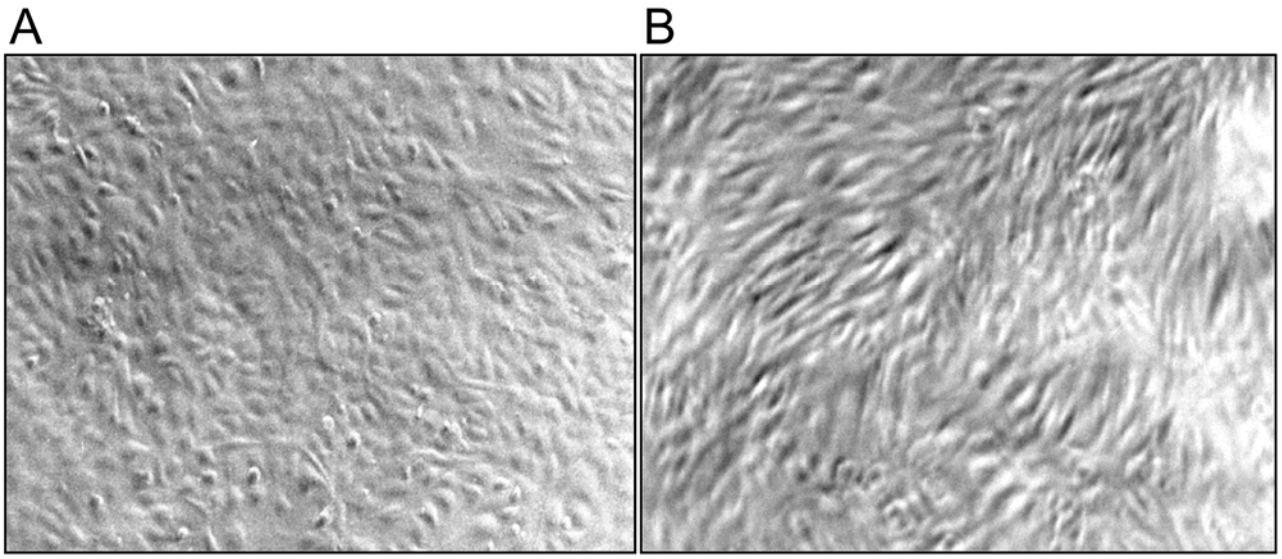


Figure 1. Morphogenic effect of PMA on HUVECs. A- Control (HUVECs at confluency). B- HUVECs treated with 1 µg/ml PMA for 24 hours. In these conditions, HUVECs presented a more elongated morphology than control cells.

Table 1. Proteins identified by MALDI-TOF/TOF differentially expressed in PMA-treated and non treated cell samples (PMA/Control); first set of experiments.

Identification	SwissProt accession n°	Protein score	Number of used peptides	Theoretical pI/Mr (kDa)	Main function and location	Spot number
Neutral alpha-glucosidase AB precursor	Q14697	497/497	33/35	5.74/106.8	Glycolysis, cytosol	144
Stress-70 protein, mitochondrial precursor	P38646	660/595	31/29	5.87/73.6	Chaperone, mitochondria	238
150 kDa Oxygen-regulated protein precursor	Q9Y4L1	174/68	13/14	5.16/111.3	Chaperone, endoplasmic reticulum	609
Isocitrate dehydrogenase (NADP)	O75874	68/71	9/7	6.53/46.6	Malic cycle, cytosol	609
Human serum albumin precursor	P02768	107/107	11/14	5.92/69.3	Plasma protein	1263
Bovine serum albumin precursor	P02769	874/822	34/33	5.82/69.2	Plasma protein	1263

taminated HUVEC preparation since foetal calf serum contains this protein. Nevertheless this particular cell type was shown to up take albumin in a more or less specific manner [32]. The three main interesting identifications were α-glucosidase, 70 kDa heat-shock protein (HSP-70) and 150 kDa oxygen-regulated protein (ORP-150), all enhanced after PMA-treatment with statistically significant increments ($p<0.05$) (Figure 2).

In a second set of 2-DE experiments, we compared 3 times of incubation (14h, 24h and 48h) with or without 1 µg/ml PMA; two 2D gels we realized for each time and condition, thus six gels in total. Gels from PMA-treated cells were compared to their controls. Other variations than in the first set of experiments were shown, but we only retained the variations in spots well matched on at least five gels. Finally the statistical significance was reached for 10 spots: 6 with enhanced protein levels in PMA-treated cells, 4 with decreased protein level in these cells compared to controls. After MALDI-TOF/TOF analysis, 4 proteins were identified : (i) HSP-90β and protein-disulfide isomerase A3 (PDI-A3), both increased in PMA-treated cells in function of time; (ii) 75 kDa glucose related protein (GRP-75) and cathepsin B, both decreased in these conditions (Table 2). More precisely, cathepsin B was expressed at low level in PMA-treated cells but was progressively more expressed in function of time reaching cathepsin B level in control cells. GRP-75 was always expressed at high level in control cells in comparison of PMA-treated cells. Conversely, HSP-90β and PDI-A3 were always expressed at high level in PMA-treated cells in comparison of controls (Figure 3).

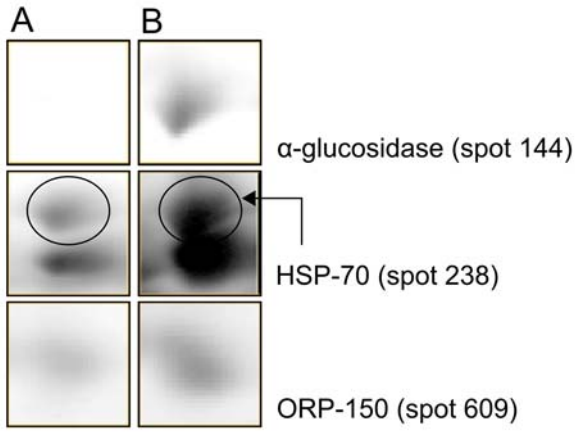


Figure 2. Protein identifications from the first set of experiments. Magnification of the 3 spots with interesting identified proteins differentially expressed in HUVECs treated with 1 µg/ml PMA for 24 hours (B) compared with control cells (not treated by PMA)(A).

4. Discussion

In our study, PMA has shown a toxic effect from 2 µg/ml even when this concentration showed more drastic morphogenic effect mimicking angiogenesis. In many studies on HUVECs, PMA was used in far more little concentrations, typically between 0.01 and 0.1 µg/ml, and often for incubations under 24 hours [7, 14, 33]. For the proteomics studies, we chose a concentration of PMA leading to a typical morphogenic effect without toxicity, and for incubations between 14 and 48 hours, Ultroser G, used as a supplement medium, was added in place of foetal calf serum. At 48 hours the morphogenic modifications were always higher than at 24 or 14 hours. Used at higher concentration, PMA was toxic for HUVECs.

In the first proteomics study, PMA was used at 1 µg/ml for 24 hours; on 2D gels, after bio-informatics analysis, 8 spots showed an enhanced level and 7 a decreased level after PMA treatment; all these modifications were statistically significant. Three interesting proteins were identified by mass spectrometry using peptide mass fingerprinting and peptide fragmentation, all enhanced after incubation of the cells with

Table 2. Proteins identified by MALDI-TOF/TOF differentially expressed in function of time in PMA-treated and non treated cell samples (PMA/Control); second set of experiments (kinetic study).

Identification	SwissProt accession n°	Protein score	Number of used peptides	Theoretical pI/Mr (kDa)	Main function and location	Spot number
Glucose regulated protein-75	P48721	68	19	5.87/73.6	Chaperone, cytosol/mitochondria	1297
Protein disulfide isomerase A3	P11598	412	37	5.88/56.9	Chaperone, Endoplasmic reticulum	1438
Heat shock protein 90 beta	P34058	185	18	4.97/83.2	Chaperone, cytosol	1726
Cathepsin B	P00787	55	3	5.88/37.7	Protease, lysosome	2783

PMA, i.e. α-glucosidase, HSP-70 and ORP-150. α-Glucosidase is a glycolytic enzyme able to liberate glucose from oligosides and polyosides, in particular the α-1,4 bound in maltose and in fragments from shark or glycogen, also the α-1,6 bound of the last glucose at glycogen branching that enables the complete hydrolysis of stock glycogen in cells. The increase in α-glucosidase after PMA treatment could be an adaptation to a defect in energy, since glucose can enhance ATP production as well as in aerobic as in anaerobic conditions. HSP-70 is a mitochondrial protein with enhanced expression normally related to a response to a stress, as well as heat shock, hypoxia, deprivation in glucose, and oxidative stress [34, 35]. This heat shock protein acts as a chaperone for a number of proteins with various functions but essential for cell survival [36]. Moreover, VEGF was shown to increase the expression of many chaperones and HSPs in HUVECs, in particular the HSPs-70 p5 and p8, in parallel of the over expression of structural proteins related to cell migration such as myosin, actin and α-tubulin [37]. ORP-150 is also a chaperone but located in endoplasmic reticulum (ER); it shares 91 % of homology with glucose-related protein 170 (GRP-170) and contains ATP-binding motif. ORP-150 expression is induced by hypoxia, 2-deoxyglucose, osmotic shock and tunikamycin [38, 39]. Its chaperone activity seems specific for vascular endothelial growth factor (VEGF) that exerts pro-angiogenic and anti-apoptotic properties [40]. In particular, ORP-150 allows VEGF secretion from ER in response to hypoxia. Moreover, when ORP-150 synthesis is inhibited, VEGF is restricted to ER canal [41]. In a number of cell types, ORP-150 appears as essential for cell survival, in particular when cells are deprived in ATP or its sources such as oxygen and glucose.

In the kinetic proteomics study, four other proteins were identified, two were enhanced in PMA-treated cells (HSP-90 β and PDI-A3), and two were decreased in PMA-treated cells (GRP-75 and cathepsin B), all differentially expressed in function of time. Unfortunately, the 3 proteins identified in the first study were not further identified in the kinectic study because the corresponding spots were not found in at least 5 gels for statistical comparison. But, the 4 proteins identified in the second study have also interesting functions. HSP-90 is a chaperone playing important functions in protein maturation, in particular for proteins controlling transcription and intercellular signalling. It also exerts anti-apoptotic properties and could be regulated by mitogenic agents. For example, PMA was shown to induce HSP-90β expression in Jurkat T lymphocytes [42]. Moreover, VEGF can induce the association between HSP-90 and Akt leading to both the activation and the over expression of NO-synthase in endothelial cells [43]. PDI-A3 (or Erp57) is a ubiquitous protein expressed in endoplasmic reticulum and belongs to the large family of the PDIs indispensable for protein folding by disulfide bridge isomerisation. In endoplasmic reticulum, Erp57 is specialized for the folding of glycoproteins, in particular when complexed to chaperones such as calnexin and calreticulin; its expression is enhanced by stress stimuli [44]. GRP-75 (or mortalin), is a chaperone belonging to the HSP-70 family, in particular for the transport of proteins into the mitochondria; it could be anti-apoptotic and is over expressed in a number of cancer cell lines [45, 46]. Cathepsin B is a ubiquitous lysosomal cysteine-protease often incriminated in cancer, in particular for tumour dissemination after degradation of the extra-cellular matrix. The relation between this protease and angiogenesis is ambiguous as sometimes it was enhanced in the angiogenic phenotype of endothelial cells. For example, in a study on HUVECs, the elevation of cathepsin B activity was associated to the formation of tubular structures such as new vessels, and in other studies cathepsin B was associated to angiogenesis inhibition with the over expression of endostatin [47-49].

5. Concluding Remarks

Our data show that a phorbol ester induces a number of protein expressions involved in multiple and intricate pathways for promoting a phenotype ensuring cell survival of endothelial cells and cell migration for new vessel formation.

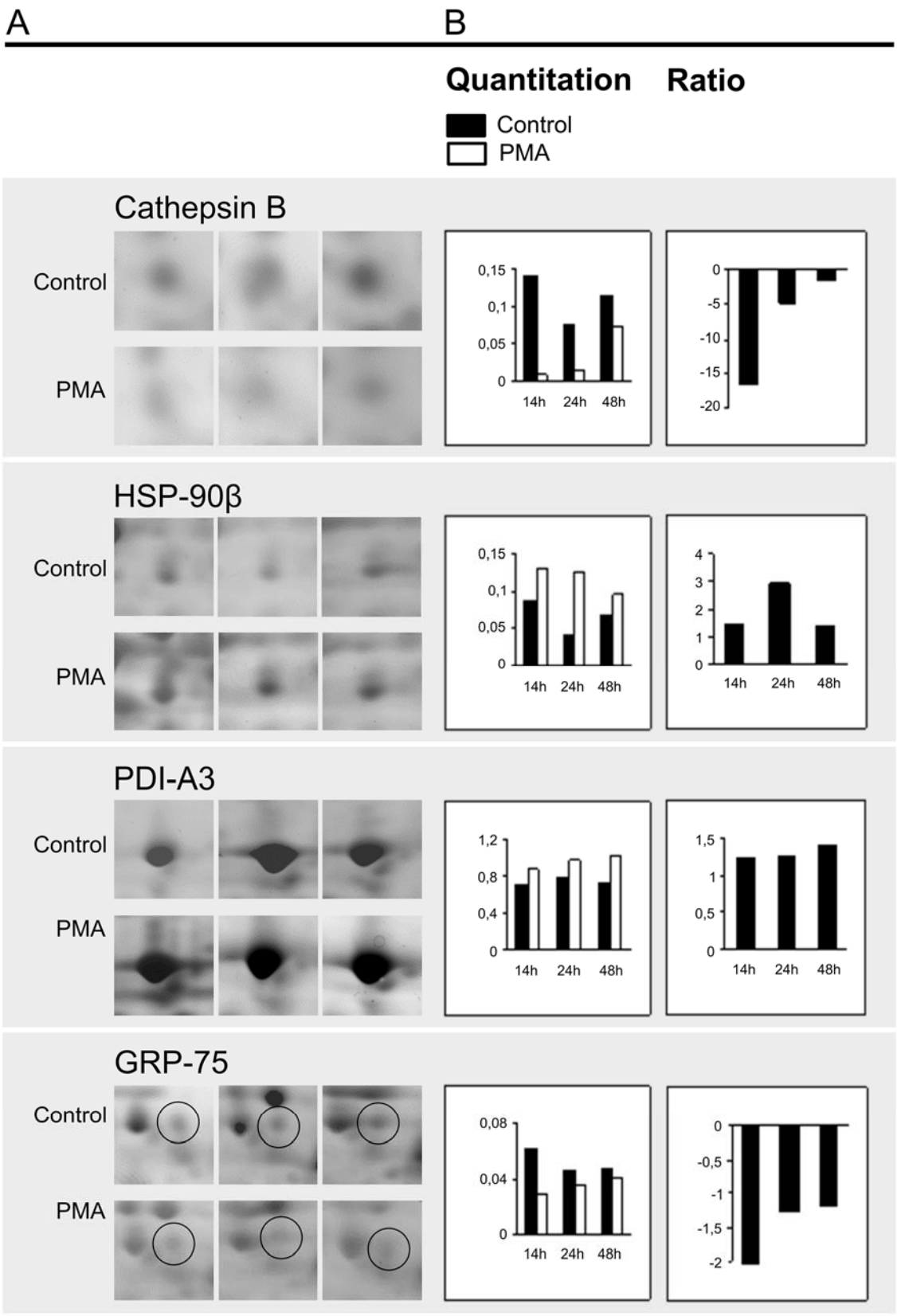


Figure 3. Protein identifications from the second set of experiments. A- Magnification of the 4 spots with interesting identified proteins differentially expressed in function of time in HUVECs treated with 1 µg/ml PMA compared with control cells (not treated by PMA). B- Histograms showing significant differences for these proteins with either enhanced or decreased levels after PMA-treatment ($p < 0.05$).

Competing interests

The authors declare no conflicts of interest for this study.

References

[1] P. Carmeliet, Nature, 438 (2005) 932-936. doi: 10.1038/nature04478

[2] D.J. Hicklin, L.M. Ellis, J Clin Oncol, 23 (2005) 1011-1027. doi: 10.1200/JCO.2005.06.081

[3] F. Hillen, A.W. Griffioen, Cancer Metastasis Rev, 26 (2007) 489-502. doi: 10.1007/s10555-007-9094-7

[4] M. Castagna, Y. Takai, K. Kaibuchi, K. Sano, U. Kikkawa, Y. Nishizuka, J Biol Chem, 257 (1982) 7847-7851.

[5] P. Mattila, M.L. Majuri, S. Tiisala, R. Renkonen, Life Sci, 55 (1994) 1253-1260. doi: 10.1016/0024-3205(94)90063-9

[6] S. Yamamura, P.R. Nelson, K.C. Kent, J Surg Res, 63 (1996) 349-354. doi: 10.1006/jsre.1996.0274

[7] C.J. Taylor, K. Motamed, B. Lilly, Angiogenesis, 9 (2006) 39 -51. doi: 10.1007/s10456-006-9028-y

[8] C.H. Hsieh, C.S. Rau, S.F. Jeng, C.J. Lin, Y.C. Chen, C.J. Wu, T.H. Lu, C.H. Lu, W.N. Chang, Exp Cell Res, 316 (2010) 1119-1126. doi: 10.1016/j.yexcr.2009.11.013

[9] R. Montesano, L. Orci, J Cell Physiol, 130 (1987) 284-291. doi: 10.1002/jcp.1041300215

[10] J.H. Reinders, R.C. Vervoorn, C.L. Verweij, J.A. van Mourik, P.G. de Groot, J Cell Physiol, 133 (1987) 79-87. doi: 10.1002/jcp.1041330110

[11] D. Ron, M.G. Kazanietz, FASEB J, 13 (1999) 1658-1676.

[12] H. Xu, P. Czerwinski, M. Hortmann, H.Y. Sohn, U. Forstermann, H. Li, Cardiovasc Res, 78 (2008) 349-355. doi: 10.1093/cvr/cvm085

[13] A.A. Ucuzian, H.P. Greisler, World J Surg, 31 (2007) 654-663. doi: 10.1007/s00268-006-0763-4

[14] N. Ilan, S. Mahooti, J.A. Madri, J Cell Sci, 111 (Pt 24) (1998) 3621-3631.

[15] M. Eyries, M. Agrapart, A. Alonso, F. Soubrier, Circ Res, 91 (2002) 899-906. doi: 10.1161/01.RES.0000042703.39845.B4

[16] S. Eligini, S.S. Barbieri, V. Cavalca, M. Camera, M. Brambilla, M. De Franceschi, E. Tremoli, S. Colli, Cardiovasc Res, 65 (2005) 683-693. doi: 10.1016/j.cardiores.2004.10.024

[17] A. Toker, Front Biosci, 3 (1998) D1134-1147.

[18] W.S. Liu, C.A. Heckman, Cell Signal, 10 (1998) 529-542. doi: 10.1016/S0898-6568(98)

[19] F. Tatin, C. Varon, E. Genot, V. Moreau, J Cell Sci, 119 (2006) 769-781. doi: 10.1242/jcs.02787

[20] A. Wang, M. Nomura, S. Patan, J.A. Ware, Circ Res, 90 (2002) 609-616. doi: 10.1161/01.RES.0000012503.30315.E8

[21] O.H. Lee, Y.M. Kim, Y.M. Lee, E.J. Moon, D.J. Lee, J.H. Kim, K.W. Kim, Y.G. Kwon, Biochem Biophys Res Commun, 264 (1999) 743-750. doi: 10.1006/bbrc.1999.1586

[22] K.R. Johnson, K.P. Becker, M.M. Facchinetti, Y.A. Hannun, L.M. Obeid, J Biol Chem, 277 (2002) 35257-35262. doi: 10.1074/jbc.M203033200

[23] B. Thompson, N. Ancellin, S.M. Fernandez, T. Hla, R.I. Sha'afi, Prostaglandins Other Lipid Mediat, 80 (2006) 15-27. doi: 10.1016/j.prostaglandins.2006.03.002

[24] C. Wong, Z.G. Jin, J Biol Chem, 280 (2005) 33262-33269. doi: 10.1074/jbc.M503198200

[25] E. Scoditti, M. Massaro, M.A. Carluccio, A. Distante, C. Storelli, R. De Caterina, Cardiovasc Res, 86 (2010) 302-310. doi: 10.1093/cvr/cvp400

[26] M. Powerski, D. Henrich, A. Sander, A. Teiler, I. Marzi, Artif Cells Blood Substit Immobil Biotechnol, 39 (2011) 214 -222. doi: 10.3109/10731199.2011.555838

[27] B. Baudin, A. Bruneel, N. Bosselut, M. Vaubourdolle, Nat Protoc, 2 (2007) 481-485. doi: 10.1038/nprot.2007.54

[28] A. Bruneel, V. Labas, A. Mailloux, S. Sharma, J. Vinh, M. Vaubourdolle, B. Baudin, Proteomics, 3 (2003) 714-723. doi: 10.1002/pmic.200300409

[29] A. Bruneel, V. Labas, A. Mailloux, S. Sharma, N. Royer, J. Vinh, P. Pernet, M. Vaubourdolle, B. Baudin, Proteomics, 5 (2005) 3876-3884. doi: 10.1002/pmic.200401239

[30] K. Le Moguen, H. Lincet, E. Deslandes, M. Hubert-Roux, C. Lange, L. Poulain, P. Gauduchon, B. Baudin, Proteomics, 6 (2006) 5183-5192. doi: 10.1002/pmic.200500925

[31] A. Ishisaki, H. Hayashi, A.J. Li, T. Imamura, J Biol Chem, 278 (2003) 1303-1309. doi: 10.1074/jbc.M207329200

[32] T.A. John, S.M. Vogel, C. Tiruppathi, A.B. Malik, R.D. Minshall, Am J Physiol Lung Cell Mol Physiol, 284 (2003) L187-196. doi: 10.1152/ajplung.00152.2002

[33] R.P. Schleimer, B.K. Rutledge, J Immunol, 136 (1986) 649-654.

[34] L. Jornot, A.F. Junod, Am J Respir Cell Mol Biol, 6 (1992) 107-115. doi: 10.1165/ajrcmb/6.1.107

[35] D. Dreher, J.R. Vargas, D.F. Hochstrasser, A.F. Junod, Electrophoresis, 16 (1995) 1205-1214. doi: 10.1002/elps.11501601201

[36] E.A. Craig, J.S. Weissman, A.L. Horwich, Cell, 78 (1994) 365-372. doi: 10.1016/0092-8674(94)

[37] Z. Pawlowska, P. Baranska, H. Jerczynska, W. Koziolkiewicz, C.S. Cierniewski, Proteomics, 5 (2005) 1217-1227. doi: 10.1002/pmic.200400983

[38] J. Ikeda, S. Kaneda, K. Kuwabara, S. Ogawa, T. Kobayashi, M. Matsumoto, T. Yura, H. Yanagi, Biochem Biophys Res Commun, 230 (1997) 94-99. doi: 10.1006/bbrc.1996.5890

[39] Y. Bando, Y. Tsukamoto, T. Katayama, K. Ozawa, Y. Kitao, O. Hori, D.M. Stern, A. Yamauchi, S. Ogawa, FASEB J, 18 (2004) 1401-1403. doi: 10.1096/fj.03-1161fje

[40] K. Ozawa, Y. Tsukamoto, O. Hori, Y. Kitao, H. Yanagi, D.M. Stern, S. Ogawa, Cancer Res, 61 (2001) 4206-4213.

[41] K. Ozawa, T. Kondo, O. Hori, Y. Kitao, D.M. Stern, W. Eisenmenger, S. Ogawa, T. Ohshima, J Clin Invest, 108 (2001) 41-50. doi: 10.1172/JCI11772

[42] J.M. Wu, L. Xiao, X.K. Cheng, L.X. Cui, N.H. Wu, Y.F. Shen, J Biol Chem, 278 (2003) 51143-51149. doi: 10.1074/jbc.M305537200

[43] J. Fontana, Circulation Research, 90 (2002) 866-873. doi: 10.1161/01.res.0000016837.26733.be

[44] P. Maattanen, G. Kozlov, K. Gehring, D.Y. Thomas, Biochem Cell Biol, 84 (2006) 881-889. doi: 10.1139/o06-186

[45] C.C. Deocaris, S.C. Kaul, R. Wadhwa, Cell Stress & Chaperones, 11 (2006) 116. doi: 10.1379/csc-144r.1

[46] A.M. Czarnecka, C. Campanella, G. Zummo, F. Cappello, Cancer Biol Ther, 5 (2006) 714-720.

[47] S. Caserman, T.T. Lah, Int J Biol Markers, 19 (2004) 120-129.

[48] E. Im, A. Venkatakrishnan, A. Kazlauskas, Mol Biol Cell, 16 (2005) 3488-3500. doi: 10.1091/mbc.E04-11-1029

[49] A. Premzl, V. Turk, J. Kos, J Cell Biochem, 97 (2006) 1230-1240. doi: 10.1002/jcb.20720

JOURNAL OF INTEGRATED OMICS
A METHODOLOGICAL JOURNAL
HTTP://WWW.JIOMICS.COM

ORIGINAL ARTICLE | DOI: 10.5584/jiomics.v3i2.138

Cellular Protein/Peptide Expression Profiles (PEPs): an alternative approach for easy identification of cyanobacterial species

Fred Wang-Fat Lee[1,2], Kin-Chung Ho[2], Daniel Yun-Lam Mak[1], Junrong Liang[3], Changping Chen[3], Yahui Gao[3], Samuel Chun-Lap Lo*[1]

[1]The Proteomic Task Force, Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hong Kong; [2]School of Science and Technology, The Open University of Hong Kong, Hong Kong; [3]Key Laboratory of the Ministry of Education for Coastal and Wetland Ecosystems, and School of Life Sciences, Xiamen University, Xiamen 361005, China.

Received: 02 May 2013 **Accepted:** 10 October 2013 **Available Online**: 23 December 2013

ABSTRACT

Cyanobacterial harmful algal blooms (CyanoHABs) are recognized as an expanding and serious global problem that threatens human health. Timely and accurate identification of cyanobacteria is of vital importance for public health. Morphologic characteristics of cyanobacteria have been used for classical taxonomic studies and identification purposes. However, misidentification may occur either due to subjective judgment by the operators or inability to recognize natural variations of morphotypes. To circumvent problems of morphology-based identification methods, we reported previously a rapid and simple method for the identification of dinoflagellates using protein/peptide expression profiles (PEPs) of whole cell protein extracts generated by MALDI-TOF-MS (Lee FWF *et. al*., 2008). In the present study, we applied this method in the identification of harmful cyanobacteria. Our results showed that various species of the cyanobacteria can be easily distinguished from each other using their PEPs.

Keywords: Cyanobacteria; Harmful Algal Bloom; Identification; MALDI-TOF-MS; Protein Expression Profiles (PEPs).

Abbreviations

CHCA: α-cyano-4-hydroxycinnamic acid; **DHB**: 2,5-dihydroxybenzoic acid; **HABs**: Harmful Algal Blooms; **PEPs**: Protein Expression Profiles; **SA**: sinapinic acid.

1. Introduction

Cyanobacteria, also known as blue-green algae, play an important role as primary producers of the food web [1, 2]. Most of them are single-celled organisms that can be found in almost every conceivable environment, from lakes, ponds, rivers brackish and marine waters, bare rock to soil throughout the world. Cyanobacteria can also grow extensively, resulting in harmful algal blooms (HABs) that can cause serious negative impacts on human health and the aquatic ecosystems. Some of the cyanobacteria produce powerful toxins including neurotoxic, hepatotoxic, dermatotoxic, as well as other bioactive compounds. Blooms of these toxic cyanobacteria pose additional threat if the blooms occur in reservoirs and other drinking water sources [3-5]. These cyanotoxins are responsible for fresh and brackish water intoxication and the intoxication of animals has been widely reported around the world [6]. Scientists generally name the blooms caused by cyanobacteria as CyanoHABs. These blooms have dramatically increased in recent decades all over the world. The recent freshwater blue-green algal bloom that occurred in Taihu (Lake Tai), Jiangsu Province, China (starting end of May 2007) is an example showing how serious the effects can be. A population of more than two million people living around Lake Tai was affected by the bloom because Taihu was the city's sole water supply, leaving them without drink-

***Corresponding author:** Samuel Chun-Lap Lo. The Proteomic Task Force, Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hong Kong. Tel: (852) 3400 8669. Fax: (852) 2364 9932. Email-Address: samuel.chun-lap.lo@polyu.edu.hk

ing water for at least a week [7]. Truck-loads of bottled drinking water had to be sent from surrounding provinces for temporary relief. At the initial stages of the bloom, there was additional fear on whether the causative agent was toxic or not. Hence, fast and accurate cyanobacterial species identification is important for surveillance and possibly prevention and control of CyanoHABs.

Cyanobacteria are a morphologically diverse group of organisms ranging from unicellular, colonial to filamentous forms. The class of Cyanophyceae includes 150 genera and around 2000 species. They are now placed within the group Eubacteria in the phylogenetic taxonomy [8]. Traditionally, identification and classification of cyanobacteria has been based on their morphological features such as cell size, shape and arrangement, pigment coloration and the presence of characters such as gas vacuoles and a sheath [9]. However, these types of identification are time-consuming, laborious and a high-level of expertise is required for identifying and distinguishing morphological characters for identification purposes. More than 50% of the known strains identified using morphological-based strategies were estimated to be misidentified [9]. The lack of distinguishing morphological features in some strains/species, the changes and variations of some diagnostic characters during different growth or environmental conditions, as well as the subjective judgment by the operators can lead to errors, resulting in incorrect species identification. Limitations of the morphology-based species identification method implicated a pressing need for the development of more reliable and rapid methods for the identification of cyanobacteria. Nowadays, DNA technologies have been routinely applied in areas of species identification of microorganism including cyanobacteria. For example, one of the common approaches is the analysis of DNA sequence similarities of 16S rRNA gene [10] and 16S-23S rRNA internal transcribed spacer (ITS) [11]. However, these PCR-based methods are labor-intensive as well as difficult to meet demands of high-throughput and especially in cases of previously unknown cyanobacterial strains.

Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) can be used to analyze the protein expression profile of a cell. This method has emerged as a new technology for species identification for various microorganisms and has been extensively reviewed [12, 13]. The MALDI-based identification method is praised as objective, fast, simple and reliable. It usually requires minimal amounts of biological material and is suitable for high-throughput routine analysis. Therefore, it has great potential for applications in clinical microbiology and environmental monitoring [13, 14]. The central idea of this method is to generate PEPs from intact unicellular cells, or cell lysates. Despite the fact that only ionized proteins/peptides that are in abundance in the cells of interest are detected in the MALDI-TOF-MS, the PEPs generated are usually genus-, species- and sometimes even strain-specific (Dworzanski and Snyder, 2005; Fenselau and Demirev, 2001). It is possible to identify the species within a few minutes [12, 13, 15]. Furthermore, Mellmann et al. (2008), showed that the MALDI-TOF-MS based identification method is more robust than the 16S rRNA gene sequencing method for species identification of non-fermenting bacteria. It is because the MALDI-TOF-MS method can provide differentiating information even when that 16S rRNA- based identification fails [15]. Although the MALDI-TOF-MS based method has been applied to a wide range of microorganism, for example bacteria [15-19]; parasites [20] and fungi [21-24], the successful application of protein-expression-profiles (PEPs) for species identification of blue-green algae has never been reported.

Our group has previously shown the successful application of using PEPs generated by MALDI-TOF-MS for the identification of different dinoflagellates species (one of the major HAB causative agents) [25]. To further extend its application, in this study, we evaluated the use of MALDI-TOF-MS based methodology for rapid identification of different cyanobacterial species/strains. We had further adopted and simplified the sample preparatory procedures for cyanobacterial investigations. Besides being more effective, this protocol is a refined version when compared to the one our group reported earlier [25]. The eventual establishment of a common protocol for all HAB causative agents will facilitate the development of automatic systems for environmental surveillance purposes. Lastly, identification of some cyanobacterial strains based on MALDI-TOF-MS detection of their low-mass toxins presented has been reported previously [26]. However, to our knowledge, this is the first report to demonstrate the potential use of protein expression profiles (PEPs) generated by MALDI-TOF-MS for rapid and objective identification of cyanobacteria.

2. Material and Methods

2.1 *Cyanobacterial species and strains*

Cyanobacteria species used in this study are listed in Table 1. Four isolates of *Synechococcus sp*. were isolated by Prof. Gao Ya-Hui of The Xiamen University. The other species, including *Lyngbya aestuarii* (CCMP473), two isolates of *Planktothrix agardhii* (CCMP600 and CCMP601), *Arthrospira platensis* (CCMP1295), *Oscillatoria sp.* (CCMP1731), *Cylindrospermopsis raciborskii* (CCMP1973) and *Anabaena sp.* (CCMP2066) were purchased from The National Center for Marine Algae and Microbiota (NCMA, previously Provasoli-Guillard National Center for Culture of Marine Phytoplankton (CCMP)). NCMA had extensively verified identities of the cyanobacterial cultures before they were sold while identities of other cultures were confirmed by DNA sequences of 16S-23S rRNA ITS.

2.2 *Culture conditions*

Primarily, f/2 or DY-V media were used for culturing the cyanobacteria. Stock cultures of all cyanobacteria were kept at exponential growth phase by transferring to new medium

Table 1. Cyanobacteria species and strains used in this study

Species/strains	Collection site
Anabaena sp. (CCMP2066)#	Red River, Trollwood Park, Fargo, North Dakota, USA
Arthrospira platensis (CCMP1295)#	Yunnan Province ,China
Cylindrospermopsis raciborskii (CCMP1973)#	Florida, USA
Lyngbya aestuarii (CCMP473)#	Woods Hole, Massachusetts, USA
Oscillatoria sp. (CCMP1731)#	Isla do Sol, Cape Verde Islands, Africa
Planktothrix agardhii (CCMP600)#	Lake Kolbotnvatn, Akershus, Norway
Planktothrix agardhii (CCMP601)#	Lake Kolbotnvatn, Akershus, Norway
Synechococcus sp. (FACHB-460)*	Wuhan, China
Synechococcus sp. (FACHB-562) *	Wuhan, China
Synechococcus sp. (FACHB-940) *	Wuhan, China
Synechococcus sp. (FACHB-7820) *	Wuhan, China

CCMP: The Provasoli-Guillard National Centre for Culture of Marine Phytoplankton (Current name NCMA: National Center for Marine Algae and Microbiota).
FACHB: Freshwater Algae Culture Collection of the Institute of Hydrobiology.
*The species were kindly provided by our Chinese partners in Xiamen University and were identified by the DNA sequences of 16S-23S internal transcribed spacer (ITS) region (see supplementary information, Figure S1).
According to the information provided, species has been identified by taxonomists.

every two week in a ratio of 1:10 (v/v). Vegetative cells from cultures in mid- or late-exponential phase of growth were inoculated into freshly prepared culture medium. Possible contamination of the culture was monitored by regular microscopic examination. The cultures were grown at 20 °C under a 12-h light:12-h dark cycle at a light intensity of 120 μE lx $m^{-1}s^{-1}$ provided by cool white fluorescent tubes in a Conviron growth chamber (Model EF7).

2.3 *16S-23S rRNA ITS sequencing*

DNA extraction was performed according to the procedures reported previously [25]. 16S-23S rRNA ITS region was amplified from the extracted DNA using the cyanobacterial 16S-specific primer CSIF: 5' GYCACGCCCGAAGTCRTTAC 3' and 23S-specific primer ULR: 5' CCTCTGTGTGCCWAGGTATC 3' [27]. PCR were performed under conditions: 95℃ 5min; 35 cycles of 94℃ 45s, 50℃ 45s and 72℃ 2min; 72℃ 10min. PCR products were cloned into pGEM-T easy vectors (Promega, USA) prior to DNA sequencing. DNA sequencing of all cloned plasmids were performed by commercial facilities using traditional dideoxy-methodolgy.

2.4 *Sample preparation to obtain PEPs using MALDI-TOF MS*

Cyanobacterial samples were prepared using methodology as described previously with minor modifications [25] and the workflow is shown in Figure 1. Briefly, exponential 100-200 ml cultures with 10^5 cells were collected by centrifugation (1500 x g for 15 min at room temperature). Cell pellets were re-suspended in 100 μL 0.1% trifluoroacetic acid (TFA) (Aldrich, USA). These cells were broken by quick sonication (a total of 2 min with short pulses of 10 s each) on ice. Cell debris was removed by centrifugation at 13,000 x g at room temperature for 5 min. Inorganic salts in the samples were cleaned up by absorbing the proteins onto C-18 zip tips (Millipore, USA) according to the manufacturer user manual. Proteins were eluted from the zip tip with 1-2 μL 0.1% TFA in 70% acetonitrile. Eluted proteins solutions were mixed with matrix solution in a ratio of 1:1 (v/v). Matrix solution contained saturated sinapinic acid (SA) in 1:1 (v/v) 0.1% TFA/acetonitrile. 1 μL of the resulting mixtures were then spotted onto a mass spectrometer target plate (MTP AnchorChip™ 600/384 T F) (Bruker, Germany).

2.5 *MALDI-TOF MS analysis*

Proteins expression profiles (PEPs) of all samples being studied were obtained with a MALDI-TOF spectrometer (Autoflex III Smartbeam, Bruker, Germany) in linear mode at an accelerating voltage of 20 kV by using a 300 ns delay time and over a mass range of 2000–16000 Da. For each sample, spectra from 500 laser shots at several different positions were combined to generate a mass spectrum. The mass spectra were calibrated externally using Protein Calibration Standard I (Bruker, Germany) and were used to provide a miminium mass accuracy of at least 1 part in 3000. The calibrant mixture contains insulin (5734.51 Da),

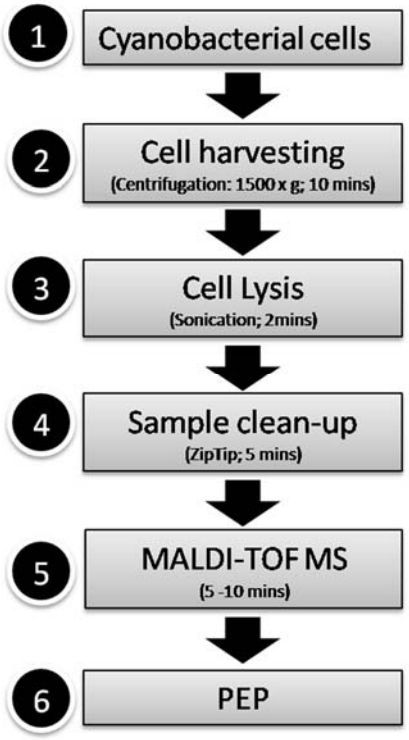


Figure 1. Summarized workflow of the HAB species sample preparation for MALDI-TOF-MS analysis. Time spent on MALDI sample preaprations include sample spotting and crystallization as well, indicating the short turn around time of results.

ubiquitin (8565.76 Da), cytochrome c (12360.97 Da) and myoglobin (16952.3 Da). Internal calibration was performed by adding purified insulin (5734.51 Da) and ubiquitin (8565.76 Da) directly into the sample mixture. Depending on the range of peak mass ions of interest, the combined strategy of calibration ensured the mass accuracy of the spectrum to be within 0.5-2 Da. Fresh calibration was performed for different samples and for different individual experiments. Prior to data analysis, each spectrum was baseline corrected and smoothed according to the TopHat and Savitzky-Golay smoothing algorithm. Normalized spectra were transferred to the software "Flex Analysis" ver3.0 (Bruker, Germany) for automated peak extraction and analysis as described in the user manual. PEPs shown in the manuscript were the representative data obtained from a triplicate analysis of 3 different batches of the samples.

3. Results and Discussion

3.1 *Sample preparation for MALDI-TOF-MS analysis*

The methodology described for sample preparation is instrumental for MALDI-TOF-MS analysis because it determines whether a high-quality spectrum of the PEP can be obtained. There are several factors that are most critical in affecting the quality of spectra obtained. These include types of matrix, extraction solvent and sample preparation procedures such as sample clean-up. We had adopted the workflow of the methodology we described previously [25] for application in cyanobacterial samples (Figure 1). In order to have appropriate conditions leading to good-quality mass spectra, we have tested these three factors in this study.

The choice of matrix is important in MALDI-TOF-MS detection because it can promote ionization of the analytes. The three most commonly used matrices are: α-cyano-4-hydroxycinnamic acid (CHCA), 2,5-dihydroxybenzoic acid (DHB) and 3,5-dimethoxy-4-hydroxycinnamic acid/ sinapinic acid (SA). These matrices were screened for compatibility with the analysis of cyanobacterial samples. Among the three matrices tested, only samples with SA show a detectable and high-quality mass spectrum (data not shown). With this matrix, it was possible to observe the ion peaks within the mass window between 2000-20000 Da. Previous reports highlighted that most of the ion peaks observed were also within this mass range [13, 25]. However, for the species detected in the present study, no significant peaks were observed in between 16000–20000 Da. Therefore, results with mass window of 2000-16000 Da was displayed in this study. Signals of other two matrices (i.e CHCA and DHB) were in low resolution with high background noise in the corresponding mass range. The findings were comparable to our previous study [25]. Hence these matrices are inappropriate for MALDI-TOF-MS analysis of cyanobacterial samples.

The conditions of protein extraction can also affect the appearance and quality of the mass spectra. One of the explanations is the solubility of proteins in a given solvent varies according to their degrees of *pI* and hydrophobicity [28]. Acid is used in the sample preparation to provide charges on the protein. Some studies showed that trifluoroacetic acid (TFA) gave a better signal intensity than other common extraction solvents such as formic acid [13, 23, 25, 29]. This solvent was shown to allow the detection of a large number of mass peaks in the mass range of interest. Therefore, this solvent was used in the present study for protein extraction. We have evaluated different concentrations of TFA used for optimal protein extraction, including 0.1%, 1% and 5 %. Among the concentrations of TFA studied, 0.1% TFA gave the best result in term of signal intensity (data not shown). The result was also comparable to our previous work on dinoflagellates [25].

Interferences such as salts and pigments that are present in the sample can also affect spectrum quality. Of these, salt is one of the major concerns as it would suppress or modify the desorption / ionization proprieties of proteins so that some ions are detected preferentially to others [30]. Therefore, in order to obtain a high quality MS spectrum, it is important to obtain a clean and desalted sample prior to mass spectrometry. For sample absorption with C-18 ziptips was found to be a simple and effective method for sample clean-up in dinoflagellate samples [25]. Subsequently, we evaluated the same method for desalting the cyanobacterial samples

and found that much more peaks were obtained with the desalted sample than that of the sample without desalting (data not shown). One of our aims in this study was to determine the simplest sample preparation procedure that is applicable for various types of HAB causative agents, including dinoflagellates and cyanobacteria (blue-green algae). Based on the results in the present and previous studies, we have summarized the sample preparation procedure and workflow (Figure 1). We believe the procedures described are generally applicable in MALDI-TOF-MS-based species identification in the studies of HABs.

3.2 *MALDI analysis of cyanobacterial species/strains*

To investigate the discriminatory ability of different cyanobacteria species by MALDI-TOF-MS, a total of eleven cyanobacterial strains, representing seven species were analyzed (Table 1). They are all environmentally important species and most of them, including *Anabaena* sp, *Cylindrospermopsis* sp, *Lyngbya* sp, *Oscillatoria* sp, *Planktothrix* sp and *Synechococcus* sp, are responsible for CyanoHABs. All seven reference species tested in the study showed unique PEPs in a mass range between 2000 and 16000 Da (Figure 2). For each PEP, the number of mass ion peaks observed is around 10-15. MALDI-TOF-MS-based identification of other microorganisms showed similar numbers of discrete mass ions observed in the same mass range [13]. Beside the unique PEPs observed from different species, a number of unique and consistently occurring species-specific ions can also be annotated from the spectra. Typically, 5 to 10 mass ions are reported to be sufficient to discriminate bacteria at the species level [14]. Although the species-specific ions found in the present study, based on the mass ion peaks observed in the mass range of 2000–16000 Da, cannot be regarded as biomarkers for the corresponding species until more PEPs of all existing cyanobacteria are analyzed in the future, we clearly demonstrated that either the PEPs or the individual specific ions can be used unambiguously for the identification of different cyanobacteria. Furthermore, whether a sample contains toxin can be determined if biomarker ions specific for toxin production can be found and used. Otherwise, other techniques, such as HPLC, are needed for analysis of toxin production. Nonetheless, based on known characteristics of toxic species/strains, the PEP technology can be used as an initial screening tool for high-throughput.

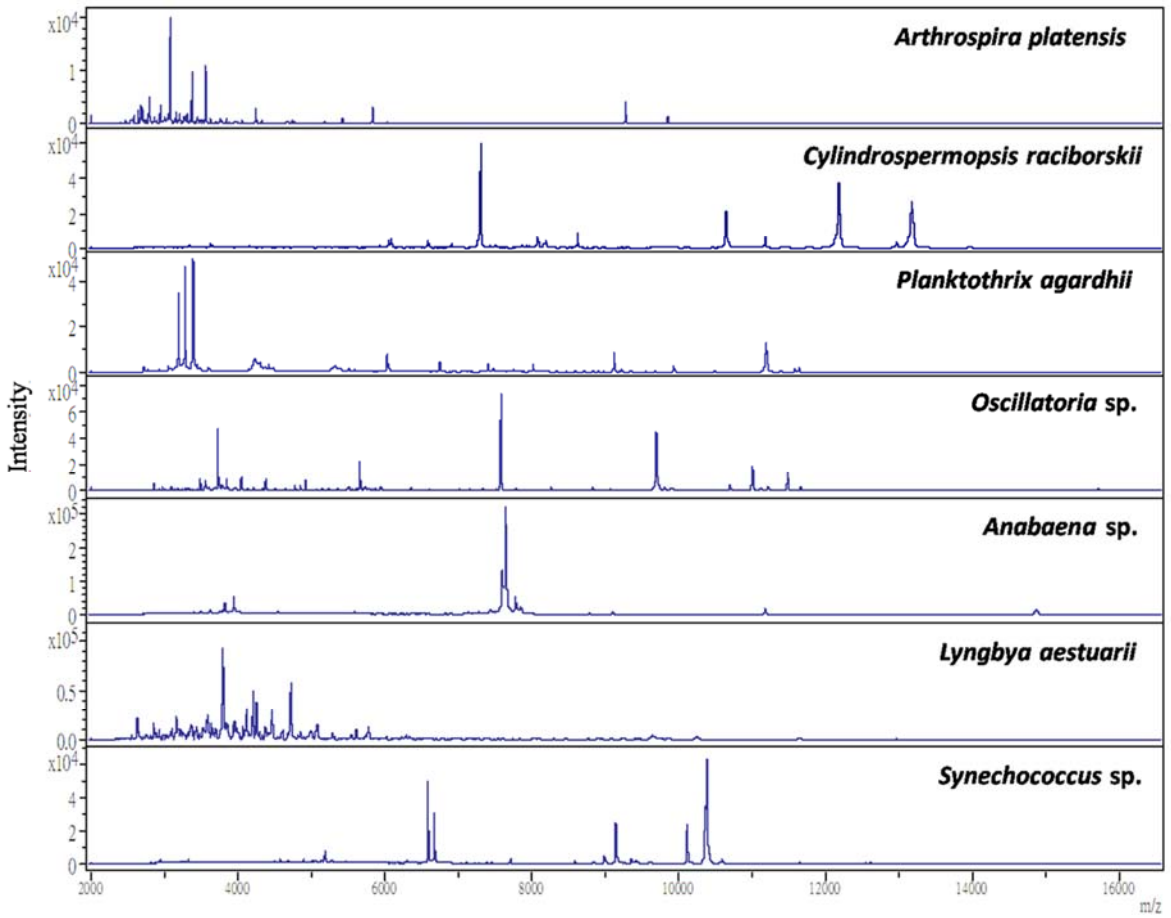


Figure 2. MALDI-TOF-MS protein expression profiles of cyanobacterial species in different genus (*Planktothrix agardhii* (CCMP600) and *Synechococcus sp.* (FACHB-940) were used in the experiment).

In order to test whether the method can differentiate different cyanobacteria at levels below the species level, strains of *Synechococcus sp* and *Planktothrix agardhii* were used as examples for MALDI-TOF-MS analysis under identical conditions (Figure 3). These strains are difficult to distinguish from each other using the traditional morphological -based identification method and even molecular gene sequencing. For instance, identification of *Synechococcus* species is unclear and there are numerous limitations with the current taxonomy of *Synechococcus* [31]. Recently, 16S-23S rRNA ITS sequence has been shown to be variable in different strains of cyanobacteria and thus can be used as a reliable identification tool [32]. 16S-23S rRNA ITS region of the four strains of *Synechococcus spp.* were sequenced (see supporting information, Figure S1). Identities of the strains were confirmed by GenBank BLAST searches. Except at most with two nucleotides, the ITS sequences of the four strains were nearly identical to each other. This result suggested that the ITS sequence alone may not sufficient to differentiate the polymorphism of the four *Synechococcus* strains. On the other hand, the PEP of the four strains obtained from the MALDI analysis display high levels of overall similarity (Figure 3A). They shared most of the common peaks (e.g "A" peaks). However, some distinguishing strains-specific ions ("B" peaks) could be easily pinpointed from the spectra (Figure 3B). These "B" peaks can only be observed from the spectra of specific *Synechococcus* strains, and their corresponding spectrum is highly similar to each other. For example, peak with ~6888 Da can only be found in FACHB-7820). Based on this finding, it was suggested that these two strains which may belong to the same group, which is different from the other two strains (FACHB-562 and FACHB-7820). Similar result was observed in the MALDI-TOF-MS analysis of two *Planktothrix agardhii* strains (Figure 3C). PEPs of the two strains were highly similar to each other, but some strain-specific biomarkers were found for each individual strain ("C" peaks). Therefore, our results demonstrated that the mass spectral data (including the PEPs and the bi-

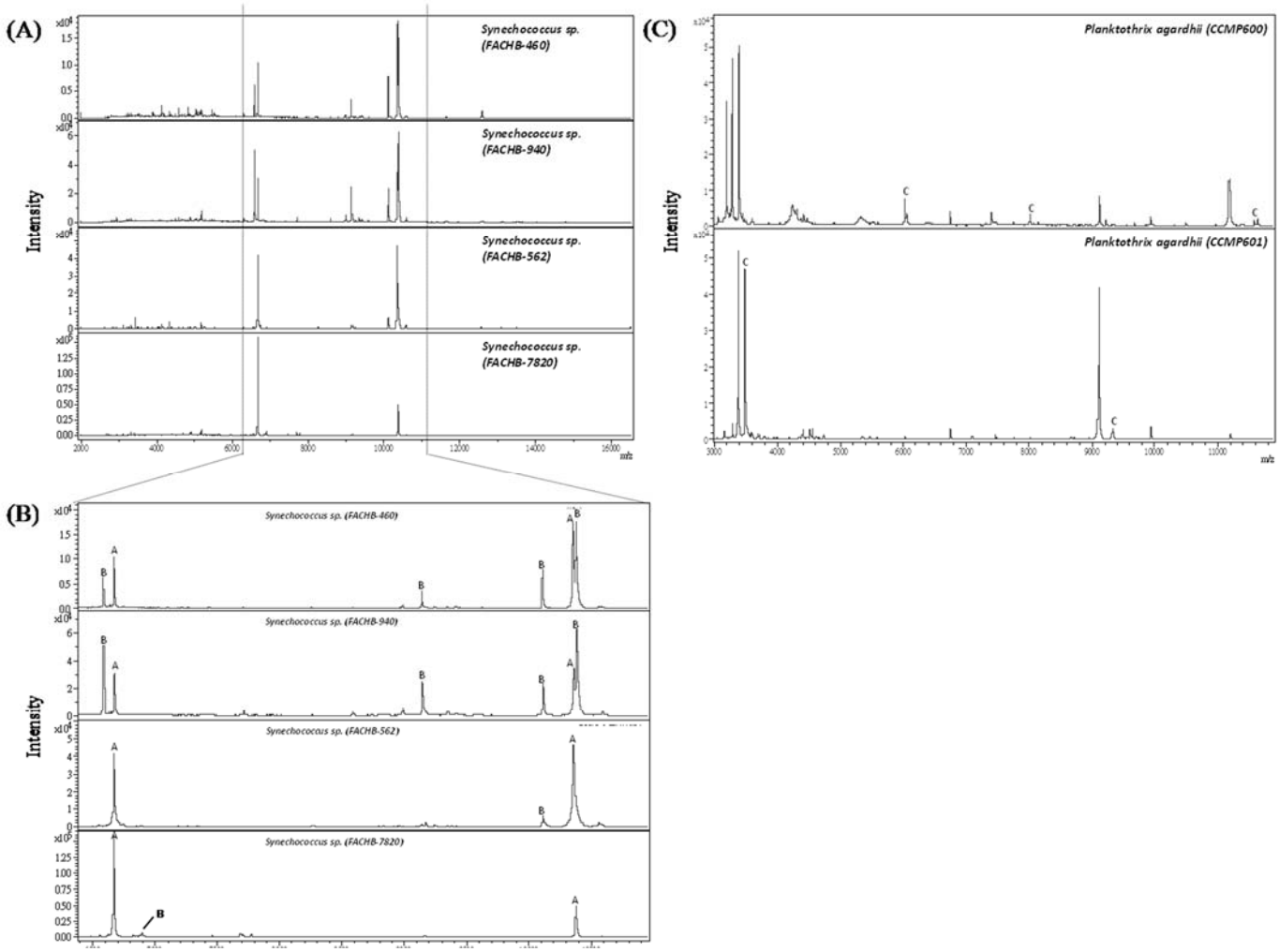


Figure 3. MALDI-TOF-MS protein expression profiles of cyanobacterial strains including (A) 4 strains of *Synechococcus sp*. and (C) 2 strains of *Planktothrix agardhii*. (B) The enlarged region of the *Synechococcus* strains spectra ranged from 6500-11000 m/z. Peaks labeled with "A", "B" and "C" representing the common peaks of all the 4 *Synechococcus* strains, the respective signature peaks of the corresponding *Synechococcus* strains and the respective signature peaks of the corresponding *Planktothrix* strains respectively.

omarkers) of the strains are not only sufficient to distinguish cyanobacteria of different genera and species but also potentially to differentiate different strains.

3.3 *Reproducibility and PEPs under different conditions*

The PEPs presented in Figure 2 and 3 were representatives of three different batches of the same samples. With the sample preparation and MALDI-TOF-MS analysis performed under the same conditions, the acquired PEPs are identical between different individual batches of samples. On the other hand, growth parameters, such as different growth media and growth phase of the organism, were reported to be important for the reproducible generation of the mass spectra and should therefore be carefully controlled [33, 34]. With the aim of achieving reproducible and reliable identification of the species/strains, we compared the PEPs of two different species (*Oscillatoria sp*. and *Synechococcus sp*.) grown under different culture media (Figure 4). Under all culture media used with a defined condition, the spectra from both pair of samples showed a high degree of uniformity. It seems that the PEPs obtained were not affected by different cultivation media used. Growth age is another important factor for the MALDI-TOF-MS based identification of cyanobacteria. Interestingly, intensity of one of the "signature" peaks (~6661 Da, indicated by arrow in Figure 4) of the *Synechococcus* sp was found to be greatly diminished if very old cultures of cells (>25 days) were used for the study (Data not shown). Relationship between the culture cells in dead growth phase and the peak intensity remains to be elucidated in further studies. However, similar phenomenon was observed in previous study in the application of PEPs method to the field samples [35]. In order to determine the PEPs of the cells in different growth phases, MALDI spectra of two cyanobacterial species of different growth time were recorded (Figure 5). However, the patterns and number of peaks obtained from different growth ages varies from each other. Some peak mass ions were observed all the time in the exponential and stationary growth phases, and others appeared transiently in only one of these growth phases. For example, the number and patterns of peaks observed from 2000-7000 Da in the spectra of *Oscillatoria* species in different growth ages were different (Figure 5A). The protein expression changes may be explained by the alterations of metabolic activities related to different culture ages of the cells. However, some peak mass ions were constantly expressed and observed in

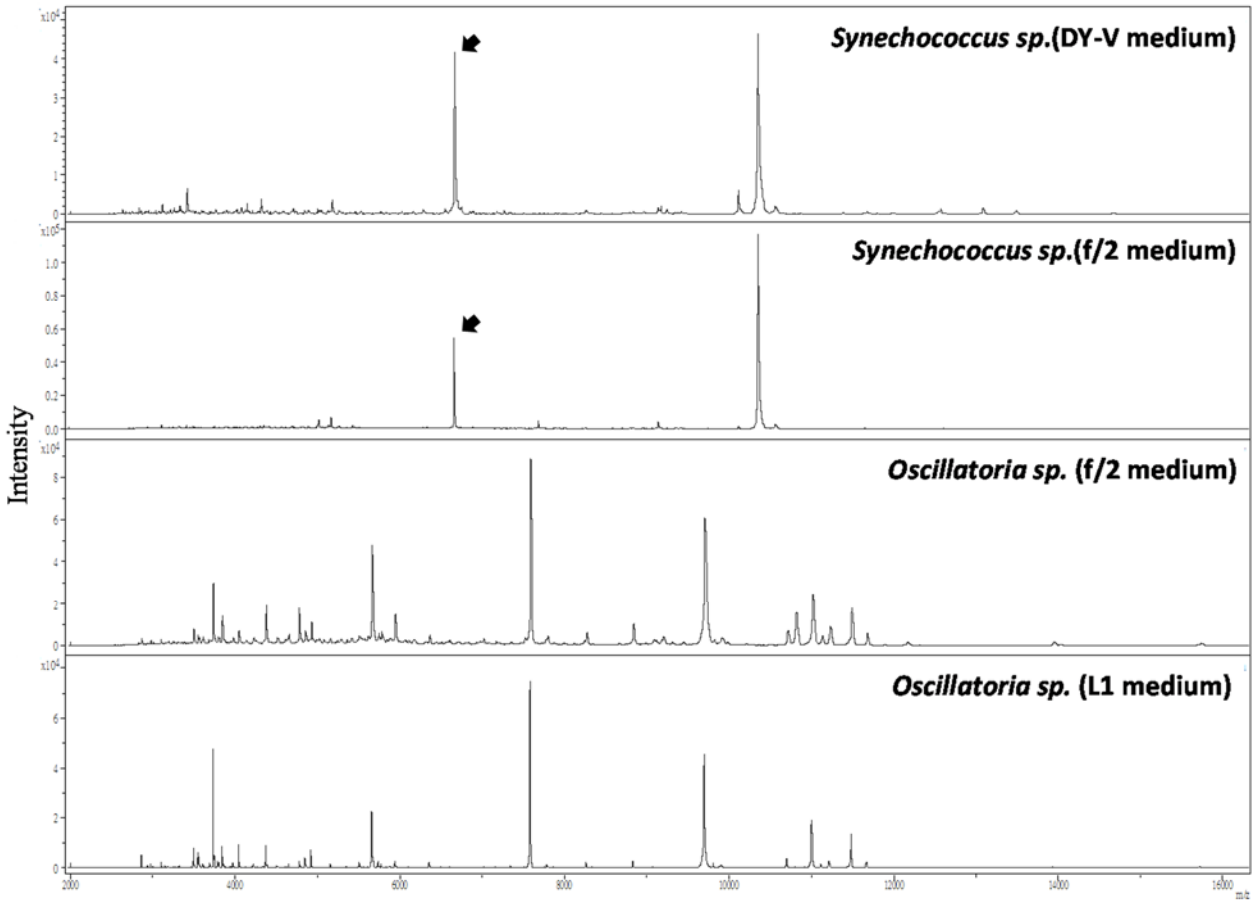


Figure 4. MALDI-TOF-MS protein expression profiles of cyanobacterial species grown in different cultivate media. (7-days cultures of *Oscillatoria sp* (CCMP1731) and *Synechococcus sp.* (FACHB-562) were used in the experiment). Arrow indicates the peak would greatly deduce in its intensity if old cultures of *Synechococcus* cells (i.e >25 days) were used for the experiments.

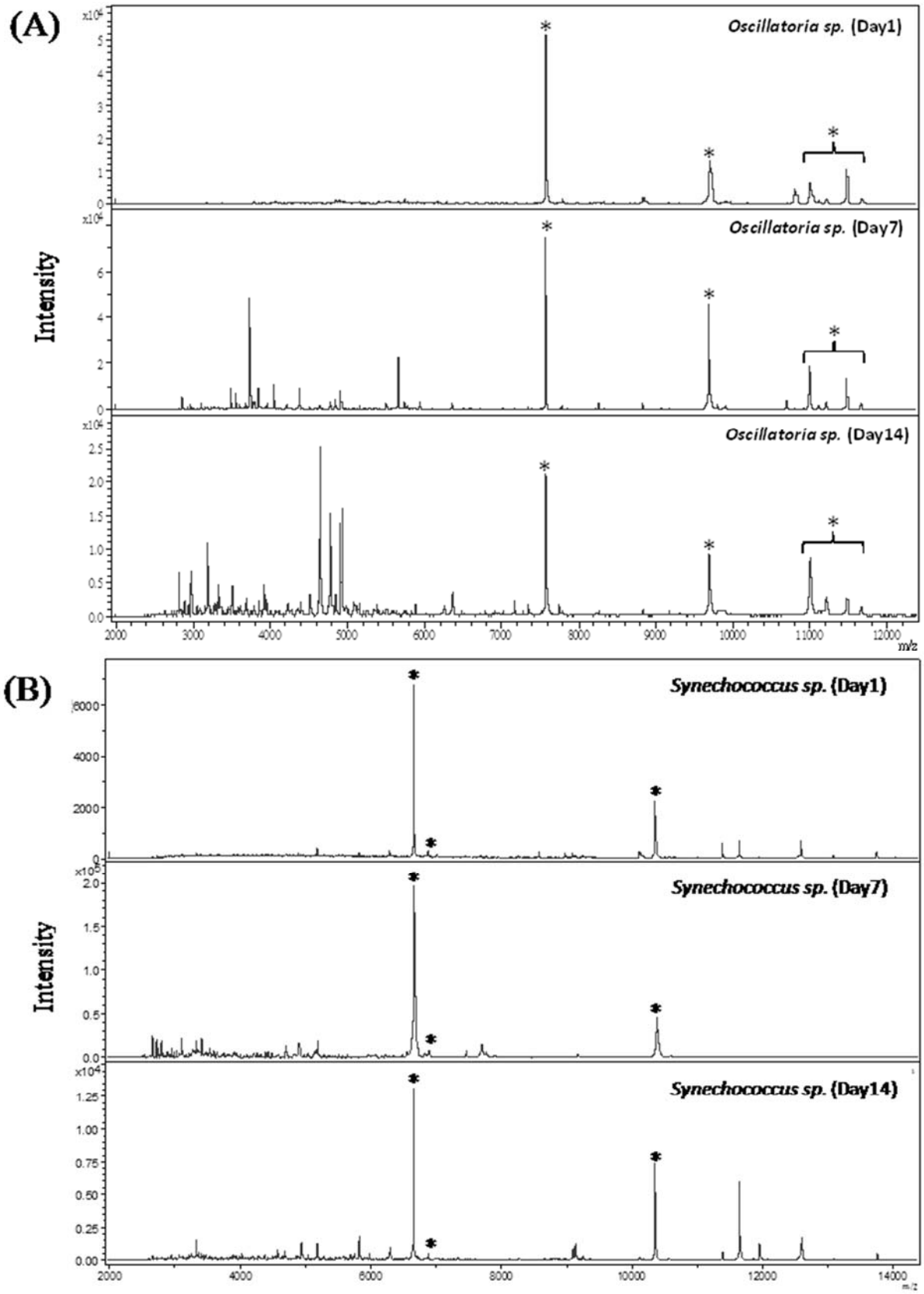


Figure 5. MALDI-TOF-MS protein expression profiles of cyanobacterial species (A) *Oscillatoria sp* (CCMP1731) and (B) *Synechococcus sp* (FACHB-7820),in different growth ages. Peaks labeled with asterisks are constantly expressed protein mass observed in the spectra irrespective of the growth ages.

the spectra irrespective of the growth ages of the cells (Figure 5, marked with asterisks). These sets of peak mass ions can be used as the "representative peaks" for the tentative identification of a specific species/strain. Whether these peaks can be used as the biomarkers should be subjected to more comprehensive studies of PEPs of other species in future studies. Further, Tao and co-authors (2004) has developed an interesting and efficient numerical method for the differentiation of microorganism irrespective of their growth stages [36]. This is a statistics-based algorithm to assign weight factors to individual mass ions observed in the mass spectra of different growth stages and the combined weight factors were used for species identification with satisfactory accuracy. Therefore, these distinguishing species-specific ions together with the established statistical algorithm will form an unambiguous reliable basis for cyanobacteria identification. Furthermore, with the aid of a commercially available software, such as "BioTyper" (from Bruker), automated species identification based on statistical analysis can be achieved.

Previously, we had demonstrated that it is possible to identify individual species in a mixed population of dinoflagellates based on species-specific signature ions in the spectrum [25]. In most of the cases of algal blooms, usually one to two algal species would predominate, usually with one succeed by another. Therefore, in this study, we evaluated whether it is also possible to differentiate different species/strains of cyanobacteria from a mixed culture comprising two dominatant species in 1:1 ratio (Figure 6). Two different species/strains, *Oscillatoria sp* and *Cylindrospermopsis raciborskii* (species of different genus) (Figure 6A); *Synechococcus sp* FACHB-460 and FACHB-7820 (strains of same species) (Figure 6B) were grown in a mixed culture (in ratio: ~1:1). The mixed cultures were analyzed by MALDI-TOF-MS with the same condition described previously. Our results showed that the species-specific (Figure 6A) or the strain-specific (Figure 6B) signature peak mass ions can be easily identified within the mass spectra of the mixed cultures even by simple visual inspection. The results are compatible to our previous data [25]. In Figure 6A, a set of signature peaks corresponding to *Oscillatoria sp* ("O" peaks) and *Cylindrospermopsis raciborskii* ("C" peaks) can be easily observed from the spectrum of the mixed culture. Although the signature peaks arise from the experiments cannot be proved to be species specific at this point in time, we would like to demonstrate the idea that strains from a mixed population comprising two domintant species can be identified based on the strain-specific peak mass ions. For instance, the two strains of *Synechococcus sp*, FACHB-460 and FACHB-7820, which are difficult to be distinguished from each other based on their morphology and even their 16S-23S rRNA ITS sequences, can be distinguished from each other based on their PEPs.

4. Concluding Remarks

Besides applicable to dinoflagellates [25], results of our present study extended our horizon that PEPs obtained from MALDI-TOF-MS are also applicable for identification of cyanobacteria species. To our knowledge, this is the first report to show the successful application of PEPs obtained by MALDI-TOF-MS for the identification of cyanobacteria. We believe that the methodology currently being developed can also be widely applied in other HAB causative agents, such as diatoms. Operation of this methodology is easy and would allow identification of various HAB species within hours. In the near future, further expansion of this MALDI-TOF-MS based methodology in generating more reference PEPs of various types of HAB causative agents and standardize the specific conditions employed will be very useful for rapid HAB identification. Given the advance development of the MALDI-TOF-MS for application in the field of proteomics and with the aid of bioinformatics examination such as the MALDI BioTyper software (Bruker, Germany), we believe this MALDI-TOF-MS approach can potentially become part of a standard platform for HAB species identification.

5. Supplementary material

Supplementary data and information is available at: http://www.jiomics.com/index.php/jio/rt/suppFiles/138/

Figure S1 - The 16S-23S internal transcribed spacer (ITS) sequences of 4 strains of Synechococcus sp. Nucleotides highlighted with dotted lines represent the differences among the 4 strains.

Table S1- Peak mass ions for the corresponding MS spectra from cyanobacterial species and strains used in this study.

Acknowledgements

This research is partly supported by a grant from the NSCF -RGC jointed research project scheme (A/C: N_PolyU 502/08) and a grant from ECF (A/C: ECF 02/2007).

References

[1] S.W. Chisholm, R.J. Olson, E.R. Zettler, R. Goericke, J.B. Waterbury, N.A. Welschmeyer, Nature, 334 (1988) 340-343. doi: 10.1038/334340a0

[2] J.B. Waterbury, S.W. Watson, R.R.L. Guillard, L.E. Brand, Nature, 277 (1979) 293-294. doi: 10.1038/277293a0

[3] W. Carmichael, Adv Exp Med Biol, 619 (2008) 1-912. doi: 10.1007/978-0-387-75865-7_4

[4] W.W. Carmichael, J Appl Bacteriol, 72 (1992) 445-459. doi: 10.1111/j.1365-2672.1992.tb01858.x

[5] W.W. Carmichael, Human and Ecological Risk Assessment, 7 (2001) 1393-1407. doi: 10.1080/20018091095087

[6] S. Pitois, M.H. Jackson, B.J. Wood, J Environ Health, 64 (2001) 25-32.

[7] B. Qin, G. Zhu, G. Gao, Y. Zhang, W. Li, H.W. Paerl, W.W. Carmichael, Environ Manage, 45 (2010) 105-112. doi: 10.1007/s00267-009-9393-6

[8] T.N. Duy, P.K. Lam, G.R. Shaw, D.W. Connell, Rev Environ

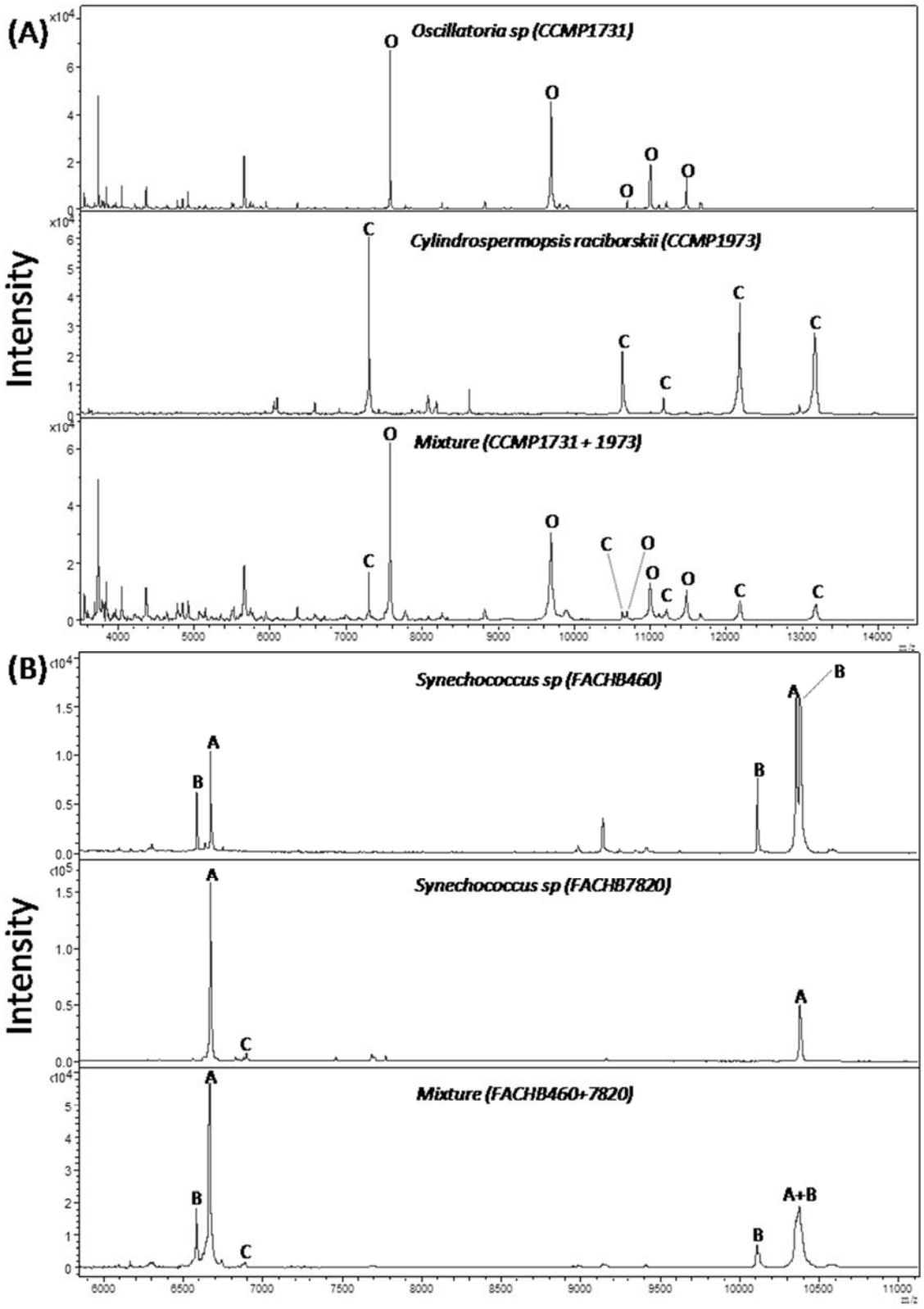


Figure 6. MALDI-TOF-MS protein expression profiles of mixed populations of cyanobacterial species. (A) PEP profiles of *Oscillatoria sp*, *Cylindrospermopsis raciborskii* and mixture of both species. Peaks labeled with "O" and "C" representing the signature peaks corresponding to *Oscillatoria* sp and *Cylindrospermopsis raciborskii* respectively. (B) PEP profiles of *Synechococcus* strains FACHB-460, FACHB-7820 and mixture of both strains. Peaks labeled with "A", "B" and "C" representing the common peaks shared by both strains, the signature peaks corresponding to FACHB-460 and the signature peak corresponding to FACHB-7820 respectively.

Contam Toxicol, 163 (2000) 113-185. doi: 10.1007/978-1-4757-6429-1_3

[9] J. Komarek, K. Anagnostidis, Archiv Fur Hydrobiologie, 82 (1989) 247-345.

[10] U. Nubel, F. Garcia-Pichel, G. Muyzer, Appl Environ Microbiol, 63 (1997) 3327-3332.

[11] E. Valerio, L. Chambel, S. Paulino, N. Faria, P. Pereira, R. Tenreiro, Microbiology, 155 (2009) 642-656. doi: 10.1099/mic.0.022848-0

[12] J.P. Dworzanski, A.P. Snyder, Expert Rev Proteomics, 2 (2005) 863-878. doi: 10.1586/14789450.2.6.863

[13] C. Fenselau, P.A. Demirev, Mass Spectrom Rev, 20 (2001) 157-171. doi: 10.1002/mas.10004

[14] R. Dieckmann, R. Helmuth, M. Erhard, B. Malorny, Appl Environ Microbiol, 74 (2008) 7767-7778. doi: 10.1128/AEM.01402-08

[15] A. Mellmann, J. Cloud, T. Maier, U. Keckevoet, I. Ramminger, P. Iwen, J. Dunn, G. Hall, D. Wilson, P. Lasala, M. Kostrzewa, D. Harmsen, J Clin Microbiol, 46 (2008) 1946-1954. doi: 10.1128/JCM.00157-08

[16] M.J. Donohue, A.W. Smallwood, S. Pfaller, M. Rodgers, J.A. Shoemaker, J Microbiol Methods, 65 (2006) 380-389. doi: 10.1016/j.mimet.2005.08.005

[17] T. Krishnamurthy, P.L. Ross, Rapid Commun Mass Spectrom, 10 (1996) 1992-1996. doi: 10.1002/(SICI)1097-0231(199612)10:15<1992::AID-RCM789>3.0.CO;2-V

[18] T. Krishnamurthy, P.L. Ross, U. Rajamani, Rapid Commun Mass Spectrom, 10 (1996) 883-888. doi: 10.1002/(SICI)1097-0231(19960610)10:8<883::AID-RCM594>3.0.CO;2-V

[19] V. Ryzhov, Y. Hathout, C. Fenselau, Appl Environ Microbiol, 66 (2000) 3828-3834. doi: 10.1128/AEM.66.9.3828-3834.2000

[20] M.L. Magnuson, J.H. Owens, C.A. Kelty, Appl Environ Microbiol, 66 (2000) 4720-4724. doi: 10.1128/AEM.66.11.4720-4724.2000

[21] B. Amiri-Eliasi, C. Fenselau, Anal Chem, 73 (2001) 5228-5231. doi: 10.1021/ac010651t

[22] T.Y. Li, B.H. Liu, Y.C. Chen, Rapid Commun Mass Spectrom, 14 (2000) 2393-2400. doi: 10.1002/1097-0231(20001230)14:24<2393::AID-RCM178>3.0.CO;2-9

[23] S.M. Mandal, B.R. Pati, A.K. Ghosh, A.K. Das, Eur J Mass Spectrom (Chichester, Eng), 13 (2007) 165-171. doi: 10.1255/ejms.842

[24] K.J. Welham, M.A. Domin, K. Johnson, L. Jones, D.S. Ashton, Rapid Commun Mass Spectrom, 14 (2000) 307-310. doi: 10.1002/(SICI)1097-0231(20000315)14:5<307::AID-RCM823>3.0.CO;2-3

[25] F.W.F. Lee, K.C. Ho, S.C.L. Lo, Harmful Algae, 7 (2008) 551-559. doi: DOI 10.1016/j.hal.2007.12.001

[26] M. Erhard, H. von Dohren, P. Jungblut, Nature Biotechnology, 15 (1997) 906-909. doi: 10.1038/nbt0997-906

[27] I. Janse, M. Meima, W.E. Kardinaal, G. Zwart, Appl Environ Microbiol, 69 (2003) 6634-6643. doi: 10.1128/AEM.69.11.6634-6643.2003

[28] V. Ryzhov, C. Fenselau, Anal Chem, 73 (2001) 746-750. doi: 10.1021/ac0008791

[29] V. Ruelle, B. El Moualij, W. Zorzi, P. Ledent, E.D. Pauw, Rapid Commun Mass Spectrom, 18 (2004) 2013-2019. doi: 10.1002/rcm.1584

[30] Z. Wang, L. Russon, L. Li, D.C. Roser, S.R. Long, Rapid Commun Mass Spectrom, 12 (1998) 456-464. doi: 10.1002/(SICI)1097-0231(19980430)12:8<456::AID-RCM177>3.0.CO;2-U

[31] B.R. Robertson, N. Tezuka, M.M. Watanabe, Int J Syst Evol Microbiol, 51 (2001) 861-871. doi: 10.1099/00207713-51-3-861

[32] J. Premanandh, B. Priya, I. Teneva, B. Dzhambazov, D. Prabaharan, L. Uma, J Microbiol, 44 (2006) 607-616.

[33] R.J. Arnold, J.A. Karty, A.D. Ellington, J.P. Reilly, Anal Chem, 71 (1999) 1990-1996. doi: 10.1021/ac981196c

[34] N. Valentine, S. Wunschel, D. Wunschel, C. Petersen, K. Wahl, Appl Environ Microbiol, 71 (2005) 58-64. doi: 10.1128/AEM.71.1.58-64.2005

[35] F.W.F. Lee, K.C. Ho, Y.L. Mak, S.C.L. Lo, Harmful Algae, 12 (2011) 1-10. doi: 10.1016/j.hal.2011.08.007

[36] L. Tao, X. Yu, A.P. Snyder, L. Li, Anal Chem, 76 (2004) 6609-6617. doi: 10.1021/ac049391g