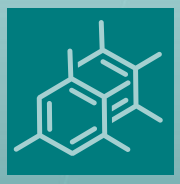


molecules

Mass Spectrometric Proteomics

Edited by
Paolo Iadarola

Printed Edition of the Special Issue Published in *Molecules*

www.mdpi.com/journal/molecules

MDPI

Mass Spectrometric Proteomics

Mass Spectrometric Proteomics

Special Issue Editor

Paolo Iadarola

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade

Special Issue Editor
Paolo Iadarola
Universita degli Studi di Pavia
Italy

Editorial Office
MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Molecules* (ISSN 1420-3049) from 2018 to 2019 (available at: https://www.mdpi.com/journal/molecules/special_issues/mass_spectrometric_proteomics)

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* **Year**, *Article Number*, Page Range.

ISBN 978-3-03897-826-8 (Pbk)
ISBN 978-3-03897-827-5 (PDF)

Contents

About the Special Issue Editor

Paolo Iadarola is Associate Professor of Methods in Biochemistry, at the Department of Biology and Biotechnologies of the University of Pavia. He has been working since 1977 on purification and structural analysis of enzymes from different sources. In particular, he has been involved in the study of proteolytic enzymes involved in the degradation process of connective tissue matrix. In this context he developed new methodologies whose application allowed to determine both the amount of biochemical markers (Desmosine and Isodesmosine) in urine of patients affected by Chronic Obstructive Pulmonary Disease (COPD) and the activity of proteolytic enzymes in sputa of these patients. The relationship between the extent of markers excreted and severity of the disease has also been drawn. He is presently involved in the development of 2-DE and LC-MS techniques for the study of proteomics of biological fluids. Aim of this research is the identification of potential biomarkers of severe diseases including COPD and Nasu-Hakola disease. He is also involved in the proteomic analysis of salivary glands and ovaries from the tick Ixodes ricinus. He is author of about 150 publications (H-Index: 21) on peer reviewed journals.

Preface to "Mass Spectrometric Proteomics"

The growing in complexity of biological samples observed over the last 20 years resulted in the emersion of sophisticated liquid chromatographic-mass spectrometric techniques that widened the application of proteomics to a variety of novel samples. These applications span from the characterization of complex protein networks and the detection of their perturbations to large-scale protein analysis and the identification of proteins which may clarify molecular mechanisms that govern cellular processes. Despite the undoubted potential of proteomics, collection of data has often proven difficult due to obstacles originated from the inherent complexity of protein profiles under investigation. Considerable technological advancements have led to the birth of innovative approaches helpful to surmount these hurdles. Attractive strategies toward the increase of the understanding of human diseases and other specific applications are discussed in this Special Issue of Molecules.

Paolo Iadarola
Special Issue Editor

molecules

MDPI

Editorial

Special Issue: Mass Spectrometric Proteomics

Paolo Iadarola

Department of Biology and Biotechnologies "L.Spallanzani", University of Pavia, 27100 Pavia, Italy; piadarol@unipv.it

Academic Editor: Derek J. McPhee
Received: 19 March 2019; Accepted: 20 March 2019; Published: 21 March 2019

The term "Proteomics" refers to the characterization of the proteome, that is, all proteins present in a biological system. Because the protein content of an organism changes in response to many conditions, the proteome dynamism helps investigation on differences in protein states between different conditions. Identifying all proteins present in biological systems is a difficult task and a crucial role in enabling the analysis of proteomes is played by mass spectrometry (MS). Given the complexity of biological systems, high performance liquid chromatography-based (HPLC) separations are currently applied to separate the mixture of proteins before analysis. After these extensive procedures, confident identifications may be obtained by MS and tandem MS (MS/MS) either on intact proteins (top-down) or on peptides obtained after enzymatic digestion (bottom-up). Changes in protein abundances between different biological states may be obtained by several approaches, with or without stable isotopic labeling.

In this Special Issue of *Molecules* the reader will find a topical selection of both research and review articles which bring together different sophisticated methodological strategies with applications in protein profiling of organs, in the study of post translational tissue modifications or in the investigation of the molecular mechanisms behind human disorders.

Mauri and co-workers [1] open the section of seven research articles with a study that describes the application of multidimensional protein identification technology (MudPIT) to discover the urinary proteomes of head and neck squamous cell carcinoma (HNSCC) and thyroid cancer patients. They analyzed urine samples from these subjects to identify protein profiles prior and after infusion with ^{10}Boron and neutron capture therapy (BNCT) during surgery. The results of this study allowed the identification of several inflammation- and cancer-related proteins, which could be potential tumor biomarkers. In particular, a reduction of three proteins (Galectin-3 binding protein, CD44, and osteopontin) that were involved in inflammation was observed after treatment. Analysis of the urinary proteome during and after therapeutic interventions made it clear that this fluid reflects changes that are induced by several diseases, including different types of cancer.

Another interesting study from the same research group aimed at exploring the potentiality of shotgun proteomic platforms in the characterization of the status and the evolution of metabolic pathways during ex vivo lung perfusion (EVLP), an emerging procedure that allows organ preservation in lung transplantation. The application of a nanoLC–MS/MS system to the proteome analysis of lung tissues from three experimental groups (native, pre- and post-EVLP) of an optimized rat model allowed identification of potential EVLP-related biomarkers. Given the promising findings provided by this work, future perspectives will attempt to confirm these data with the aim of translating the experimental research to human specimens [2].

The huge potential of sophisticated technologies in the proteomic area is shown by two original articles.

The report by Delphine and coworkers [3] demonstrates the utility of quadrupole time-of-flight (Q–ToF) hybrid mass spectrometer equipped with electrospray ionisation source (ESI). Their study aimed at optimising tandem MS analysis by testing the effect of some of the parameters concerning the

in-source collision-induced dissociation (IS-CID) procedure in combination or not with conventional CID. A total of 11 MS/MS methods were assessed on samples bearing increased complexity, including individual milk protein standards, mixed protein standards and cow's raw milk samples from different breeds. Results of top-down sequencing from the nine most abundant proteoforms of caseins, alpha-lactalbumin and beta-lactoglubulins were compared. Nine MS/MS methods achieved more than 70% sequence coverage overall to distinguish between allelic proteoforms varying only by one or two amino acids.

Zhang and colleagues [4] used ultra-performance liquid chromatography/Orbitrap Fusion mass-spectrometry (UPLC/MS/MS) and chemometrics to study *Ophiocordyceps sinensis*, an important fungal drug used in Chinese medicine. This approach allowed identification of a number of marker tryptic peptides in the wild *O. sinensis* fruiting body and in various commercially available mycelium fermented powders. This represented the first extensive study on the authentication of *O. sinensis* and cultured *Ophiocordyceps* mycelia.

The isobaric tags for relative and absolute quantitation (iTRAQ) technique was used by Su and coworkers [5] to investigate the oviduct of *Rana chensinensis* (another important traditional Chinese medicine resource widely used in the treatment of asthenia after sickness or delivery, deficiency in vigor, palpitation and insomnia), which, in contrast to the breeding period, significantly expands during prehibernation. To explain this phenomenon at the molecular level, the protein expression profiles of *Rana chensinensis* oviduct during these two conditions were analyzed. Among the 4479 proteins identified, 312 were up- or down-regulated between these two periods. The application of gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses allowed them to understand that this distinctive physiological phenomenon of *Rana chensinensis* oviduct was mainly involved in extracellular matrix–receptor interaction, metabolic pathways, and focal adhesion.

The interest of Li and co-authors [6] was focused on the study of protein lysine acetylation (PKA), a key post-translational modification involved in the regulation of various biological processes in rice. They applied an MS-based label-free approach to perform a quantitative analysis of PKA proteins in rice callus, root, leaf and panicle. The identification of 890 PKA proteins, which covered 1536 sites, allowed construction of a tissue atlas of rice acetylome and provided an overall view of the acetylation events in rice tissues.

The targeted proteome analysis performed by Tokumaru and co-workers [7] aimed at investigating the SigE dependent-regulation of central metabolism in *Synechocystis sp.* PCC 6803, a model cyanobacteria. In their study, they compared the protein abundance profiles among the wild type, a SigE deletion mutant (DSigE), and a SigE over-expression (SigEox) strain. The protein profile showed that SigE plays a pivotal role as a positive regulator of oxidative pentose phosphate pathway (OxPPP) activity and NADPH reproduction. The results also suggested that SigE over-expression increases GdhA abundance, which is involved in the nitrose assimilation pathway using NADPH and downregulates the proteins involved in photosynthesis.

The Special Issue is concluded by a compilation of review articles on different topics.

Given the relevance of isobaric labeling reagents for quantitative isobaric derivatization strategies in proteomics, Bachor and colleagues [8] review the trends in the design of new isobaric markers. The development of selective methods for introducing into a peptide a tag that increases the multiplexicity of markers and the sensitivity of measurement and lowers the cost of synthesis is currently a topic of great interest in proteomics. Different types of isobaric reagents are used in quantitative proteomics, and their chemistry and advantages offered by their application are clearly presented in this article.

In recent years, platinum-based anticancer drugs (especially cisplatin) have played an important role in the clinical chemotherapy of cancer. Due to the adverse effects and acquired resistance of these drugs, efforts have been made to exploit novel anticancer metallodrugs and unveil the molecular mechanism of anticancer activity and drug resistance. The article by Jia and co-workers [9] is focused

on the identification, by MS-based quantitative strategies, of proteins which specifically respond or bind to metal-based anticancer drugs and on the elucidation of their mechanisms of action.

Finally, Zdenek Perutka and Marek Šebela [10] review the cleavage properties of bovine pseudotrypsin, a trypsin proteoform, in the digestion of protein samples with a different complexity. While cleaving peptide bonds predominantly at the conventional trypsin cleavage sites, pseudotrypsin shows non-specific cleavages (mostly after the aromatic residues of Tyr and Phe) which are not expected to occur for the major trypsin forms. The valuable information provided by this peculiar activity could be of great utility in common proteomics protocols.

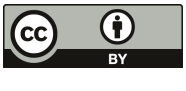
I hope that the content of this Special Issue meets the expectations of readers and would like to thank all authors for their excellent contributions.

Acknowledgments: The author would like to thank Maddalena Cagnone in Biochemistry at the Department of Molecular Medicine, University of Pavia, Italy, for fruitful and stimulating discussions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ferrari, E.; Wittig, A.; Basilico, F.; Rossi, R.; De Palma, A.; Di Silvestre, D.; Sauerwein, W.A.; Mauri, P.L. Urinary Proteomics Profiles Are Useful for Detection of Cancer Biomarkers and Changes Induced by Therapeutic Procedures. *Molecules* **2019**, *24*, 794. [CrossRef] [PubMed]
2. Roffia, V.; De Palma, A.; Lonati, C.; Di Silvestre, D.; Rossi, R.; Mantero, M.; Gatti, S.; Dondossola, D.; Valenza, F.; Mauri, P.; Blasi, F. Proteome Investigation of Rat Lungs Subjected to Ex Vivo Perfusion (EVLP). *Molecules* **2018**, *23*, 3061. [CrossRef] [PubMed]
3. Vincent, D.; Mertens, D.; Rochfort, S. Optimisation of Milk Protein Top-Down Sequencing Using In-Source Collision-Induced Dissociation in the Maxis Quadrupole Time-of-Flight Mass Spectrometer. *Molecules* **2018**, *23*, 2777. [CrossRef] [PubMed]
4. Zhang, P.; Li, S.; Li, J.; Wei, F.; Cheng, X.; Zhang, G.; Ma, S.; Liu, B. Identification of *Ophiocordyceps sinensis* and Its Artificially Cultured *Ophiocordyceps* Mycelia by Ultra-Performance Liquid Chromatography/Orbitrap Fusion Mass Spectrometry and Chemometrics. *Molecules* **2018**, *23*, 1013. [CrossRef] [PubMed]
5. Su, H.; Zhang, H.; Wei, X.; Pan, D.; Jing, L.; Zhao, D.; Zhao, Y.; Qi, B. Comparative Proteomic Analysis of *Rana chensinensis* Oviduct. *Molecules* **2018**, *23*, 1384. [CrossRef] [PubMed]
6. Li, Z.; Wang, Y.; Bello, B.K.; Ajadi, A.A.; Tong, X.; Chang, Y.; Zhang, J. Construction of a Quantitative Acetylomic Tissue Atlas in Rice (*Oryza sativa* L.). *Molecules* **2018**, *23*, 2843. [CrossRef] [PubMed]
7. Tokumaru, Y.; Uebayashi, K.; Toyoshima, M.; Osanai, T.; Matsuda, F.; Shimizu, H. Comparative Targeted Proteomics of the Central Metabolism and Photosystems in SigE Mutant Strains of *Synechocystis* sp. PCC 6803. *Molecules* **2018**, *23*, 1051. [CrossRef] [PubMed]
8. Bąchor, R.; Waliczek, M.; Stefanowicz, P.; Szewczuk, Z. Trends in the Design of New Isobaric Labeling Reagents for Quantitative Proteomics. *Molecules* **2019**, *24*, 701. [CrossRef] [PubMed]
9. Jia, S.; Wang, R.; Wu, K.; Jiang, H.; Du, Z. Elucidation of the Mechanism of Action for Metal Based Anticancer Drugs by Mass Spectrometry-Based Quantitative Proteomics. *Molecules* **2019**, *24*, 581. [CrossRef] [PubMed]
10. Perutka, Z.; Šebela, M. Pseudotrypsin: A Little-Known Trypsin Proteoform. *Molecules* **2018**, *23*, 2637. [CrossRef] [PubMed]



molecules

MDPI

Article

Urinary Proteomics Profiles Are Useful for Detection of Cancer Biomarkers and Changes Induced by Therapeutic Procedures

Emanuele Ferrari [1], Andrea Wittig [2], Fabrizio Basilico [1,†], Rossana Rossi [1], Antonella De Palma [1], Dario Di Silvestre [1], Wolfgang A.G. Sauerwein [3] and Pier Luigi Mauri [1,4,*]

1 Proteomics and Metabolomics Unit, Institute for Biomedical Technologies (ITB-CNR), 20090 Segrate (MI), Italy; emanuele.ferrari@itb.cnr.it (E.F.); rossana.rossi@itb.cnr.it (R.R.); antonella.depalma@itb.cnr.it (A.D.P.); dario.disilvestre@itb.cnr.it (D.D.S.)
2 Dept. of Radiotherapy and Radiation Oncology, University Hospital Jena, 07743 Jena, Germany; andrea.wittig@med.uni-jena.de
3 NCTeam, Strahlenklinik, Universitätsklinikum Essen, 45122 Essen, Germany; wolfgang.sauerwein@uni-due.de
4 Istituto di Scienze della Vita, Scuola Superiore Sant'Anna, 56127 Pisa, Italy
* Correspondence: pierluigi.mauri@itb.cnr.it; Tel.: +39-02-264226728
† The author was deceased.

Academic Editor: Paolo Iadarola
Received: 18 January 2019; Accepted: 19 February 2019; Published: 22 February 2019

Abstract: Boron neutron capture therapy (BNCT) is a binary cancer treatment modality where two different agents (^{10}B and thermal neutrons) have to be present to produce an effect. A dedicated trial design is necessary for early clinical trials. The concentration of ^{10}B in tissues is an accepted surrogate to predict BNCT effects on tissues. Tissue, blood, and urines were sampled after infusion of two different boron carriers, namely BSH and BPA in the frame of the European Organisation for Research and Treatment of Cancer (EORTC) trial 11001. In this study, urine samples were used to identify protein profiles prior and after drug infusion during surgery. Here, an approach that is based on the mass spectrometry (MS)-based proteomic analysis of urine samples from head and neck squamous cell carcinoma (HNSCC) and thyroid cancer patients is presented. This method allowed the identification of several inflammation- and cancer-related proteins, which could serve as tumor biomarkers. In addition, changes in the urinary proteome during and after therapeutic interventions were detected. In particular, a reduction of three proteins that were involved in inflammation has been observed: Galectin-3 Binding Protein, CD44, and osteopontin. The present work represents a proof of principle to follow proteasome changes during complex treatments based on urine samples.

Keywords: urine; thyroid cancer; squamous cell cancer of head and neck; BNCT; boron; proteomics; LC-MS; MudPIT

1. Introduction

Boron neutron capture therapy (BNCT) is based on the high cross section of the non-radioactive isotope boron-10 for capturing thermal neutrons, leading to the nuclear reaction $^{10}B(n,\alpha)^{7}Li$ [1]. The resulting high linear energy transfer (LET) particles have a very short range in tissues, limiting the damages to cells containing ^{10}B [2]. For a successful tumor treatment, ^{10}B has to be selectively delivered to tumor cells through specific boron-containing compounds. In the frame of the research project "Therapeutic strategies for Boron Neutron Capture Therapy (BNCT): Boron imaging" (financed by the European Commission QLK3-CT-1999-01067), several early clinical trials under the auspices of the European Organisation for Research and Treatment of Cancer (EORTC) were performed [3].

The investigation that is presented here is based on urine samples that were collected in the EORTC trial 11001 and urine samples from heavily smoking (>25 cigarettes/day) volunteers.

BNCT has already proven to be a promising tool in the treatment of a number of cancer, including head and neck squamous cell carcinoma (HNSCC) [4–7]. BNCT has been explored as an alternative therapy to the currently employed surgery, chemotherapy, or radiotherapy, bearing the advantages of less application (one or two doses) and the ability to maintain intact the oro-facial structures and functions [8]. Although no clinical application has been performed, BNCT has been discussed for the treatment of thyroid cancer in patients not responding to standard therapies [9,10].

To date, a very limited number of papers have investigated the molecular effects of boron compounds and BNCT. Molecular studies were mainly based on the monitoring of Boron-containing compounds in cell lines [11,12], tissue [13], plasma [14,15], and urine [16,17]. As far as we know, no extensive genomic or transcriptomic studies have been performed to evaluate the changes in the transcription that is induced by this treatment. A preliminary study focused on the DNA damage induced by irradiation while using a rat tumor graft model [18]; this work followed the levels of some proteins by means of western blotting and showed an upregulation of High mobility group box 1 (HMGB1), a nuclear protein involved in necrosis and inflammation processes, which was proposed as an early diagnostic marker. Surely, for the study of the effects of BNCT, proteomics is a promising tool, which has already been applied in an in vitro study on the effects of mercaptoundecahydrododecaborate (BSH) on phospholipid hydroperoxide glutathione peroxidase [19]; in this case, a gel-based approach was used to separate the native protein from the protein that was covalently bound to BSH, which were then characterized by means of liquid chromatography coupled to mass spectrometry (LC-MS). A more comprehensive application of the gel-based proteomics coupled to MALDI-TOF identification was reported in a study where a human oral squamous carcinoma cell line was treated with boronophenylalanine (BPA) and then irradiated with thermal neutrons [20]. Changes in the levels of 29 proteins were observed and they were mainly related to vesicle regulation, mRNA processing, and transcription.

In the last years, proteomics technologies considerably improved and a gel-free approach, which was mainly based on LC-MS (so called shotgun or MS-based proteomics), became the gold standard methodology to investigate the proteome profiles, increasing both the number of identified proteins and their quantitative analysis. In the present work, we took advantage of the availability of both urine samples, which were previously used for pharmacokinetic investigations, and of a gel- and label-free proteomics facility. We analyzed urine samples by means of multidimensional protein identification technology (MudPIT) to describe the urinary proteomes of HNSCC and thyroid cancer patients. Moreover, the effects on urinary proteome due to boron infusion (as BPA or BSH) in cancer patients were also investigated.

2. Results

2.1. Proteome Profiles of Urine

The multidimensional separation of peptides and the MS-based proteomic approach that was used in this work allowed for the characterization of proteome profiles of urine from healthy subjects and tumor patients. Thus, it was possible to compare healthy subjects and tumor patients with a simple and convenient shotgun protocol, which was based on liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). Good repeatability between technical replicates was observed (Figure S1). The complete list of proteins for each sample is reported in the supplementary materials (Table S1). When considering all of samples, 365 proteins were identified and 62.2% of them have more than one peptide. For the healthy group, 172 proteins were detected, while for HNSCC and thyroid cancer patients, 253 and 228 proteins were found, respectively. The three groups share 110 proteins, while the two groups of tumor patients share almost 60% of their proteins (Figure S2). An analysis of the Uniprot Keywords showed that 219 out of 365 proteins have a signal peptide sequence, indicating that

they are destined to the endoplasmic reticulum and possibly secreted. When considering the groups separately, the analysis showed that about two-thirds of each group's protein have a signal peptide. The Uniprot keywords also showed that 84 out of 365 proteins have a transmembrane domain, with no significant differences between healthy subjects and tumor patients.

The proteins identified for each group of patients were plotted on a two-dimensions map with MAProMa software, using theoretical molecular weight (MW) and isoelectrical point (pI) (Figures S3–S5). Figure 1 reports the two-dimensional map that was obtained by plotting all identified proteins. These maps show the ability of the shotgun approach to identify proteins in a wide range of MW and pI.

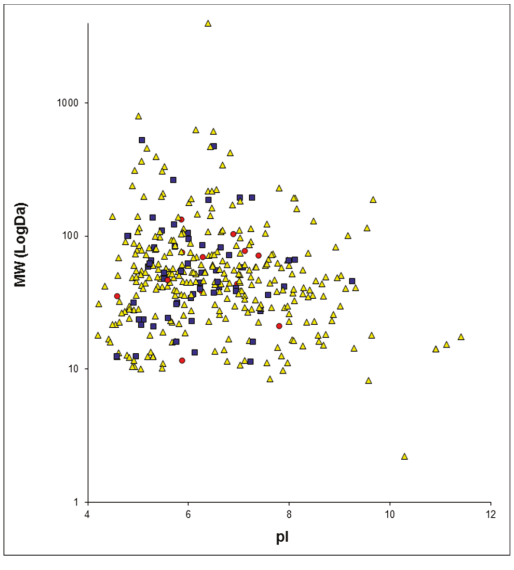


Figure 1. Two-dimensions map of the identified proteins from all analyzed urinary samples plotted with MAProMa software, according to their theoretical pI and MW (in Log scale). Proteins identified by 1 Peptide Spectrum Match (PSM) are reported as yellow triangles, those with PSMs between 2 and 4 as blue squares, and those with a number of PSMs over or equal to 5 as red circles.

An analysis of the enriched Gene Ontology (GO) biological process terms that was performed with DAVID confirms a qualitative similarity in the protein composition of urine between the three groups of subjects (Figure 2). In particular, the terms enriched in both tumor groups were the same, with little difference in the number of proteins for each term. It has to be noted that three of these terms were not found to be enriched in the urine of healthy subjects: phagocytosis, engulfment (GO:0006911), cell adhesion (GO:0007155), and Fc-gamma receptor signaling pathway that is involved in phagocytosis (GO:0038096).

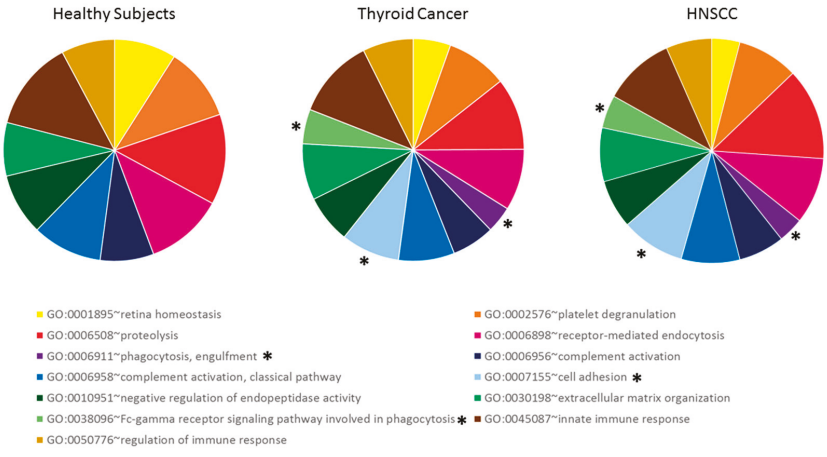


Figure 2. Enrichment in Gene Ontology (GO) biological process terms, using DAVID for each considered group. * denotes GO Terms enriched only in tumor patients' proteome but not in healthy subjects.

2.2. *Differentially Secreted Proteins in Urine from Tumor Patients*

A quantitative approach was used, since there were no evident qualitative differences in the proteic composition of urine for the three groups. In this study, a label-free quantification has been applied, considering the total number of peptides (Peptide Spectrum Matches—PSMs, also called Spectral Count—SpC) for each protein. Indeed, this quantitative analysis led to the identification of significant alterations in the amount of single proteins that were secreted in the urine, when comparing healthy subjects with tumor patients, using DAVE and DCI algorithms [21]. Proteins with highly significant differences (i.e., $|DAVE| > 0.2$ and $|DCI| > 10$) are reported in Figure 3 and in Table S2, together with fold changes and G test values.

By comparing the healthy subjects with those that were affected by HNSCC, 14 proteins were found to be more abundant in urine from tumor patients, as indicated by the positive values of DAVE and DCI (Figure 3), while the levels of six proteins were higher in control subjects. Analogously, the analysis of urine from thyroid cancer patients led to the identification of 17 differentially secreted proteins in comparison to healthy volunteers. Among these proteins, 13 were overrepresented in affected subjects. It is interesting to note that the two types of tumors taken in account share similar highly secreted proteins, and some of them show significant differences in both cases. Nine proteins were higher in both tumors as compared to healthy patients, while three were lower: keratin, type II cytoskeletal 1 (KRT1), serotransferrin (TF), and alpha-1-antitrypsin (SERPINA1). The urinary proteins of the two tumor-affected groups were compared with the same method in order to identify the differences between them. Indeed, 21 proteins were differentially secreted, eight proteins were overrepresented in HNSCC patients, and 13 were overrepresented in thyroid cancer patients (Table S3).

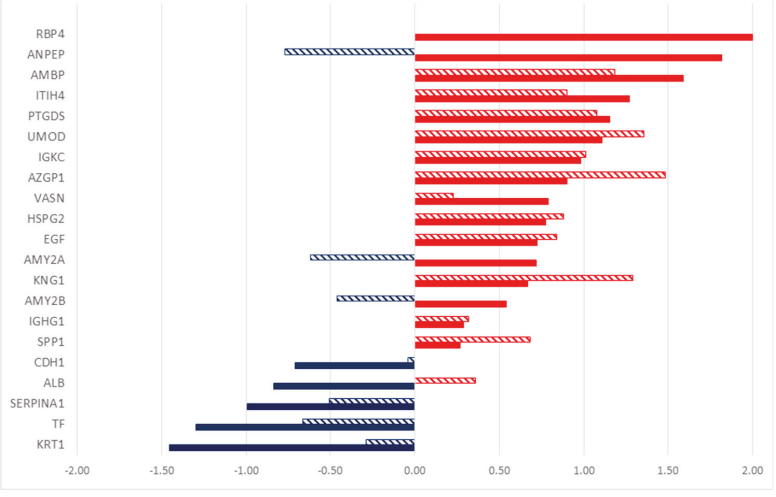


Figure 3. DAVE values of the significantly different levels of proteins found in urinary proteome of head and neck squamous cell carcinoma (HNSCC) patients (full bars) and thyroid cancer patients (dashed bars) in comparison to the healthy subjects. Red/positive and blue/negative bars correspond respectively to up- and down-secreted proteins in tumor patients compared to healthy subjects.

In addition, the enriched GO biological process terms were searched for the proteins that were altered by the presence of tumor (Figure 4), and they showed similarity in the enriched terms between the two tumor entities. Only two terms were found to be enriched in HNSCC and not in thyroid cancer differentially secreted proteins: ERK1 and ERK2 cascade (GO:0070371) and B cell receptor signaling pathway (GO:0050853). Three terms were enriched exclusively in differentially secreted proteins of thyroid cancer patients: angiogenesis (GO:0001525), extracellular matrix disassembly (GO:0022617), and extracellular matrix organization (GO:0030198). A comparison with the GO term that was enriched in the whole urinary proteome profiles (Figure 2) showed that they share six GO biological process terms with the ones enriched in the differentially secreted proteins; these terms are reported in Figure 4 as slices with stripes pattern). The remaining terms were not enriched in urinary proteome profiles, and they are reported in Figure 4 with a colored solid fill: angiogenesis (GO:0001525), phagocytosis, recognition (GO:0006910), acute-phase response (GO:0006953), extracellular matrix disassembly (GO:0022617), lipoprotein metabolic process (GO:0042157), positive regulation of B cell activation (GO:0050871), ERK1 and ERK2 cascade (GO:0070371), and B cell receptor signaling pathway (GO:0050853).

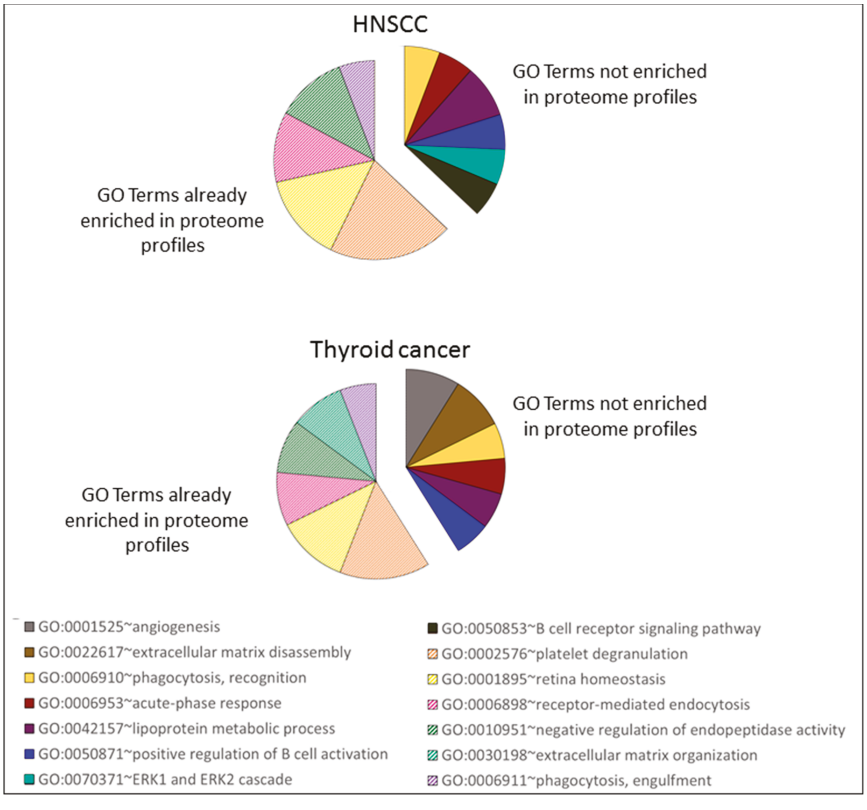


Figure 4. Enrichment in GO biological process terms, as performed with DAVID on the significantly different secreted proteins list reported in Table S2 and Figure 3 for HNSCC and thyroid cancer patients, as compared to healthy subjects. Each slice is proportional to the number of proteins identified for each GO biological process terms. The slices with the stripes pattern were already found to be enriched in the analysis that was performed on the total urinary proteome of HNSCC and thyroid cancer patients (see Figure 2).

2.3. Proteome Profiles in Urine to Monitor Effects of Interventions

2.3.1. Infusion of a HNSCC Patient with BPA

The first step in the BNCT therapy involves boron infusion before neutron irradiation; in this context, we investigated the effect of boron on urinary proteome. Specifically, the effect of the two boron-containing compounds BPA and BSH was investigated. One patient with a HNSCC was infused with BPA. Using flow-injection electrospray ionization mass spectrometry (FI-ESI-MS), the BPA-concentration in urine of the patient was measured before and after BPA-infusion (Figure 5) [16]. To evaluate the molecular effects of the compound on urine-secreted proteins, MudPIT proteomics analysis was performed on a urine sample that was taken before BPA-infusion and on a sample taken at the time point, when the urine concentration of BPA resulted in the highest. This time point equals 5 h after starting the BPA infusion, 6:45 h after starting anesthesia, and 6 h after starting the surgery (Figure 5).

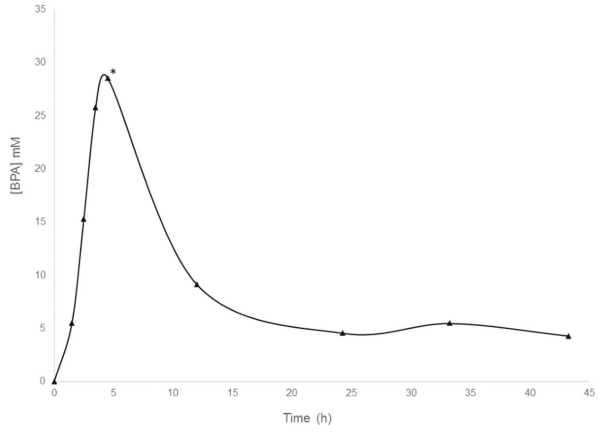


Figure 5. Concentration of boronophenylalanine (BPA) in the urine of the HNSCC patient analyzed. BPA infusion started at 0 min and ended at 1 h. * indicates the time point of urine sampling for proteomic analysis.

This proteomic analysis allowed for a preliminary evaluation of the combined effects of BPA infusion during surgery under anesthesia on the urinary proteome. Significant changes in the levels of urinary proteins were identified using the MAProMa platform, which showed variations in the urinary level of 29 proteins (Figure 6 and Table S4). The majority (22) of protein levels decreased after BPA infusion (as indicated by negative values of DAVE and DCI). Among the proteins that decreased after the intervention were some that showed higher levels in the urine of HNSCC patients as compared to healthy subjects, namely Inter-alpha-trypsin inhibitor heavy chain H4 (ITIH4), Pancreatic alpha-amylase (AMY2A), osteopontin (abbreviated as OPN–gene name SPP1), uromodulin (UMOD), aminopeptidase N (ANPEP), pro-epidermal growth factor (EGF), and vasorin (VASN). Two proteins higher in the tumor patients, as compared to healthy volunteers, resulted in additional increase after the infusion of BPA: Retinol-binding protein 4 (RBP4) and Protein AMBP (AMBP).

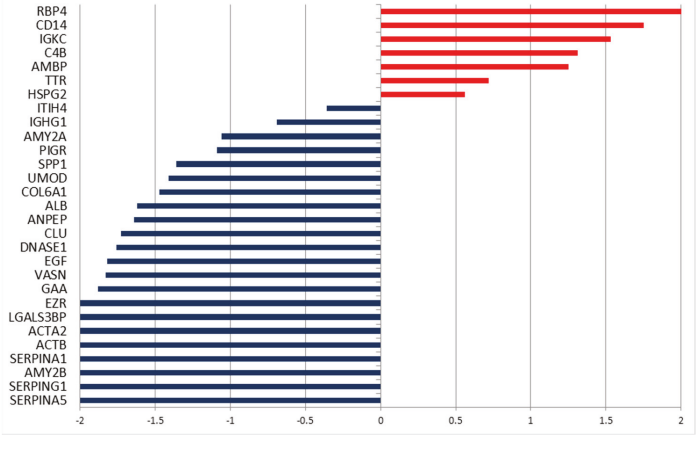


Figure 6. DAVE values of the significantly different secreted proteins that were found in urinary proteome of a patient with HNSCC, before and 5 h after the infusion of BPA and during surgery under anesthesia. The positive values indicate proteins increased after infusion, while negative values indicate proteins decreased after infusion.

The GO analysis of enriched biological process terms (Table S5) showed that there are some terms that were affected by the infusion with BPA. Some of GO terms were found to be enriched only after infusion: positive regulation of protein ubiquitination involved in ubiquitin-dependent protein catabolic process (GO:2000060), leukocyte cell-cell adhesion (GO:0007159), retinol metabolic process (GO:0042572), retinoid metabolic process (GO:0001523), and carbohydrate metabolic process (GO:0005975).

2.3.2. Infusion of a Thyroid Cancer Patient with BSH

In a patient with thyroid cancer, we investigated the changes in the urinary proteome in urine samples that were taken before BSH infusion and 10 h after starting the infusion. This time point equals 30 min after starting anesthesia and just before starting surgery. Again, we identified the differential levels of proteins (Figure 7 and Table S6). Twenty-one proteins changed their levels in urine after BSH infusion. Only two proteins that were more abundant in the urine of tumor patients in comparison to healthy volunteers decreased: uromodulin and OPN. Interestingly, there was an increase of serotransferrin (TF) after BSH infusion, while levels of this protein were lower in tumor patients as compared to healthy subjects.

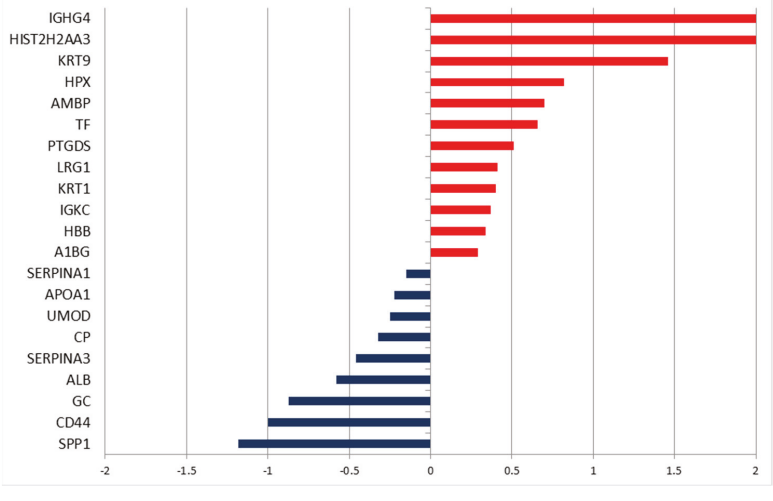


Figure 7. DAVE values of the significantly different secreted proteins found in urinary proteome of a patient affected by thyroid cancer, before and 10 h after the infusion with BSH and at the start of surgery under anesthesia. The positive values indicate proteins increased after infusion, while negative values indicate proteins decreased after infusion.

For this patient, a total number of 24 GO biological process terms were enriched, when considering the differentially secreted proteins after BSH-infusion (Table S7). Eleven GO terms were specifically enriched in relation to BSH infusion: Defense response to bacterium (GO:0042742), inflammatory response (GO:0006954), phagocytosis, engulfment (GO:0006911), B cell receptor signaling pathway (GO:0050853), cellular oxidant detoxification (GO:0098869), cellular iron ion homeostasis (GO:0006879), vitamin transport (GO:0051180), oxygen transport (GO:0015671), establishment of skin barrier (GO:0061436), hydrogen peroxide catabolic process (GO:0042744), and positive regulation of cell death (GO:0010942).

2.4. Reduction of Proteins Related to Inflammation after Infusion: galectin 3 Binding Protein, CD44, and OPN

In both treated patients, regardless of the boron compound infused, a reduction in three proteins has been observed: galectin 3 Binding Protein (LGALS3BP), CD44, and OPN (SPP1) (Figure 8). These glycoproteins are involved in inflammatory processes and possibly tumor progression [22–24]. All three proteins were reduced or even absent in urine that was sampled after boron compound infusion and during anesthesia and surgery.

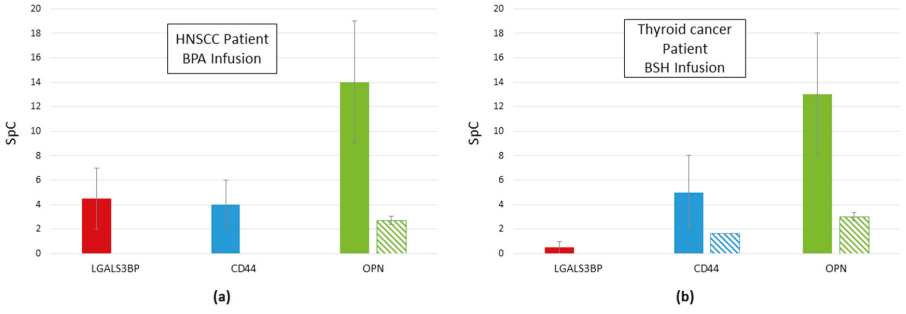


Figure 8. (**a**) Total amount of LGALS3BP (red fill), CD44 (blue fill), and osteopontin (OPN) (SPP1, green fill) found in urine of the HNSCC patient infused with BPA before (solid fill) and 5 h after infusion (diagonal lines fill). (**b**) Total amount of LGALS3BP (red fill), CD44 (blue fill), and OPN (SPP1, green fill) found in urine of the thyroid cancer patient infused with BSH before (solid fill) and 10 h after infusion (diagonal lines fill). Data are expressed as SpC ± SEM.

3. Discussion

In this work, we present the potential of LC-MS-based proteomics for monitoring treatment, using an easily collectable biological fluid, such as urine. Protein profiles from urine samples allowed for the discovery of potential tumor biomarkers by using a label-free quantitative comparison between healthy subjects and HNSCC or thyroid cancer patients to find proteins that were differentially secreted in urine from tumor patients. The urinary proteomes of cancer patients showed similarity in their overall composition, as confirmed by the GO biological process enriched terms. In particular, "cell adhesion" term was found to be enriched in patients but not in healthy volunteers, suggesting that this biological process is disrupted by the presence of tumors and it is reflected in urine protein composition.

In both groups of patients with tumors, a great increase in uromodulin (UMOD) is interesting to note, as this is known to be the most abundant in urine [25], as confirmed in our data for all subjects. Although the total amount of uromodulin is related to the disease, it is hard to explain this increase with respect to the healthy subjects, since the role of uromodulin is still not completely clear in both physiological and pathological conditions. It is possible that uromodulin is related to an inflammatory state that is observed in tumor patients, since its involvement in inflammatory processes in kidney injury has been demonstrated [26].

Indeed, some of the proteins with higher levels in both groups are involved in inflammatory processes. Among these, one of the most interesting is ITIH4, which is involved in inflammatory responses to trauma [27]. Several groups have also demonstrated the potential of ITIH4 as a predictive biomarker in a number of tumor entities [28–30]. In our investigation, the ITIH4 levels are elevated in the urine of patients with HNSCC and in patients with thyroid cancer, but it is significantly higher in the urine of HNSCC patients as compared to the urine of thyroid cancer patients.

Two immunoglobulins were found to be excreted in urine from tumor-affected subjects to a higher extent, with no significant differences between the two tumor entities: Ig gamma-1 chain C region (IGHG1), and Ig kappa chain C region (IGKC). In particular, IGKC show a very distinctive increase, which was similar for both of the tumor entities examined. Proteins that are secreted in urine clearly

reflect what happens in the organism and, in particular, the inflammation processes in progress. OPN (SPP1) is an interesting protein that is involved in inflammatory processes and it resulted higher in urine samples of tumor patients. Many cells, including the cells of the immune system but also tumor cells, secrete it. Specifically, OPN increases in breast, colon, gastric, liver, lung, and prostate cancers [31]. It interacts with the CD44 receptor and it is considered to be one of the most promising prognostic biomarkers for cancer [32].

The presence of inflammatory processes in tumor patient seems to be also confirmed by the GO biological process enrichment analysis of differentially secreted proteins. Indeed, the enrichment of GO terms, such as acute-phase response (GO:0006953), positive regulation of B cell activation (GO:0050871), and B cell receptor signaling pathway (GO:0050853), indicate that the inflammatory processes are in progress and they are reflected by changes in urine proteome composition.

Obviously, inflammation-related proteins are not sufficient for the diagnosis of cancer, since they could be linked to additional conditions. Indeed, among the other proteins found to be excreted to an enhanced extent, some have been already related to cancer and could therefore be considered to be cancer biomarkers. As an example, aminopeptidase N (ANPEP) has been associated with the growth of different human cancers and has been suggested as a suitable target for antineoplastic treatment approaches [33]. In particular, an increase in aminopeptidase N activity in HNSCC has already been observed [34]. Vasorin (VASN) is a type I transmembrane protein that plays an important role in tumor development and vasculogenesis; it is a potential serum biomarker of hepatocellular carcinoma and it may be a drug target for its treatment [35]. Pancreatic alpha-amylase (AMY2A) has been related to cell adhesion, growth, and the invasion of cancer cells and it has been proposed as a candidate urine biomarker for pancreatic cancer [36]. Zinc-alpha-2-glycoprotein (AZGP1) is a protein that is known to have a role in different tumor entities. An overexpression of zinc-alpha-2-glycoprotein has been observed in a proteomic analysis of saliva and serum from patients that are diagnosed with HNSCC [37].

Some of the proteins that were identified as significantly higher in the urine of tumor patients as compared to healthy subjects have been already proposed as possible biomarkers, although no cancer-related functions have been identified so far. For example, the protein AMBP (AMBP), which shows high levels in urine from patients with both HNSCC and thyroid cancer, has been found to also be secreted in urine from the bladder [38] and prostate cancer patients [39] and in other biofluids (such as cerebrospinal fluid and serum) of such patients [40,41].

The effect of tumor could also lead to a loss of function and therefore to a decrease in particular proteins. As an example, the reduction in serotransferrin (TF) observed is possibly related to a dysfunction in iron homeostasis. This could lead to further damage of the tissues following cancer progression. Additionally, the reduction of cadherin 1 (CDH1), a growth and invasion suppressor, is an interesting observation [42].

These data show that LC-MS proteomic analysis of urine could be a suitable and convenient method for the characterization of different types of cancer. In particular, it is possible to simultaneously follow several urine-secreted proteins that were altered by the presence of a tumor, thus creating a tool more reliable than an approach based on the changes of exclusively one single protein. Furthermore, specific proteins in urine could be followed for differentiating one cancer from another. As an example, in this study, we showed that retinol-binding protein 4 (RBP4) was exclusively found in the urine of patients with HNSCC but not thyroid cancer, indicating that this protein could be a specific HNSCC biomarker. RBP4 is an adipokine that is mainly produced by adipose tissue and its serum levels in medullary thyroid carcinoma patients are not significantly different from those that are found in healthy individuals [43]. On the contrary, RBP4 was recently reported as a potential biomarker of oral squamous cell carcinoma, in combination with clusterin, haptoglobin, complement C3c, and proapolipoprotein A1 [44].

The method presented has been proven to also be useful for monitoring the effects at a molecular level of two different boron compounds, used for BNCT: BPA and BSH. The infusion of patients with

either one of the boron compounds induces changes in the proteins that were secreted in the urine. In particular, we observed a reduction in proteins that were involved in inflammation, such as GALS3BP and CD44. Additionally, OPN was decreased after boron infusion, both by BPA and BSH. OPN is a secreted protein that is overexpressed in several human carcinomas. It has been implicated in different mechanisms of tumor progression: cell proliferation, survival, drug resistance, invasion through its role in intercellular communication, and tumor microenvironment formation [45]. Moreover, high levels of OPN have been found in cerebrospinal fluid and serum of cancer patients [46] and it is considered a possible target for cancer therapy [23]. It is to be noted that the role in inflammation of OPN is possibly mediated by CD44 isoforms [47,48], suggesting that a specific inflammatory pathway could be the target of BSH or BPA.

It has already been reported that other boron-containing compounds are able to reduce inflammation both in animals and in humans [49], suggesting a beneficial effect of boron as a regulator of the inflammatory reactions with a possible role in cancer and other diseases [50]. Our proteomics data suggest that BPA and BSH could also reduce tumor-derived inflammation, thus possibly helping the treatment. Certainly, further experiments are needed to confirm this hypothesis and to find other inflammation-related proteins that are reduced by boron infusion. Of note, in the setting of this feasibility investigation, the effects of drugs infused for anesthesia, and/or surgery itself might influence the proteasome of the urine as well; thus, the observed effects cannot be solely attributed to the boron compounds.

To date, very few studies investigated the effects of BNCT on the proteome. Although BNCT is a binary therapy, up to now, the existing studies analyze the effects of both boron infusion and neutron irradiation, but not the effect of either component of this treatment modality. In very few cases, the effect of neutron irradiation with and without the infusion of a boron compound were evaluated, such as the unique proteomics investigation after BNCT, as performed by Sato et al. [20]. Unfortunately, their data cannot be compared with those that were obtained in the present work, because the authors analyzed squamous cell carcinoma SAS cells in vitro irradiated with a thermal neutron beam with and without BPA. The authors used a 2D gel approach and characterized approximately 20 proteins that were mainly related to the endoplasmic reticulum-localized lymphoid-restricted protein.

The results of our study demonstrate the ability of shotgun proteomics based on the MudPIT approach to identify a great number of potential protein biomarkers in urine. This is of growing interest, since it is becoming clear that urinary proteome reflects changes that are induced by several diseases, including different types of cancer. Furthermore, urine could be preferred to blood or cerebrospinal fluid because of the easier and less invasive procedure of collection. The presented data must be considered as a feasibility evaluation due to low patients' numbers and urine samples that are not systematically drawn for this purpose. Obviously, larger patient sets will also help to discover the influence on the urinary proteome of other factors, such as age, gender, medication, or diet. The proteomic analysis of urine is non-invasive and fast method, not only for tumor diagnosis, but also for following the effects of treatment at the molecular level. Indeed, in the future, urine and possibly other biofluids will be analyzed during and after the treatment in order to follow and possibly verify the changes that are induced by therapy. In addition, urinary exosomes will be studied, since it seems that biomarkers are enriched in secreted vesicles, facilitating the identification of disease-specific proteins [51]. A forthcoming application of this approach could also be the identification of biomarkers that are useful in monitoring the BNCT treatment and potentially predicting the response of patients to treatment, allowing for the application of the best therapy for each person in the pursuit of personalized medicine.

Finally, although, in BNCT trials, the molecular effects due to boron-containing compounds were never systematically evaluated, our data indicate the proteome analysis of urine samples as a specific and informative method, and it may be a potential mirror of disease. These results suggest that future investigations with this method should evaluate the effects of boron compound infusion. The present work represents a proof of principle to investigate the BNCT treatment effects; of course, presented

data should be considered as a preliminary evaluation and they need to be confirmed by further analyses on larger patient sets through a multicenter collaboration.

4. Materials and Methods

4.1. Patients Recruitment and Urine Collection

The samples were collected from patients who participated in the clinical trial EORTC 11001, "^{10}B-uptake in different tumors using the boron compounds BSH and BPA". This study investigates the delivery of ^{10}B to certain tumor entities by the boron compounds BSH and BPA. The study procedures included intravenous infusion of BSH and BPA prior to a planned surgical procedure that was necessary for tumor treatment and sample collection, but no radiotherapy with BNCT. All of the participants gave written informed consent prior to inclusion. The evaluation presented here became possible because all of the participants agreed that the samples taken might also be investigated by innovative procedures that were not yet available at the time of the sampling. The Protocol Review Committee of the EORTC and the Ethics Committee of the Medical Faculty of the University Duisburg-Essen approved the trial. The complete clinical characteristics of patients are reported in Table S8.

In addition to sample collection within the EORTC trial 11001, urine samples were collected after written informed consent from heavily smoking (>25 cigarettes/day) healthy staff members of University Hospital Essen, who volunteered to participate. Healthy volunteers did not receive any medication or study the compounds. The age of the healthy volunteers were in the same range of the tumor-affected subjects.

All of the analyzed samples from cancer patients are reported in Table 1.

Table 1. Analyzed cancer patients' age at the time of urine collection and tumor diagnosis.

Subject	Age	Diagnosis
P12	45.3	Squamous cell carcinoma of the head and neck region
P21	50.6	Squamous cell carcinoma of the head and neck region
P20	68.5	Squamous cell carcinoma of the head and neck region
P22	53.1	Squamous cell carcinoma of the head and neck region
P11	79.6	Thyroid carcinoma
P14	34.6	Papillary thyroid carcinoma
P16	34.0	Papillary thyroid carcinoma

4.2. Boron Compounds and Urine Collection

BSH and BPA were purchased from KATCHEM Ltd. (Praha, Czech Republic). The quality of the study medication was strictly controlled, including an examination of the identity of compounds by infra-red spectroscopy, monitoring of purity by high pressure liquid chromatography, and test for pyrogenicity. The enrichment of ^{10}B was $\geq$99.6% in both compounds and it was tested with prompt gamma ray spectroscopy (PGRA) and inductively coupled plasma-atomic emission spectroscopy (ICP-AES). The injection-solutions were prepared according to standard operating procedures that were established for the EORTC trials 11961, 11001, and 11011 [52,53].

To improve solubility, BPA (100 mg/kg) was infused as a BPA–fructose complex, within 1 h. Urine samples were taken before and up to 34 h after the start of infusion. Urine samples were immediately frozen at −20 °C until analysis.

BSH (50 mg/kg) was dissolved in saline and infused within 1 h. The urine samples were taken before the start of infusion and at 9.5 h after the start of the BSH infusion. Samples were immediately frozen at −20 °C until analysis.

In the past, urine samples from patients that were infused by boron-containing compounds were investigated to determine compound pharmacokinetics [16]. We took advantages of the availability of these samples and the shotgun proteomics approach to describe the related proteome profiles. Specifically, urine samples from four patients with HNSCC and three patients with thyroid cancer,

which were collected before infusion of the boron-containing compounds, were analyzed and compared to the healthy volunteers. Urine samples from two patients (one patient with HNSCC and one patient with thyroid cancer) were analyzed before compound infusion and after compound infusion in order to evaluate the eventual changes in the urinary proteome that is caused by the infusion of the boron-containing drugs BPA and BSH, as well as surgery and anesthesia.

4.3. Urine Preparation for Proteomics Analysis

For each sample, 600 µL of urine were centrifuged at 2000 g for 10 min. Supernatant was collected and filtered twice with the Microcon YM10 system to eliminate salts. The protein concentration was assayed using the SPN^{TM}–Protein Assay kit (G-Biosciences, St. Louis, MO, USA).

4.4. Enzymatic Digestion of Protein Samples

Sequencing grade modified trypsin (Promega, Madison, WI, USA) was added to 50 µL of conditioned medium containing 1 µg protein at a 1:50 enzyme:protein ratio (*w:w*) in 100 mM ammonium bicarbonate, pH 8.0, and then incubated at 37 °C overnight. The reaction was stopped by acidification with trifluoroacetic acid. A second aliquot of each sample was digested with pepsin (Sigma-Aldrich, Milan, Italy) at 1:50 enzyme:protein ratio (*w:w*) in 100 mM ammonium acetate pH 3.0 at room temperature for 4 h and then immediately analyzed. Ten microliters of the peptide mixture were directly injected into the 2DC-MS/MS.

4.5. Two-Dimensional Capillary Chromatography-Tandem Mass Spectrometry (2DC-MS/MS) Analysis

Ten microliters of the peptide mixtures, as obtained from the digestion of the protein samples, were analyzed by means of two-dimensional microchromatography coupled with an ion trap mass spectrometer, using the ProteomeX system (ThermoElectron, San Josè, CA, USA) that was equipped with Bioworks 3.1 as graphical interface for data handling. Peptide mixtures were first separated by means of strong cation-exchange chromatography (Biobasic-SCX column, 5 µm, 0.3 ID × 150 mm, ThermoHypersil, Bellofonte, PA, USA) using seven steps of increasing ammonium chloride concentration (0, 50, 100, 150, 200, 300, and 600 mM). Each salt step was directly loaded onto the reversed phase column (Biobasic-C18, 0.180 ID × 100 mm, ThermoHypersil, Bellofonte, PA) and then separated with an acetonitrile gradient: eluent A, 0.1% formic acid in water; eluent B, 0.1% formic acid in acetonitrile; the gradient profile was 5% B for 3 min, followed by 5 to 50% B within 40 min. Peptides that were eluted from the C18 column were sent directly to an ion trap LCQXP mass spectrometer through an ESI ion source interface that was equipped with a metal needle (10 µm ID). The heated capillary was held at 160 °C, ion spray 3.2 kV, and capillary voltage 67 V. Spectra were acquired in positive mode (in the range of 400–1600 *m/z*) using dynamic exclusion for MS/MS analysis (collision energy 35%).

4.6. Data Processing of MS Results

All data generated were searched using the Sequest HT search engine contained in the Proteome Discoverer software, version 2.1 (Thermo Scientific, Waltham, MA, USA). The experimental MS/MS spectra were correlated to tryptic peptide sequences by comparison with the theoretical mass spectra that were obtained by in silico digestion of the Uniprot human proteome database (70,726 entries), downloaded in January 2017 (www.uniprot.org). The following criteria were used for the identification of peptide sequences and related proteins: trypsin as enzyme, three missed cleavages per peptide, mass tolerances of ±1 Da for precursor ions, and ±1 Da for fragment ions. Percolator node was used with a target-decoy strategy to give final false discovery rates (FDR) at Peptide Spectrum Match (PSM) level of 0.01 (strict) based on q-values, when considering a maximum deltaCN of 0.05 [54]. Only peptides with high confidence, minimum peptide length of six amino acids, and rank 1 were considered. Protein grouping and strict parsimony principle were applied. The MS data have been

deposited to the ProteomeXchange Consortium via the PRIDE partner repository [55] and they are available upon request to the corresponding author.

The output data that were obtained from SEQUEST software were treated with the MAProMa (Multidimensional Algorithm Protein Map) in-house algorithm for a comparison of the protein lists, evaluation of relative abundances, and plotting virtual 2D maps [56]. A label-free approach has been used, counting the peptide spectrum matches (PSMs) for each identified protein [21]. Fold changes were calculated as natural logarithm of the SpC (SpC1/SpC2) [57]; the natural logarithm of the proteins identified exclusively in one of the two compared groups was set to ± 100. G test was performed for the differentially expressed protein, as already reported for label-free quantification [58].

Venn diagrams were created with Venny (http://bioinfogp.cnb.csic.es/tools/venny/index.html).

4.7. GO Terms Enrichment

GO terms enrichment was performed with DAVID 6.8 (https://david.ncifcrf.gov/), using Uniprot accessions [59]. Biological process terms were considered to be enriched when the p-value and Bonferroni test results were below 0.0001 and at least 10 gene were found for the term.

Supplementary Materials: The following are available online. Figure S1: Comparison between two different replicates of patient 20, Figure S2: Two dimensions map of healthy subjects urine proteome profiles, Figure S3: Two dimensions map of HNSCC patients urine proteome profiles, Figure S4: Two dimensions map of thyroid cancer patients urine proteome profiles, Figure S5: Venn diagram of proteins distribution among the three groups, Table S1: List of all proteins identified in the samples analyzed, Table S2: DAVE and DCI values, fold changes, and G test values of the proteins differentially secreted in urine from healthy subjects versus those from cancer patients, Table S3: DAVE and DCI values, fold changes, and G test values of the proteins differentially secreted in urine from thyroid cancer patients versus those from HNSCC patients, Table S4: DAVE and DCI values of the proteins differentially secreted in urine from a patient affected by HNSCC, before and 5 h after the infusion with BPA, Table S5: GO biological process terms enriched in differentially expressed proteins for HNSCC patient treated with BPA, Table S6: DAVE and DCI values of the proteins differentially secreted in urine from a patient affected by thyroid cancer, before and 4 h after the infusion with BSH, Table S7: GO biological process terms enriched in differentially expressed proteins for thyroid cancer patient treated with BSH, Table S8: Subjects analyzed with complete clinical characteristics.

Author Contributions: Conceptualization, W.A.G.S. and P.L.M.; methodology, A.D.P. and R.R.; formal analysis, E.F. and D.D.S.; investigation, F.B., A.D.P. and R.R.; resources, A.W. and W.A.G.S.; data curation, E.F. and D.D.S.; writing—original draft preparation, E.F. and P.L.M.; supervision, W.A.G.S. and P.L.M.; project administration, A.W., W.A.G.S. and P.L.M.; funding acquisition, A.W., W.A.G.S. and P.L.M.

Funding: This work was partially supported by the Italian CNR FLAGSHIPs Project and AMANDA project, funded by the MIUR "InterOmics" and CNR-Lombardy Agreement, respectively. The BNCT clinical trial was financially supported by the European Commission under contract no QLK3-CT-1999-01067 (FP5) and partly under contract number 005045 (FP6). E.F. was supported by a Fondazione Umberto Veronesi Fellowship.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Moss, R.L. Critical review, with an optimistic outlook, on Boron Neutron Capture Therapy (BNCT). *Appl. Radiat. Isot.* **2014**, *88*, 2–11. [CrossRef] [PubMed]
2. Miyatake, S.-I.; Kawabata, S.; Hiramatsu, R.; Kuroiwa, T.; Suzuki, M.; Kondo, N.; Ono, K. Boron Neutron Capture Therapy for Malignant Brain Tumors. *Neurol. Med. Chir. (Tokyo)* **2016**, *56*, 361–371. [CrossRef] [PubMed]
3. Sauerwein, W.; Zurlo, A. The EORTC Boron Neutron Capture Therapy (BNCT) Group: achievements and future projects. *Eur. J. Cancer* **2002**, *38*, 31–34. [CrossRef]
4. Kankaanranta, L.; Seppälä, T.; Koivunoro, H.; Saarilahti, K.; Atula, T.; Collan, J.; Salli, E.; Kortesniemi, M.; Uusi-Simola, J.; Välimäki, P.; et al. Boron neutron capture therapy in the treatment of locally recurred head-and-neck cancer: final analysis of a phase I/II trial. *Int. J. Radiat. Oncol. Biol. Phys.* **2012**, *82*, e67–e75. [CrossRef] [PubMed]
5. Suzuki, M.; Kato, I.; Aihara, T.; Hiratsuka, J.; Yoshimura, K.; Niimi, M.; Kimura, Y.; Ariyoshi, Y.; Haginomori, S.; Sakurai, Y.; et al. Boron neutron capture therapy outcomes for advanced or recurrent head and neck cancer. *J. Radiat. Res. (Tokyo)* **2014**, *55*, 146–153. [CrossRef] [PubMed]

6. Aihara, T.; Morita, N.; Kamitani, N.; Kumada, H.; Ono, K.; Hiratsuka, J.; Harada, T. BNCT for advanced or recurrent head and neck cancer. *Appl. Radiat. Isot.* **2014**, *88*, 12–15. [CrossRef] [PubMed]
7. Lim, D.; Quah, D.S.; Leech, M.; Marignol, L. Clinical potential of boron neutron capture therapy for locally recurrent inoperable previously irradiated head and neck cancer. *Appl. Radiat. Isot.* **2015**, *106*, 237–241. [CrossRef]
8. Kato, I.; Fujita, Y.; Maruhashi, A.; Kumada, H.; Ohmae, M.; Kirihata, M.; Imahori, Y.; Suzuki, M.; Sakrai, Y.; Sumi, T.; et al. Effectiveness of boron neutron capture therapy for recurrent head and neck malignancies. *Appl. Radiat. Isot. Data Instrum. Methods Use Agric. Ind. Med.* **2009**, *67*, S37–S42. [CrossRef]
9. Wittig, A.; Collette, L.; Moss, R.; Sauerwein, W.A. Early clinical trial concept for boron neutron capture therapy: a critical assessment of the EORTC trial 11001. *Appl. Radiat. Isot. Data Instrum. Methods Use Agric. Ind. Med.* **2009**, *67*, S59–S62. [CrossRef]
10. Dagrosa, A.; Carpano, M.; Perona, M.; Thomasz, L.; Nievas, S.; Cabrini, R.; Juvenal, G.; Pisarev, M. Studies for the application of boron neutron capture therapy to the treatment of differentiated thyroid cancer. *Appl. Radiat. Isot.* **2011**, *69*, 1752–1755. [CrossRef]
11. Yoshida, F.; Matsumura, A.; Shibata, Y.; Yamamoto, T.; Nakauchi, H.; Okumura, M.; Nose, T. Cell cycle dependence of boron uptake from two boron compounds used for clinical neutron capture therapy. *Cancer Lett.* **2002**, *187*, 135–141. [CrossRef]
12. Carpano, M.; Perona, M.; Rodriguez, C.; Nievas, S.; Olivera, M.; Santa Cruz, G.A.; Brandizzi, D.; Cabrini, R.; Pisarev, M.; Juvenal, G.J.; et al. Experimental Studies of Boronophenylalanine ((10)BPA) Biodistribution for the Individual Application of Boron Neutron Capture Therapy (BNCT) for Malignant Melanoma Treatment. *Int. J. Radiat. Oncol. Biol. Phys.* **2015**, *93*, 344–352. [CrossRef] [PubMed]
13. Koivunoro, H.; Hippeläinen, E.; Auterinen, I.; Kankaanranta, L.; Kulvik, M.; Laakso, J.; Seppälä, T.; Savolainen, S.; Joensuu, H. Biokinetic analysis of tissue boron (10B) concentrations of glioma patients treated with BNCT in Finland. *Appl. Radiat. Isot.* **2015**, *106*, 189–194. [CrossRef] [PubMed]
14. Svantesson, E.; Capala, J.; Markides, K.E.; Pettersson, J. Determination of boron-containing compounds in urine and blood plasma from boron neutron capture therapy patients. The importance of using coupled techniques. *Anal. Chem.* **2002**, *74*, 5358–5363. [CrossRef] [PubMed]
15. Kulvik, M.; Kallio, M.; Laakso, J.; Vähätalo, J.; Hermans, R.; Järviluoma, E.; Paetau, A.; Rasilainen, M.; Ruokonen, I.; Seppälä, M.; et al. Biodistribution of boron after intravenous 4-dihydroxyborylphenylalanine-fructose (BPA-F) infusion in meningioma and schwannoma patients: A feasibility study for boron neutron capture therapy. *Appl. Radiat. Isot. Data Instrum. Methods Use Agric. Ind. Med.* **2015**, *106*, 207–212. [CrossRef] [PubMed]
16. Basilico, F.; Sauerwein, W.; Pozzi, F.; Wittig, A.; Moss, R.; Mauri, P.L. Analysis of 10B antitumoral compounds by means of flow-injection into ESI-MS/MS. *J. Mass Spectrom. JMS* **2005**, *40*, 1546–1549. [CrossRef] [PubMed]
17. Bendel, P.; Wittig, A.; Basilico, F.; Mauri, P.L.; Sauerwein, W. Metabolism of borono-phenylalanine-fructose complex (BPA-fr) and borocaptate sodium (BSH) in cancer patients–results from EORTC trial 11001. *J. Pharm. Biomed. Anal.* **2010**, *51*, 284–287. [CrossRef] [PubMed]
18. Masutani, M.; Baiseitov, D.; Itoh, T.; Hirai, T.; Berikkhanova, K.; Murakami, Y.; Zhumadilov, Z.; Imahori, Y.; Hoshi, M.; Itami, J. Histological and biochemical analysis of DNA damage after BNCT in rat model. *Appl. Radiat. Isot.* **2014**, *88*, 104–108. [CrossRef]
19. Mauri, P.L.; Basilico, F. Proteomic Investigations for Boron Neutron Capture Therapy. In *Neutron Capture Therapy: Principles and Applications*; Sauerwein, W., Wittig, A., Moss, R., Nakagawa, Y., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2012; pp. 189–200. ISBN 978-3-642-31334-9.
20. Sato, A.; Itoh, T.; Imamichi, S.; Kikuhara, S.; Fujimori, H.; Hirai, T.; Saito, S.; Sakurai, Y.; Tanaka, H.; Nakamura, H.; et al. Proteomic analysis of cellular response induced by boron neutron capture reaction in human squamous cell carcinoma SAS cells. *Appl. Radiat. Isot. Data Instrum. Methods Use Agric. Ind. Med.* **2015**, *106*, 213–219. [CrossRef]
21. Mauri, P.; Scigelova, M. Multidimensional protein identification technology for clinical proteomic analysis. *Clin. Chem. Lab. Med.* **2009**, *47*, 636–646. [CrossRef]
22. Orian-Rousseau, V.; Ponta, H. Perspectives of CD44 targeting therapies. *Arch. Toxicol.* **2015**, *89*, 3–14. [CrossRef]

23. Castello, L.M.; Raineri, D.; Salmi, L.; Clemente, N.; Vaschetto, R.; Quaglia, M.; Garzaro, M.; Gentilli, S.; Navalesi, P.; Cantaluppi, V.; et al. Osteopontin at the Crossroads of Inflammation and Tumor Progression. *Mediators Inflamm.* **2017**, *2017*, 4049098. [CrossRef] [PubMed]
24. Loimaranta, V.; Hepojoki, J.; Laaksoaho, O.; Pulliainen, A.T. Galectin-3-binding protein: A multitask glycoprotein with innate immunity functions in viral and bacterial infections. *J. Leukoc. Biol.* **2018**, *104*, 777–786. [CrossRef] [PubMed]
25. Devuyst, O.; Olinger, E.; Rampoldi, L. Uromodulin: from physiology to rare and complex kidney disorders. *Nat. Rev. Nephrol.* **2017**, *13*, 525–544. [CrossRef]
26. El-Achkar, T.M.; Wu, X.-R. Uromodulin in kidney injury: an instigator, bystander, or protector? *Am. J. Kidney Dis. Off. J. Natl. Kidney Found.* **2012**, *59*, 452–461. [CrossRef] [PubMed]
27. Kashyap, R.S.; Nayak, A.R.; Deshpande, P.S.; Kabra, D.; Purohit, H.J.; Taori, G.M.; Daginawala, H.F. Inter-alpha-trypsin inhibitor heavy chain 4 is a novel marker of acute ischemic stroke. *Clin. Chim. Acta Int. J. Clin. Chem.* **2009**, *402*, 160–163. [CrossRef]
28. Opstal-van Winden, A.W.J.; Krop, E.J.M.; Kåredal, M.H.; Gast, M.-C.W.; Lindh, C.H.; Jeppsson, M.C.; Jönsson, B.A.G.; Grobbee, D.E.; Peeters, P.H.M.; Beijnen, J.H.; et al. Searching for early breast cancer biomarkers by serum protein profiling of pre-diagnostic serum; a nested case-control study. *BMC Cancer* **2011**, *11*, 381. [CrossRef] [PubMed]
29. Davalieva, K.; Kiprijanovska, S.; Komina, S.; Petrusevska, G.; Zografska, N.C.; Polenakovic, M. Proteomics analysis of urine reveals acute phase response proteins as candidate diagnostic biomarkers for prostate cancer. *Proteome Sci.* **2015**, *13*, 2. [CrossRef]
30. Li, X.; Li, B.; Li, B.; Guo, T.; Sun, Z.; Li, X.; Chen, L.; Chen, W.; Chen, P.; Mao, Y.; et al. ITIH4: Effective Serum Marker, Early Warning and Diagnosis, Hepatocellular Carcinoma. *Pathol. Oncol. Res. POR* **2018**, *24*, 663–670. [CrossRef]
31. Icer, M.A.; Gezmen-Karadag, M. The multiple functions and mechanisms of osteopontin. *Clin. Biochem.* **2018**, *59*, 17–24. [CrossRef]
32. Wolff, D.; Greinix, H.; Lee, S.J.; Gooley, T.; Paczesny, S.; Pavletic, S.; Hakim, F.; Malard, F.; Jagasia, M.; Lawitschka, A.; et al. Biomarkers in chronic graft-versus-host disease: quo vadis? *Bone Marrow Transplant.* **2018**, *53*, 832–837. [CrossRef] [PubMed]
33. Wickström, M.; Larsson, R.; Nygren, P.; Gullbo, J. Aminopeptidase N (CD13) as a target for cancer chemotherapy. *Cancer Sci.* **2011**, *102*, 501–508. [CrossRef] [PubMed]
34. Pérez, I.; Varona, A.; Blanco, L.; Gil, J.; Santaolalla, F.; Zabala, A.; Ibarguen, A.M.; Irazusta, J.; Larrinaga, G. Increased APN/CD13 and acid aminopeptidase activities in head and neck squamous cell carcinoma. *Head Neck* **2009**, *31*, 1335–1340. [CrossRef] [PubMed]
35. Li, S.; Li, H.; Yang, X.; Wang, W.; Huang, A.; Li, J.; Qin, X.; Li, F.; Lu, G.; Ding, H.; et al. Vasorin is a potential serum biomarker and drug target of hepatocarcinoma screened by subtractive-EMSA-SELEX to clinic patient serum. *Oncotarget* **2015**, *6*, 10045–10059. [CrossRef] [PubMed]
36. Davalieva, K.; Kiprijanovska, S.; Maleva Kostovska, I.; Stavridis, S.; Stankov, O.; Komina, S.; Petrusevska, G.; Polenakovic, M. Comparative Proteomics Analysis of Urine Reveals Down-Regulation of Acute Phase Response Signaling and LXR/RXR Activation Pathways in Prostate Cancer. *Proteomes* **2017**, *6*, 1. [CrossRef] [PubMed]
37. Vidotto, A.; Henrique, T.; Raposo, L.S.; Maniglia, J.V.; Tajara, E.H. Salivary and serum proteomics in head and neck carcinomas: before and after surgery and radiotherapy. *Cancer Biomark. Sect. Dis. Markers* **2010**, *8*, 95–107. [CrossRef] [PubMed]
38. Lei, T.; Zhao, X.; Jin, S.; Meng, Q.; Zhou, H.; Zhang, M. Discovery of potential bladder cancer biomarkers by comparative urine proteomics and analysis. *Clin. Genitourin. Cancer* **2013**, *11*, 56–62. [CrossRef]
39. Fujita, K.; Kume, H.; Matsuzaki, K.; Kawashima, A.; Ujike, T.; Nagahara, A.; Uemura, M.; Miyagawa, Y.; Tomonaga, T.; Nonomura, N. Proteomic analysis of urinary extracellular vesicles from high Gleason score prostate cancer. *Sci. Rep.* **2017**, *7*, 42961. [CrossRef]
40. Saraswat, M.; Joenväärä, S.; Seppänen, H.; Mustonen, H.; Haglund, C.; Renkonen, R. Comparative proteomic profiling of the serum differentiates pancreatic cancer from chronic pancreatitis. *Cancer Med.* **2017**, *6*, 1738–1751. [CrossRef]

41. Gong, Y.; Pu, W.; Jin, H.; Yang, P.; Zeng, H.; Wang, Y.; Pang, F.; Ma, X. Quantitative proteomics of CSF reveals potential predicted biomarkers for extranodal NK-/T-cell lymphoma of nasal-type with ethmoidal sinus metastasis. *Life Sci.* **2018**, *198*, 94–98. [CrossRef]
42. Roy, F. van Beyond E-cadherin: roles of other cadherin superfamily members in cancer. *Nat. Rev. Cancer* **2014**, *14*, 121–134. [PubMed]
43. Jabbari, S.; Hedayati, M.; Yaghmaei, P.; Parivar, K. Medullary Thyroid Carcinoma–Circulating Status of Vaspin and Retinol Binding Protein-4 in Iranian Patients. *Asian Pac. J. Cancer Prev. APJCP* **2015**, *16*, 6507–6512. [CrossRef]
44. Chen, Y.; Azman, S.N.; Kerishnan, J.P.; Zain, R.B.; Chen, Y.N.; Wong, Y.-L.; Gopinath, S.C.B. Identification of host-immune response protein candidates in the sera of human oral squamous cell carcinoma patients. *PloS ONE* **2014**, *9*, e109012. [CrossRef] [PubMed]
45. Shevde, L.A.; Samant, R.S. Role of osteopontin in the pathophysiology of cancer. *Matrix Biol. J. Int. Soc. Matrix Biol.* **2014**, *37*, 131–141. [CrossRef]
46. Karpinsky, G.; Fatyga, A.; Krawczyk, M.A.; Chamera, M.; Sande, N.; Szmyd, D.; Izycka-Swieszewska, E.; Bien, E. Osteopontin: its potential role in cancer of children and young adults. *Biomark. Med.* **2017**, *11*, 389–402. [CrossRef] [PubMed]
47. Marcondes, M.C.G.; Poling, M.; Watry, D.D.; Hall, D.; Fox, H.S. In vivo osteopontin-induced macrophage accumulation is dependent on CD44 expression. *Cell. Immunol.* **2008**, *254*, 56–62. [CrossRef]
48. Wittig, B.M.; Sabat, R.; Holzlöhner, P.; Witte-Händel, E.; Heilmann, K.; Witte, K.; Triebus, J.; Tzankov, A.; Laman, J.D.; Bokemeyer, B.; et al. Absence of specific alternatively spliced exon of CD44 in macrophages prevents colitis. *Mucosal Immunol.* **2018**, *11*, 846–860. [CrossRef]
49. Nielsen, F.H. Update on human health effects of boron. *J. Trace Elem. Med. Biol. Organ Soc. Miner. Trace Elem. GMS* **2014**, *28*, 383–387. [CrossRef]
50. Routray, I.; Ali, S. Boron Induces Lymphocyte Proliferation and Modulates the Priming Effects of Lipopolysaccharide on Macrophages. *PloS One* **2016**, *11*, e0150607. [CrossRef]
51. Ferrari, E.; De Palma, A.; Mauri, P. Emerging MS-based platforms for the characterization of tumor-derived exosomes isolated from human biofluids: challenges and promises of MudPIT. *Expert Rev. Proteomics* **2017**, *14*, 757–767. [CrossRef]
52. Wittig, A.; Malago, M.; Collette, L.; Huiskamp, R.; Bührmann, S.; Nievaart, V.; Kaiser, G.M.; Jöckel, K.-H.; Schmid, K.W.; Ortmann, U.; et al. Uptake of two 10B-compounds in liver metastases of colorectal adenocarcinoma for extracorporeal irradiation with boron neutron capture therapy (EORTC Trial 11001). *Int. J. Cancer* **2008**, *122*, 1164–1171. [CrossRef] [PubMed]
53. Wittig, A.; Collette, L.; Appelman, K.; Bührmann, S.; Jäckel, M.C.; Jöckel, K.-H.; Schmid, K.W.; Ortmann, U.; Moss, R.; Sauerwein, W.A.G. EORTC trial 11001: distribution of two 10B-compounds in patients with squamous cell carcinoma of head and neck, a translational research/phase 1 trial. *J. Cell. Mol. Med.* **2009**, *13*, 1653–1665. [CrossRef] [PubMed]
54. Käll, L.; Canterbury, J.D.; Weston, J.; Noble, W.S.; MacCoss, M.J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **2007**, *4*, 923–925. [CrossRef] [PubMed]
55. Vizcaíno, J.A.; Csordas, A.; del-Toro, N.; Dianes, J.A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **2016**, *44*, D447–D456. [CrossRef] [PubMed]
56. Mauri, P.; Dehò, G. Chapter 6 A Proteomic Approach to the Analysis of RNA Degradosome Composition in Escherichia coli. In *Methods in Enzymology*; RNA Turnover in Bacteria, Archaea and Organelles; Academic Press: Cambridge, MA, USA, 2008; Volume 447, pp. 99–117.
57. Vigani, G.; Di Silvestre, D.; Agresta, A.M.; Donnini, S.; Mauri, P.; Gehl, C.; Bittner, F.; Murgia, I. Molybdenum and iron mutually impact their homeostasis in cucumber (Cucumis sativus) plants. *New Phytol.* **2017**, *213*, 1222–1241. [CrossRef] [PubMed]
58. Zhang, B.; VerBerkmoes, N.C.; Langston, M.A.; Uberbacher, E.; Hettich, R.L.; Samatova, N.F. Detecting differential and correlated protein expression in label-free shotgun proteomics. *J. Proteome Res.* **2006**, *5*, 2909–2918. [CrossRef] [PubMed]

59. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44–57. [CrossRef]

Sample Availability: Samples of the compounds are not available from the authors.

molecules

MDPI

Article

Proteome Investigation of Rat Lungs Subjected to Ex Vivo Perfusion (EVLP)

Valentina Roffia [1,2,†], Antonella De Palma [2,†], Caterina Lonati [3], Dario Di Silvestre [2], Rossana Rossi [2], Marco Mantero [4], Stefano Gatti [1], Daniele Dondossola [5], Franco Valenza [6], Pierluigi Mauri [2,*] and Francesco Blasi [4]

1 Center for Preclinical Research, Fondazione IRCCS Ca' Granda-Ospedale Maggiore Policlinico, 20122 Milan, Italy; valentina.roffia@itb.cnr.it (V.R.); stefano.gatti@policlinico.mi.it (S.G.)
2 Proteomics and Metabolomics Unit, Institute for Biomedical Technologies (ITB-CNR), 20090 Segrate (MI), Italy; antonella.depalma@itb.cnr.it (A.D.P.); dario.disilvestre@itb.cnr.it (D.D.S.); rossana.rossi@itb.cnr.it (R.R.)
3 Department of Anesthesia and Critical Care, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, 20122 Milan, Italy; caterina.lonati@gmail.com
4 Department of Pathophysiology and Transplantation, Università degli Studi di Milano, and Internal Medicine Department, Respiratory Unit and Adult Cystic Fibrosis Center, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, 20122 Milano, Italy; marco.mantero@unimi.it (M.M.); francesco.blasi@unimi.it (F.B.)
5 Surgery and Liver Transplant Center, Fondazione IRCCS Ca' Granda-Ospedale Maggiore Policlinico, 20122 Milan, Italy; dondossola.daniele@gmail.com
6 Department of Pathophysiology and Transplantation, University of Milan, 20122 Milan, Italy; franco.valenza@unimi.it
* Correspondence: pierluigi.mauri@itb.cnr.it; Tel.: +39-02-2642-2736
† These authors contributed equally to this work.

Received: 15 September 2018; Accepted: 21 November 2018; Published: 22 November 2018

Abstract: Ex vivo lung perfusion (EVLP) is an emerging procedure that allows organ preservation, assessment and reconditioning, increasing the number of marginal donor lungs for transplantation. However, physiological and airflow measurements are unable to unveil the molecular mechanisms responsible of EVLP beneficial effects on lung graft and monitor the proper course of the treatment. Thus, it is urgent to find specific biomarkers that possess these requirements but also accurate and reliable techniques that identify them. The purpose of this study is to give an overview on the potentiality of shotgun proteomic platforms in characterizing the status and the evolution of metabolic pathways during EVLP in order to find new potential EVLP-related biomarkers. A nanoLC-MS/MS system was applied to the proteome analysis of lung tissues from an optimized rat model in three experimental groups: native, pre- and post-EVLP. Technical and biological repeatability were evaluated and, together with clustering analysis, underlined the good quality of data produced. In-house software and bioinformatics tools allowed the label-free extraction of differentially expressed proteins among the three examined conditions and the network visualization of the pathways mainly involved. These promising findings encourage further proteomic investigations of the molecular mechanisms behind EVLP procedure.

Keywords: proteomics; Ex Vivo Lung Perfusion (EVLP); nanoLC-MS/MS; transplantation

1. Introduction

Ex vivo lung perfusion (EVLP) is an emerging procedure proposed for organ preservation and reconditioning which can increase the number of lungs available for transplantation without compromising its success [1]. In fact, it permits the evaluation and the treatment of lungs of doubtful

function, such as ones from donors after cardiac death or from high risk donors, not only expanding the pool of possible transplants but also improving patient outcomes [2,3].

This derives from the fact that organs, previously considered unsuitable for transplantation, thanks to EVLP are now safely used and show similar or even better performances than those of lungs transplanted immediately after procurement. In these circumstances, EVLP application could represent a benefit for all donor lungs prior to implantation (so that it could be inserted in routine transplant protocols), but the evidence-based medical data collected so far are still insufficient [4] and there are different EVLP systems and application procedures adopted worldwide [5–7]. These weaknesses emerged on the basis of lung function measurements, but further research and tools are needed to obtain a deeper comprehension of the biological actions exerted by EVLP not only at the functional but also at the molecular level.

Accordingly, there is the need to find clinical and molecular targets that could provide evidence to recommend a specific protocol, to monitor the proper course of treatment and predict donor lung performance during EVLP procedure and after transplantation. Despite the emergence of EVLP technology, currently available methods employed for the evaluation of this procedure, are limited to airway and pulmonary measurements and therefore inadequate in assessing EVLP procedure monitoring and its molecular effects on lung graft [8]. In addition, the same respiratory diseases, for which lung transplantation is the only possible therapy, are a class of highly heterogeneous and multifactorial disorders that often make early screening, diagnosis, prognosis, disease progression, and treatment effect challenging with the diagnostic techniques applied so far. In fact, traditional methods based on reactive medicine (physiological examinations, medical history and airflow measurements) are not sufficient to fully satisfy these requirements.

All these factors contributed to the urgency in finding new specific molecular and clinical biomarkers that could be considered as objective diagnostic parameters by clinicians. In the new era of medicine [8,9], biomarkers play an essential role, because they summarize a precise biological state and provide quantifiable and comparable features that represent indicator of normal state, pathogenic processes or therapeutic response [10]. Certainly, thanks also to rapid development of next generation sequencing techniques (NGS) genomics approach is the first to be widely applied for the identification of genes as molecular biomarkers of diseases and for the selection of personalized patient therapies [11,12]. However, the genome only provides theoretical information of the protein levels while the proteome describes its actual content, which ultimately determines the phenotype. For this reason, genomic/transcriptomic investigations are recently giving way to proteomic approaches for a deepened and a more precise characterization of biological processes related to different clinical phenotypes and for the molecular monitoring of markers involved in respiratory diseases, such as COPD [13,14], cystic fibrosis [15], rhinitis [16], and the effect of anti-IgE therapy [17].

Solid organ transplantation represents an interesting challenge to test the potential of the proteomics approach and the first attempts to find new potential EVLP-related biomarkers were made by analyzing the perfusate protein expression during EVLP with multiplex immunoassays [18–20]. In these studies, the attention was exclusively focused on a specific class of proteins involved in inflammatory processes related to the transplant procedure.

Indeed, it is important to evaluate if these processes and other lung biological responses are effectively due to EVLP-induced changes and not to reconditioning mechanisms that could occur during the ex vivo perfusion procedure. To this end, our coauthors of Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico of Milan developed a rat model [21] that reproduces a prolonged ex vivo perfusion on uninjured lungs without the use of anti-inflammatory drugs. With gene expression studies on lung tissues and the assessment of mediators and metabolites on perfusate, they demonstrated that EVLP broadly affects the transcriptional profile and signaling pathways under monitoring [22].

In this context, the present work represents another step forward in our understanding of lung biological response under ex vivo perfusion, employing the same rat model and taking advantage of innovative proteomic technologies. In fact, advances in proteomic approaches, especially those

which are mass spectrometry (MS)-based [23–27], could provide new tools to overcome the lack of sensitivity and specificity of traditional diagnostic techniques, and to detect more than few proteins simultaneously in a great dynamic range of concentration and in a wide variety of bio specimens [17,28].

Our main focus is to develop a shotgun label-free platform based on a nano-chip LC system combined to an Orbitrap mass spectrometer benchtop as proof of principle to characterize the complete proteome of frozen lung tissues derived from optimized EVLP rat models [21] stratified in three experimental conditions: lung biopsies collected from animals in resting conditions (native group), lung biopsies taken after death induction, in situ cold perfusion and lung procurement (pre-EVLP group), and at the end of ex vivo perfusion (post-EVLP group). An added value to this approach is given by the application of in-house software and powerful bioinformatics tools that facilitate large scale analyses, allowing the detection and validation of differentially expressed proteins as potential biomarkers of biological EVLP-induced states and the identification of metabolic pathways in which they are involved [29–33]. Finally, the comparison of data obtained from frozen and formalin-fixed paraffin-embedded (FFPE) tissues is also performed with the aim to point out all the potential of the platform proposed, considering the wide availability of those tissues in clinical biobanks [34].

2. Results

The application of a shotgun label-free platform, based on a nanochip LC system combined to a hybrid quadrupole Orbitrap mass spectrometer, allowed the direct analysis of lung samples obtained from a reproducible EVLP animal model and in three distinct conditions: native, pre-EVLP and post-EVLP. Specifically, each experimental condition was evaluated using five animals (biological replicates) and separately analyzed in duplicate (technical replicates), for a total of thirty nanoLC-MS/MS runs from which it was possible to identify proteins and perform label-free semiquantitative evaluation, cluster and network analyses (Figure 1).

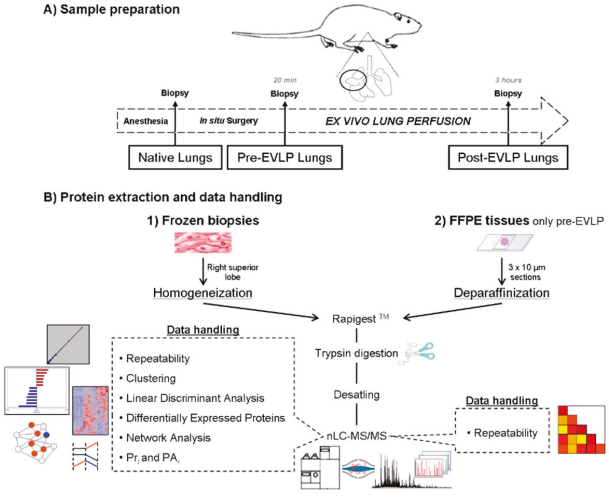


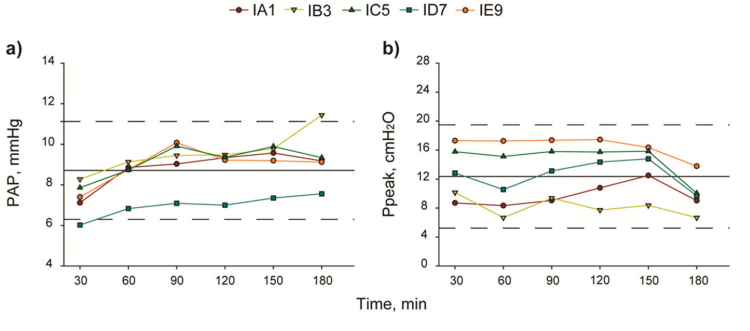
Figure 1. Experimental pipeline. Applied workflow for proteomic investigation of frozen and formalin-fixed paraffin-embedded (FFPE) lung tissues derived from an optimized ex vivo lung perfusion (EVLP) rat model. (**A**) Three distinct conditions (n = 5 for each group), native (lungs procured from animals in resting conditions), pre-EVLP (lungs subjected to in situ cold flushing and then procured), and post-EVLP (lungs subjected to in situ cold flushing, procurement and then to ex vivo perfusion/ventilation), were analyzed. (**B**) After protein extraction and digestion, the obtained peptide mixture was analyzed by a nanoLC-MS/MS system in order to allow protein identification and label-free semiquantitative data handling, such as repeatability, extraction of differentially expressed and discriminant proteins and systems biology analysis.

2.1. Surgical Procedures and Physiological Assessments

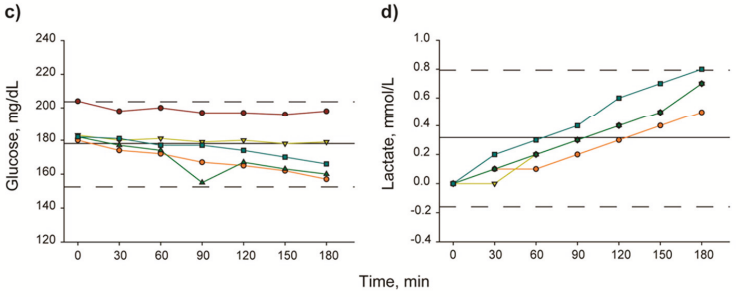
As shown in Supplementary Table S1, mean time intervals of warm and cold ischemia were similar among animals of the post-EVLP group. Conversely, agonal phase timespan was significantly longer in the rat IE9 compared to other animals of the same experimental group (Grubb's test, $p < 0.05$).

Lung functionality was evaluated through assessment of hemodynamic and respiratory parameters and measurement of the perfusate composition throughout the EVLP procedure. Pulmonary artery pressure (PAP) significantly increased ($p < 0.003$), whereas peak inspiratory pressure (Ppeak) decreased following recruitment maneuvers ($p < 0.001$). In addition, gas analysis showed that, together with a decrease in glucose concentration ($p = 0.003$), there was a rise of lactate up to 0.7 mmol/L ($p < 0.001$) (Figure 2).

Finally, the lung edema index of the Post-EVLP group was not significantly different compared to the pre-EVLP and the native groups ($p = 0.094$).



	30	60	90	120	150	180	p-value
PAP, mmHg	7.40[6.57-8.07]	8.75[7.78-9.00]	9.45[8.06-10.00][b]	9.34[8.11-9.42]	9.57[8.27-9.84][b]	9.18[8.34-10.39][b]	0.003
Ppeak, cmH_20	12.95±1.62	11.58±2.01	12.94±1.67	13.20±1.76	13.57±1.46	9.84±1.15[b,c,d,e]	0.001



	0	30	60	90	120	150	180	p-value
Gluc, mg/dL	187±6	182±5	182±6	175±9[a]	177±7	175±8[a]	174±10[a]	0.003
Lac, mmol/L	0.0[0.0-0.0]	0.1[0.0-0.1]	0.2[0.1-0.2]	0.3[0.2-0.3]	0.4[0.3-0.4]	0.5[0.4-0.5][a]	0.7[0.6-0.7][a, b]	<0.001

Figure 2. Evaluation of lung function during EVLP procedure. (**a**) PAP, significantly increased ($p < 0.003$), and (**b**) Ppeak, decreased following recruitment maneuvers ($p < 0.001$). (**c**) Concentration of glucose (from 187 ± 6 mg/dL at 0 min to 174 ± 10 at 180 min, $p = 0.003$) and (**d**) lactate (0.6–0.7 mmol/L at 180 min, $p < 0.001$) significantly changed during reperfusion. PAP, pulmonary artery pressure; Ppeak, peak inspiratory pressure. One-way repeated measures ANOVA; Tukey's multiple comparison test. $p < 0.05$: a = vs. 0 min; b = vs. 30 min; c = vs. 90 min; d = vs. 120 min; e = vs. 150 min.

2.2. Protein Profiles of Frozen Tissue Lung Lobes

All of the thirty protein lists of biological and technical replicate analyses were aligned through an in-house Multidimensional Algorithm Protein Map (MAProMa) tool [29,30] and grouped by calculating, for each distinct protein, the identification frequency and the average Spectral Counts (SpC*) and Score (Score*), which are useful parameters for relative abundance protein assessment in label-free quantitative approaches. In fact, SpC represents the total number of MS/MS spectra assigned to each protein and, consequently, it reflects protein relative abundance in each analyzed condition, while Score value is another semi quantitative parameter calculated by the integration of several search engine variables. Overall, 1895 distinct proteins were identified: 1588, 1616 and 1330 for native, pre- and post-EVLP, respectively. A complete list of the identified proteins is reported in Supplementary Table S2. Figure S1a) shows the proportional Venn diagram of specific proteins for each condition. In particular, 1160 proteins were identified in all conditions, while 205 proteins were identified in both native and pre-EVLP, 75 proteins were common to pre- and post-EVLP, and 39 proteins were identified in both native and post-EVLP conditions. Finally, a total of 184, 176 and 56 distinct proteins were exclusively detected in native, pre- and post-EVLP, respectively. Using the MAProMa software it was also possible to plot each experimental condition list on a 2D virtual map, according to the theoretical Molecular Weight (MW) and Isoelectric point (pI) of identified proteins. An example is shown in Figure S1b), where the protein list of a post-EVLP sample was mapped. For each protein, a color/shape code was used according to the confidence of identification.

2.3. Data Handling

2.3.1. Repeatability

Technical and biological repeatability of the method was tested by plotting SpC values of lists previously aligned with MAProMa; specifically, each technical replicate showed linear correlation (R2) and slope close to the theoretical value of 1 (Figure 3 and Supplementary Table S3). Figure 4a reports the technical repeatability of a sample from the native group. Similarly, the biological repeatability was verified (Supplementary Table S3 and Figure 3). As shown in Figure 4b, the average SpC of sample IIIM21 (X axis) was compared with that of IIIN23 (Y axis), both belonging to the post-EVLP condition: R2 = 0.946 and y = 0.962. These findings demonstrated the good repeatability of the proteomic approach adopted, because the reduced oscillations of SpC values are a robust starting point for subsequently label-free investigations on protein profiles. Furthermore, similar results were also reported for all the analyzed samples (data not shown), except for sample IE9 (belonging to the post-EVLP condition) which shows a lower biological repeatability compared to the other samples in the same condition. An example of low biological repeatability with sample ID7 was reported in Figure 4c. For this reason sample, IE9 was further investigated and only 30 proteins (out of 1187) resulted significantly scattered in IE9 (compared to the other post-EVLP samples); in particular, these nonlinear proteins were mainly related to plasma proteins. Of note, deleting the plasma related proteins identified in the IE9 sample the correlation with the other post-EVLP samples improved (Figure 4d). For these reasons, the IE9 sample was taken into account for label-free quantitation and network analysis. The proteomics finding in the IE9 sample were in agreement with its different agonal phase as reported in Section 2.1. The different behaviors of the IE9 sample (both biological and molecular) could be correlated to possible lesions due to the supplemental manipulation (double cannulation required) that occurred to IE9 during the surgical procedure.

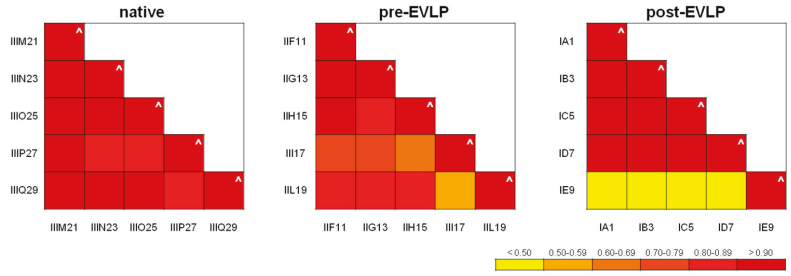


Figure 3. Technical and biological repeatability of proteomic analysis. Diagrams of correlation coefficient (R2) values of technical and biological replicates. Each R2 value of technical and biological replicates was assigned a color code, and the corresponding ranges associated to a chromatic scale were reported. The technical R2 value (^) was obtained comparing the spectral count (SpC) values from two runs of the same sample. Similarly, plotting the average SpC* values of two samples belonging to the same condition the biological R2 value was calculated. The numerical linear regression values (R2) for each comparison are reported in Supplementary Table S3.

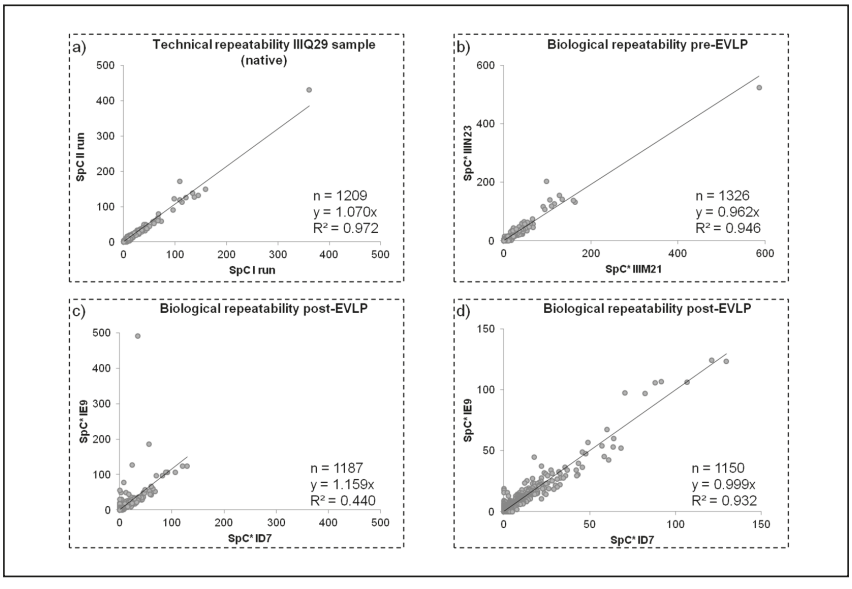


Figure 4. Examples of technical and biological repeatability of proteomic analysis. (**a**) Typical technical repeatability (IIIQ29 sample, native group). (**b**) Biological repeatability between two native samples (IIIM21 and IIIN23). (**c**) Biological repeatability of IE9 sample compared to ID7 (post-EVLP samples). (**d**) Biological repeatability of IE9 sample compared to ID7 without the scattered proteins (mainly plasma proteins). SpC*: Average spectral count value.

2.3.2. Clustering and Linear Discriminant Analyses

Protein lists obtained by the proteomics analyses of the three experimental conditions analyzed were processed by Linear Discriminant Analysis (LDA) [35] and 451 proteins with F ratio > 3.4 and p-value < 0.05 were selected as descriptors of features able to correctly discriminate the three groups (Supplementary Table S4). An oriented dendogram was obtained using all the average SpC protein values of the fifteen animals analyzed. As shown in Figure 5, all post-EVLP samples resulted segregated into a group and separated from native and pre-EVLP samples. Of note, observing the post-EVLP

grouping more in detail, the IE9 sample, while remaining within the same cluster, shows a slightly different trend than the other animals in the group and is localized on a separate branch, much closer to the pre-EVLP and native ones.

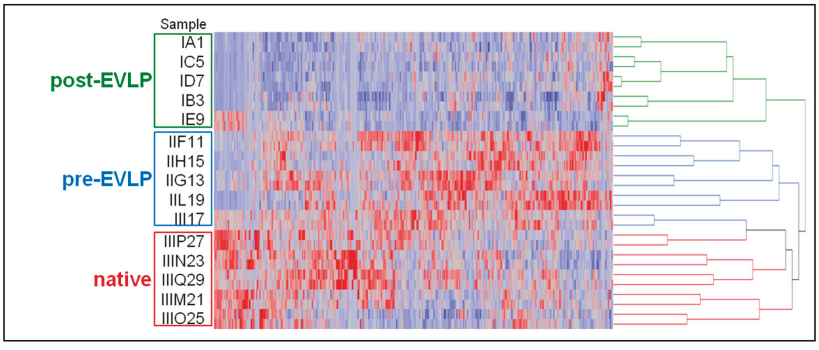


Figure 5. Hierarchical clustering obtained by Linear Discriminant Analysis (LDA). Using the average SpC values from all conditions, the post-EVLP samples were correctly segregated and 451 descriptor proteins by LDA were identified. Of note, IE9 sample, although remaining within the correct subgroup, is localized on a separate branch. The complete list of descriptor proteins extracted by LDA is reported in Supplementary Table S4. filtering criteria.

2.3.3. Differential Analysis

Independently from cluster analysis, a label-free evaluation of differentially expressed proteins was also performed using two parameters available in MAProMa software and comparing the average SpC values obtained for each identified protein in the three examined conditions. These parameters are DAve (Differential Average) and DCI (Differential Coefficient Index) representing the ratio and the confidence in differential expression, respectively, of a protein between two samples. They allowed the identification of 239 distinct differentially expressed proteins (DEP), using the filters described in Material and Methods section (Section 4.4.3). In particular, 31 and 77 proteins resulted respectively down- and upregulated in pre-EVLP compared to native. On the other hand, 180 resulted more abundant in pre-EVLP compared to post-EVLP, and only three proteins were upregulated in post-EVLP. Finally, 135 and 11 were down- and upregulated, respectively, in post-EVLP compared to native. Noteworthy, the combination of the data obtained by the differential evaluation with LDA and DEP analysis allowed the identification of 499 proteins, among which, 190 distinct proteins resulted common to the two methods applied. In fact, about 80% of DEP were confirmed by LDA. The complete list of differentially expressed proteins is reported in Supplementary Table S4, where proteins derived from the same gene were grouped.

2.4. Network Analysis

Proteins selected by LDA and DEP analyses were used to reconstruct a Protein–Protein Interaction (PPI) network of *Rattus norvegicus* involving 487 nodes and 7806 interactions for the investigation of the functional relationships among groups of proteins identified in our proteomic approach. Considered nodes were grouped into 25 distinct subnetworks based on their function and/or subcellular localization. These subnetworks were grouped into five distinct main groups: extracellular matrix, translation, cytoskeleton, cell cycle, and metabolism. The proteins displayed on the network are represented by a chromatic scale that reflects the relative abundance (SpC) of their expression. By means of Cytoscape [36] it is also possible to upload to the same metabolic network the results derived from the label-free differential analysis of pairwise comparison: native vs. pre-EVLP (Supplementary Figure S2a), pre-EVLP vs. post-EVLP (Supplementary Figure S2b), and native vs. post-EVLP (Supplementary Figure S2c). The majority of subnetworks resulted mainly up- and

downregulated in pre- and post-EVLP, respectively, compared to the native condition. Of note, some subnetworks, such as redox, PDIA (Protein Folding Drives Disulfide Formation), actin- and myosin-related proteins and vesicle-mediated transport seem unchanged in native vs. post-EVLP comparison; on the contrary, these subnetworks resulted up regulated in pre-EVLP.

To further investigate more the modifications of proteome in response to the EVLP treatment, it was calculated for each protein its variations (modules) due to perturbation (native vs. pre-EVLP) and treatment (pre-EVLP vs. post-EVLP). Theoretically, if the two modules present similar values but they have opposite sign, the protein recovers its level to the reference (in our case native condition). The extraction of so-called "recovered proteins" was simplified by calculating the Proteome Remodeling index (PRi) for each protein using an unbiased procedure. In this way, we extracted 104 proteins, and 32 had a statistically significant result (ANOVA $\leq$ 0.05, Supplementary Table S5). Proteome Activated index (PAi) was calculated using baseline and treatment modules to extract proteins which unchanged between native and pre-EVLP conditions (perturbation module), but resulted to be activated or inhibited (up- or downregulated, respectively) with treatment (post-EVLP condition). In this case 168 proteins were found, of which 67 resulted statistically significant by ANOVA ($\leq$0.05) (Supplementary Table S6). Of note, only two proteins resulted increased with the EVLP treatment. These extracted proteins were uploaded in the PPI network previously prepared and four subnetworks resulted to be enriched (Figure 6 and Figure S3): myosin-related proteins, actin-related proteins, and vesicle-mediated transport (belonging to cytoskeleton), and redox-processes (metabolism). Similar results were obtained by means of ln[FoldChange] (Supplementary Tables S5 and S6).

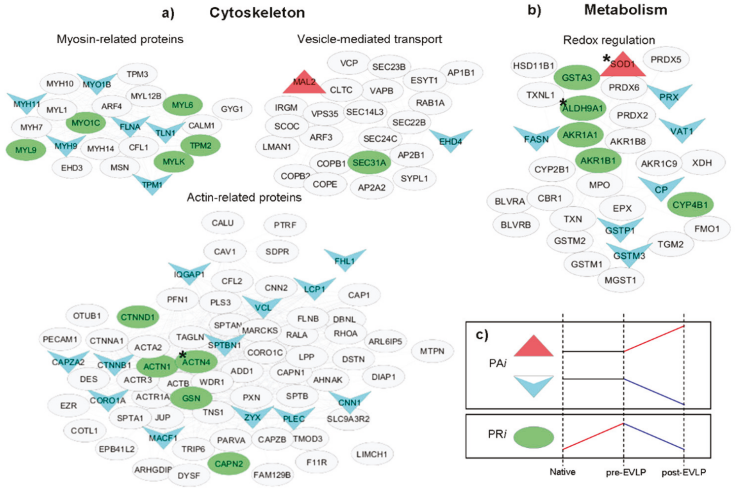


Figure 6. Main pathways involved in the proteome remodeling induced by the EVLP procedure. Enriched subnetworks based on recovered and activated proteins after ANOVA (<0.05) and Tukey's test (<0.05; * <0.08). These statistically significant proteins were previously extracted using an unbiased procedure that evaluated simultaneously the changes in the proteome in the three conditions: native, pre-EVLP, and post-EVLP. Recovered proteins, marked as a green/circle, were extracted using the Proteome Remodeling index (PRi); while activated proteins marked as red/triangle and blue/arrow for up- and downregulated in post-EVLP, respectively, were extracted using Proteome Activated index (PAi). The application of PRi and PAi filters highlights an enrichment in (**a**) vesicle-mediated transport, myosin-, and actin-related proteins, belonging to cytoskeleton and (**b**) redox regulation, belonging to metabolism. The blank nodes identify the protein descriptors resulted outside the applied PRi and PAi filters. (**c**) Summary of considered trends of remodeled proteins. The complete lists of PRi and PAi extracted proteins are reported in Supplementary Tables S5 and S6.

2.5. Comparison between Frozen and FFPE Tissue Lung Lobes

Shotgun proteomics approach was also employed to compare frozen and FFPE lung tissues. In particular, three FFPE samples from the pre-EVLP condition were analyzed. Using MAProMa software the three FFPE resulting proteomes were aligned with the corresponding frozen ones. The resulting twelve nanoLC-MS/MS runs allowed the identification of 1764 distinct proteins: 1458 and 1351 for frozen and FFPE, respectively (Table S7). As reported in Figure 7a more than 77% (1044) of FFPE proteins were confirmed in frozen specimens. Plotting the normalized average SpC values between the two datasets both R2 and slope were close to the theoretical value of 1 (Figure 7b). Finally, through the Protein ANalysis THrough Evolutionary Relationships (PANTHER) classification system [33] it was possible to classify proteins based on their molecular function. As shown in Figure 7c the distribution of proteins between the FFPE and frozen tissues was nearly identical, suggesting that the differences in proteins observed in the two datasets were not biased toward a specific molecular function.

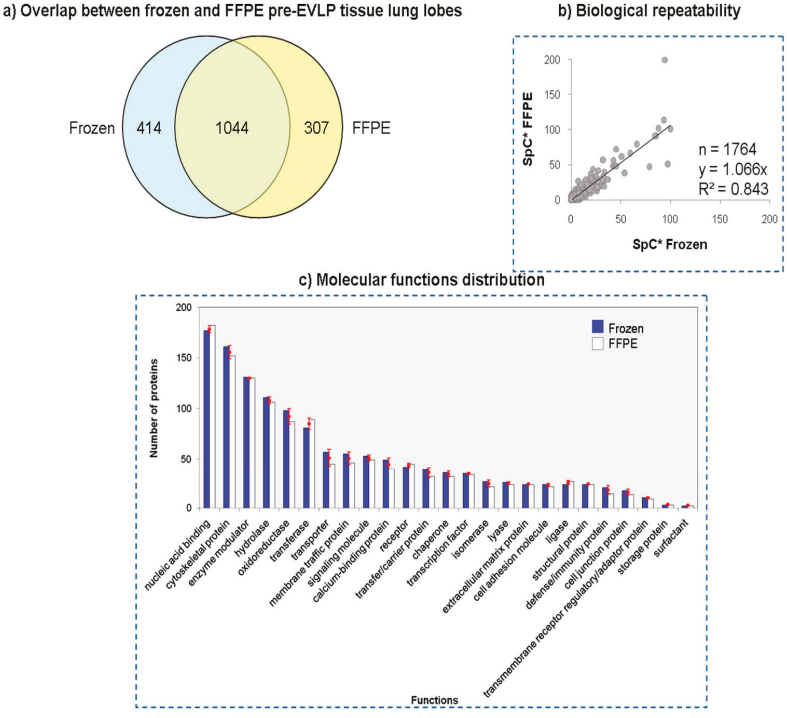


Figure 7. Comparison between frozen and FFPE pre-EVLP lung lobes. (**a**) Protein overlapping between frozen and FFPE in three pre-EVLP analyzed samples (IIL11, IIG13 and IIH15). (**b**) Plot of frozen vs. FFPE samples using normalized average spectral count (SpC*) values. (**c**) Distribution of proteins between the frozen and FFPE according to their molecular function using Protein ANalysis THrough Evolutionary Relationships (PANTHER) classification system. For each molecular function, reported as percentage of proteins for the two tissues (blue/solid colour for frozen and blank for FFPE datasets), the difference frozen vs. FFPE was <0.9%.

3. Discussion

Here, we employed a rat model and recent advances in MS-based proteomics to generate a proteomic system biology view of the EVLP procedure. The EVLP procedure cannot be solely evaluated on the basis of physiological measurements. In fact, analyses at the molecular level are mandatory to find potential protein markers for a better understanding of mechanisms by which EVLP

is ruled [18–20]. In light of this our approach represents a first attempt in this direction, with the intent to develop a shotgun label-free platform which could be helpful in extracting informative features from complex biospecimens in large scale analyses useful to correlate molecular expressions to biological processes. In particular, a nanoLC system based on chip technology was combined to a next-generation mass spectrometer with improved resolution and accurate mass detection in order to investigate protein extracts obtained from frozen lung tissues in three experimental groups (native, pre-EVLP and post-EVLP). This high throughput system was completely automatized and allowed the simultaneous identification of thousands of proteins with a wide range of concentrations and without limits of pI and MW [17,28] (see Figure 1).

Moreover, the large amount of data produced were properly processed with dedicated, commercial or in-house, bioinformatic tools using parameters provided by the employed search engines [29–33]. In our case, data handling was greatly simplified by means of MAProMa software [29,30] since it allowed the automatic alignment of different protein lists, presenting the results in a user-friendly format and highlighting differentially expressed proteins. In addition, data dimensional reduction with support vector machines was also applied to extract informative features useful for the correct classification and sample grouping using only information correlated to them and their relationships [35]. In this respect, the hierarchical clustering analysis performed by LDA and conducted on the fifteen average protein lists obtained by the three examined conditions, permitted to discern the differences among them and to correctly group samples in the three distinct categories (Figure 5). These findings indicate the high reproducibility and statistical reliability of the proposed approach. An additional confirmation derived from the technical and biological repeatability results obtained comparing SpC values of protein lists of the same sample injected in replicate and of different animals belonging to the same experimental group (Figure 3). In this case, our platform was able to point out differences in the IE9 sample with respect to the other biological replicates of the post-EVLP group. Specifically, the IE9 sample showed a very low biological repeatability in all the comparisons with animals of the same experimental group (Figure 3c). With the aim of further understanding the IE9 molecular variability, the protein profiles of this animal and its biological replicates were investigated. The results of differential analysis showed a similar pattern of IE9 with the other post-EVLP animals except for only a few proteins that resulted up regulated and essentially related to blood components. These findings were in agreement with the longer agonal phase of this animal herein reported and could be correlated to possible lesions due to the double cannulation incurred during the surgical procedure. However, this did not compromise the final functional outcome of the animal, which was not excluded from the study, but with its unique profile revealed the potentiality of our platform as confirmed by clustering analysis (Figure 5). In contrast to the inability of functional measurements to detect even small differences, clustering correctly maintained IE9 within the post-EVLP group but underlined its incomplete alignment with the other components of the same group, providing in this way a deeper and more accurate data evaluation strategy.

Another interesting feature that could be supported by our approach is the possibility to conduct a label-free quantitative investigation for the identification of potential biomarkers. Given the increasing popularity of EVLP treatment, the search for new molecular targets which could be employed in the future as objective parameters to recommend a protocol over others, to monitor the entire procedure and to predict lung performance is urgent. With this in mind, our study represents a first effort in this direction, as far as we know. In particular, LDA analysis showed the post-EVLP group completely segregated from the pre-EVLP and native conditions and provided the selection of 451 proteins whose trends were able to discriminate the three groups (Figure 5). Using an independent methodology, supplied by the application of DAve and DCI parameters included in MAProMa software [29,30], 246 distinct differentially expressed proteins were found with an overlap of 80% with those previously extracted by LDA.

At this point, the differentially expressed proteins resulting from both these computational tools were uploaded in a network built using Cytoscape [32] and with which it was possible to

visualize protein–protein interactions and metabolic pathways mainly involved. As reported in the Results section, the differentially expressed proteins were network visualizable with a color code that represented the DAve values resulting by the pairwise comparison of the three conditions (Supplementary Figure S2a–c). What emerged from a first observation was that EVLP treatment involved a downregulation of a great number of proteins reported in the metabolic pathways. If, from a functional point of view, both EVLP and native lungs seemed to achieve the same results in term of graft acceptability criteria, the study of molecular profiles suggests that this take place in different ways with the intervention of different pathways. Certainly, the experimental data permitted to exclude that this phenomenon was attributable to sample degradation due to a different processing time that the EVLP treatment required. In fact, the total number of identified proteins in the three conditions presented similar orders of magnitude and only 13% of the distinct identified proteins showed a differential expression (Supplementary Table S1).

Moreover, by means of proteome remodeling indexes, PR*i* and PA*i*, proteomic changes were simultaneously evaluated in pre-EVLP and in post-EVLP, where proteome expression in the latter seems to return to the reference condition (native group). After EVLP treatment, approximately 32 proteins resulted statistically significant in rebalancing the proteome to the reference condition of native lungs. Based on these findings it was possible to speculate that restored proteins might be correlated to the reconditioning effect exerted by EVLP. Of note, these proteins were always down regulated in post-EVLP compared to pre-EVLP; specifically, they resulted mainly involved in cytoskeleton (vesicle-mediated transport and myosin- and actin-related proteins) and in redox regulation (metabolism) (Figure 6). Another interesting class of proteins extracted by calculating the Proteome Activated index (PA*i*) did not show differential expression between native and pre-EVLP conditions, but increased or decreased their abundance after EVLP treatment. In particular only Mal2 and SOD1 increased their level, while 65 proteins decreased their level in post-EVLP compared to both native and pre-EVLP groups. Similarly to PR*i*-extracted proteins, those PA*i*-extracted resulted being mainly involved in cytoskeleton (vesicle-mediated transport and myosin- and actin-related proteins) and in redox regulation (metabolism).

To better understand these data and hypothesize possible mechanisms that govern the activation or suppression of metabolic pathways, we need to keep in mind what happens at a biochemical level during ex vivo lung perfusion. During the entire EVLP procedure, the lungs are subjected to a first phase of oxygen breakdown, for cold flushing and recruitment of organ (cold ischemia), and to a subsequent reperfusion by ex vivo perfusion/ventilation. The reduction of O_2 causes alterations in cellular metabolism and the switch from aerobic to anaerobic metabolism [37] with the subsequent reduction of intracellular ATP. Alteration of cell pH and generation of reactive oxygen species (ROS) [38,39] due to increase of Na^+ [40,41] and Ca^{2+} ions [42,43] are significant events that also occur. In addition, changes in cytoskeleton composition are signaled in endothelial cells with the activation of adhesion proteins [44] and others one involved in the maintenance of flexibility and adaptability of cytoskeleton during hypoxia [45].

As stated before, all these effects attributable to EVLP on different biomolecular pathways are easily observable in networks built with proteins highlighted for their differential expression and extracted by MAProMa, LDA and remodeling indexes. In particular, it is interesting to note that, among upregulated proteins in the post-EVLP condition, there are some belonging to the extracellular matrix (ECM), constituents of the basement membrane, and involved in inflammatory process and its regulation. Their secretion is stimulated by IL-1β and TNF-α [46,47] in order to repair the basement membrane, regenerating ECM around pulmonary fibroblasts and reconstituting epithelial tissue integrity after lung injury [47]. Of note, our coauthors of the Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico of Milan, in a work just published [22] and based on studies of gene expression on the same animal model samples, observed that IL-1β and TNF-α resulted upregulated in EVLP, due to lung injury resulting in the induction of ischemia by the surgical procedure, and activated the NF-kB complex. The occurrence of an inflammatory response together with other alterations at

the proteomic level in lungs, in a rat model designed without the use of anti-inflammatory drugs that could falsify results, is to be considered another small step towards the discovery of molecular mechanisms modulated by EVLP.

If on the one hand during riperfusion the activation of inflammatory processes leads to the expression of pro-inflammatory mediators (like IL-1β and TNF-α) [48,49] on the other hand stimulates the release of reactive oxygen species (ROS) [50–52]. Together with the involvement of redox regulation and heat shock response in our networks, the observed SOD1 increase during the EVLP treatment has just the purpose to decrease ROS levels and to neutralize ROS-dependent damage. This seems to perform a protective action to contrast the inflammatory effect triggered by the procedure itself and raises the possibility that EVLP donor lungs could be less immunogenic than standard ones. The regulation of described proteins can be considered in all respects an integral part of the mechanism of action of the EVLP treatment and it may explain at least in part its efficacy in restoring lung function, or, as suggested by Lonati et al. [22], the resistance to damage caused by ischemia/reperfusion occurrence after transplantation. Thus, these findings add new key information to the mechanism behind potential EVLP beneficial effects and encourage additional research to implement EVLP therapeutic interventions. If this is the case, unveiling the modes of action of EVLP on lung phenotype and its monitoring at the molecular level, may be crucial not only for improving the transplant procedure but also in providing a powerful tool in translational medicine and in the advancement of pulmonary physiology. In addition, the present study can provide useful information to implement ex vivo perfusion procedure of other solid organs.

In order to give a more complete view of our workflow potential, we performed a preliminary proteomic characterization of FFPE tissues. In particular, we analyzed three FFPE samples belonging to the pre-EVLP condition in order to compare their protein profiles with corresponding frozen tissues. Despite the low overlap between FFPE and frozen datasets (around 60%), R^2 and slope values plotting the normalized average SpC values of the two tissues resulted being close to 1 (Figure 7). This suggests that the differences in the two pools were only apparent and probably related to the effects of fixation. In fact, as proof of this, the molecular function distribution of the identified proteins between frozen and FFPE tissues was nearly identical and not biased toward a specific molecular function. The last finding opens up the possibility to work with FFPE tissues and to connect proteomic outputs with information on clinical outcomes.

Obviously, these results are still to be considered as preliminary and must be thoroughly investigated; however, these results offer the possibility to generate with a proteomic platform an overview of what happens at the molecular level during the entire transplant procedure with the ability to track the pattern of target proteins. In this regard, our approach facilitates the extraction of signature proteins from large datasets and their distribution in multiple pathways and quantitative traceability without the need of labeling, make these proteins ideal targets in the biomarker validation phase. In fact, the ongoing developments of MS-based proteomic technologies combined to increasingly powerful bioinformatics tools are accelerating their employment as a valid alternative to traditional assay methods for a number of features including: increased sensitivity, cost containment and simultaneous recording of multiple biomarkers.

Of course, our study represents a proof of principle with which we want to prove the potential of a shotgun proteomic platform in these investigations. However, animal model and sample size limit drawing any further conclusions on proteins and pathways discussed above. In the near future, it will be necessary to increase the cohort of study and translate to human specimens the experimental research in order to confirm the findings derived from animal models and to overcome possible drawbacks due to the incomplete overlap of data.

Another interesting aspect will be represented by the possibility of using other biospecimens, in addition to lung tissues, for minimizing invasiveness without compromising diagnostic accuracy. In this respect, exosomes and extracellular vesicles (EVs) are attracting great interest. As reported by Vallabhajosyula et al. [53] in a study on the perfusate from EVLP lungs, understanding EVs biology

may have important diagnostic and therapeutic implications in pulmonary medicine, expanding even more the potentiality of the EVLP model.

4. Materials and Methods

4.1. Experimental Proteomic Pipeline

Figure 1 shows a schematic overview of the protocol adopted to prepare frozen and FFPE tissue lung lobs prior to the nanoLC-MS/MS analysis. The protocol consists of different parts that will be described in detail below.

4.2. Surgical Procedure and Frozen and FFPE Tissue Collection

4.2.1. Animals and Experimental Groups

The experiments were performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health, at the Center for Preclinical Investigation, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. The experimental protocol was approved by the Italian Institute of Health (permit number: 1082-15).

Adult Sprague–Dawley male rats (Charles River, Calco, Lecco, Italy) weighing 270 to 330 g were housed in a ventilated cage system (Tecniplast S.p.A., Varese, Italy) at 22 ± 1 °C, 55 ± 5% humidity, on a 12 h dark/light cycle, and were allowed free access to rat chow feed and water ad libitum.

Rat were randomly assigned to one of the following experimental groups (n = 5 in each group). Following sacrifice of rats under general anesthesia, lungs were subjected to in situ cold flushing and were then retrieved (pre-EVLP group). In separate experiments, after in situ cold flushing, isolated lungs underwent ex vivo perfusion/ventilation for 180 min (post-EVLP group). Finally, in a third group lung biopsies were taken from animals in resting conditions (native group). Samples of perfusion fluid were collected at time intervals throughout EVLP procedure for gas analysis. Tissue biopsies were harvested to investigate lung edema and proteome. Specifically, right superior lobes were used for lung edema evaluation, right inferior lobes were snap-frozen and stored at −80 °C for proteomic analysis, and median lobes were added to 4% formalin at room temperature.

4.2.2. Ex Vivo Lung Perfusion (EVLP)

Anesthesia, in situ flushing and lung procurement were performed according to our previous work. Briefly, after sacrifice, lungs were flushed with ice-cold Perfadex® (XVIVO Perfusion AB, Gotebörg, Sweden). Thereafter, the heart-lungs block was procured and connected to an isolated lung perfusion system (Hugo Sachs Elektronik, Harvard Apparatus, GmbH, March-Hugsstetten, Germany) primed with in-house acellular solution with extracellular electrolyte composition [21]. No treatment with corticosteroids or other anti-inflammatory drugs was performed.

The protocol for rat ex vivo lung perfusion included a reperfusion phase (from 0 to 40 min) during which the lungs were gradually re-warmed, ventilated and perfused in order to avoid hemodynamic stress, and a steady-state phase (from 40 to 180 min). Specifically, initial pulmonary artery flow was set at 20% of target value and was then augmented over 40 min. Lung temperature was enhanced up to 37 °C in a 25 min time-span. Volume-controlled ventilation was started when the lungs reached a normothermic state, with respiratory rate of 35 bpm and pulmonary end expiratory pressure (PEEP) at 3 cmH2O. Initial tidal volume was set at 5 ml/kg and was increased up to 7 ml/kg in 10 min.

4.2.3. Lung Function

- Pulmonary artery and airway pressure assessment.

  PAP and Ppeak were continuously recorded, and digitally stored via an amplifier (data acquisition software Colligo Elekton, Milan, Italy).

- Gas analysis and determination of glucose and lactate concentration.

  Gas analysis, glucose, and lactate concentration evaluation were performed on perfusate samples using a gas analyzer (ABL 800 FLEX, A. De Mori Strumenti, Milan, Italy).

- Lung edema index.

  At the end of the procedure, the right superior lobe was weighed (KERN & SOHN GmbH, Balingen, Germany), before (wet) and after (dry) storage at 50 °C for 24 h (LTE Scientific, Oldham, UK). Wet-to-dry ratio (W/D) was then calculated and used as index for pulmonary edema.

4.3. Protein Extraction

4.3.1. Frozen Lobes

For proteomic analysis of frozen samples, each lobe was suspended in 300 µL 0.1 M H_4HCO_3 pH 7.9 buffer with 10% acetonitrile (Sigma-Aldrich Inc., St Louis, MO, USA) and homogenized in ice. The protein concentration was assayed using dotMETRIC™ (G-Biosciences, St. Louis, MO, USA) and the membrane proteins were solubilized by adding Rapigest™ SF reagent (Waters Co, Milford, MA, USA) at the final concentration of 0.33% (w/v). The resulting suspensions were incubated under stirring at 100 °C for 5 min. The digestion was carried out on 50 ± 0.5 µg proteins of each sample by adding Sequencing Grade Modified Trypsin (Promega Inc., Madison, WI, USA) at an enzyme/substrate ratio of 1:50 (w/w) overnight at 37 °C. An additional aliquot of 2 µg of trypsin (1:25 w/w) was added in the morning, and the digestion continued for 4 h. Moreover, the addition of 0.5% Trifluoroacetic acid (TFA) (Sigma-Aldrich Inc., St Louis, MO, USA) stopped the enzymatic reaction, and a subsequent incubation at 37 °C for 45 min completed the RapiGest acid hydrolysis [54]. The water immiscible degradation products were removed by centrifugation at 13,000 rpm for 10 min. Finally, the tryptic digest mixture was desalted using PepClean™ C-18 spin column (Pierce, Rockford, Il, USA) and was resuspended in 0.1% formic acid (Sigma-Aldrich Inc., St. Louis, MO, USA) in water. Water was deionized by passing through a Direct Q (Millipore, Bedford, MA, USA).

4.3.2. Formalin-Fixed Paraffin Embedded (FFPE) tissues

Paraffin-embedded tissue sections were dewaxed by means of xylene deparaffinization. Briefly, each lung biopsy was cut up to obtain five 10 µm-sections from the same block. Slices were placed in a 1.5 mL collection tube and added to xylene (Sigma-Aldrich Inc., St Louis, MO, USA). Following 10 min-incubation at room temperature, samples were centrifuged at 13,000 rpm for 2 min. Supernatants were discarded and pellets underwent two additional xylene washes. Next, samples were rehydrated using increasing ethanol dilutions (100%, 96%, and 70% ethanol, Sigma-Aldrich Inc., St Louis, MO, USA). Dry pellets were stored at −80 °C.

The dewaxed tissues from 3 pre-EVLP samples (IIF11, IIG13 and IIH15), were solubilized in Rapigest™ SF reagent at the final concentration of 0.25% (w/v) and incubated under stirring, first at 100 °C for 20 min and then sequentially at 80 °C for 2 h. Then the protein concentration was assayed using dotMETRIC™, and 24 ± 2.1 µg of proteins from each sample was digested by Sequencing Grade Modified Trypsin at an enzyme/substrate ratio of 1:25 (w/w) overnight at 37 °C. An additional aliquot of 0.5 µg of trypsin (1:12,5 w/w) was added in the morning, and the digestion continued for 4 h. The enzymatic digestion was stopped with the addition of 0.5% TFA, and a subsequent incubation at 37 °C for 45 min completed the RapiGest acid hydrolysis [54]. The water immiscible degradation products were removed by centrifugation at 13,000 rpm for 10 min. Finally, the tryptic digest mixture was desalted using PepClean™ C-18 spin column and was resuspended in 0.1% formic acid in water.

4.4. Proteomic Analysis

4.4.1. Liquid Chromatography

Trypsin digested mixture were analyzed using Eksigent nanoLC-Ultra® 2D System (Eksigent, part of AB SCIEX, Dublin, CA, USA) combined with cHiPLC®-nanoflex system (Eksigent) in trap-elute mode. Briefly, samples were first loaded on the cHiPLC trap (200 μm × 500 μm ChromXP C18-CL, 3 μm, 120 Å) and washed with the loading pump running in isocratic mode with 0.1% formic acid in water for 10 min at a flow rate of 3 μL/min. The automatic switching of cHiPLC ten-port valve then eluted the trapped mixture on a nano cHiPLC column (75 μm × 15 cm ChromXP C18-CL, 3 μm, 120 Å) through a 65 minute gradient of 5 to 40% of eluent B (eluent A, 0.1% formic acid in water; eluent B, 0.1% formic acid in acetonitrile) at a flow rate of 300 nL/min. Trap and column were maintained at 35 °C for retention time stability.

4.4.2. Mass Spectrometry

MS spectra were acquired using a QExactive mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA), equipped with an EASY-Spray ion source (Thermo Fisher Scientific). Easy spray was achieved using an EASY-Spray Emitter (Dionex Benelux BV, Amsterdam, The Netherlands) (nanoflow 7 μm ID Transfer Line 20 μm × 50 cm) held to 2.1 kV, while the ion transfer capillary was held at 220 °C. Full mass spectra were recorded in positive ion mode over a 400 to 1600 m/z range with a resolution setting of 70,000 FWHM (Full Width at Half Maximum) (@ m/z 200) with 1 microscan per second. Each full scan was followed by 7 MS/MS events, acquired at a resolution of 17,500 FWHM, sequentially generated in a data dependent manner on the top seven most abundant isotope patterns with charge $\geq$2, selected with an isolation window of 2 m/z from the survey scan, fragmented by higher energy collisional dissociation (HCD) with normalized collision energies of 30 and dynamically excluded for 30 sec. The maximum ion injection times for the survey scan and the MS/MS scans were 50 and 200 ms and the ion target values were set to 10^6 and 10^5, respectively.

The MS data have been deposited to the ProteomeXchange Consortium via the PRIDE [55] partner repository with the dataset identifier PXD006692.

4.4.3. Data Handling

All data generated were processed using the SEQUEST search engine [56] of BioWorks platform (version 3.3.1, University of Washington, licensed to ThermoFinnigan Corp., San Jose, CA, USA). The experimental MS/MS spectra were correlated to tryptic peptide sequences by comparison with theoretical MS/MS spectra obtained by in silico digestion of the Rattus norvegicus protein database (about 79609 entries), downloaded (January 2014) from the National Centre for Biotechnology Information (NCBI) website (www.ncbi.nlm.nih.gov). This allowed the identification of peptide sequences and related proteins. The confidence of protein identification, particularly when using data from a single peptide, was assured by applying stringent filtering criteria: the MS/MS tolerance was set to 1.00 Da; the precursor ion tolerance was set to 50 ppm (part-per million); and the intensity threshold was set to 100. Moreover, searches were performed with no enzyme. Finally, to assign a final score to the proteins, the SEQUEST output data were filtered by setting the peptide probability to 1×10^{-3}, the chosen minimum correlation score values (Xcorr) were 1.5, 2.0, 2.5, and 3.0 for single-, double-, triple-, and quadruple-charged ions, respectively, and the consensus score higher than 10 (FDR < 0.01). The output data obtained from the SEQUEST software were aligned, compared and grouped by calculating the average Spectral Count (SpC*) and Score values for each distinct protein using an *in-house* algorithm, namely, the Multidimensional Algorithm Protein Map (MAProMa) [29,30].

4.4.4. Label-Free Differential Analysis

Differential analysis was performed by comparing the average Spectral count (SpC*) values of each protein, obtained from native, pre-EVLP and post-EVLP lungs, using Dave and DCI algorithm

inserted in the MAProMa software. DAve is an index of the relative ratio between the two compared protein lists and DCI is an index to evaluate the confidence of DAve. Most confident uprepresented proteins in lung tissues showed a DAve $\geq$ +0.2 and a DCI $\geq$ +10; most confident downrepresented proteins showed a DAve ≤ -0.2 and DCI ≤ -10.

4.4.5. Hierarchical Clustering and Linear Discriminant Analysis

Biological replicates were evaluated using an unsupervised learning method, such as hierarchical clustering (HC) [57] using in-house R-scripts, which was based on the XlsReadWrite, clue, and clValid libraries (http://cran.r-project.org). In particular, the Euclidean distance metric was applied, and an Agglomerative Coefficient was calculated (Agglomerative Nesting—AGNES). AGNES, which as additional information, yields the AC that measures the amount of clustering structures found. AC is a dimensionless quantity, which could vary between 0 (bad structure) and 1 (very clear structuring). In order to perform multivariate analysis, all protein lists were processed using Linear Discriminant Analysis (LDA) [35]. This analysis was applied by using a common covariance matrix for all the groups and the Mahalanobis distance [58] from each point to each group's multivariate mean (proteins derived from the same gene were grouped). The identifying features (proteins) that discriminated the analyzed samples were considered to be those with the largest F ratio (>3.4) and the smallest *p*-value (<0.05). Specifically, the F ratio represented the model mean square divided by the error mean square, whereas the p-value indicated the probability of obtaining an F value greater than that calculated if, in reality, there was no difference between the population group means.

4.4.6. Network Analysis

Starting from the list of experimentally identified proteins, the corresponding Rattus norvegicus Protein-Protein Interaction (PPI) network was extracted. By means of a Cytoscape plug-in, STRING 8 database (http://string.embl.de/) [36], known interactions were retrieved from several databases such as Prolink, DIP, KEGG, BIND and others. All types of interactions were retrieved from each repository without applying a *p*-value threshold. Protein–DNA, protein–RNA, protein–metabolite, and protein–drug interactions, if present in the data sets, were removed, as were duplicates and self-interactions. PPI were examined using Cytoscape 3.2.1 (http://www.Cytoscape.org) [32] by applying a score >0.15. Protein–DNA, protein–RNA, protein–metabolite, and protein–drug interactions, if present in the data sets, were removed, as were duplicates and self-interactions. PPI were processed using Cytoscape 3.2.1 (http://www.cytoscape.org/) [32]. Proteins that were not mapped or were mapped as isolated components were not considered in the analysis. Bingo 2.44 [59], a Cytoscape plug-in, was used to emphasize subnetworks based on functionally organized GO terms, and the MCODE plug-in [60] was used to cluster subnetworks based on their topology and, specifically, by considering densely connected regions. Further selection of nodes was made by means of evaluation of baseline, perturbation and treatment modules, using Dave or ln[FoldChange] values to calculate PR*i* and PA*i* parameters.

4.4.7. Proteome Recovery and Proteome Activation Index

For each protein it was evaluated its variations (modules) due to perturbation (native vs. pre-EVLP) and/or treatment (pre-EVLP vs. post-EVLP); if the two modules present similar values but opposite sign, the protein recovers its level to the reference (in our case native condition). The extraction of so-called "recovered proteins" was simplified by calculating the Proteome Remodeling index (PR*i*) for each protein using an unbiased procedure. The PR*i* is calculated by the following formula,

$$\mathrm{PR}_i = \frac{\mathrm{M(b-p)_J}}{\mathrm{M(p-t)_J}}$$

where J: each identified protein. M(b-p)$_J$: Perturbation Module = Difference native vs. pre-EVLP (b: baseline; p: perturbed); differential abundance of specific protein J comparing baseline vs. perturbed (pre-EVLP) conditions. M(p-t)$_J$: Treatment Module = Difference pre-EVLP vs. post-EVLP (t:treated); differential abundance of specific protein J comparing perturbed (pre-EVLP) vs. treated (post-EVLP) conditions.

Theoretically, if protein level after treatment remodeled to reference (native) the two modules (perturbation and treatment) should have similar value and opposite sign; then the PR*i* will be negative and close to unit (-1). In our case, we considered proteins with a PR*i* in the range -0.5 to -2.

Also, using a variant of PR index, called Proteome Activated index (PA*i*), it was extracted unchanged proteins between native and pre-EVLP conditions (perturbation module), but activated or inhibited (up- or downregulated, respectively) with treatment (post-EVLP condition). The related formula, based on baseline and treatment modules, is

$$\mathrm{PA}_i = \frac{\mathrm{M(t-b)_J}}{\mathrm{M(p-t)_J}} \sim -1$$

M(t $-$ b)$_J$: Baseline Module = Difference post-EVLP vs. native (t: treated condition, post-EVLP); differential abundance of specific protein J comparing treated (post-EVLP) vs. baseline conditions.

Differential abundance of each module may be expressed as DAVE value, from MAProMa algorithm or ln[fold change]). However, because in some case the fold-change calculation returns non-sense values (such as n/0 or 0/n), we preferred to use DAVE value to calculate modules. Moreover, MAProMa platform permits the filtration by absolute variation, using DCI algorithm, excluding very low expressed protein (very low spectral count, multiple DCI $< |5|$) confusingly with noise.

4.5. Statistics

Sample size was determined considering a statistical test power of 0.80 and an alpha value of 0.05. Results indicated that a sample size of 15 animals (n = 5 per group) would enable to detect a minimum difference in protein expression of 0.35 with an expected standard deviation of 0.15.

To detect outliers, Grubb's test was applied for each parameter. A probability value <0.05 was considered significant. Baseline characteristics and trends of PAP, Ppeak, glucose, and lactate during ex vivo perfusion were analyzed by means of one-way analysis of variance (ANOVA) followed by Tukey's multiple comparison test. All analyses were performed using Sigma Stat 11.0 dedicated software (Systat Software Inc., San Jose, CA, USA).

Identified protein were evaluated by LDA (F ratio > 3.4 and a p-value < 0.05) [61] and MAProMa (confidence thresholds were set as DAve $\geq |0.2|$ and a DCI $\geq |10|$) platforms.

Finally, proteins extracted by PR*i* and PA*i* algorithms were statistically evaluated by ANOVA and Tukey's test.

Supplementary Materials: The supplementary materials are available online.

Author Contributions: Conceptualization, S.G., F.V. and P.M.; Data Curation, C.L., A.D.P. and V.R.; Formal analysis, V.R. and D.D.; Funding acquisition, F.B. and P.M.; Investigation, V.R. and S.G.; Methodology, V.R., A.D.P. and R.R.; Software, V.R., A.D.P. and D.D.S.; Supervision, S.G. and P.M.; Validation, C.L. and M.M.; Writing—original draft, A.D.P. and P.M.

Funding: This work was partially supported by the Italian CNR FLAGSHIPs Project and AMANDA project, funded by the MIUR "InterOmics" and CNR-Lombardy Agreement, respectively; also, 5x1000 MIUR 2014, "Laboratory animal welfare in preclinical research".

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Briot, R.; Gennai, S.; Maignan, M.; Souilamas, R.; Pison, C. Ex vivo lung graft perfusion. *Anaesth. Crit. Care Pain Med.* **2016**, *35*, 123–131. [CrossRef] [PubMed]

2. Andreasson, A.S.I.; Dark, J.H.; Fisher, A.J. Ex vivo lung perfusion in clinical lung transplantation—State of the art. *Eur. J. Cardio-Thorac. Surg.* **2014**, *46*, 779–788. [CrossRef] [PubMed]
3. Popov, A.F.; Sabashnikov, A.; Patil, N.P.; Zeriouh, M.; Mohite, P.N.; Zych, B.; Saez, D.G.; Schmack, B.; Ruhparwar, A.; Dohmen, P.M.; et al. Ex vivo lung perfusion—State of the art in lung donor pool expansion. *Med. Sci. Monit. Basic Res.* **2015**, *21*, 9–14. [PubMed]
4. Levin, K.; Kotecha, S.; Westall, G.; Snell, G. How can we improve the quality of transplantable lungs? *Expert Rev. Respir. Med.* **2016**, *10*, 1–7. [CrossRef] [PubMed]
5. Valenza, F.; Rosso, L.; Coppola, S.; Froio, S.; Palleschi, A.; Tosi, D.; Mendogni, P.; Salice, V.; Ruggeri, G.M.; Fumagalli, J.; et al. Ex vivo lung perfusion to improve donor lung function and increase the number of organs available for transplantation. *Transpl. Int.* **2014**, *27*, 553–561. [CrossRef] [PubMed]
6. Van Raemdonck, D.; Neyrinck, A.; Cypel, M.; Keshavjee, S. *Ex-vivo* lung perfusion. *Transpl. Int.* **2015**, *28*, 643–656. [CrossRef] [PubMed]
7. Makdisi, G.; Makdisi, T.; Jarmi, T.; Caldeira, C.C. Ex vivo lung perfusion review of a revolutionary technology. *Ann. Transl. Med.* **2017**, *5*, 343–349. [CrossRef] [PubMed]
8. Hood, L.; Friend, S.H. Predictive, personalized, preventive, participatory (P4) cancer medicine. *Nat. Rev. Clin. Oncol.* **2011**, *8*, 184–187. [CrossRef] [PubMed]
9. Jameson, J.L.; Longo, D.L. Precision medicine–personalized, problematic, and promising. *N. Engl. J. Med.* **2015**, *372*, 2229–2234. [CrossRef] [PubMed]
10. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* **2001**, *69*, 89–95. [CrossRef] [PubMed]
11. Corvol, H.; Thompson, K.E.; Tabary, O.; le Rouzic, P.; Guillot, L. Translating the genetics of cystic fibrosis to personalized medicine. *Transl Res.* **2016**, *168*, 40–49. [CrossRef] [PubMed]
12. Herazo-Maya, J.D.; Kaminski, N. Personalized medicine: Applying 'omics' to lung fibrosis. *Biomark Med.* **2012**, *6*, 529–540. [CrossRef] [PubMed]
13. Pang, M.; Bai, X.Y.; Li, Y.; Bai, J.Z.; Yuan, L.R.; Ren, S.A.; Hu, X.Y.; Zhang, X.R.; Yu, B.F.; Guo, R.; et al. Label-free LC-MS/MS shotgun proteomics to investigate the anti-inflammatory effect of rCC16. *Mol. Med. Rep.* **2016**, *1*, 4496–4504. [CrossRef] [PubMed]
14. Rossi, R.; De Palma, A.; Benazzi, L.; Riccio, A.M.; Canonica, G.W.; Mauri, P. Biomarker discovery in asthma and COPD by proteomic approaches. *Proteomics Clin. Appl.* **2014**, *8*, 901–915. [CrossRef] [PubMed]
15. Rauniyar, N.; Gupta, V.; Balch, W.E.; Yates, J.R. Quantitative proteomic profiling reveals differentially regulated proteins in cystic fibrosis cells. *J. Proteome Res.* **2014**, *13*, 4668–4675. [CrossRef] [PubMed]
16. Lässer, C.; O'Neil, S.E.; Shelke, G.V.; Sihlbom, C.; Hansson, S.F.; Gho, Y.S.; Lundbäck, B.; Lötvall, J. Exosomes in the nose induce immune cell trafficking and harbour an altered protein cargo in chronic airway inflammation. *J. Transl. Med.* **2016**, *14*, 181–194. [CrossRef] [PubMed]
17. Mauri, P.; Riccio, A.M.; Rossi, R.; Di Silvestre, D.; Benazzi, L.; De Ferrari, L.; Dal Negro, R.W.; Holgate, S.T.; Canonica, G.W. Proteomics of bronchial biopsies: Galectin-3 as a predictive biomarker of airway remodelling modulation in omalizumab-treated severe asthma patients. *Immunol. Lett.* **2014**, *162*, 2–10. [CrossRef] [PubMed]
18. Mohamed, M.S.A. Ex Vivo Lung Perfusion and Transplant: State of the Art and View to the Future. *Exp. Clin. Transplant.* **2015**, *13*, 493–499. [PubMed]
19. Mohamed, M.S.A. Translational Insights on Lung Transplantation: Learning from Immunology. *Iran. J. Immunol.* **2015**, *12*, 156–165. [PubMed]
20. Reeb, J.; Cypel, M. Ex vivo lung perfusion. *Clin. Transplant.* **2016**, *30*, 183–194. [CrossRef] [PubMed]
21. Bassani, G.A.; Lonati, C.; Brambilla, D.; Rapido, F.; Valenza, F.; Gatti, S. Ex Vivo Lung Perfusion in the Rat: Detailed Procedure and Videos. *PLoS ONE* **2016**, *11*, e0167898. [CrossRef] [PubMed]
22. Lonati, C.; Bassani, G.A.; Brambilla, D.; Leonardi, P.; Carlin, A.; Faversani, A.; Gatti, S.; Valenza, F. Influence of ex vivo perfusion on the biomolecular profile of rat lungs. *FASEB J.* **2018**, *32*, 5532–5549. [CrossRef] [PubMed]
23. Link, A.J.; Eng, J.; Schieltz, D.M.; Carmack, E.; Mize, G.J.; Morris, D.R.; Garvik, B.M.; Yates, J.R. Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **1999**, *17*, 676–682. [CrossRef] [PubMed]
24. Wolters, D.A.; Washburn, M.P.; Yates, J.R. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **2001**, *73*, 5683–5690. [CrossRef] [PubMed]

25. Mauri, P.; Scarpa, A.; Nascimbeni, A.C.; Benazzi, L.; Parmagnani, E.; Mafficini, A.; Della Peruta, M.; Bassi, C.; Miyazaki, K.; Sorio, C. Identification of proteins released by pancreatic cancer cells by multidimensional protein identification technology: A strategy for identification of novel cancer markers. *FASEB J.* **2005**, *19*, 1125–1127. [CrossRef] [PubMed]
26. McDonald, W.H.; Yates, J.R. Shotgun proteomics: Integrating technologies to answer biological questions. *Curr. Opin. Mol. Ther.* **2003**, *5*, 302–309. [PubMed]
27. Chait, BT. Chemistry. Mass spectrometry: Bottom-up or top-down? *Science* **2006**, *314*, 65–66. [CrossRef] [PubMed]
28. Brambilla, F.; Lavatelli, F.; Di Silvestre, D.; Valentini, V.; Rossi, R.; Palladini, G.; Obici, L.; Verga, L.; Mauri, P.; Merlini, G. Reliable typing of systemic amyloidoses through proteomic analysis of subcutaneous adipose tissue. *Blood* **2012**, *119*, 1844–1847. [CrossRef] [PubMed]
29. Mauri, P.; Dehò, G. A proteomic approach to the analysis of RNA degradosome composition in Escherichia coli. *Methods Enzymol.* **2008**, *447*, 99–117. [PubMed]
30. Comunian, C.; Rusconi, F.; De Palma, A.; Brunetti, P.; Catalucci, D.; Mauri, P.L. A comparative MudPIT analysis identifies different expression profiles in heart compartments. *Proteomics* **2011**, *11*, 2320–2328. [CrossRef] [PubMed]
31. Di Silvestre, D.; Brunetti, P.; Vella, D.; Brambilla, F.; De Palma, A.; Mauri, P. Automated Extraction of Proteotypic Peptides by Shotgun Proteomic Experiments: A New Computational Tool and Two Actual Cases. *Curr. Biotechnol.* **2015**, *4*, 39–45. [CrossRef]
32. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef] [PubMed]
33. Thomas, P.D.; Campbell, M.J.; Kejariwal, A.; Mi, H.; Karlak, B.; Daverman, R.; Diemer, K.; Muruganujan, A.; Narechania, A. PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res.* **2003**, *13*, 2129–2141. [CrossRef] [PubMed]
34. Gustafsson, O.J.; Arentz, G.; Hoffmann, P. Proteomic developments in the analysis of formalin-fixed tissue. *Biochim. Biophys. Acta* **2015**, 559–580. [CrossRef] [PubMed]
35. Hilario, M.; Kalousis, A. Approaches to dimensionality reduction in proteomic biomarker studies. *Brief. Bioinform.* **2008**, *9*, 102–118. [CrossRef] [PubMed]
36. Pawłowski, K.; Muszewska, A.; Lenart, A.; Szczepińska, T.; Godzik, A.; Grynberg, M. A widespread peroxiredoxin-like domain present in tumor suppression- and progression-implicated proteins. *BMC Genom.* **2010**, *11*, 590. [CrossRef] [PubMed]
37. Kosieradzki, M.; Rowiński, W. Ischemia/reperfusion injury in kidney transplantation: Mechanisms and prevention. *Transplant. Proc.* **2008**, *40*, 3279–3288. [CrossRef] [PubMed]
38. Gustafsson, A.B.; Gottlieb, R.A. Hearth mitochondria: Gates of life and death. *Cardiovasc Res.* **2008**, *77*, 334–343. [CrossRef] [PubMed]
39. Peng, T.I.; Jou, M.J. Oxidative stress caused by mitochondrial calcium overload. *Ann. N. Y. Acad. Sci.* **2010**, 183–188. [CrossRef] [PubMed]
40. Kako, K.; Kato, M.; Matsuoka, T.; Mustapha, A. Depression of membrane-bound Na+-K+-ATPase activity induced by free radicals and by ischemia of kidney. *Am. J. Physiol.* **1988**, *254*, C330–C337. [CrossRef] [PubMed]
41. Kato, M.; Kako, K.J. Effects of N-(2-mercaptopropionyl)glycine on ischemic-reperfused dog kidney in vivo and membrane preparation in vitro. *Mol. Cell. Biochem.* **1987**, *78*, 151–159. [CrossRef] [PubMed]
42. Roberts, B.N.; Christini, D.J. NHE inhibition does not improve Na(+) or Ca(2+) overload during reperfusion: Using modeling to illuminate the mechanisms underlying a therapeutic failure. *PLoS Comput. Biol.* **2011**, *7*, e1002241. [CrossRef] [PubMed]
43. Sanada, S.; Komuro, I.; Kitakaze, M. Pathophysiology of myocardial reperfusion injury: Preconditioning, postconditioning, and translational aspects of protective measures. *Am. J. Physiol Heart Circ. Physiol.* **2011**, *301*, 1723–1741. [CrossRef] [PubMed]
44. Ostergaard, L.; Simonsen, U.; Eskildsen-Helmond, Y.; Vorum, H.; Uldbjerg, N.; Honoré, B.; Mulvany, M.J. Proteomics reveals lowering oxygen alters cytoskeletal and endoplasmatic stress proteins in human endothelial cells. *Proteomics* **2009**, *9*, 4457–4467. [CrossRef] [PubMed]

45. Veith, C.; Schmitt, S.; Veit, F.; Dahal, B.K.; Wilhelm, J.; Klepetko, W.; Marta, G.; Seeger, W.; Schermuly, R.T.; Grimminger, F.; et al. Cofilin, a hypoxia-regulated protein in murine lungs identified by 2DE: Role of the cytoskeletal protein cofilin in pulmonary hypertension. *Proteomics* **2013**, *13*, 75–88. [CrossRef] [PubMed]
46. Elias, J.A.; Freundlich, B.; Adams, S.; Rosenbloom, J. Regulation of human lung fibroblast collagen production by recombinant interleukin-1, tumor necrosis factor, and interferon-gamma. *Ann. N. Y. Acad. Sci.* **1990**, *580*, 233–244. [CrossRef] [PubMed]
47. Furuyama, A.; Hosokawa, T.; Mochitate, K. Interleukin-1beta and tumor necrosis factor-alpha have opposite effects on fibroblasts and epithelial cells during basement membrane formation. *Matrix Biol.* **2008**, *27*, 429–440. [CrossRef] [PubMed]
48. Rahman, I.; Marwick, J.; Kirkham, P. Redox modulation of chromatin remodeling: Impact on histone acetylation and deacetylation, NF-kappaB and pro-inflammatory gene expression. *Biochem. Pharmacol.* **2004**, *68*, 1255–1267. [CrossRef] [PubMed]
49. Chiang, C.H.; Pai, H.I.; Liu, S.L. Ventilator-induced lung injury (VILI) promotes ischemia/reperfusion lung injury (I/R) and NF-kappaB antibody attenuates both injuries. *Resuscitation* **2008**, *79*, 147–154. [CrossRef] [PubMed]
50. Eckenhoff, R.G.; Dodia, C.; Tan, Z.; Fisher, A.B. Oxygen-dependent reperfusion injury in the isolated rat lung. *J. Appl. Physiol.* **1992**, *72*, 1454–1460. [CrossRef] [PubMed]
51. Puskas, J.D.; Hirai, T.; Christie, N.; Mayer, E.; Slutsky, A.S.; Patterson, G.A. Reliable thirty-hour lung preservation by donor lung hyperinflation. *J. Thorac. Cardiovasc. Surg.* **1992**, *104*, 1075–1083. [PubMed]
52. Date, H.; Matsumura, A.; Manchester, J.K.; Obo, H.; Lima, O.; Cooper, J.M.; Sundaresan, S.; Lowry, O.H.; Cooper, J.D. Evaluation of lung metabolism during successful twenty-four-hour canine lung preservation. *J. Thorac. Cardiovasc. Surg.* **1993**, *105*, 480–491. [PubMed]
53. Vallabhajosyula, P.; Korutla, L.; Habertheuer, A.; Reddy, S.; Schaufler, C.; Lasky, J.; Diamond, J.; Cantu, E. Ex Vivo Lung Perfusion Model to Study Pulmonary Tissue Extracellular Microvesicle Profiles. *Ann. Thorac. Surg.* **2017**, *103*, 1758–1766. [CrossRef] [PubMed]
54. Nomura, E.; Katsuta, K.; Ueda, T.; Toriyama, M.; Mori, T.; Inagaki, N. Acid-labile surfactant improves in-sodium dodecyl sulfate polyacrylamide gel protein digestion for matrix-assisted laser desorption/ionization mass spectrometric peptide mapping. *J. Mass Spectrom.* **2004**, *39*, 202–207. [CrossRef] [PubMed]
55. Vizcaíno, J.A.; Csordas, A.; del-Toro, N.; Dianes, J.A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; et al. 2016 update of the PRIDE database and related tools. *Nucleic Acids Res.* **2016**, *44*, D447–D456. [CrossRef] [PubMed]
56. Eng, J.K.; McCormack, A.L.; Yates, J.R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989. [CrossRef]
57. Zhao, Y.; Karypis, G. Data clustering in life sciences. *Mol. Biotechnol.* **2005**, *31*, 55–80. [CrossRef]
58. Jain, A.K.; Murty, M.N.; Flynn, P.J. Data Clustering: A Review. *ACM Comput. Surv.* **1999**, *31*, 264–323. [CrossRef]
59. Maere, S.; Heymans, K.; Kuiper, M. BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **2005**, *21*, 3448–3449. [CrossRef] [PubMed]
60. Bader, G.D.; Hogue, C.W.V. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **2003**, *4*, 2. [CrossRef]
61. Di Silvestre, D.; Brambilla, F.; Scardoni, G.; Brunetti, P.; Motta, S.; Matteucci, M.; Laudanna, C.; Recchia, F.A.; Lionetti, V.; Mauri, P. Proteomics-based network analysis characterizes biological processes and pathways activated by preconditioned mesenchymal stem cells in cardiac repair mechanisms. *Biochim. Biophys. Acta* **2017**, *1861*, 1190–1199. [CrossRef] [PubMed]

Sample Availability: Not available.

molecules

MDPI

Article

Construction of a Quantitative Acetylomic Tissue Atlas in Rice (*Oryza sativa* L.)

Zhiyong Li [1], Yifeng Wang [1], Babatunde Kazeem Bello [1], Abolore Adijat Ajadi [1], Xiaohong Tong [1], Yuxiao Chang [2,*] and Jian Zhang [1,*]

[1] State Key Lab of Rice Biology, China National Rice Research Institute, Hangzhou 311400, China; lzhy1418@163.com (Z.L.); wangyifeng@caas.cn (Y.W.); tunlapa2k3@yahoo.com (B.K.B.); threetriplea@yahoo.com (A.A.A.); tongxiaohong@caas.cn (X.T.)

[2] Agricultural Genomes Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China

* Correspondence: Changyuxiao@caas.cn (Y.C.); zhangjian@caas.cn (J.Z.); Tel.: +86-571-6337-0277 (J.Z.)

Received: 9 October 2018; Accepted: 31 October 2018; Published: 1 November 2018

Abstract: PKA (protein lysine acetylation) is a key post-translational modification involved in the regulation of various biological processes in rice. So far, rice acetylome data is very limited due to the highly-dynamic pattern of protein expression and PKA modification. In this study, we performed a comprehensive quantitative acetylome profile on four typical rice tissues, i.e., the callus, root, leaf, and panicle, by using a mass spectrometry (MS)-based, label-free approach. The identification of 1536 acetylsites on 1454 acetylpeptides from 890 acetylproteins represented one of the largest acetylome datasets on rice. A total of 1445 peptides on 887 proteins were differentially acetylated, and are extensively involved in protein translation, chloroplast development, and photosynthesis, flowering and pollen fertility, and root meristem activity, indicating the important roles of PKA in rice tissue development and functions. The current study provides an overall view of the acetylation events in rice tissues, as well as clues to reveal the function of PKA proteins in physiologically-relevant tissues.

Keywords: Rice (*Oryza sativa* L.); protein lysine acetylation; proteome; tissue atlas; post-translational modification

1. Introduction

PKA (Protein lysine acetylation) refers to the substitution of an acetyl group for an active hydrogen atom on the lysine residues of a protein. Three types of proteins are required to catalyze the reversible PKA reaction. As acetylation "writers", lysine acetyltransferases (KATs) catalyze the addition of acetyl groups from acetyl-coenzyme A (acetyl-CoA) to proteins, whereas the reversible deacetylation process is conducted by the "erasers" enzyme lysine deacetylases (KDACs). Proteins containing conserved bromodomain (BRD) or YEATS domain were found to play the roles of PKA "readers" (acetyllysine binders), as they can selectively interact with acetylated proteins [1]. Since its first discovery on histones in over 50 years ago, PKA has been implicated for the functionality of their target proteins in various cellular processes [2]. Histone acetylation has been associated with chromatin remodeling and transcription activation, because a negatively-charged acetyl group could neutralize the positive charges of lysine residues, which weakens the interaction of the histone with negatively charged DNA, and consequently leads to a more relaxed chromatin structure for transcription. Conversely, the reversible histone deacetylation usually results in a tighter interaction with DNA, leading to chromatin condensation and transcription repression [3]. Recently, knowledge regarding PKA has been extended to non-histone proteins, particularly key metabolic enzymes related to glycolysis, tricarboxylic acid

(TCA) cycle in different organisms, as well as photosynthesis in plants [4–7]. The status and intensities of PKA may impose profound effects on the function of non-histone proteins by altering their enzyme activity, cellular compartment localization, protein-nucleotide/protein-protein interaction, and protein stability [3,8]. For example, the inhibition of PKA on tumor protein p53 is believed to be a cause of cervical cancer in human [9]. In Arabidopsis, Lee et al. (2015) revealed that the *P. syringae* type-III effector HopZ3, which is a YopJ type acetyltransferase, suppresses plant immune system by acetylating multiple members of the RPM1 immune complex and its triggering effectors [10].

As the first step toward understanding PKA, identification of PKA sites and dynamics is crucial. Aided by the technologies of acetylpeptides immune affinity purification and nano-HPLC/MS/MS, Kim et al. (2006) reported the first proteome-wide profiling of PKA in HeLa cells and mouse liver mitochondria. This screening identified 388 PKA sites on 195 proteins, which dramatically extended the known inventory of in vivo acetylation sites and substrates [11]. So far, the information of 111253 PKA sites on 33025 PKA proteins from various species have been deposited into the PLMD (Protein Lysine Modifications Database, http://plmd.biocuckoo.org/download.php) [12]. In comparison with the tremendous progress achieved in human, mouse, fungi, and bacterium, PKA identification in plants is lagging behind. Until 2011, Finkemeier et al. reported the first plant acetylomic analysis in the dicot model species Arabidopsis. They revealed the extensive involvement of PKA in regulating the activity of central metabolic enzymes such as Rubisco, phosphoglycerate kinase, glyceraldehyde 3-phosphate dehydrogenase, and malate dehydrogenase [13]. Nevertheless, only two reports are presently available on Arabidopsis, which profiled a total of 398 PKA sites on 251 proteins from suspension cells and young seedlings [13,14]. Similar works have also been done on grape fruits, pea seedlings, soybean developing seeds, wheat leaf, strawberry leaf, grass leaf, potato tuber, and spruce somatic embryo, but only yielded the identification of less than 7000 PKA sites in total [6,15–21].

Rice (*Oryza sativa* L.) is one of the most important food crops, as it serves as a staple food for over half of the global population. On the other hand, rice is also a model species for biological research due to its relatively small genome size, released genome sequence, ample genetic resources, as well as the co-linearity with other grasses [22]. PKA has been recognized as a key mechanism in regulating photosynthesis and metabolism in plants. For example, acetylation of the Lhcb1 and Lhcb2 proteins appear to be involved in determining the LHC attachment to PSII complexes. Higher PKA level on Lhcbs impaired their binding ablity to PSII to form the PSII-LHCII supercomplexes [23]. This also intrigued researchers to explore PKAs in rice. So far, at least 8 independent cases have been reported regarding the profiling of PKA sites, peptides, and proteins in rice seedlings, reproductive organs, and leaves under oxidative stress, which yielded the identification of over 8599 PKA sites and 4990 proteins [7,24–30]. Despite efforts being made on rice, the previous studies only explored the tip of the iceberg of rice acetylome, given that rice genome contains over 56,000 protein coding genes [31]. Due to the spatial- and temporal-pattern of protein expression, investigating PKA in various tissues has been considered as an effective way to expand the coverage of rice acetylome. In this study, we performed a quantitative, MS-based identification of PKA proteins in four rice tissues, i.e., the callus, root, leaf, and panicle, and obtained 890 PKA proteins covering 1536 sites, which allowed us to construct a tissue atlas of rice acetylome. The results obtained from this work aimed to provide an overall view of the acetylation events in rice tissues, as well as clues to reveal the function of PKA proteins in physiologically-relevant tissues.

2. Results

2.1. Profiling the Acetylsites and Acetylproteins on Various Rice Tissues

PKA is a highly transient and reversible post-translational modification. To achieve an overview of PKAs in rice, we conducted a comprehensive, quantitative acetylomic profiling in 4 tissues of Nipponbare (*Oryza sativa* L. *ssp japonica*), i.e., callus, leaves, panicles, and roots (Figure 1A–C), by employing a label-free, MS-based approach. Prior to the MS identification, we performed western

blot analysis on the tissue proteins using a pan anti-acetylation antibody. The protein acetylation levels and patterns varied divergently from tissues, and the majority of the acetylated proteins are non-histone proteins with a molecular weight over 25 KD (Figure 1D,E). MS identification yielded the data of 1164, 1105, 921, and 428 acetylsites on 1089, 1034, 848, and 385 acetylpeptides, which represented 682, 664, 547, and 263 acetylproteins from the callus, leaves, panicles and roots, respectively (Figure 2A, Tables 1 and S1). One hundred and seventy acetylproteins were commonly profiled in all the four tissues, while 97, 93, 27, and 44 acetylproteins were specifically detected in the callus, leaves, panicles, and roots, respectively (Figure 2B). By integrating the data from four tissues and removing redundancies, a total of 1454 acetylpeptides from 890 acetylproteins, covering 1536 acetylsites, were obtained. The mass errors for those acetylpeptides were near 0, with the majority less than 5 ppm, implying a high level of accuracy of the data (Figure S1). The length of most of the identified acetylpeptides ranged from 7–22 amino acids with a few longer exceptions that even reached up to 32 amino acids (AAs) (Figure 2C). Over 60% of the acetylpeptides carried only 1 acetylsite, around 20% of the acetylpeptides covered 2 or 3 acetylsites, while the remaining 10% acetylpeptides had more than 3 acetylsites (Figure 2D).

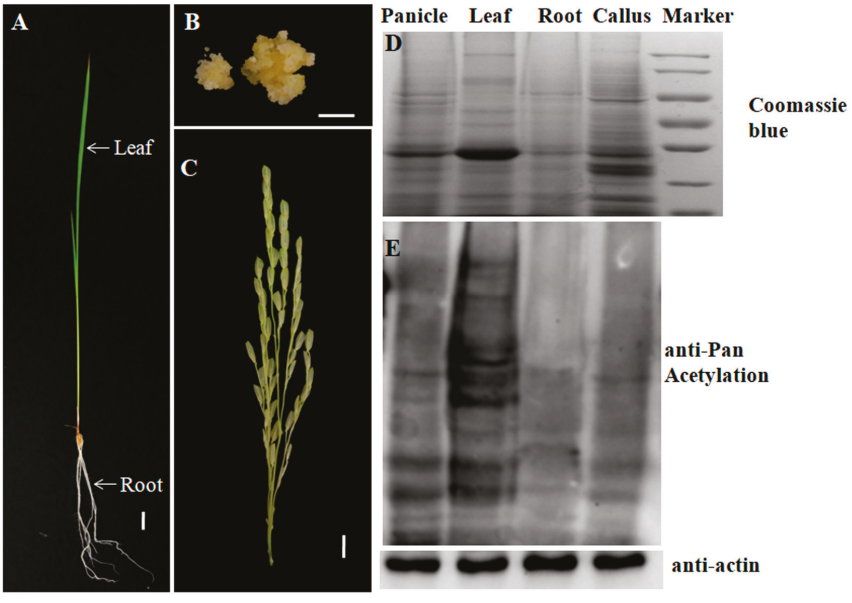


Figure 1. Tissue morphologies of different rice tissues tested in this study. (**A**) leaf and root; (**B**) callus; (**C**) mature panicle. Scale bar = 1 cm; (**D**) Total proteins of different tissues were resolved by SDS-PAGE and stained by Coomassie brilliant blue (CBB) and (**E**) Western blot analysis of the acetylation dynamics in different rice tissues by using anti-acetyl lysine antibodies. Equal amount of proteins (20 μg) were used. Anti-actin was used as an internal control for normalization [32].

Table 1. Some of the selected examples of the identified acetylproteins.

Identification				Intensity on Average			
Protein Accession	**Protein Names**	**Modified Sequence**	**Protein Annotation**	**Cal**	**Lea**	**Pan**	**Root**
Q9XJ60	MADS50	_LEALETYK(ac)R_	MADS-box transcription factor 50			1	
Q10N21	APX1	_PLVEK(ac)YAADEK_	L-ascorbate peroxidase 1, cytosolic	1.9	1.93	1.32	0.5
Q8W3D9	PORB	_ELLADLTSSDYPSK(ac)R_	Protochlorophyllide reductase B		1		
Q69RJ0	Fd-GOGAT1	_TDILK(ac)AK_	Ferredoxin-dependent glutamate synthase	0.38	1.83		
P30298	SUS2	_IYEK(ac)YTWK_	Sucrose synthase 2	0.78	3.83	0.82	1.41
Q0IMS5	WSL12	_NVVHGSDSPDNGK(ac)R_	Nucleoside diphosphate kinase	0.82	2.27		
Q2RAK2	OsPK1	_VFNQDLYFK(ac)R_	Pyruvate kinase	2.41		0.6	1.46
Q5JK84	API5	_DFLLK(ac)PELLR_	DEAD-box ATP-dependent RNA helicase			1.57	
P31924	SUS1	_IEEK(ac)YTWK_	Sucrose synthase 1	1.48	0.91	2.1	1.09
Q93X08	UGP	_TNPSNPSIELGPEFK(ac)K_	UDP-glucose pyrophosphorylase protein	1.38	2.21	0.79	1.56
A3C6G9	GDCSH	_YTK(ac)HCEEEDAH_	Glycine cleavage system H protein		0.97	0.03	
Q53LQ0	PDIL1-1	_SPEDATNLIDDK(ac)K_	Protein disulfide isomerase-like 1-1	0.55	2.84		1.09
Q0D9D0	SBE1	_CLIEK(ac)HEGGLEEFSK_	Glycoside Hydrolase	1.29	0.78	0.83	
Q43009	SUS3	_AEK(ac)HLAGITADTPYSEFHHR_	Sucrose synthase 3		2.08		
Q5Z8Y4	USP	_THGAISEFVNPK(ac)YTDSTK_	UDP-sugar pyrophosphorylase	1.79	3.59	1.07	
Q69V57	OsAld-Y	_(ac)SAFVGK(ac)YADELIK_	Fructose-bisphosphate aldolase	2.1	1.32	0.34	
Q8W1L6	AIM1	_YTK(ac)HCEEEDAH_	Peroxisomal fatty acid beta-oxidation	5.16	2.57	1.24	
Q7GD79	RAN2	_LTYK(ac)NVPTWHR_	GTP-binding nuclear protein Ran-2	0.96	1.8	1.16	

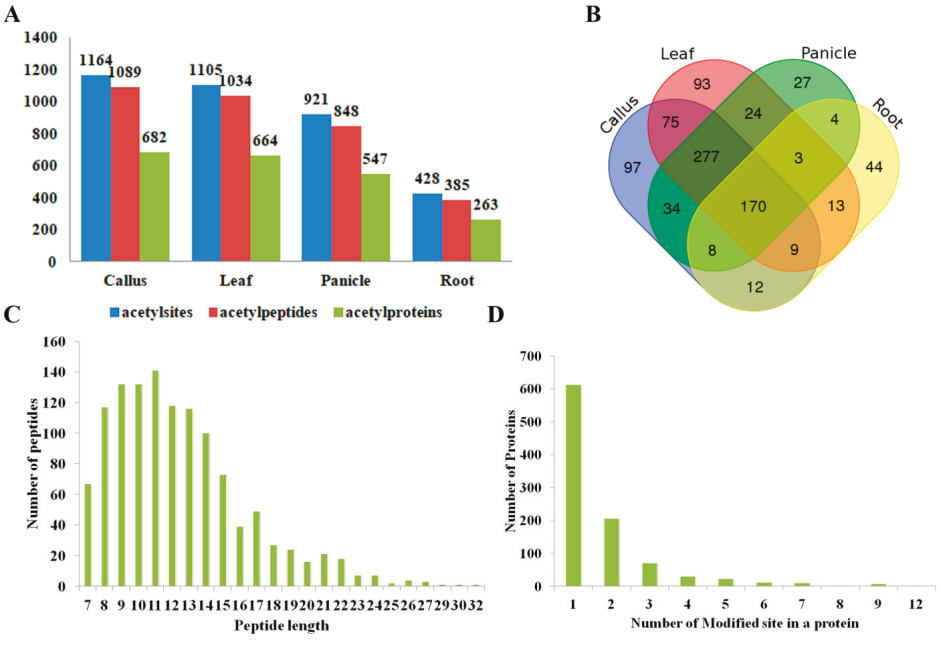
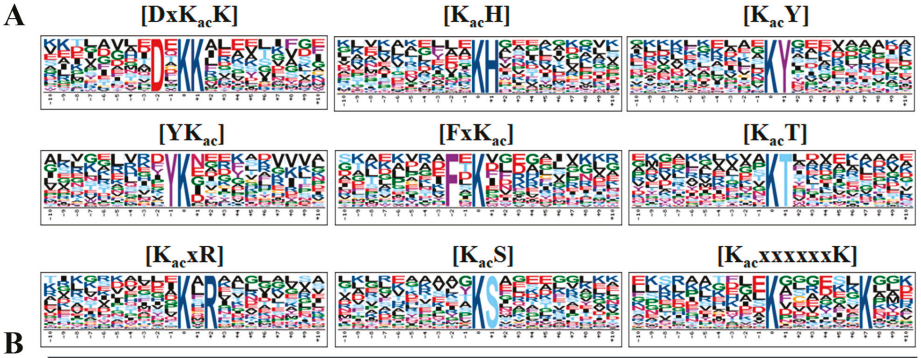


Figure 2. (**A**) The counts of acetylsites, acetylpeptides and acetylproteins in the callus, leaf, panicle, and root, respectively; (**B**) Venn diagram showing the overlap of our identified acetylproteins in the callus, leaf, panicle, and root, respectively; (**C**) Numbers of each identified peptide length; (**D**) Numbers of each identified modified site in a protein.

2.2. Motif Analysis of the Acetylsite Flanking Sequences

Employing the motif-X algorithm (http://motif-x.med.harvard.edu/) [33], we searched the potential conserved motifs in the context of 21 AA-long sequences centered by the acetylsites. Nine of the most enriched motifs are depicted in Figure 3A (Fold increase > 2.5, P < 0.01). Positively-charged residues like K and H are greatly favored by PKA, as we found the top three enriched consensus motifs were [DxK$_{ac}$K,K$_{ac}$H] and [K$_{ac}$K]. There were 132, 118, and 73 acetylsites located in [K$_{ac}$xR,K$_{ac}$S] and [K$_{ac}$T] motifs respectively. Meanwhile, motifs [YK$_{ac}$] [FxK$_{ac}$] [KxxxxxxK$_{ac}$] also showed over 2.5 fold enrichment when compared with the rice proteome background. In addition to the conserved motifs, we calculated the frequency of each AA type in the positions flanking the acetylsites. As revealed in Figure 3B, the distribution of AAs displayed strong bias in certain positions. For example, there was significantly higher chance of finding a H, K, R, or Y in the +1 position to acetylsites, while residues like M, I, G, and E were barely detected in the same position.



Amino	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	K	1	2	3	4	5	6	7	8	9	10
A	1.01	-0.08	0.05	-0.10	1.77	0.58	0.07	0.19	5.86	5.44	0.00	-2.47	-0.02	0.45	0.77	0.95	0.05	1.07	2.66	0.29	1.71
C	-0.50	-1.54	-2.63	-0.63	-0.85	-0.68	0.27	0.30	2.04	-1.13	0.00	-0.39	-0.14	-0.03	-0.78	-0.91	-1.31	-0.38	-2.18	-1.75	-0.19
D	-0.10	0.23	-0.13	-0.78	-0.67	1.59	-0.43	0.74	6.16	0.63	0.00	-5.34	-1.45	0.47	1.08	0.31	-0.80	-0.56	0.12	0.57	-0.02
E	-0.09	0.82	0.45	1.44	1.94	-0.05	4.21	2.74	2.17	9.86	0.00	-11.83	-0.32	1.49	1.92	2.18	-0.43	0.72	-0.14	-0.28	3.64
F	-0.70	-0.41	-0.16	0.19	0.55	0.22	-1.87	3.05	4.53	4.86	0.00	0.70	-0.10	-0.07	0.00	0.00	0.35	0.41	0.74	0.10	-0.19
G	2.03	-0.65	0.16	0.42	-0.39	0.29	0.83	0.46	0.21	2.27	0.00	-9.77	0.00	0.20	1.82	0.59	1.10	1.98	0.52	1.59	0.54
H	-0.57	-0.09	-0.03	0.87	0.81	-0.35	3.38	1.57	1.33	-3.24	0.00	-40.42	0.59	0.99	-0.64	0.20	-0.16	-1.39	-0.51	-0.85	-0.03
I	-0.25	0.28	-0.94	0.22	-0.59	0.21	-0.35	2.67	1.26	0.46	0.00	-10.12	0.08	-0.19	0.31	-0.09	2.30	0.75	-0.76	0.00	-0.06
K	2.29	15.84	7.62	5.32	1.74	0.85	-0.87	-3.87	-15.40	-29.81	0.00	10.48	20.96	1.00	2.71	1.26	2.23	7.68	2.29	5.00	2.65
L	-0.71	-0.57	-0.70	0.44	-0.95	-0.57	0.06	-0.08	0.29	-0.25	0.00	-12.41	-0.75	-0.41	-0.85	-1.14	0.35	-1.86	-0.06	0.00	-1.64
M	0.20	-0.56	0.20	-0.96	-0.63	-0.95	-1.58	-1.15	-1.50	-3.98	0.00	-3.20	-0.09	-0.45	-1.27	-1.33	-0.85	-2.28	-0.03	-0.46	-0.35
N	-0.07	-0.48	-1.39	-0.05	0.29	0.05	-0.22	0.31	-0.99	-0.92	0.00	-0.98	-0.59	-0.52	-0.19	-0.16	-1.10	-0.26	-0.13	-1.09	0.25
P	-0.24	-0.30	-1.58	-1.54	-1.52	-0.16	-0.46	0.11	-1.76	0.04	0.00	-4.37	-2.41	-1.91	-1.29	-1.65	-0.63	0.00	-0.77	-0.63	-0.72
Q	-0.48	-0.88	-0.22	-1.23	-0.54	0.16	1.14	-0.47	-1.73	0.53	0.00	-1.76	-1.32	-0.25	-0.57	-1.00	-0.29	-0.75	-0.57	-1.58	-0.47
R	0.41	0.41	0.34	0.96	-0.35	-1.79	-3.37	-11.52	-16.12	-33.22	0.00	30.12	10.10	4.35	-0.10	0.47	0.32	-0.29	-0.14	-0.14	-0.39
S	-4.36	-1.91	-0.34	-2.89	-0.76	-0.87	-0.90	-3.75	-3.85	-0.72	0.00	2.69	-6.53	-3.38	-0.61	-1.25	-2.47	-2.33	-2.50	-1.02	-2.07
T	0.63	-0.06	0.48	-0.06	1.41	1.18	0.72	-0.73	-1.14	1.07	0.00	0.04	-5.02	0.42	-0.44	-0.11	0.49	0.06	0.22	0.16	-0.31
V	0.37	0.12	1.37	-0.31	0.87	1.61	1.20	0.44	0.91	1.55	0.00	13.39	-1.49	-1.03	-0.56	0.99	0.63	-0.02	0.95	2.72	0.71
W	-1.95	-1.06	-1.87	-0.23	-2.30	-2.73	-0.27	-0.62	-1.08	-2.17	0.00	-0.51	-0.16	-1.06	-1.91	-2.32	-2.04	-0.77	-1.06	0.00	-0.77
Y	2.28	-0.30	0.15	-0.26	0.39	0.35	0.34	7.32	7.57	12.50	0.00	53.15	2.50	0.19	0.85	1.84	1.65	0.36	1.44	-0.72	0.12

Figure 3. (**A**) Acetylation sequence motifs and conservation of acetylation sites in identified callus, leaf, panicle acetylproteins; and (**B**) Heat map of the amino acid composition of the acetylation sites, showing the frequency of different amino acids surrounding the acetylated lysine (K).

2.3. Functional Features of Rice Acetylproteins

GO functional classification of all the acetylproteins were investigated in the terms of "biological process", "cellular component", and "molecular function". For the "biological process" category, the largest two types of acetylprotein was associated with "metabolic process" and "cellular process", which accounted for over half of the identified acetylproteins, suggesting a predominant role of PKA in metabolism. The remaining acetylproteins were related to processes such as "response to stimulus", "single-organism process", and "cellular component organization or biogenesis" (Figure 4A). Within the category of "cellular component", 30%, 24%, and 14% of the acetylproteins were annotated as "cell",

"organelle", and "membrane", respectively. Meanwhile, we found that a small portion of the proteins were relevant to "macromolecular complex", "cell junction", and so on (Figure 4B). Consistent with the highly-presented "metabolic process", the "molecular function" of the majority of the acetylproteins was on "binding" (47%) and "catalytic activity" (40%), while the remaining were likely to be involved in "structural molecule activity" and "transporter activity" (Figure 4C).

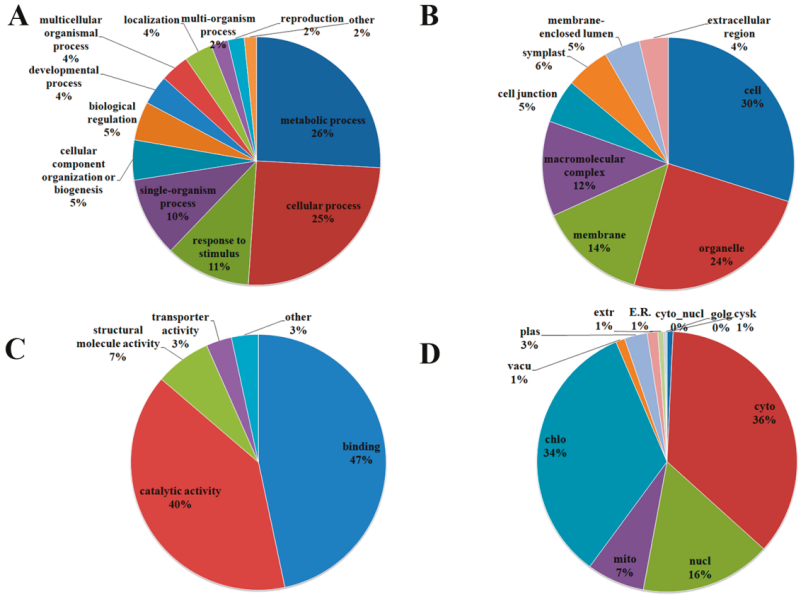


Figure 4. GO analysis of differentially-acetylated proteins in terms of: biological process (**A**); molecular function (**B**); cellular component (**C**); subcellular location (**D**); and acetylation intensity, respectively. The abbreviations in (**D**) represent the following. E.R.: endoplasmic reticulum; extr: extracellular matrix; cyto: cytoplasm; mito: mitochondrial; chlo: chloroplast; vacu: vacuole; cysk: cytoplasmic skeleton; cyto nucl: cytoplasm nuclear.

Eukaryotic cells are comprised of several membrane-bound subcellular compartments, such as the nucleus, cytoplasm, and mitochondria, where functional proteins play distinct roles [34]. The prediction of the subcellular localization helps to understand the function of the acetylproteins. We revealed that most of the acetylproteins were located to the cytosol (36%) and chloroplast (34%) (Figure 4D). Nuclear proteins such as histones and transcription factors are regarded as potential regulators of gene transcriptions. Sixteen percent of our identified acetylproteins were localized to nuclear, indicating that PKA is extensively involved in gene regulation. Some of the acetylproteins were predicted to be in mitochondria (7%), plasma membrane (3%), cytoskeleton (2%), and vacuole (1%) (Figure 4D).

Functional domains on a protein usually indicated its function in biological processes. In this study, we performed a PFAM functional domain enrichment assay to address the functional domain features of rice acetylproteins (Figure 5A). As a result, the most enriched proteins are core histones and histone-fold proteins, indicating PKA occurrs extensively on histones. Other enriched functional domains include NAD(P)-binding domain, single hybrid motif, aldolase-type TIM barrel, enhancer of polycomb-like and groEL-like equatorial domains, which are extensively involved in metabolism, and transcriptional and translational regulations.

We further conducted KEGG pathway enrichment analysis of the acetylproteins. The top 5 over-presented pathways are carbon metabolism, glycolysis, TCA cycle, biosynthesis of amino acides, and carbon fixation in photosynthetic organisms, which is in line with the functional domain analysis result that PKA occurs extensively on metabolism-related proteins (Figure 5B).

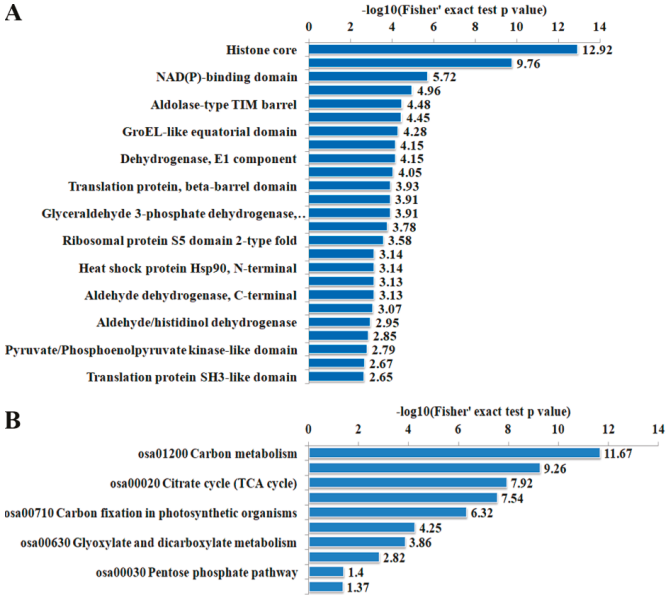


Figure 5. Protein domain enrichment analysis (**A**) and KEGG pathway enrichment analysis (**B**) proteins identified acetylproteins in this study.

2.4. Differentially Acetylated (DA) Peptides among Rice Tissues

Employing a label-free approach, we quantified the acetylation intensity in each peptide, and identified 1445 DA peptides on 887 proteins (Table S2). These included 125, 113, 35, and 53 peptides which were specifically acetylated in the callus, leaf, panicle, and root, respectively. In addition, we found 27 peptides were commonly acetylated in all the four tested tissues, with no significant variations in the acetylation intensity level, indicating the potential use as an internal control for PKA quantification experiments among rice tissues. There were also 82 acetylpeptides that were constitutively acetylated with significantly different intensities in the four tissues. The acetylation intensities of each DA proteins in various tissues were further visualized in a heatmap (Figure 6). The 1445 DA peptides were primarily divided into 12 clusters; each represented a divergent acetylation pattern from the others. For example, DA peptides in cluster 5 are predominantly acetylated in leaf, while cluster 9 contains peptides that are predominantly acetylated in the callus. In cluster 8, the peptides are highly acetylated in both the callus and leaf, suggesting the PKA on these peptides may regulate pathways that are common to the two tissues.

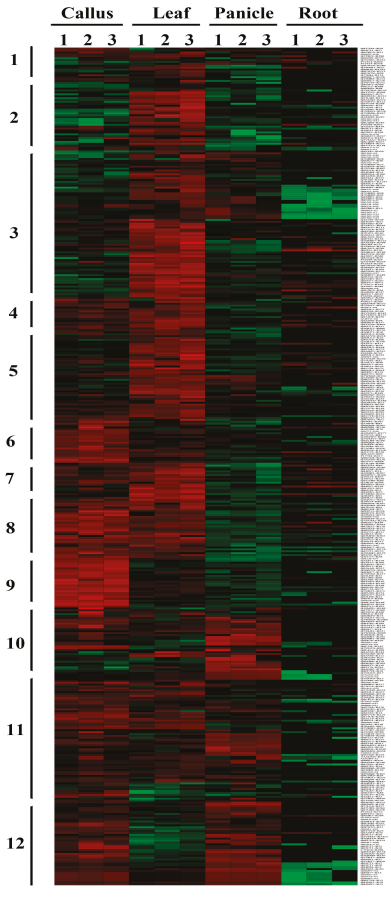


Figure 6. Hierarchical clustering analysis of the DA proteins in the callus, leaf, panicle, and root; Color bar at the bottom represents the log 2 acetylation site quantitation values. Green, black, and red indicate the low, medium, and high acetylation intensity, respectively.

2.5. Protein-Protein Interaction (PPI) Analysis of Acetylproteins

Using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins version 10.0; http://string-db.org/) and Cytoscape software (Cytoscape 3.7.0, Bethesda, Rockville, MD, USA), we constructed protein-protein interaction (PPI) networks of the specifically-acetylated proteins in the four tissues for a better understanding of the possible relationships among them. The callus acetylproteins form a profound PPI network containing 98 nodes (proteins) and 532 edges (interaction-ships) (Figure 7A; Table S3). In this network, we identified 10 ribosomal proteins, including 40S ribosomal protein S24, S9-2, 60S ribosomal protein L27, L36, as well as some ribosome recycling factors, which may form a ribosomal complex and take charge of the protein translation in callus. For the leaf acetylproteins, we acquired a network with 85 nodes and 465 edges (Figure 7B). Photosynthesis-related proteins were involved in this network. For examples, three chlorophyll A–B binding proteins, which serve as light receptors in the light-harvesting complex (LHC), may interact with each other. The network also contains carbohydrates assimilation proteins such as sucrose synthases and exo-beta-glucanase. In addition, we revealed an interaction module of gra(t)-OsAld-OsPORB, in which all members are functionally related to chloroplast development. In the network of panicle-acetylated proteins,

an actin protein was associated with rice flowering date regulator OsMADS50 and pollen fertility regulator OsAIP1, suggesting that PKA on these proteins plays important roles in flower development (Figure 7C). The network identified from root-acetylated proteins is majorly implicated in pentose metabolism, as a couple of 6-phosphogluconate dehydrogenases, dihydroxy-acid dehydratases, and glutamate dehydrogenases were involved in it (Figure 7D).

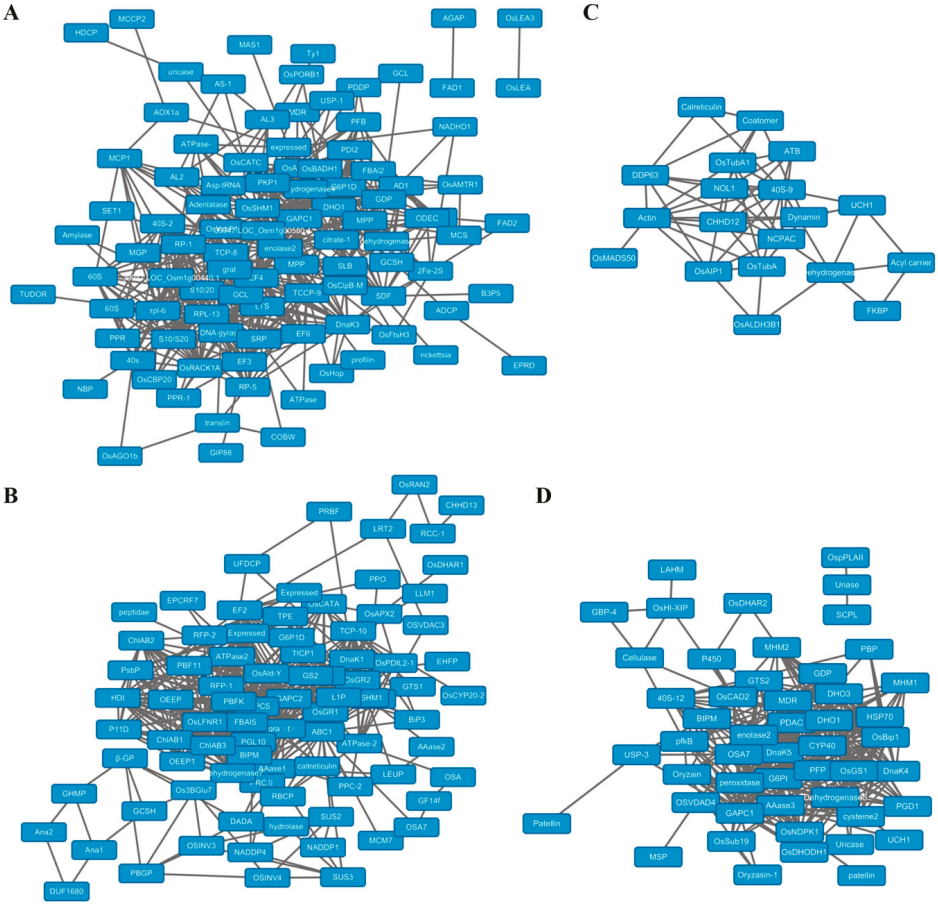


Figure 7. Protein–protein interaction (PPI) network of DA proteins identified in this study. (**A**) callus; (**B**) leaf; (**C**) mature panicle, and (**D**) root.

3. Discussion

3.1. PKA Profiling in Rice

Given the essential regulatory roles of PKA in plant growth and development, acetylomic profiling has become a hot topic in plant research. To date, at least 8 PKA identification cases have been reported in rice, as well as many in other plant species (Table 2). Due to its easy availability, vegetative tissues like leaf or whole seedlings have been repeatedly used for acetylomic identification. Xiong et al. (2016) profiled a total of 716 rice seedling PKA proteins, which were implicated in glyoxylate and dicarboxylate metabolism, carbon metabolism, and photosynthesis pathways [7]. Later, Xue et al. (2017) further identified 866 acetylproteins by using the same seedling tissue. However, only 21.3%

of the acetylproteins were overlapped with the previous study [29]. The authors deduced that the difference in identification is due to the physiological status or growth stages of the samples used. Zhou et al. (2018) also mapped 1024 acetylproteins from rice leaves under oxidative stress [30]. In rice germinating embryo, 699 PKA sites on 389 proteins were found to be acetylated, including 144 sites that were simultaneously modified by succinylation [26]. In an effort to discover the PKA in rice anthers in the stage of meiosis, Li et al. (2018) identified a total of 1354 PKA sites in 676 proteins, which are mostly related to chromatin silencing, protein folding, and fatty acid biosynthetic processes [27]. There were also 44 and 692 lysine acetylated proteins that were identified from suspension cells and grain-filling stage seeds [24,28]. With a special focus on the seed development, Wang et al. (2017) carried out a quantitative acetylproteomic study on pistil and early developing seeds. A total of 972 acetylproteins harboring 1817 acetylsites were identified, and 268 acetylproteins were differentially acetylated in the three developing stages, providing novel insight into PKA in rice seed development [25]. Nevertheless, rice acetylome is extremely complicated, largely due to the dynamic expression pattern of proteins in various tissues and physiological conditions. To gain a more complete coverage of the rice acetylome, the current study quantitatively profiled a total of 1454 acetylpeptides from 890 acetylproteins, covering 1536 acetylsites from callus, leaf, panicle, and root, which represented 4 major tissue types in rice. In comparison with the previous data, around 85% of our identified acetylprotiens were overlapped with other studies. A few examples are MADS50 (Q9XJ60), SUS1 (P31924), and USP (Q5Z8Y4). Meanwhile, we also found that 139 acetylproteins are novel, which were majorly identified from previously untested tissues such as the callus and panicle (Table S1). Our dataset greatly expanded the rice acetylome inventory, considering that callus and panicle are two unexplored tissues in terms of acetylome profiling. In contrast to previous studies that used the whole seedling for profiling, our dataset separated the tissues into leaf and root, thus presenting more detailed information of the spatial-distribution of the acetylproteins. Moreover, the quantified protein acetylation intensities also provided cues for functional characterization of the acetylproteins in the development of the corresponding tissues.

Table 2. Summary of the reported acetylome identification cases in rice and other plant species.

Species	Tissue	Treatment	Acetylsites	Acetylproteins	Reference
Oryza sativa	Callus, root, leaf and panicle		1536	890	This study
Oryza sativa	Seed		1003	692	[28]
Oryza sativa	Pistil and developing seeds		1817	972	[25]
Oryza sativa	Suspension cells		60	44	[24]
Oryza sativa	Young seedling		1337	716	[7]
Oryza sativa	Leaf	Oxidative stress	1669	1024	[30]
Oryza sativa	Seedling		1353	866	[29]
Oryza sativa	Anther		1354	676	[27]
Oryza sativa	Seed	Imbibition	699	389	[26]
Zea may	Seedling	Fungus	2791	912	[35]
Triticum aestivum	Leaf		416	277	[6]
Glycine max	Seed		400	245	[17]
Arabidopsis	Mitochondria of leaf		243	120	[14]
Arabidopsis	Chloroplast of leaf		91	74	[13]
Arabidopsis	Leaf	Deacetylase inhibitor	2057	1022	[36]
Arabidopsis	Root, leaf, shoot, flower and silique		64	57	[23]

3.2. Acetylproteins in Rice Callus

Rice callus is a mass of unorganized parenchyma cells which is capable of being regenerated in an individual plant. Callus cells have robust metabolisms and protein expressions to maintain their capability for differentiation. Ribosome is a complex molecular machine known as the translational apparatus for protein synthesis. Ribosome consists of two major components: small ribosomal subunits

for RNA recognition, and large ribosomal subunits for the synthesis of polypeptide chain using amino acid substrates. Early in the 1970s, PKA was found to occur on rabbit ribosomal proteins, and was implicated in the formation of the initiation complex during translation [37]. In yeast, protein N-terminal acetylation of ribosomal proteins by *N*-acetyltransferase is necessary to maintain protein synthesis [38]. It seems that PKA on ribosomal proteins might be a conserved mechanism, as we found that 41 ribosomal proteins were acetylated in rice callus. Interestingly, twelve of these proteins were predominantly acetylated in callus but not in the other tested tissues, including 40S ribosomal proteins S4 (Q0E4Q0), S9 (Q2R1J8), S20 (Q10P27), and S24 (Q0DBK8), as well as 60S ribosomal proteins L2 (P92812), L21 (Q10RZ3), L27 (Q7XC31), and L36 (Q6L510), suggesting their essential roles in protein expression regulation in callus (Table S1). In addition to the ribosomal proteins, OASA1 (Q94GF1), which is a key enzyme for the synthesis of indole-3-acetic acid, and OsGH3.8 (Q0D4Z6), which is involved in the auxin homeostasis, were both more acetylated in callus than in other tissues [39,40]. Thus, the auxin level may subject to PKA regulation in callus.

3.3. PKA on Chloroplast Development and Photosynthesis Proteins in Leaf

Photosynthesis majorly occurs in the chloroplast of leaf, where the photosynthetic pigment chlorophyll captures and converts the light energy into the energy storage molecules ATP and NAPDH, which could be further converted into chemical energy in the form of carbohydrates, via the famous Calvin cycle. A number of key regulators of chloroplast development have been identified as acetylproteins in the current study. Fd-GOGAT1 (Q69RJ0) is a ferredoxin-dependent glutamate synthase which is known as a key enzyme in the process of inorganic nitrogen assimilation. Mutants of Fd-GOGAT1 exhibited defective chlorophyll synthesis and low photosynthesis rates, and finally led to early leaf senescence [41]. PORB (Q8W3D9) is defined as a NADPH:protochlorophyllide oxidoreductase catalyzing the photoreduction of protochlorophyllide to chlorophyllide in synthesis. Under high light conditions, PORB is essential for maintaining light-dependent chlorophyll synthesis, while disruption of the gene could turn the new leaves yellow and cause lesions [42]. WSL12 (Q0IMS5) is a chloroplast-localized, nucleoside diphosphate kinase knock-down of WLS12, which results in abnormal chloroplast and white stripes on leaves [43]. Interestingly, these three proteins were all predominantly acetylated in leaves, implying that PKA is a key switch for chloroplast development in rice.

In terms of photosynthesis, we found that numerous key enzymes were predominantly acetylated in leaf. For examples, glycine decarboxylase complex (GDC) recovers carbon following the oxygenation reaction of ribulose-1,5-bisphosphate carboxylase/oxygenase in the photorespiratory C2 cycle of C3 species [44]. LFNR (Leaf-type ferredoxin-NADP (+) oxidoreductase), an essential chloroplast enzyme, functions in the last step of photosynthetic linear electron transfer [45]. The current study revealed that rice GDCH (A3C6G9) was acetylated at the 156th lysine site, while LFNR1 (Q0DF89) harbored acetylation modifications on 79th and 188th lysine, and their acetylation intensities in leaf were significantly higher than those in other tissues. In addition, Chlorophyll A-B binding proteins (Q5ZA98, Q6Z411 and Q53N82), which are responsible for the capturing and delivering of excitation energy to photosystems [46], as well as sucrose synthases participating in starch metabolism (P31924, P30298 and Q43009), also showed similar acetylated patterns in leaf, suggesting the extensive regulatory roles of PKA on photosynthesis in leaf.

3.4. PKA May Regulate Flowering and Pollen Fertility in Rice Panicle

Initiation of panicle represents the transition of rice from vegetative growth to reproductive growth. Transcription factor OsMADS50 is a known key regulator in rice flowering [47]. Suppression of OsMADS50 significantly delayed rice flowering and increased the number of elongated internodes. Meanwhile, ectopical expression of this gene could even induce the initiation of inflorescence on the callus. OsMADS50 may be functionally antagonistic with another flowering repressor, OsMADS56, and operates upstream of the rice florigen gene *Hd3a* to control rice flowering [47,48]. We found an

acetylation modification of OsMADS50 on the 107th lysine site, which is located in the conserved DNA binding domain MADS box (17–135 in protein sequence). It would be interesting to explore the effects of acetylation on the binding ability of OsMADS50 to its target site in future study.

Programmed cell death (PCD) of tapetum cells in anthers is critical for the proper development of male gametophytes in rice. Apoptosis inhibitor 5 (API5, Q5JK84) encoding a putative antiapoptosis protein promotes the degeneration of the tapetum by interacting with two DEAD-box ATP-dependent RNA helicases, AIP1 and AIP2. *api5* mutants were fully sterile due to the inhibited tapetal PCD process, suggesting their key roles in the development of male gametophytes [49]. Our MS identification results showed that lysine 59, 144, and 161 were acetylated on API5. However, the three acetylsites displayed various acetylation patterns in tissues. The acetylation intensities on lysine 59 and 144 are significantly higher than in other tissues, whereas lysine 161 has the highest acetylation level in leaves, although it is also moderately acetylated in the panicles. Therefore, the indication is that PKA on lysine 59 and 144, other than on lysine 161, is more likely to be involved in pollen development.

3.5. Acetylation on Root Proteins

In vascular plants, the root is an important, specified organ, providing mechanic anchoring strength as well as controlling the water and nutrient uptake from soil. Root meristem activity is the key to root growth and architecture. An enzyme 3-hydroxyacyl-CoA dehydrogenase named as AIM1 (Q8W1L6) was found to be a vital regulator of root meristem activity in rice. aim1 mutant displayed shortened roots with reduced salicylic acid and ROS level. It is proposed that AIM1 mediates the salicylic acid biosynthesis to promote ROS accumulation, thereby maintaining the root meristem in an active status [50]. It is interesting to note that PKA only occurred on lysines 361 and 663 of AIM1 in the callus, leaf, and panicle, but AIM1 acetylation was not found in roots, where it is expected to be functional. The observation suggested that PKA on AIM1 may have negative effects on its function.

4. Materials and Methods

4.1. Collection and Preparation of Plant Materials

The Nipponbare (*Oryza sativa* L. *ssp jing*) plants used in this study were hydroponically-cultured or grown in the field of the China National Rice Research Institute (CNRRI). Roots were harvested from 14 days after germination of the hydroponic seedlings, leaves and panicles were collected from plants in the field at 3–5 days before heading, when the husk of spikelets were in pale color, and calluses were induced as described [51]. Three biological replicates of the callus, root, leaf, and panicle were harvested and immediately stored in liquid nitrogen before use.

4.2. Protein Extraction

Samples was ground into fine powders in liquid nitrogen, transferred to a 5 mL centrifuge tube containing lysis buffer (8 M urea, 150 mM Tris-HCl pH 8.0, 1 mM phenylmethylsulfonyl fluoride and 1× phosphoprotein protease inhibitor complex), and shaken for 30 min on ice. The extracted proteins were sheared by sonication (Biosafer 650-92 model, Scientz, Ningbo, China) on ice before centrifugation; then, the protein in the supernatant was precipitated in cold acetone for 30 min. The protein pellets were washed in 75% ethanol, and finally dissolved in PBS (Phosphate Buffer Saline) buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4). Qubit 2.0 fluorometer (Invitrogen, Carlsbad, CA, USA) was used for the quantification of the extracted total proteins.

4.3. Western Blotting

In brief, around 20 μg of each extracted total proteins was separated by 10% SDS–polyacrylamide gels, and subsequently transferred to a polyvinylidine fluoride fluoropolymer (PVDF) membrane (0.45 μm, Millipore, Darmstadt, Germany) using a Trans-Blot Turbo transfer system (Bio-Rad, Hercules, CA, USA). The TBST (10 mM Tris-HCl, 150 mM NaCl, and 0.05% Tween 20, pH 8.0) containing

5% BSA was used to block the transferred membrane for overnight at 4 °C. Acetyl lysine primary antibodies (1:1000 dilution in TBST) (ImmuneChem, Burnaby, BC, Canada) and secondary antibodies HRP (horseradish peroxidase–conjugated) (1:1000 dilution in TBST) (Beyotime Company, Shanghai, China) were used to detect the target protein bands, and lastly, visualized using the enhanced chemiluminescence (Pierce, Waltham, WA, USA).

4.4. Protein Digestion and Acetylpeptide Enrichment

Protein was treated by a sequential of 10 mM DTT (DL-Dithiothreitol) reduction for 1 h at 37 °C, 20 mM IAA (indole-3-acetic acid) alkylation for 45 min at room temperature in darkness, 100 mM NH_4CO_3 dilution and trypsin at pH 8.0 (enzyme:protein = 1:50) digestion for overnight. For acetylpeptide enrichment, the digested peptides were dissolved with NETN buffer (100 mM NaCl, 1 mM EDTA, 50 mM Tris-HCl, 0.5% NP-40, pH 8.0). The tryptic peptides were incubated with pre-washed antibody beads (PTM Biolabs, Hangzhou, China) by gently shaking at 4 °C overnight to enrich PKA peptides. Afterwards, the beads were gently washed four times with NETN buffer, and twice with ddH_2O. Finally, the bound acetylproteins were eluted with 0.1% TFA, subsequently vacuum dried using a SpeedVac (Thermo, Waltham, MA, USA), and cleaned with C18 ZipTips (Millipore, Darmstadt, Germany) before LC–MS/MS analysis.

4.5. LC–MS/MS Analyses

The procedures were done as described by Wang et al. [25]. In brief, peptides were dissolved in 0.1% formic acid, and directly loaded onto a reversed-phase pre-column (Thermo, Waltham, MA, USA.). Peptide separation was performed using a reversed-phase analytical column (Thermo, Waltham, MA, USA.). The gradient for MS analysis was described as follows: starting from 6 to 22% solvent B (0.1% FA in 98% ACN) for 24 min, then 22 to 40% for 8 min, 80% over 2 min, and lastly at 80% for 5 min at a constant flow rate of 300 μL/min on an EASY-nLC 1000 UPLC system coupled with a Q ExactiveTM mass spectrometer (Thermo, Waltham, MA, USA) over a mass range of 350–1800 m/z with a resolution of 7000. Subsequently, raw data were processed for acetylpeptide identification and acetylsite quantification with MaxQuant search engine against the uniprot_Oryza sativa database. In the MaxQuant searches for carbamidomethylation on Cys, oxidation on Met, Acetylation on Lys protein-N term was set as variable modification, and 4 missed cleavages on trypsin/P were allowed. Peptide mass error was set at 10 ppm for precursor ions and 0.02 Da for fragment ions; the minimum peptide length was set at 7. The false discovery rates (FDR) were set to <1%, and the minimum score for modified peptides was set >40 for peptide identification for all searches.

Supplementary Materials: The supplementary materials are available online.

Author Contributions: Z.L., Y.W., B.K.B., A.A.A. and X.T. performed the experiments and analyzed the data, J.Z. and Y.C. conceived of the project, designed and coordinated the experiments, J.Z. wrote the manuscript. All the authors read and approved the final manuscript.

Funding: This work was supported by Agricultural Sciences and Technologies Innovation Program of Chinese Academy of Agricultural Sciences (CAAS) to Rice Reproductive Developmental Biology Group, "Elite Youth" program (CAAS) to Jian Zhang and Yuxiao Chang, National Natural Science Foundation of China (grant number: 31401366), and the Shenzhen science and technology research funding (JSGG20160429104101251).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rao, R.S.; Thelen, J.J.; Miernyk, J.A. Is Lys-Nvarepsilon-acetylation the next big thing in post-translational modifications? *Trends Plant Sci.* **2014**, *19*, 550–553. [CrossRef] [PubMed]
2. Allfrey, V.G.; Mirsky, A.E. Acetylation and Methylation of Histones and their Possible Role in the Regulation of RNA Synthesis. *Proc. Natl. Acad. Sci. USA* **1964**, *51*, 786–794. [CrossRef] [PubMed]
3. Gil, J.; Ramirez-Torres, A.; Encarnacion-Guevara, S. Lysine acetylation and cancer: A proteomics perspective. *J. Proteom.* **2017**, *150*, 297–309. [CrossRef] [PubMed]

4. Batta, K.; Das, C.; Gadad, S.; Shandilya, J.; Kundu, T.K. Reversible acetylation of non histone proteins: Role in cellular function and disease. *Sub-Cell. Biochem.* **2007**, *41*, 193–212.
5. Spange, S.; Wagner, T.; Heinzel, T.; Krämer, O.H. Acetylation of non-histone proteins modulates cellular signalling at multiple levels. *Int. J. Biochem. Cell Biol.* **2009**, *41*, 185–198. [CrossRef] [PubMed]
6. Zhang, Y.; Song, L.; Liang, W.; Mu, P.; Wang, S.; Lin, Q. Comprehensive profiling of lysine acetylproteome analysis reveals diverse functions of lysine acetylation in common wheat. *Sci. Rep.* **2016**, *6*, 21069. [CrossRef] [PubMed]
7. Xiong, Y.; Peng, X.; Cheng, Z.; Liu, W.; Wang, G.L. A comprehensive catalog of the lysine-acetylation targets in rice (Oryza sativa) based on proteomic analyses. *J. Proteom.* **2016**, *138*, 20–29. [CrossRef] [PubMed]
8. Singh, B.; Zhang, G.; Hwa, Y.L.; Li, J.; Dowdy, S.C.; Jiang, S.W. Nonhistone protein acetylation as cancer therapy targets. *Expert Rev. Anticancer Ther.* **2010**, *10*, 935–954. [CrossRef] [PubMed]
9. Zimmermann, H.; Degenkolbe, R.; Bernard, H.U.; O'Connor, M.J. The human papillomavirus type 16 E6 oncoprotein can down-regulate p53 activity by targeting the transcriptional coactivator CBP/p300. *J. Virol.* **1999**, *73*, 6209–6219. [PubMed]
10. Lee, J.; Manning, A.J.; Wolfgeher, D.; Jelenska, J.; Cavanaugh, K.A.; Xu, H.; Fernandez, S.M.; Michelmore, R.W.; Kron, S.J.; Greenberg, J.T. Acetylation of an NB-LRR Plant Immune-Effector Complex Suppresses Immunity. *Cell Rep.* **2015**, *13*, 1670–1682. [CrossRef] [PubMed]
11. Kim, S.C.; Sprung, R.; Chen, Y.; Xu, Y.; Ball, H.; Pei, J.; Cheng, T.; Kho, Y.; Xiao, H.; Xiao, L. Substrate and Functional Diversity of Lysine Acetylation Revealed by a Proteomics Survey. *Mol. Cell* **2006**, *23*, 607–618. [CrossRef] [PubMed]
12. Xu, H.; Zhou, J.; Lin, S.; Deng, W.; Zhang, Y.; Xue, Y. PLMD: An updated data resource of protein lysine modifications. *J. Genet. Genom.* **2017**, *44*, 243–250. [CrossRef] [PubMed]
13. Finkemeier, I.; Laxa, M.; Miguet, L.; Howden, A.J.; Sweetlove, L.J. Proteins of diverse function and subcellular location are lysine acetylated in Arabidopsis. *Plant Physiol.* **2011**, *155*, 1779–1790. [CrossRef] [PubMed]
14. Konig, A.C.; Hartl, M.; Boersema, P.J.; Mann, M.; Finkemeier, I. The mitochondrial lysine acetylome of Arabidopsis. *Mitochondrion* **2014**, *19*, 252–260. [CrossRef] [PubMed]
15. Melo-Braga, M.N.; Verano-Braga, T.; Leon, I.R.; Antonacci, D.; Nogueira, F.C.; Thelen, J.J.; Larsen, M.R.; Palmisano, G. Modulation of protein phosphorylation, *N*-glycosylation and Lys-acetylation in grape (Vitis vinifera) mesocarp and exocarp owing to Lobesia botrana infection. *Mol. Cell. Proteom.* **2012**, *11*, 945–956. [CrossRef] [PubMed]
16. Smith-Hammond, C.L.; Hoyos, E.; Miernyk, J.A. The pea seedling mitochondrial Nepsilon-lysine acetylome. *Mitochondrion* **2014**, *19*, 154–165. [CrossRef] [PubMed]
17. Smith-Hammond, C.L.; Swatek, K.N.; Johnston, M.L.; Thelen, J.J.; Miernyk, J.A. Initial description of the developing soybean seed protein Lys-N(epsilon)-acetylome. *J. Proteom.* **2014**, *96*, 56–66. [CrossRef] [PubMed]
18. Fang, X.; Chen, W.; Zhao, Y.; Ruan, S.; Zhang, H.; Yan, C.; Jin, L.; Cao, L.; Zhu, J.; Ma, H.; et al. Global analysis of lysine acetylation in strawberry leaves. *Front. Plant Sci.* **2015**, *6*, 739. [CrossRef] [PubMed]
19. Salvato, F.; Havelund, J.F.; Chen, M.; Rao, R.S.; Rogowska-Wrzesinska, A.; Jensen, O.N.; Gang, D.R.; Thelen, J.; Moller, I.M. The potato tuber mitochondrial proteome. *Plant Physiol.* **2014**, *164*, 637–653. [CrossRef] [PubMed]
20. Xia, Y.; Jing, D.; Kong, L.; Zhang, J.; Yang, F.; Zhang, H.; Wang, J.; Zhang, S. Global Lysine Acetylome Analysis of Desiccated Somatic Embryos of Picea asperata. *Front. Plant Sci.* **2016**, *7*, 1927. [CrossRef] [PubMed]
21. Zhen, S.; Deng, X.; Wang, J.; Zhu, G.; Cao, H.; Yuan, L.; Yan, Y. First Comprehensive Proteome Analyses of Lysine Acetylation and Succinylation in Seedling Leaves of *Brachypodium distachyon* L. *Sci. Rep.* **2016**, *6*, 31576. [CrossRef] [PubMed]
22. Zhang, J.; Guo, D.; Chang, Y.; You, C.; Li, X.; Dai, X.; Weng, Q.; Zhang, J.; Chen, G.; Li, X.; et al. Non-random distribution of T-DNA insertions at various levels of the genome hierarchy as revealed by analyzing 13,804 T-DNA flanking sequences from an enhancer-trap mutant library. *Plant J.* **2007**, *49*, 947–959. [CrossRef] [PubMed]
23. Wu, X.; Huber, S.C. Lysine Acetylation Is a Widespread Protein Modification for Diverse Proteins in Arabidopsis. *Plant Physiol.* **2011**, *155*, 1769–1778. [CrossRef] [PubMed]
24. Nallamilli, B.R.; Edelmann, M.J.; Zhong, X.; Tan, F.; Mujahid, H.; Zhang, J.; Nanduri, B.; Peng, Z. Global analysis of lysine acetylation suggests the involvement of protein acetylation in diverse biological processes in rice (Oryza sativa). *PLoS ONE* **2014**, *9*, e89283. [CrossRef] [PubMed]

25. Wang, Y.; Hou, Y.; Qiu, J.; Li, Z.; Zhao, J.; Tong, X.; Zhang, J. A Quantitative Acetylomic Analysis of Early Seed Development in Rice (*Oryza sativa* L.). *Int. J. Mol. Sci.* **2017**, *18*, 249–265.
26. He, D.; Wang, Q.; Ming, L.; Damaris, R.N.; Yi, X.; Cheng, Z.; Yang, P. Global Proteome Analyses of Lysine Acetylation and Succinylation Reveal the Widespread Involvement of both Modification in Metabolism in the Embryo of Germinating Rice Seed. *J. Proteome Res.* **2016**, *15*, 879–890. [CrossRef] [PubMed]
27. Li, X.; Ye, J.; Ma, H.; Lu, P. Proteomic analysis of lysine acetylation provides strong evidence for acetylated proteins involved in plant meiosis and tapetum function. *Plant J.* **2018**, *93*, 142–154. [CrossRef] [PubMed]
28. Meng, X.; Lv, Y.; Mujahid, H.; Edelmann, M.J.; Zhao, H.; Peng, X.; Peng, Z. Proteome-wide lysine acetylation identification in developing rice (*Oryza sativa*) seeds and protein co-modification by acetylation, succinylation, ubiquitination, and phosphorylation. *Biochim. Biophys. Acta* **2017**, *1866*, 451–463. [CrossRef] [PubMed]
29. Xue, C.; Liu, S.; Chen, C.; Zhu, J.; Yang, X.; Yong, Z.; Rui, G.; Liu, X.; Gong, Z. Global Proteome Analysis Links Lysine Acetylation to Diverse Functions in *Oryza sativa*. *Proteomics* **2017**, *18*. [CrossRef]
30. Zhou, H.; Finkemeier, I.; Guan, W.; Tossounian, M.A.; Wei, B.; Young, D.; Huang, J.; Messens, J.; Yang, X.; Zhu, J. Oxidative stress-triggered interactions between the succinyl- and acetyl-proteomes of rice leaves. *Plant Cell Environ.* **2018**, *41*, 1139–1153. [CrossRef] [PubMed]
31. Kawahara, Y.; De la Bastide, M.; Hamilton, J.P.; Kanamori, H.; McCombie, W.R.; Ouyang, S.; Schwartz, D.C.; Tanaka, T.; Wu, J.; Zhou, S.; et al. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice* **2013**, *6*, 4. [CrossRef] [PubMed]
32. Szklarczyk, D.; Franceschini, A.; Wyder, S.; Forslund, K.; Heller, D.; Huerta-Cepas, J.; Simonovic, M.; Roth, A.; Santos, A.; Tsafou, K.P.; et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **2015**, *43*, D447–D452. [CrossRef] [PubMed]
33. Chou, M.F.; Schwartz, D. Biological sequence motif discovery using *motif-x*. *Curr. Protoc. Bioinform.* **2011**, 15–24. [CrossRef]
34. Qiu, J.; Hou, Y.; Tong, X.; Wang, Y.; Lin, H.; Liu, Q.; Zhang, W.; Li, Z.; Nallamilli, B.R.; Zhang, J. Quantitative phosphoproteomic analysis of early seed development in rice (*Oryza sativa* L.). *Plant Mol. Biol.* **2016**, *90*, 249–265. [CrossRef] [PubMed]
35. Walley, J.W.; Shen, Z.; Mcreynolds, M.R.; Schmelz, E.A.; Briggs, S.P. Fungal-induced protein hyperacetylation in maize identified by acetylome profiling. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 210–215. [CrossRef] [PubMed]
36. Hartl, M.; Füßl, M.; Boersema, P.J.; Jost, J.O.; Kramer, K.; Bakirbas, A.; Sindlinger, J.; Plöchinger, M.; Leister, D.; Uhrig, G. Lysine acetylome profiling uncovers novel histone deacetylase substrate proteins in *Arabidopsis*. *Mol. Syst. Biol.* **2017**, *13*. [CrossRef] [PubMed]
37. Liew, C.C.; Yip, C.C. Acetylation of reticulocyte ribosomal proteins at time of protein biosynthesis. *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 2988–2991. [CrossRef] [PubMed]
38. Kamita, M.; Kimura, Y.; Ino, Y.; Kamp, R.M.; Polevoda, B.; Sherman, F.; Hirano, H. N(α)-Acetylation of yeast ribosomal proteins and its effect on protein synthesis. *J. Proteom.* **2011**, *74*, 431–441. [CrossRef] [PubMed]
39. Tozawa, Y.; Hasegawa, H.; Terakawa, T.; Wakasa, K. Characterization of rice anthranilate synthase alpha-subunit genes OASA1 and OASA2. Tryptophan accumulation in transgenic rice expressing a feedback-insensitive mutant of OASA1. *Plant Physiol.* **2001**, *126*, 1493–1506. [CrossRef] [PubMed]
40. Yadav, S.R.; Khanday, I.; Majhi, B.B.; Veluthambi, K.; Vijayraghavan, U. Auxin-Responsive OsMGH3, a Common Downstream Target of OsMADS1 and OsMADS6, Controls Rice Floret Fertility. *Plant Cell Physiol.* **2011**, *52*, 2123–2135. [CrossRef] [PubMed]
41. Sun, L.; Wang, Y.; Liu, L.; Wang, C.; Gan, T.; Zhang, Z.; Wang, Y.; Wang, D.; Niu, M.; Long, W. Isolation and characterization of a spotted leaf 32 mutant with early leaf senescence and enhanced defense response in rice. *Sci. Rep.* **2017**, *7*. [CrossRef] [PubMed]
42. Sakuraba, Y.; Rahman, M.L.; Cho, S.H.; Kim, Y.S.; Koh, H.J.; Yoo, S.C.; Paek, N.C. The rice faded green leaf locus encodes protochlorophyllide oxidoreductaseB and is essential for chlorophyll synthesis under high light conditions. *Plant J.* **2013**, *74*, 122–133. [CrossRef] [PubMed]
43. Ye, W.; Hu, S.; Wu, L.; Ge, C.; Cui, Y.; Chen, P.; Wang, X.; Xu, J.; Ren, D.; Dong, G. White stripe leaf 12 (WSL_{12}), encoding a nucleoside diphosphate kinase 2 ($OsNDPK_2$), regulates chloroplast development and abiotic stress response in rice (*Oryza sativa* L.). *Mol. Breed.* **2016**, *36*. [CrossRef] [PubMed]

44. Lin, H.; Karki, S.; Coe, R.A.; Bagha, S.; Khoshravesh, R.; Balahadia, C.P.; Ver, S.J.; Tapia, R.; Israel, W.K.; Montecillo, F. Targeted Knockdown of GDCH in Rice Leads to a Photorespiratory Deficient Phenotype Useful as a Building Block for C4 Rice. *Plant Cell Physiol.* **2016**, *57*. [CrossRef] [PubMed]
45. Yang, C.; Hu, H.; Ren, H.; Kong, Y.; Lin, H.; Guo, J.; Wang, L.; He, Y.; Ding, X.; Grabsztunowicz, M. LIGHT-INDUCED $RICE_1$ Regulates Light-Dependent Attachment of LEAF-TYPE FERREDOXIN-NADP + OXIDOREDUCTASE to the Thylakoid Membrane in Rice and Arabidopsis. *Plant Cell* **2016**, *28*, 712–728. [CrossRef] [PubMed]
46. Magorzata, A.; Krzysztof, G.; Robert, L.; Wojciech, G.; Przemyslaw, C.M.; Sebastian, S.; Weronika, S.; Rienk, V.G.; Grzegorz, J. Excitation energy transfer and charge separation are affected in *Arabidopsis thaliana* mutants lacking light-harvesting chlorophyll *a/b* binding protein Lhcb3. *J. Photochem. Photobiol. B* **2015**, *153*, 423–428.
47. Lee, S.; Kim, J.; Han, J.J.; Han, M.J.; An, G. Functional analyses of the flowering time gene OsMADS50, the putative SUPPRESSOR OF OVEREXPRESSION OF CO 1/AGAMOUS-LIKE 20 (SOC1/AGL20) ortholog in rice. *Plant J. Cell Mol. Biol.* **2010**, *38*, 754–764. [CrossRef] [PubMed]
48. Cho, L.H. OsMADS50 and OsMADS56 function antagonistically in regulating long day (LD)-dependent flowering in rice. *Plant Cell Environ.* **2010**, *32*, 1412–1427.
49. Li, X.; Gao, X.; Wei, Y.; Deng, L.; Ouyang, Y.; Chen, G.; Li, X.; Zhang, Q.; Wu, C. Rice APOPTOSIS $INHIBITOR_5$ coupled with two DEAD-box adenosine 5′-triphosphate-dependent RNA helicases regulates tapetum degeneration. *Plant Cell* **2011**, *23*, 1416–1434. [CrossRef] [PubMed]
50. Xu, L.; Zhao, H.; Ruan, W.; Deng, M.; Wang, F.; Peng, J.; Luo, J.; Chen, Z.; Yi, K. ABNORMAL INFLORESCENCE $MERISTEM_1$ Functions in Salicylic Acid Biosynthesis to Maintain Proper Reactive Oxygen Species Levels for Root Meristem Activity in Rice. *Plant Cell* **2017**, *29*. [CrossRef] [PubMed]
51. Hiei, Y.; Komari, T.; Kubo, T. Transformation of rice mediated by Agrobacterium tumefaciens. *Plant Mol. Biol.* **1997**, *35*, 205–218. [CrossRef] [PubMed]

Sample Availability: Not available.

molecules

MDPI

Article

Optimisation of Milk Protein Top-Down Sequencing Using In-Source Collision-Induced Dissociation in the Maxis Quadrupole Time-of-Flight Mass Spectrometer

Delphine Vincent [1,*], Dominik Mertens [2] and Simone Rochfort [1,3]

1 Department of Economic Development, Jobs, Transport and Resources, AgriBio Centre, Bundoora, Victoria 3083, Australia; simone.rochfort@ecodev.vic.gov.au
2 Genedata AG, 4016 Basel, Switzerland; Dominik.Mertens@genedata.com
3 School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia
* Correspondence: delphine.vincent@ecodev.vic.gov.au; Tel.: +61-3-9032-7116

Academic Editor: Paolo Iadarola
Received: 8 October 2018; Accepted: 25 October 2018; Published: 26 October 2018

Abstract: Top-down sequencing in proteomics has come of age owing to continuous progress in LC-MS. With their high resolution and broad mass range, Quadrupole Time-of-Flight (Q-ToF) hybrid mass spectrometers equipped with electrospray ionisation source and tandem MS capability by collision-induced dissociation (CID) can be employed to analyse intact proteins and retrieve primary sequence information. To our knowledge, top-down proteomics methods with Q-ToF have only been evaluated using samples of relatively low complexity. Furthermore, the in-source CID (IS-CID) capability of Q-ToF instruments has been under-utilised. This study aimed at optimising top-down sequencing of intact milk proteins to achieve the greatest sequence coverage possible from samples of increasing complexity, assessed using nine known proteins. Eleven MS/MS methods varying in their IS-CID and conventional CID parameters were tested on individual and mixed protein standards as well as raw milk samples. Top-down sequencing results from the nine most abundant proteoforms of caseins, alpha-lactalbumin and beta-lactoglubulins were compared. Nine MS/MS methods achieved more than 70% sequence coverage overall to distinguish between allelic proteoforms, varying only by one or two amino acids. The optimal methods utilised IS-CID at low energy. This experiment demonstrates the utility of Q-ToF systems for top-down proteomics and that IS-CID could be more frequently employed.

Keywords: top-down proteomics; HPLC-ESI-Q-TOF MS; cow's milk; whey proteins and caseins; tandem MS

1. Introduction

Top-down proteomics, a term invented by Kelleher and colleagues 20 years ago [1], describes the analysis of intact proteins, either in their native form or more often in a denatured state, which allows for a characterisation of proteoforms as comprehensively as possible. Coined in 2014, the term proteoform "designates all of the different molecular forms in which the protein product of a single gene can be found, encompassing all forms of genetic variation, alternative splicing of RNA transcripts, and post-translational modifications (PTMs)" [2]. The analysis of intact proteins is now always performed using mass spectrometry (MS) and technical progresses in top-down proteomics are tightly linked to improvements made on mass analysers. The most important technological advance was the coupling of the soft ionisation technique, electrospray ionisation (ESI), to a mass spectrometer and the production of gas phase ions from large molecules [3], which led to its first application to

intact proteins [4,5]. ESI generates multiply-charged protein ions of low m/z, thereby allowing the analysis of very large molecules, even on an MS platform with a limited mass range. By applying a deconvolution algorithm [6], the resulting complex multi-peak spectrum can be converted into a single peak corresponding to the true molecular weight of the molecule. Further developments in intact protein analysis quickly followed with the introduction of complex multistage mass analysers of high resolution such as linear ion trap (LIT), quadrupole time-of-flight (Q-ToF), Fourier transform (FT) ion cyclotron resonance, and orbitrap instruments.

Today, various methods of ion fragmentation are available, such as collision-induced dissociation (CID, also called collisionally activated dissociation CAD), in-source CID (IS-CID, also known as up-front CID, cone-voltage CID, and nozzle-skimmer dissociation), higher-energy collisional dissociation (HCD), electron capture dissociation (ECD), and electron transfer dissociation (ETD) (for review [7]). Different types of fragmentation modes yield different information about the structure and composition of the analyte. Tandem mass spectrometers that carry out CID to generate product ions from precursor ions have proven extremely useful for the identification and characterisation of proteins from a complex mixture. CID was first described by McLafferty and Bryce in 1967 [8] and Jennings in 1968 [9]. Parent ions collide with neutral gas atoms or molecules (typically helium, nitrogen or argon), which result in the formation of b- and y-type ions. The efficacy of a CID experiment will depend on the relative translational energy of the ion and target, the nature of the target, the number of collisions that is likely to take place, and the m/z window of the instrument. Initial top-down sequencing experiments exploited low-energy CID methods to induce protein fragmentation. However, for large molecules, CID does not produce a fragmentation pattern comprehensive enough to fully characterise proteins, but rather produces enough fragments or sequence tags to identify the protein. Additionally, if the protein contains PTMs, low-energy CID most likely will not be sufficient to localise the modified site or the PTM may be the preferred site of cleavage [10,11]. IS-CID is a proven albeit seldom utilised fragmentation mode allowing single-stage instrumentation, such as single quadrupole (Q) or ToF mass analysers, to produce spectra similar to those obtained with far more expensive hybrid instruments. In this method, invented by Katta and colleagues in 1991 [12], CID is carried out within the ion source in the high-pressure region between the capillary exit and the skimmer entrance to the Q mass spectrometer. Because there is no prior selection of the precursor ion, this does not qualify as a bona fide MS/MS experiment. By manipulating lens voltages that channel ions from the source to the mass analyser, relatively low-energy ions formed in the atmospheric pressure region of the source collide with residual background gas, usually nitrogen, in the transition region. As a result, excited ions can undergo unimolecular decomposition to produce fragment ions. As for CID, only b- and y-type ions are observed. Tandem spectra from protonated peptides produced by IS-CID or low-energy CID are comparable [13,14]. In-source dissociation has also been demonstrated for whole protein ions [15]. The time scale for IS-CID is on the order of a few hundred microseconds to a few milliseconds, which is much faster than that of CID. IS-CID fragmentation can be reproducible provided ion source parameters, such as temperature, pressure, voltage, and sample purity are tightly controlled [16].

Kelleher in 2004 [17] rigorously defines top-down proteomics as a multistep process whereby the molecular weights (MWs) of intact proteoforms are accurately measured using a high-resolution mass analyser in combination with a direct fragmentation of the protein ions using tandem MS. Top-down sequencing of intact proteins were initially performed using triple quadrupole (QQQ) instruments and IS-CID [5]. In 1990, Loo and colleagues characterised a 14-kDa bovine ribonuclease A in its native and reduced forms using CID [18]. Soon after, higher mass resolution was achieved by applying FT-MS to equine cytochrome c, porcine albumin, thioredoxin and ubiquitin [19], myoglobin [20], and carbonic anhydrase [15]. A hybrid Q-ToF instrument, also called Q-ToF [21], was invented in 1996 by Morris and colleagues [22]. Q-ToF mass spectrometers combine the quadrupole one (Q1) in which ion precursors are funnelled through and selected, the quadrupole collision cell (q) of a QQQ in which selected precursor ions are fragmented usually through CID and more recently ETD, and a reflector ToF

detector in which the fragment ions are mass analysed. Spectra obtained in both full-scan (MS1) and MS/MS (MS2) modes exhibit clean spectra due to orthogonal pulsing into the ToF section, high mass accuracy along with stable isotopic resolution across the m/z range, and high fidelity meaning accurate isotopic ratios. This permits the determination of charge states and unambiguous assignment of the mono-isotopic signal of the intact molecules. This leads to accurate prediction of AA sequence and successful identification of peptides and proteins of medium MW via database searches by tightening the search parameters and augmenting the confidence in the results. Q-ToF instruments perform well for quantitative analyses [23] and for the identification of PTMs [24]. Q-ToF were demonstrated to outperform QQQ mass spectrometers owing to both their enhanced sensitivity and resolution in the region of one to two orders of magnitude [22]. This initial observation was further validated on phosphopeptides; the higher resolving power of the Q-ToF improved the selectivity and sensitivity of parent ions, thus minimising interference from other product ions and maximising sequencing results [25].

The first use of LC-ESI-Q-ToF MS in top-down sequencing of intact proteins was reported by Nemeth-Cawley and Rouse in 2002 [26], who used CID to fragment ion precursors from eight known protein standards, ranging from 5 to 66 kDa. Peptide sequence tags thus obtained led to the unambiguous identification of the analysed proteins, along with the characterisation of disulphide bonds and glycosylation. While their method was successful, the authors noted some limitations: first, it could not yet be applied to complex protein mixtures without prior separation; secondly, at a low concentration, the signal-to-noise ratio was affected, thus necessitating longer MS/MS acquisition times. Based on their experience, the following year, the same group applied their validated method to characterise a recombinant immunoglobulin gamma-1 (IgG-1) fusion protein [27]. The top-down strategy allowed them to demonstrate that the recombinant protein was expressed as a full-length form as well as N- and C-processed truncated proteoforms [27]. In 2004, Ginter and colleagues exploited both IS-CID and traditional CID mode of a Q-ToF mass spectrometer, thereby achieving pseudo-MS3 levels, to top-down sequence seven known proteins spanning from 11 to 66 kDa [28]. Individual sequence tags of 10 to 26 AAs were retrieved from both N- and C-termini, and unambiguous identification of protein standards was achieved. IS-CID on its own was used to identify various purified recombinant proteins using ESI-Q-ToF MS [29]; fragmentation efficiency depended on cone voltage and y-ions formed predominantly by cleavage on the C-terminal side of nonpolar residues. An unknown 4.9 kDa recombinant peptide was completely sequenced using LC-ESI-Q-ToF MS/MS of 10,000 resolution and traditional CID fragmentation; complete y- and b-type ion series were obtained and the formation of beta-mercaptoethanol adducts was reported [30]. In 2009 Armirotti and colleagues reported 90–100% sequence coverage of horse myoglobin (17 kDa) and bovine carbonic anhydrase II (30 kDa) using LC-ESI-Q-ToF MS/MS and CID fragmentation mode [31]. They could also identify an unknown protein as superoxide dismutase (16 kDa) and locate one acetylation site. ESI CID Q-TOF MS/MS was used to identify eight intact antimicrobial peptides from Asian frog skin and locate disulphide bridges [32]. ESI-Q-ToF technology has also been employed to investigate the precise stoichiometry of protein assemblies, the interactions between subunits and the position of subunits within the complex [33].

In our laboratory, we host a maXis HD UHR-Q-ToF (60,000 resolution) with an ESI source on-line with a UHPLC 1290 Infinity Binary LC system (Agilent, Mulgrave, VIC, Australia), which we have used to develop a top-down method to analyse cow's milk proteins [23]. This method was further implemented with CID MS/MS analysis and applied to the study of UHT milk shelf life [34]. The present study aimed at optimising MS/MS analysis by testing the effect of some of the parameters pertaining to IS-CID in combination or not with conventional CID. A total of 11 MS/MS methods were assessed on samples bearing increased complexity, namely individual milk protein standards, mixed protein standards and cow's raw milk samples from Jersey or Holstein breeds. Milk proteins exhibit numerous PTMs, however in this initial study, we focused on optimising protein fragmentation to achieve the greatest sequence coverage possible. To this end, nine of the most abundant proteoforms of milk caseins and whey proteins, including genetic variants, were compared across samples and MS/MS methods whose efficacy was assessed based on protein sequence coverage alone. This necessitated developing processing workflows using the powerful Genedata Expressionist software to thoroughly explore and annotate the MS1 and MS2 files.

2. Results and Discussion

2.1. Sample Complexity

In this experiment, we wanted to test whether sample complexity would impact the quality of spectral acquisition. To this end, we prepared a set of samples displaying increasing proteome complexity: from simple individual milk protein standards, albeit varying in their purity level from 70% (α-CN) to 98% (β-CN), to a mixture of these standards, to the highly complex biological matrices raw milk samples that were obtained from two common cow breeds. Figure 1 describes the experimental design of this study.

2.1.1. UPLC-ESI-Q-ToF MS Can Handle Complex Samples

While the raw data clearly illustrate increasing complexity with more LC-MS1 peaks detected in the milk samples compared to the individual protein standard samples (Figure 2A), the latter are more complex than first anticipated, particularly the β-CN standard, which is claimed to be 98% pure.

These standards were analysed using a bottom-up shotgun approach and yielded many protein hits [35]. The α-LA standard sample (85% pure) generated by far the simplest LC-MS1 pattern, with clear elution from 14 to 19 min and distinct m/z peaks of the charged envelope (Figure 2A).

The LC-MS1 pattern of the mixed standard sample is quite similar and as complex as the patterns of milk samples owing to the fact that the individual protein standards are far from pure. Indeed, following protein mass deconvolution and the display of protein accurate masses separated by HPLC (Figure 2B), many deconvoluted peaks appear besides the expected proteins of interest. Such a display confirms that the α-LA standard is the least complex sample with only 211 deconvoluted peaks, whereas the other individual standards display between 2194 (α-CN) and 3372 peaks (β-CN).

With 4819 deconvoluted peaks, the mixed standard sample is indeed of intermediate complexity between the individual standards and the raw milk samples, which resolved 6988 and 7554 peaks in Jersey and Holstein cows, respectively. While proteins displayed some degree of co-elution during the HPLC separation, particularly in complex samples, separation along the m/z range could isotopically resolve all proteins of less than 30 kDa, thus yielding deconvoluted monoisotopic masses (also shown in [23]).

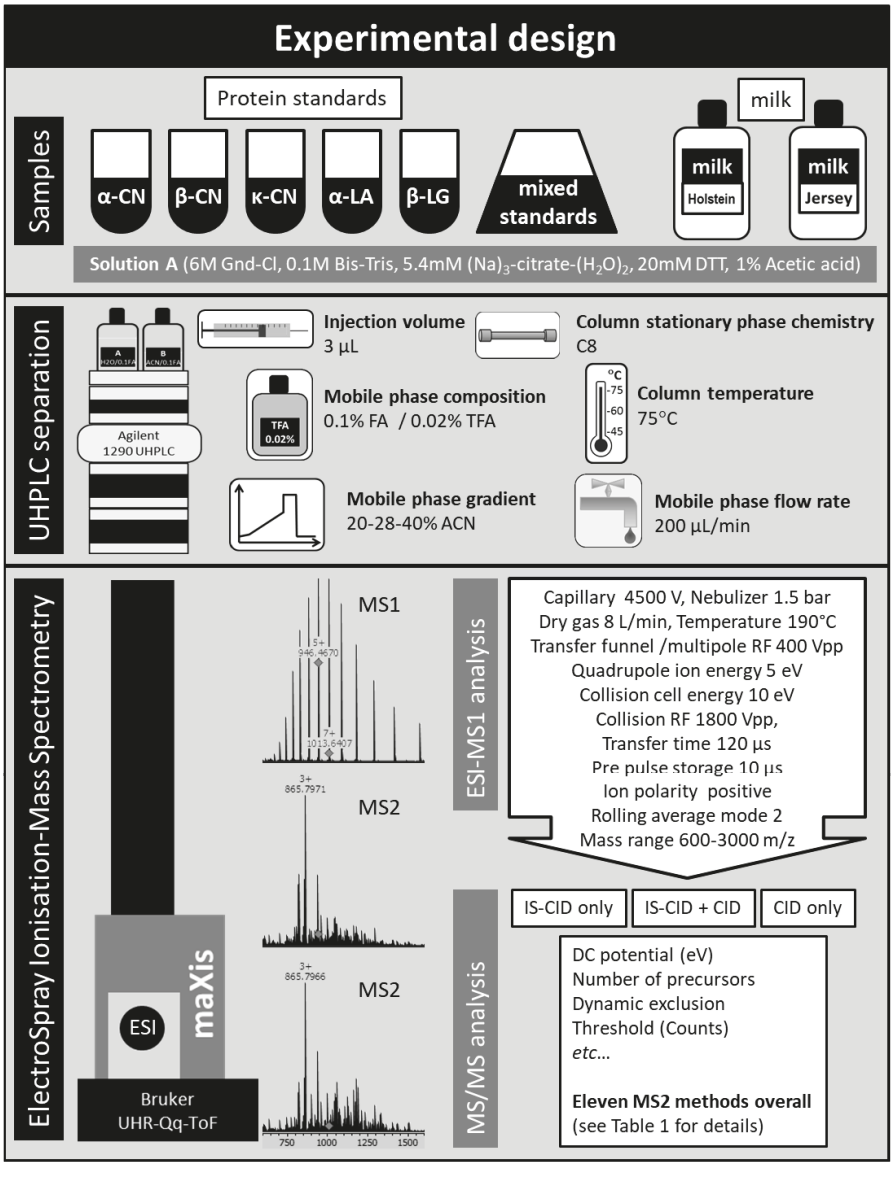


Figure 1. Schematic diagram of the experimental design of the study.

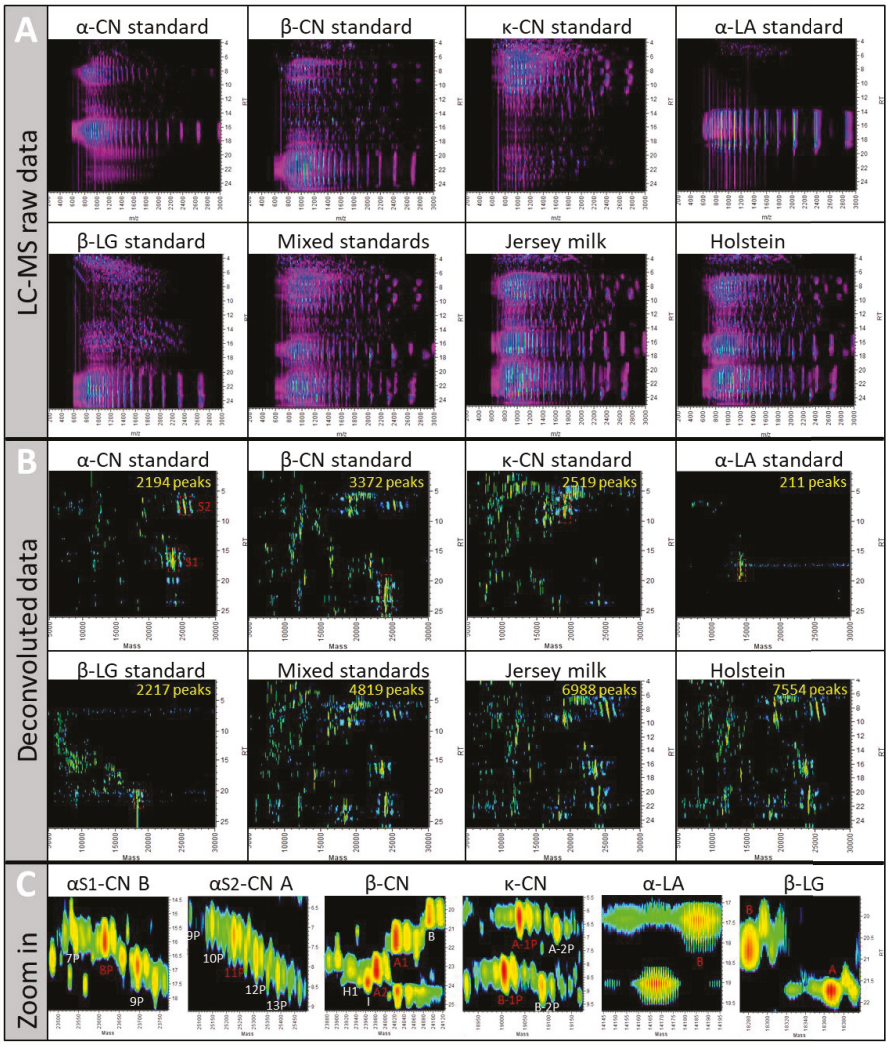


Figure 2. LC-MS maps of the different samples (individual protein standards, mixed protein standards, Jersey milk, and Holstein milk) visualised in Genedata Refiner pre- and post-deconvolution with m/z on the x axis and LC retention time in min on the y axis. (**A**) LC-MS1 unprocessed raw data. (**B**) Deconvoluted data displaying proteins of interest in the red boxed areas. The total numbers of deconvoluted peaks detected are indicated in yellow. (**C**) Close-ups of the boxed areas of the individual protein standards; the nine proteoforms highlighted in red are the proteins targeted for method validation.

2.1.2. The 11 MS/MS Methods Are Assessed Using a Reference Set of Known Proteins

Another aim of our study was to optimise the MS/MS method by modifying key parameters and devise objective tools to compare the different methods. We thought the best comparative approach would be a targeted one relying on a small reference set of known proteins. The proteins of interest are depicted in the boxed areas of Figure 2B which have been zoomed in in Figure 2C. We chose nine prominent proteoforms, some of them displaying allelic variations, in each of the individual standards, namely, α-CN type S1 variant B with 8 phosphorylations (α_{S1}-CN B-8P, 23,600.3 Da), α-CN type S2 variant A with 11 phosphorylations (α_{S2}-CN A-11P, 25,213.0 Da), β-CN variants A1 and A2 (β-CN A1 24,008.2 Da, β-CN A2 23,968.2 Da) each with five phosphorylations, κ-CN variants A and B with 1 phosphorylation (κ-CN A-1P 19,026.5 Da, κ-CN B-1P 18,993.6 Da), α-LA variant B (α-LA B 14,176.8 Da), and β-LG variants A and B (β-LG A 18,355.5 Da, β-LG B 18269.4 Da). We have previously used these proteins to optimise a quantitative LC-MS method [23]. The full primary sequence of these proteins is known, and therefore fragmentation efficiency for top-down sequencing purpose could be assessed. This is presented in the last chapter. While not included in this article, our MS/MS method also allowed us to identify various PTMs of milk most abundant proteins such as lactosylation (+324 Da), oxidation (+16 Da), glycosylation ([Hex(1)HexNAc(1)NeuAc(2)] + 947 Da), and degradation products, as reported in [34].

2.2. Effect of IS-CID and CID on Their Own or Combined on Spectral Data

In a recent study [34], we have used conventional CID (Method 2 in the present study) to identify the degradation products of milk's most abundant proteins occurring following UHT treatment and storage on the shelf. While this top-down sequencing method proved successful for the identification of small proteins and degradation products, we wanted to optimise it for the analysis of intact milk proteins. Our Q-ToF mass spectrometer allows us to perform both IS-CID and CID in a single experiment. The benefits of both have been stated in the introduction of this manuscript. In the present study, we have only slightly varied the parameters controlling CID fragmentation, choosing to operate it in a fully automated manner. We have mostly fine-tuned the parameters pertaining to IS-CID. We thus tested different energy values for IS-CID MS (ion funnel 1 exit) and IS-CID MS/MS (ion funnel 2 entrance) parameters. A total of 11 MS/MS methods were thus devised that employed either CID or IS-CID on their own or combined (Table 1).

In summary, Methods 2 and 3 employed CID only, Methods 4 and 12 employed IS-CID only, and Methods 5–11 employed IS-CID within the ESI source/ion transfer region, followed by CID within the collision cell.

Table 1. Parameters applied on the Q-TOF mass spectrometer for the 12 MS methods.

Mode	Step	Parameter	Method 1	Method 2	Method 3	Method 4	Method 5	Method 6	Method 7	Method 8	Method 9	Method 10	Method 11	Method 12
MS	Source	End plate offset	500 V	500 V	500 V	500 V	500 V	500 V	500 V	500 V	500 V	500 V	500 V	500 V
MS	Source	Capillary	4500 V	4500 V	4500 V	4500 V	4500 V	4500 V	4500 V	4500 V	4500 V	4500 V	4500 V	4500 V
MS	Source	Nebuliser	1.5 bar	1.5 bar	1.5 bar	1.5 bar	1.5 bar	1.5 bar	1.5 bar	1.5 bar	1.5 bar	1.5 bar	1.5 bar	1.5 bar
MS	Source	Dry gas	8 L/min	8 L/min	8 L/min	8 L/min	8 L/min	8 L/min	8 L/min	8 L/min	8 L/min	8 L/min	8 L/min	8 L/min
MS	Source	Temperature	190 °C	190 °C	190 °C	190 °C	190 °C	190 °C	190 °C	190 °C	190 °C	190 °C	190 °C	190 °C
MS	Source	Waste (min)	0–2.5	0–2.5	0–2.5	0–2.5	0–2.5	0–2.5	0–2.5	0–2.5	0–2.5	0–2.5	0–2.5	0–2.5
MS	Source	Source (min)	2.5–40	2.5–40	2.5–40	2.5–40	2.5–40	2.5–40	2.5–40	2.5–40	2.5–40	2.5–40	2.5–40	2.5–40
MS	Transfer	Funnel RF	400 Vpp	400 Vpp	400 Vpp	400 Vpp	400 Vpp	400 Vpp	400 Vpp	400 Vpp	400 Vpp	400 Vpp	400 Vpp	400 Vpp
MS	Transfer	IS-CID energy	0 eV	0 eV	0 eV	3 eV	30 eV	50 eV	70 eV	5 eV	0 eV	3 eV	0 eV	10 eV
MS	Transfer	Transfer multipole RF	400 Vpp	400 Vpp	400 Vpp	400 Vpp	400 Vpp	400 Vpp	400 Vpp	400 Vpp	400 Vpp	400 Vpp	400 Vpp	400 Vpp
MS	Quadrupole	Ion energy	5 eV	5 eV	5 eV	5 eV	5 eV	5 eV	5 eV	5 eV	5 eV	5 eV	5 eV	5 eV
MS	Quadrupole	Low mass	300 m/z	300 m/z	300 m/z	300 m/z	300 m/z	300 m/z	300 m/z	300 m/z	300 m/z	300 m/z	300 m/z	300 m/z
MS	Collision cell	Collision energy	10 eV	10 eV	10 eV	10 eV	10 eV	10 eV	10 eV	10 eV	10 eV	10 eV	10 eV	10 eV
MS	Collision cell	Collision RF (Vpp)	1800	1800	1800	1800	1800	1800	1800	1800	1800	1800	1800	1800
MS	Collision cell	Transfer time	120 µs	120 µs	120 µs	120 µs	120 µs	120 µs	120 µs	120 µs	120 µs	150 µs	120 µs	120 µs
MS	Collision cell	Pre-pulse storage	10 µs	10 µs	10 µs	10 µs	10 µs	10 µs	10 µs	10 µs	10 µs	15 µs	10 µs	10 µs
MS	MS1	Ion polarity	positive	positive	positive	positive	Positive	Positive	positive	positive	positive	positive	positive	positive
MS	MS1	Mass range (m/z)	600–3000	600–3000	600–3000	200–2000	200–2000	200–2000	200–2000	200–2000	200–2000	200–2000	200–2000	200–2000
MS	MS1	Summation	15,000	15,000	15,000	18,321	18,321	18,321	18,321	18,321	18,321	18,321	18,321	18,321
MS	MS1	Rolling average mode	2	2	2	2	2	2	2	2	2	2	2	2
in source	IS-CID	IS-CID	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES	YES
in source	IS-CID	IS-CID MS (funnel 1 exit)				3 eV	30 eV	50 eV	70 eV	5 eV	3 eV	3 eV	3 eV	10 eV
in source	IS-CID	IS-CID MS/MS (funnel 2 entry)				15 eV	10 eV	10 eV	10 eV	20 eV	10 eV	30 eV	10 eV	15 eV
in source	IS-CID	Acquisition time factor				1	1	1	1	1	1	1	1	1
MS/MS	MS2	Auto MS/MS mode	OFF	ON	ON	OFF	ON	ON	ON	ON	ON	ON	ON	OFF
MS/MS	CID	Acquisition	NO	YES	YES	NO	YES	YES	YES	YES	YES	YES	YES	NO
MS/MS	CID	MS/MS low (per 1000 sum.)		10^4 cts	10^4 cts		10^4 cts	10^4 cts	10^4 cts	10^4 cts	10^4 cts	10^4 cts	10^4 cts	
MS/MS	CID	MS/MS high (per 1000 sum.)		10^6 cts	10^6 cts		10^6 cts	10^6 cts	10^6 cts	10^6 cts	10^6 cts	10^6 cts	10^6 cts	
MS/MS	MS2	Preference activated	NO	NO	NO	NO	NO	NO	NO	NO	YES	NO	NO	NO

Table 1. *Cont.*

Mode	Step	Parameter	Method 1	Method 2	Method 3	Method 4	Method 5	Method 6	Method 7	Method 8	Method 9	Method 10	Method 11	Method 12
MS/MS	MS2	Charge state preferred range									7–15			
MS/MS	MS2	Exclude singly									YES			
MS/MS	Precursor ions	Exclude (*m*/*z*)		100–600	100–600		100–600	100–600	100–600	100–600	100–600	100–600	100–600	
MS/MS	Precursor ions	No. of precursors		2	2		2	2	2	3	4	4	4	
MS/MS	Threshold	Absolute (per 1000 sum.)		1153 cts	170 cts		10 cts	10 cts	10 cts	10 cts	10 cts	10 cts	10 cts	
MS/MS	Threshold	Absolute		17,302 cts	2550 cts		183 cts	183 cts	183 cts	183 cts	183 cts	183 cts	183 cts	
MS/MS	Active exclusion	ON/OFF		ON	ON		ON	ON	ON	ON	ON	ON	ON	
MS/MS	Active exclusion	Exclude after		3 spectra	3 spectra		1 spectra	1 spectra	1 spectra	1 spectra	1 spectra	1 spectra	1 spectra	
MS/MS	Active exclusion	Release after		5 min	5 min		5 min	5 min	5 min	5 min	5 min	5 min	5 min	

Figure 3 displays the LC-MS patterns for each of the 12 methods using the mixed protein standards sample.

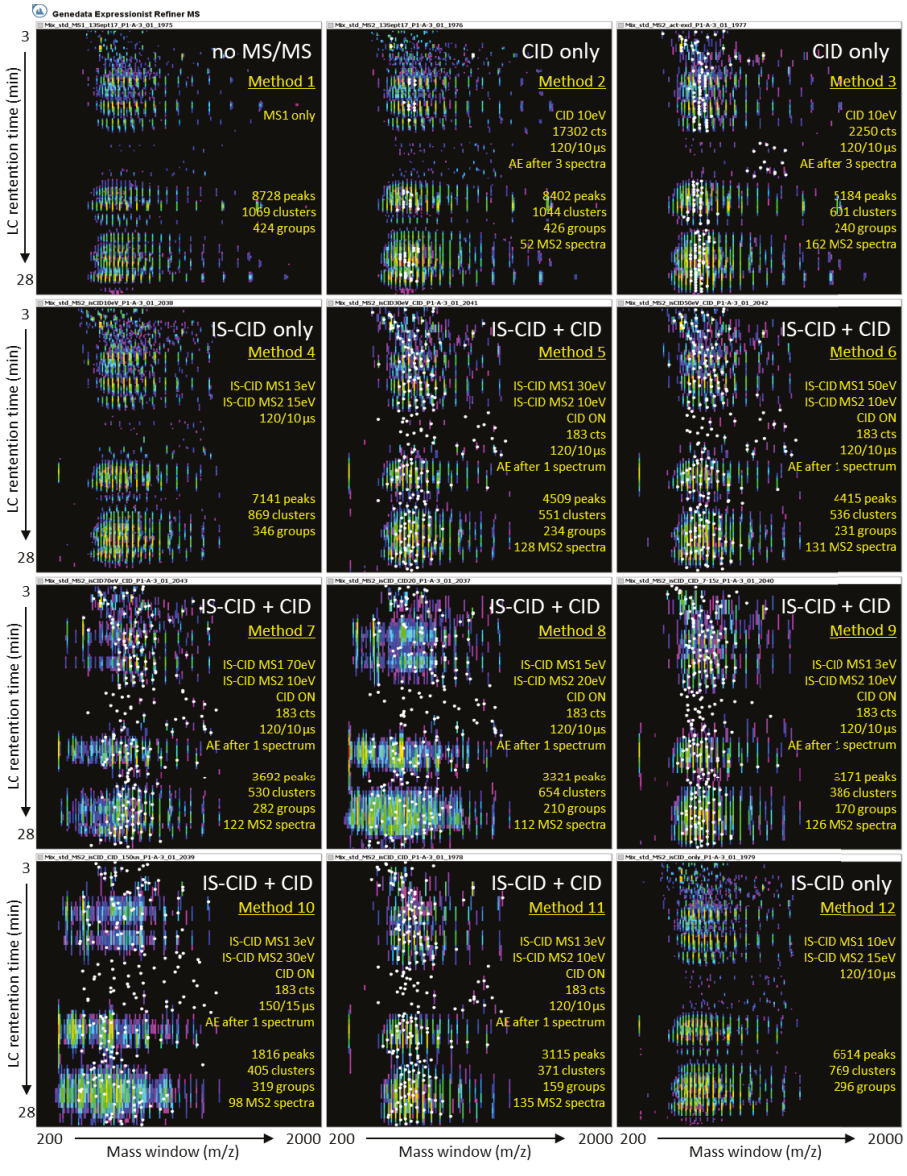


Figure 3. LC-MS maps of the mixed protein standards sample visualised in Genedata Refiner with the scanned mass range (200–2000 m/z) on the x axis and LC retention time (3–28 min) on the y axis across all 12 methods following data processing. White dots represent true MS/MS (MS2) events using conventional CID fragmentation. AE, active exclusion.

Table 2 compares all 12 methods based on LC-MS and LC-MS/MS observations

Table 2. Number of spectral peaks, clusters, groups and MS/MS spectra found in the mixed standard sample across all 12 methods using Genedata Refiner workflow.

Method	1	2	3	4	5	6	7	8	9	10	11	12
LC separation (min)	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5
LC separation (s)	1350	1350	1350	1350	1350	1350	1350	1350	1350	1350	1350	1350
Duty cycle (s)	2.7	13.7	13.7	5.6	13.7	13.7	13.8	19.1	24.6	24.6	24.6	5.5
No. of full MS1 scans	493	99	99	243	99	99	98	71	55	55	55	247
No. of precursors		2	2		2	2	2	3	4	4	4	
(a) Theor. no. of MS2 events	0	197	198	0	197	198	196	212	219	219	220	0
(b) MS1 peaks (300–3000 *m/z*)	8728	8402	5184	7141	4509	4415	3692	3321	3171	1816	3115	6614
(c) MS1 peaks (700–3000 *m/z*)	8711	8387	5184	7124	4503	4415	3599	3040	3169	1628	3113	6599
(d) MS1 peaks (300–700 *m/z*)	17	15	0	17	6	0	93	281	2	188	2	15
(d)/(b) (%)	0.2	0.2	0.0	0.2	0.1	0.0	2.5	8.5	0.1	10.4	0.1	0.2
MS1 clusters	1069	1044	601	869	551	536	530	654	386	405	371	769
MS1 groups	424	426	240	346	234	231	282	487	170	319	159	296
(e) Observed No. of MS2 spectra	0	56	170	0	196	196	196	210	218	218	218	0
(e)/(a) (%)		28	86		99	99	100	99	99	99	99	
(e)/(b) (%)		0.7	3.3		4.3	4.4	5.3	6.3	6.9	12.0	7.0	
MS2 *m/z* range		730–1636	644–1854		643–1955	703–1955	657–1966	609–1913	607–1774	609–1966	634–1928	
600–700 *m/z*			1		3		2	18	3	20	4	
700–800 *m/z*		2	2		5	5	5	14	4	19	8	
800–900 *m/z*		4	12		20	20	15	22	26	27	29	
900–1000 *m/z*		20	53		45	44	21	34	56	36	54	
1000–1100 *m/z*		28	77		53	51	47	23	63	24	57	
1100–1200 *m/z*			5		29	26	35	24	30	18	23	
1200–1300 *m/z*			2		19	21	26	26	20	29	18	
1300–1400 *m/z*			2		3	7	17	14	8	7	2	
1400–1500 *m/z*			2		2	5	4	11	2	10	5	
1500–1600 *m/z*			2		3	4	5	8	3	9	3	
1600–1700 *m/z*		2	6		4	4	7	7	1	8	7	
1700–1800 *m/z*			3		4	3	3	1	2	5	5	
1800–1900 *m/z*			3		4	4	7	7		4	1	
1900–2000 *m/z*					2	2	2	1		2	2	

2.2.1. Precursor Intensity Threshold Is a Key Parameter for CID Fragmentation

Methods 2 and 3 (CID only) only differed in the intensity threshold for precursor selection, being almost 7 times higher in Method 2 (2550 counts) relative to Method 3 (17,302 counts) (Table 1). This allows precursor ions of low signal intensity to undergo CID fragmentation. As expected, this did not affect the LC-MS1 data; however, it did result in many more MS/MS events when Method 3 was used. If we consider the mixed standard sample as an example, Method 2 resulted in 8402 ions with only 56 (0.7%) MS2 spectra, whereas Method 3 yielded 5184 peaks and 170 (3.3%) MS2 spectra (Table 2). Consequently, the top-down sequencing efficiency varied between Methods 2 and 3, as will be discussed in the following section.

Charge state preference aside, in this study we did not attempt to optimise CID conditions and took full advantage of the automatic MS/MS mode available on our instrument with default parameters as it proved efficient [34]. Using CID on its own, Nemeth-Cawley and colleagues successfully top-down sequenced unknown intact proteins purified by affinity and size-exclusion chromatography from mammalian cells using an ESI-Q-ToF MS/MS strategy [27].

2.2.2. Ion Funnel 1 Energy Has Little Impact on IS-CID Fragmentation

Methods 4 and 12 (IS-CID only) differed only in the energy applied to the ion funnel 1 exit during the IS-CID MS1 step. Method 4 applied 3 eV and Method 12 applied 10 eV at ion funnel 1 (Table 1). LC-MS patterns displayed only slight changes, with Method 4 yielding 7141 peaks, 869 clusters, and 346 groups, and Method 12 producing 6614 peaks, 769 clusters, and 296 groups (Table 2 and

Figure 3). Protein annotation by top-down sequencing was also comparable, as will be described in the next chapter.

2.2.3. Charge State Preference Does Not Influence Precursor Selection

Methods 9 and 11 (IS-CID + CID) only differed in their charge state preferred range. No charge state was specified in Method 11 (Table 1). In Method 9, a preference for the most abundant ions bearing 7-15 charges was applied to assess whether it would favour the selection of highly protonated ions for CID fragmentation. This parameter had no impact on the selection process of parent ions since both methods generated similar numbers of peaks, clusters and groups and, most importantly, an identical number of MS2 spectra (218) (Table 2). This indicated that the automatic MS2 mode of the Q-ToF aptly selected highly charged protein ions for subsequent CID fragmentation.

2.2.4. A Long Duty Cycle Minimises MS/MS Events

To maximise signal sensitivity and minimise background noise, high summation values were utilised, which resulted in duty cycles that lasted at least 2.7 s (MS1 only—Method 1) (Table 2). Furthermore, these duty cycles were extended upon MS/MS analyses in a method-specific fashion (from 5.5 s in Methods 4 and 12, 13.7 s in Methods 2–3 and 5–6, to 24.6 s in Methods 9–11), which restricted the number of full MS1 scans (Table 2). Consequently, it limited how many ions could be selected as parents for the CID process during the LC run.

Under our LC-MS/MS conditions, protein separation occurred from 2.5 to 25 min (22.5 min or 1350 s) with two to four precursors selected per duty cycle. The maximum number of MS2 events were computed and reported in Table 2. Apart from the CID-only methods, which only realised 28% (56/197 in Method 2) and 86% (170/198 in Method 3) of their potential due to high threshold requirements, all the other methods delivered 99–100% of the expected number of MS2 events (Table 2). However, these numbers are low. Instruments of high resolution offering a faster scanning capability would therefore be extremely advantageous.

In our study, the slow scanning rate was partially alleviated by resorting to IS-CID only (Methods 4 and 12), whose fast fragmentation mode allowed for shorter duty cycles (5.5 s) and up to 247 scans during the 22.5-min separation (Table 2).

2.2.5. High Energy Fragmentation Produces Lower m/z Ions

Overall, low-energy fragmentation was applied at the ion funnel 1 exit during the IS-CID MS mode (3, 5 or 10 eV), except for three methods for which high energy was applied. Methods 5–7 applied 30, 50, and 70 eV, respectively (Table 1). The consequences of high-energy IS-CID become visible upon reaching 70 eV (Method 7), which generates many more ions of less than 700 m/z, indicative of protein fragmentation. For instance, in the mixed standard sample, Method 1 (MS1 only) resulted in 8728 peaks, with the vast majority above 700 m/z (8711, 99.8%). Method 5 (30 eV) produced 4509 peaks in total with only 6 (0.1%) below 700 m/z. Conversely, Method 7 (70 eV) produced 3692 peaks in total, of which 93 (2.5%) were of less than 700 m/z (Table 2).

Similarly, low-energy fragmentation (10 or 15 eV) was applied at the ion funnel 2 entrance during the IS-CID MS/MS mode, except for two methods for which greater energy was delivered. Methods 8 and 10 applied 20 and 30 eV, respectively (Table 1). Increasing the energy level during the IS-CID MS/MS step proved even more effective at fragmenting intact proteins within the ESI source, as more spectral peaks of low m/z were created (Figure 3). For instance, in the mixed standard sample, applying 20 eV (Method 8) or 30 eV (Method 10) resulted in the proportion of peaks below 700 m/z reaching 8.5% (281/3321 peaks) and 10.4% (188/1816 peaks) for Methods 8 and 10, respectively (Table 2). This resulted in ions of lower m/z being selected as precursors for a subsequent CID process. In the mixed standard sample, only three (1.5%) MS2 spectra resulted from precursors of less than 700 m/z using Method 5, while 20 (9.2%) MS2 spectra were acquired from precursors of less than 700 m/z

using Method 10 (Table 2). However, as will be reported in the next chapter, such enhanced in-source fragmentation did not benefit the top-down sequencing output.

2.2.6. CID Versus IS-CID

While CID parameters mostly revolve around ion features (e.g., intensity threshold, charge state, number of precursors) that allow for relatively controlled and robust conditions, IS-CID parameters themselves pertain to factors operating globally within the ESI source and the ion transfer region, the most important one being the DC voltage applied, which affects signal sensitivity. Owing to its less controllable nature, the IS-CID fragmentation process is more complex and far less predictable than MS/MS results. Therefore, a prerequisite of IS-CID efficacy is good chromatographic resolution with little or no background contamination [16].

Our ESI-Q-ToF mass spectrometer establishes IS-CID fragmentation not within the ESI source per se but more accurately within the ion transfer area called funnel 2. The two funnel-staged ion transfer region separates the ions from the drying gas and solvent, as well as transfers these ions, with minimal losses, to the quadrupole stage. The first and second funnel stages are separated by a DC plate. By increasing the DC potentials of funnel 1, the ions are accelerated into funnel 2, which in turn activates IS-CID. This ingenious, albeit unpredictable, fragmentation mode was invented 30 years ago and reported to favour highly charged species at lower energetic collision (i.e., cone voltage) akin to conventional CID process [5]. As with CID, as the ions' m/z increase, the IS-CID efficacy decreases. That was confirmed in our experiment. Parent ions were predominantly sampled from the 800–1300 m/z range, where the most intense ions resolved, and only a handful of precursors arose from 1600 m/z and above (Table 2).

IS-CID in combination with conventional CID on a ESI-Q-ToF system was tested at increasing cone voltage conditions (from 45 to 90 V) by Ginter and colleagues (2004) to top-down sequence a mixture of seven protein standards; they reported that some proteins necessitated higher voltage than others for in-source fragmentation to occur [28]. In a different study where only IS-CID was applied to generate MS2 spectra using a ESI-Q-ToF instrument, source cone voltage varied from 20 to 70 V and was evaluated on a sample mixture of 13 protein standards [29]. The authors also concluded that IS-CID efficient fragmentation greatly depended on cone voltage; in their experience, 40–60 V proved the most optimal range.

2.3. Top-Down Sequencing of Milk Proteins

In the present study, a total of 11 MS/MS methods (Methods 2–12) were compared and the ultimate validation criterium was the coverage depth resulting from top-down sequencing annotation. The nine most abundant proteins known from milk were chosen to systematically assess AA sequence coverage across the 11 methods using Genedata Expressionist program. These proteoforms were α_{S1}-CN B-8P, α_{S2}-CN A-11P, β-CN A1, β-CN A2, κ-CN A-1P, κ-CN B-1P, α-LA B, β-LG A, and β-LG B. Table 3 documents the number of AAs top-down sequenced for each protein from either individual or mixed standards, or milk samples from Jerseys or Holsteins.

Results were consistent from sample to sample, particularly for the proteins that responded very well to top-down sequencing such as β-CN A2 with a coefficient of variation (CV) as low as 0.4%.

For ease of interpretation, the values reported in Table 2 were converted to percentages and displayed as histograms (Figure 4).

Table 3. Number of AAs top-down sequenced across the 11 MS2 methods from the nine most abundant milk proteins with the individual standards, mixed standards, and milk samples.

Protein	Variant	PTM	Code	Mo. Mass (Da)	Length (AAs)	RT (min)	Sample 1	Method 2	Method 3	Method 4	Method 5	Method 6	Method 7	Method 8	Method 9	Method 10	Method 11	Method 12
α_{S1}-CN	B	8P	α_{S1}-CN B 8P	23,600.3	199	15–18	I	196	196	195	194	195	190	70	195	7	194	195
α_{S2}-CN	A	11P	α_{S2}-CN A 11P	25,213.0	207	6.8–8.3	I	14	6	83	6	6	6	6	6	0	6	83
β-CN	A1	5P	β-CN A1 5P	24,008.2	209	21–22	I	204	204	197	205	205	203	197	205	95	205	195
β-CN	A2	5P	β-CN A2 5P	23,968.2	209	21.5–23	I	202	202	201	203	203	201	186	203	92	203	200
κ-CN	A	1P	κ-CN A 1P	19,026.5	169	6.5–8.5	I	160	160	165	157	157	156	119	157	31	157	165
κ-CN	B	1P	κ-CN B 1P	18,993.6	169	8.5–10	I	162	162	157	160	159	157	133	161	59	161	154
α-LA	B		α-LA B	14,176.8	123	15–18	I	98	97	84	94	100	98	90	97	86	97	83
β-LG	A		β-LG A	18,355.5	162	22–24	I	162	162	161	155	155	153	153	162	151	162	161
β-LG	B		β-LG B	18,269.4	162	20–22	I	159	158	159	153	154	149	155	159	146	159	158
α_{S1}-CN	B	8P	α_{S1}-CN B 8P	23,600.3	199	15–18	M	178	178	192	179	177	173	91	178	52	181	192
α_{S2}-CN	A	11P	α_{S2}-CN A 11P	25,213.0	207	6.8–8.3	M	121	123	150	111	116	108	52	115	8	115	151
β-CN	A1	5P	β-CN A1 5P	24,008.2	209	21–22	M	171	172	168	162	167	156	79	163	27	165	174
β-CN	A2	5P	β-CN A2 5P	23,968.2	209	21.5–23	M	201	201	201	203	201	200	182	202	43	203	201
κ-CN	A	1P	κ-CN A 1P	19,026.5	169	6.5–8.5	M	108	109	122	106	109	93	41	106	14	107	124
κ-CN	B	1P	κ-CN B 1P	18,993.6	169	8.5–10	M	122	125	112	120	120	110	33	122	10	121	116
α-LA	B		α-LA B	14,176.8	123	15–18	M	86	85	96	83	82	77	37	83	21	86	94
β-LG	A		β-LG A	18,355.5	162	22–24	M	158	157	156	158	158	157	135	157	62	157	155
β-LG	B		β-LG B	18,269.4	162	20–22	M	111	112	126	106	107	96	42	106	20	111	125
α_{S1}-CN	B	8P	α_{S1}-CN B 8P	23,600.3	199	15–18	J	194	195	198	195	195	189	137	194	97	195	198
α_{S2}-CN	A	11P	α_{S2}-CN A 11P	25,213.0	207	6.8–8.3	J	92	140	177	122	110	97	18	114	20	126	177
β-CN	A1	5P	β-CN A1 5P	24,008.2	209	21–22	J	206	206	205	206	206	205	191	206	110	206	204
β-CN	A2	5P	β-CN A2 5P	23,968.2	209	21.5–23	J	199	199	201	200	200	199	188	199	106	199	201
κ-CN	A	1P	κ-CN A 1P	19,026.5	169	6.5–8.5	J	82	132	151	121	107	104	12	110	30	119	154
κ-CN	B	1P	κ-CN B 1P	18,993.6	169	8.5–10	J	164	164	157	164	163	164	146	164	78	163	157
α-LA	B		α-LA B	14,176.8	123	15–18	J	120	120	121	120	120	111	62	120	38	120	121
β-LG	A		β-LG A	18,355.5	162	22–24	J	154	154	153	154	154	154	139	154	73	154	153
β-LG	B		β-LG B	18,269.4	162	20–22	J	158	156	156	158	156	154	134	156	76	156	155
α_{S1}-CN	B	8P	α_{S1}-CN B 8P	23,600.3	199	15–18	H	191	191	197	183	184	177	133	185	89	183	197
α_{S2}-CN	A	11P	α_{S2}-CN A 11P	25,213.0	207	6.8–8.3	H	172	181	192	179	179	172	65	178	88	180	193
β-CN	A1	5P	β-CN A1 5P	24,008.2	209	21–22	H	204	204	200	204	204	204	181	204	170	204	201
β-CN	A2	5P	β-CN A2 5P	23,968.2	209	21.5–23	H	206	207	203	207	207	207	198	207	168	207	203
κ-CN	A	1P	κ-CN A 1P	19,026.5	169	6.5–8.5	H	147	150	165	147	153	133	54	151	72	150	165
κ-CN	B	1P	κ-CN B 1P	18,993.6	169	8.5–10	H	89	97	118	91	86	82	11	89	17	95	118
α-LA	B		α-LA B	14,176.8	123	15–18	H	118	119	119	106	105	94	52	107	23	104	119
β-LG	A		β-LG A	18,355.5	162	22–24	H	157	159	156	158	158	155	136	156	97	157	155
β-LG	B		β-LG B	18,269.4	162	20–22	H	162	162	159	162	162	160	138	162	128	162	158

I, individual standard; M, mixed standard; J, Jersey milk; H, Holstein milk.

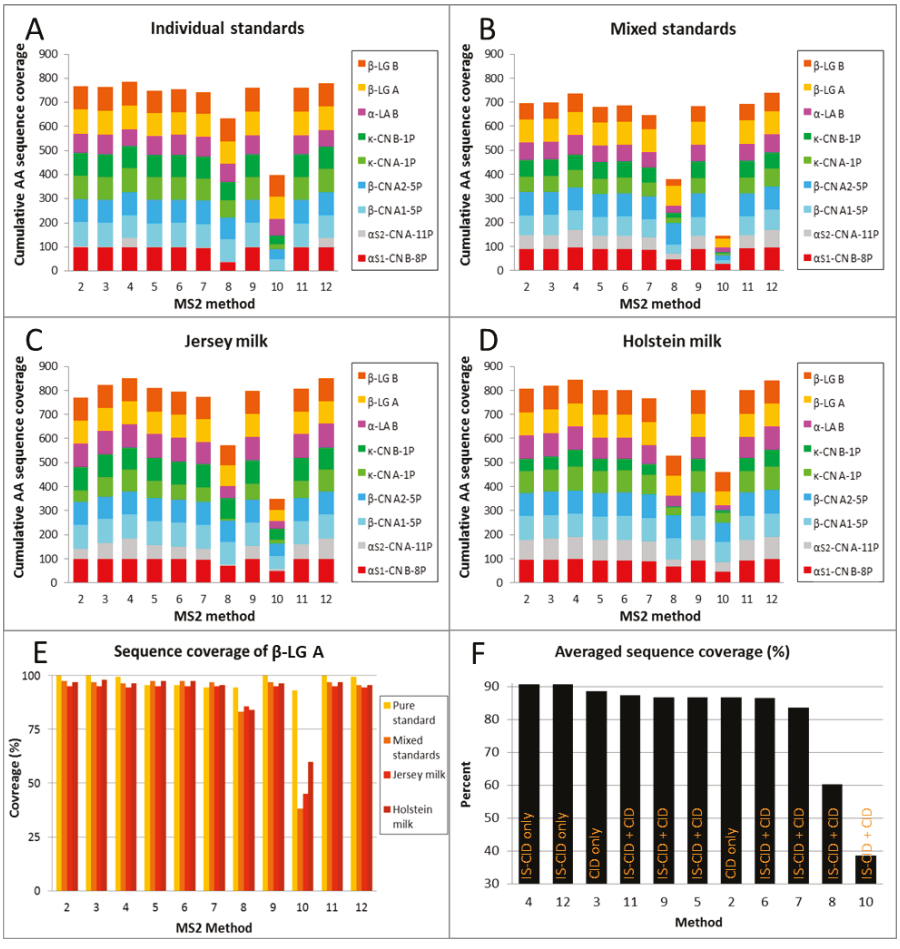


Figure 4. Comparison of the 11 MS/MS methods based on the AA sequence coverage in percent of the nine most prominent milk proteins. (**A**) Cumulative coverage in the individual standards. (**B**) Cumulative coverage in the mixed standards. (**C**) Cumulative coverage in the Jersey milk sample. (**D**) Cumulative coverage in the Holstein milk sample. (**E**) Illustration of AA sequence coverage for one particular protein, β-LG A, in increasingly complex samples. (**F**) Averaged sequence coverage for all eight proteins sorted by method efficacy; the fragmentation modes are reported as well.

2.3.1. Reproducibility and Protein Specificity

While being both protein- and method-specific, top-down sequencing results were reproducible across samples of increasing complexity, from the least (individual standards, Figure 4A) and mildly (mixed standards, Figure 4B) complex, to the most complex samples (raw milk, Figure 4C–D). The sample reproducibility was also demonstrated on one particular protein, β-LG A, which exhibited CV values spanning from 1.0% (Method 7) to 1.9% (Methods 9 and 12), omitting Methods 8 and 10, which were not only the least efficient but also the least reproducible (Figure 4E). As observed in previous studies [28,29], top-down sequencing success is protein-dependent. In our conditions, α_{S2}-CN A-10P was the least responsive, possibly due to its elevated number of phosphorylation sites, while β-LGs and β-CNs were the most responsive.

2.3.2. Low-Energy Fragmentation Drives Top-Down Sequencing Efficacy

The histograms in Figure 4 also clearly demonstrate which methods were suitable for top-down sequencing and which were not. In Figure 4F, the methods have been sorted in descending order of AA sequence coverage. Under our conditions, the best methods were Methods 4 and 12 (90.6% sequence coverage on average), closely followed by Methods 3 (88.6%), 11 (87.4%), 9 (86.8%), 5 and 2 (86.7%), 6 (86.6%), and 7 (83.6%). The worst-performing methods were 8 (60.2%) and 10 (38.6%). Based on our observations, we can conclude that the methods that resorted to IS-CID on its own with little energy applied, in combination with CID or not, were best suited to top-down sequence prominent milk proteins.

Optimal conditions in our case involved applying 3 eV (Method 4) or 10 eV (Method 12) to the ion funnel 1 exit during the IS-CID MS step and 15 eV to the ion funnel 2 entrance during the IS-CID MS/MS step. What seemed to be detrimental for our purposes was using too much energy (20 eV in Method 8 and 30 eV in Method 10) during the IS-CID MS/MS step. Moreover, we do not recommend applying an energy level higher than 10 eV during the IS-CID MS step, as was tested in Methods 5 (30 eV), 6 (50 eV), and 7 (70 eV).

2.3.3. AA Position Has No Influence on Fragmentation

In an attempt to further explain how CID-based fragmentation operated under our conditions, we aligned side by side all the AA sequences of the nine proteins of interest for each of the 11 MS/MS methods and blackened the AAs that were successfully top-down sequenced. For each protein, the sequenced AAs were counted across the methods and, as a result, we generated a score from 0 (not sequenced at all) to 11 (sequenced in all methods). Based on this scoring system, the corresponding position in the sequence was coloured red (the higher the score, the darker the shade). This is represented in Supplementary Figure S1. This display allowed for the quick visualisation of regions that remained unfragmented under our various conditions. Such recalcitrant regions were scattered along the protein sequence, suggesting that the AA position had no influence over fragmentation efficiency.

All cow's milk caseins are heavily phosphorylated. There was no visible pattern associated with phosphorylation sites (highlighted in yellow) either, apart from α_{S2}-CN A-10P, for which five phosphorylated serine residues (S at positions 7, 13, 31, 56, and 143) resisted fragmentation. PTMs were not investigated in this study. Milk proteins are also glycosylated, as has been well documented for κ-CN, for instance [36]. Glycosylation sites could account for areas difficult to fragment, but we could not test this hypothesis in the present study.

2.3.4. Hydrophobicity Affects Fragmentation

To further elucidate whether AAs themselves influenced fragmentation efficiency, we converted the AA counts from Supplementary Figure S1 into percentages, sorted them according to their highest fragmentation efficiency (meaning sequenced across all 11 MS/MS methods) and plotted them (Supplementary Figure S2). Underneath the chart we listed the physical properties attributed to AAs (ThermoFisher Scientific website) to assist us in finding a pattern.

In this representation, leucine, glycine, cysteine, tryptophan, histidine, and isoleucine displayed the best response to top-down sequencing with a success rate above 17% across 11 MS/MS methods (Supplementary Figure S2). These AAs are moderately to highly hydrophobic. Conversely, threonine, asparagine and arginine showed the lowest success rate (less than 9% across the 11 methods). Such AAs are hydrophilic. Our results suggest that AA hydrophobicity level has an impact on fragmentation efficiency.

2.3.5. High Sequence Coverage Is Critical for Allelic Variants

High sequence coverage is a prerequisite to successfully distinguishing between protein allelic variants such as β-CN A1 and A2, which only vary at position 67 (H to P); κ-CN A and B, which

vary at positions 136 (T to I) and 148 (D to A); and β-LG A and B, which vary at positions 64 (D to G) and 118 (V to A). The mixed standard sample, Method 10, and, in the case of κ-CNs, Method 8 failed to deliver such relevant information since AAs at the aforementioned positions were not sequenced (Supplementary Figure S1).

Interestingly, some genetic variants proved more amenable to top-down sequencing than others. For instance, in the mixed standard sample processed with Method 12, β-CN A2 (97%) was more thoroughly sequenced than β-CN A1 (83%). Likewise, β-LG A (96%) displayed greater sequence coverage than β-LG B (77%); this was also observed for κ-CN A (73%) and κ-CN B (68%) (Table 3 and Figure S1). In Holstein milk samples, we managed to reach 100% sequence coverage of β-LG B in six out of 11 MS/MS methods. Complete top-down sequencing was also achieved for β-LG A in individual standard samples in four MS/MS methods (Table 3).

In the early days of top-down proteomics on ESI-Q-ToF platforms, the resulting short sequence tags from C- and N-termini from small to medium-sized proteins were just long enough to unambiguously identify the protein family, but not their allelic variants [26–30]. Almost complete top-down sequencing was achieved in 2009 with 17 kDa horse myoglobin and 30 kDa bovine carbonic anhydrase II being 92% and 96% sequenced, respectively [31]. Since then and to our knowledge, Q-ToF instruments have not been employed for top-down sequencing of intact proteins from complex biological matrices because they have been superseded by more expensive platforms such as FT mass spectrometers, offering better power resolution and alternative fragmentation methods like ETD, HCD, and ECD.

3. Materials and Methods

3.1. Materials and Sample Preparation

The experimental design is schematised in Figure 1. Protein standards purchased from Sigma-Aldrich (Castle Hill, NSW, Australia) were alpha-casein (α-CN) from bovine milk (C6780-250MG, 70% pure), beta-casein (β-CN) from bovine milk (C6905-250MG, 98% pure), kappa-casein (κ-CN) from bovine milk (C0406-250MG, 70% pure), alpha-lactalbumin (α-LA) from bovine milk (L5385-25MG, 85% pure), and beta-lactoglobulin (β-LG) from bovine milk (L3908-250MG, 90% pure). These lyophilised standards were prepared as described in [23]. Briefly, they were fully solubilised at a 10 mg/mL concentration in 50% MilliQ (MerckMillipore, Bayswater, VIC, Australia) water/50% solution A (0.1 M Bis-Tris, 6 M Guanidine-HCl, 5.37 mM sodium citrate tribasic dehydrate, and 20 mM DTT). A volume of 50% acetic acid to reach 1% acetic acid final concentration was added to the standards. A 0.1-mL aliquot of the solubilised standard was transferred into a 100-µL glass insert placed in a 2-mL glass vial for immediate analysis by LC-MS. A standard mixture was prepared by mixing individual standards in the following proportions to account for various ionisation efficiency [23]: 25% α-CN, 25% κ-CN, 20% β-LG, 20% β-CN, and 10% α-LA.

Milk collection from Holstein-Friesian cows (coded H) and Jersey cows (coded J) was described in [35]. Milk sample preparation was described in [23]. Briefly, 0.5 mL of cold skim milk was transferred into a 1.5-mL tube and 0.5 mL of Solution A was added. A 0.02 mL volume of 50% acetic acid (1% acetic acid final concentration, pH 5.8) was then added. A 0.1-mL aliquot of the milk protein extract was transferred into a 100-µL glass insert placed in a 2-mL glass vial for immediate analysis by LC-MS.

3.2. HPLC Separation of Intact Proteins

The separation of intact proteins by UHPLC 1290 Infinity Binary LC system (Agilent) through a Aeris™ WIDEPORE XB-C8 (3.6 µm particle size, 200 Å pore size, 150 × 2.1 mm dimensions, C8 reverse phase core-shell silica from Phenomenex (Lane Cove, NSW, Australia) column at 75 °C was described in [23]. In brief, 3 µL of sample was injected and separated as followed: starting conditions 20% B, ramping to 28% B in 2.5 min, ramping to 40% B in 27.5 min, ramping to 99% B in 1 min and held for 4 min, lowering to 20% B in 0.1 min, equilibration at 20% B for 4.9 min. Mobile phase flow

rates was 200 µL/min. Mobile phase A contained ACN with 0.1% FA and 0.02% trifluoroacetic acid (TFA), and mobile phase B contained H2O with 0.1% FA and 0.02% TFA. The diode array detector (DAD) spectrum was acquired from 190 to 400 nm. The pressure limit was set at 600 bars.

3.3. MS1 Analysis

MS1 analysis was described in [23]. The UHPLC was on-line with a maXis HD UHR-Q-ToF (60,000 resolution) fitted with a standard ESI Apollo-source (BrukerDaltonikGmbh, Preston, VIC, Australia). To ensure mass accuracy, a Na-formate solution was infused continuously at 0.1 mL/h and the first 2.5 min of each run were used to re-calibrate masses post-acquisition. Each 40-min run was thus segmented as follows: 2.5 min to waste and the following 37.5 min to source. Capillary voltage was set at 4500 V. The nebuliser was set at 1.5 bar. The dry gas was set at 8 L/min. The dry temperature was set at 190 °C. The transfer funnel RF and multipole RF were set at 400 Vpp; no IS-CID energy was applied. The quadrupole ion energy was 5 eV, the collision cell energy was 10eV, and the collision RF 1800 Vpp. The ion cooler transfer time was 120 µs, with a prepulse storage of 10 µs and a RF of 400 Vpp. The ion polarity was positive and scan mode was MS. The rolling average mode was activated and set at 2. This MS1 method is called Method 1 hereafter.

3.4. MS2 Analyses and Top-Down Sequencing

MS/MS experiments were performed using the same LC parameters and MS1 parameters on the Q-TOF mass spectrometer as described above. A total of 11 MS2 methods were tested; their parameters are indicated in Table 1. These methods are referred to as Method 2 to Method 12 hereafter. LC-MS files were visualised using Bruker Compass DataAnalysis version 4.2. MS2 spectra were annotated using Bruker Biotools version 3.2 and SequenceEditor version 3.2. The retrieval of AA sequences was detailed in [23].

3.5. Protein Annotation in Genedata Expressionist

The data files obtained following LC-MS analysis using MS2 methods 2 to 12 were curated in the Refiner MS module of Genedata Expressionist® version 11.0 with the following parameters: (1) Conversion of MS/MS to primary MS. (2) Chromatogram chemical noise subtraction using moving average algorithm and a five-scan window. (3) Spectrum smoothing using a moving average algorithm and a five-point m/z window. (4) Chromatogram peak detection using a five-scan summation window, a minimum peak size of four scans with boundaries merge strategy and five-point maximum merge distance, a curvature-based peak detection with 70% intensity threshold and inflection points boundary determination. (5) Chromatogram isotope clustering using a 0.2 min and 20 ppm tolerances, a peptide isotope shaping with 1-8 protons, a 0.6 log-ration maximum distance, mono-isotopic computation, linear charge dependency with a minimum size ratio of 0.65. (6) Singleton filtering. (7) Peptide mapping with a 50 ppm tolerance, unspecific enzyme, three missed cleavage maximum, 10 AAs minimum length, and variable modifications, as explained hereafter. The peptide mapping was performed using a text file containing all 49 AA sequences in FASTA format of the bovine allelic variants of milk caseins, alpha-lactalbumin, and beta-lactoglobulin (Farrell et al., 2004) with the following variable modifications: pyro-Glu (N-term Q) and Phospho (ST). Some aspects of this workflow are illustrated in Figure S3.

Protein deconvolution was also performed on Method 1 files in the Refiner MS module of Genedata Expressionist® version 11.0 with the following parameters. (1) Data sweep to remove UV data. (2) Chromatogram chemical noise subtraction using moving average algorithm, with a subtraction method, a 71 scan window, and 60% quantile. (3) Chromatogram lock mass using the Na-formate ion series. (4) Retention time range restriction from 4 to 25 min. (5) Intact protein activity using the harmonic suppression deconvolution method with 0.02 Da steps and 5-30 kDa masses. (6) Intensity thresholding using a clipping method with an intensity of 50. (7) Spectrum baseline

subtraction using a 99% quantile and 30 kDa m/z window. (8) Chromatogram retention time alignment using a pairwise alignment-based tree scheme with 50-scan search interval and a gap penalty of 1.

All MS1 and MS2 files are available from the stable public repository MassIVE at the following URL: http://massive.ucsd.edu/ProteoSAFe/datasets.jsp with the accession number MSV000082070.

3.6. Validation and Computational Analyses

Top-down sequencing results were visualised and validated in Genedata Expressionist® version 11.0 Refiner MS module within the Peptide Mapping activity and all its subsidiary tabs as well as using Bruker Biotools version 3.2 and SequenceEditor version 3.2, as exemplified in Figure S3.

Top-down sequencing annotations (including b- and y-ions series) for each of the nine proteins of interest in every sample and each of the 11 MS/MS methods were exported from Refiner MS module as .csv files. The .csv files contained the list of peptides identified for each protein along with the peptide AA lengths and their position in the protein sequence they matched to. The .csv files were imported into Microsoft® Excel 2016 software for further processing. Coverage information was derived from the length and position of the identified peptides relative to their matching protein sequence to produce Table 3 and Figure 4. AA responsiveness to MS2 fragmentation was obtained using the positions of the first and last AA of the identified peptides. These analyses produced Figures S1 and S2.

4. Conclusions

In this experiment, we have fine-tuned IS-CID parameters with and without traditional CID using a LC-ESI-Q-ToF system in order to optimise the top-down sequencing of milk intact proteins. The most efficient methods utilised IS-CID on its own at low energy. Full (100%) sequence covering could be obtained on β-LGs, and overall enough coverage was achieved to distinguish between allelic proteoforms, varying only by one or two AAs. Such information is highly biologically relevant as alleles influence function and phenotypic features. In bovine milk, for instance, 12 allelic variants of beta-caseins have been identified across different breeds and populations [37]. The most common variants, A1 and A2, differ at amino acid position 67 with histidine in A1 and proline in A2 milk, as a result of a single nucleotide difference. One application of this research would be screening for the presence of the A1 variant in cow's milk samples.

Supplementary Materials: The following are available online, Figure S1: Top-down sequencing responsiveness of AAs for each of the eight most abundant proteins from the mixed standard sample, Figure S2: CID-friendly amino acids from top-down sequenced milk proteins, Figure S3: Data processing of MS/MS files using Genedata Refiner and DataAnalysis/Biotools exemplified on the Holstein milk sample processed using Method 3.

Author Contributions: Conceptualisation, D.V.; methodology, D.V. and D.M.; software, D.V. and D.M.; validation, D.V.; formal analysis, D.V.; investigation, D.V.; resources, S.R.; data curation, D.V.; writing—original draft preparation, D.V.; writing—review and editing, D.V., D.M., and S.R.; visualisation, D.V.; supervision, S.R.; project administration, D.V.; funding acquisition, S.R.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

Abbreviation	Definition
AA	Amino acid
ACN	Acetonitrile
AE	Active exclusion
CAD	Collisionally activated dissociation
CID	Collision-induced dissociation
CV	Coefficient of variation

DAD	Diode array detector
DC	Direct current
ECD	Electron capture dissociation
ESI	Electrospray ionisation
ETD	Electron transfer dissociation
eV	Electron-volt
FA	Formic acid
FT	Fourier transform
HCD	Higher-energy collisional dissociation
HPLC	High-performance liquid chromatography
IS-CID	In-source CID
kDa	KiloDalton
keV	Kiloelectron-volt
LC-MS	Liquid chromatography-mass spectrometry
LIT	Linear ion trap
m/z	Mass-to-charge ratio
MS	Mass spectrometry
MS/MS	Tandem MS
MS1	Full scan MS
MS2	Tandem MS
MW	Molecular weight
ppm	Parts per million
PTM	Post-translational modification
Q	Quadrupole
QQQ	Triple quadrupole
Q-ToF	Quadrupole time-of-flight
RF	Radiofrequency
TFA	Trifluoroacetic acid
ToF	Time-of-flight
UHT	Ultra-high temperature
V	Volt
α-LA	Alpha-lactalbumin
αS1-CN	Alpha-S1-casein
αS2-CN	Alpha-S2-casein
β-CN	Beta-casein
β-LG	Beta-lactoglobulin
κ-CN	Kappa-casein

References

1. Kelleher, N.L.; Lin, H.Y.; Valaskovic, G.A.; Aaserud, D.J.; Fridriksson, E.K.; McLafferty, F.W. Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *J. Am. Chem. Soc.* **1999**, *121*, 806–812. [CrossRef]
2. Toby, T.K.; Fornelli, L.; Kelleher, N.L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **2016**, *9*, 499–519. [CrossRef] [PubMed]
3. Dole, M.; Mack, L.L.; Hines, R.L.; Mobley, R.C.; Ferguson, L.D.; Alice, M.B. Molecular Beams of Macroions. *J. Chem. Phys.* **1968**, *49*, 2240–2249. [CrossRef]
4. Fenn, J.B.; Mann, M.; Meng, C.K.; Wong, S.F.; Whitehouse, C.M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246*, 64–71. [CrossRef] [PubMed]
5. Loo, J.A.; Udseth, H.R.; Smith, R.D. Collisional Effects on the Charge Distribution of Ions from Large Molecules, Formed by Electrospray-ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 207–210. [CrossRef]
6. Mann, M.; Meng, C.K.; Fenn, J.B.F. Interpreting Mass Spectra of Multiply Charged Ions. *Anal. Chem.* **1989**, *61*, 1702–1708. [CrossRef]

7. Jones, A.W.; Cooper, H.J. Dissociation techniques in mass spectrometry-based proteomics. *Analyst* **2011**, *136*, 3419–3429. [CrossRef] [PubMed]
8. McLafferty, F.W.; Bryce, T.A. Metastable-ion characteristics: Characterization of isomeric molecules. *Chem. Commun.* **1967**, *0*, 1215–1217. [CrossRef]
9. Jennings, K.R. Collision-induced decompositions of aromatic molecular ions. *Int. J. Mass Spectrom. Ion. Phys.* **1968**, *1*, 227–235. [CrossRef]
10. Garcia, B.A. What does the future hold for Top Down mass spectrometry? *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 193–202. [CrossRef] [PubMed]
11. Pamreddy, A.; Panyala, N.R. Top-down proteomics: Applications, recent developments and perspectives. *J. Appl.'Bioanal.* **2016**, *2*, 52–75. [CrossRef]
12. Katta, V.; Chowdhury, S.K.; Chait, B.T. Use of a single-quadrupole mass spectrometer for collision-induced dissociation studies of multiply charged peptide ions produced by electrospray ionization. *Anal. Chem.* **1991**, *63*, 174–178. [CrossRef] [PubMed]
13. Bure, C.; Lange, C. Comparison of dissociation of ions in an electrospray source, or a collision cell in tandem mass spectrometry. *Curr. Org. Chem.* **2003**, *7*, 1613–1624. [CrossRef]
14. Van Dongen, W.D.; van Wijk, J.I.; Green, B.N.; Heerma, W.; Haverkamp, J. Comparison between collision induced dissociation of electrosprayed protonated peptides in the up-front source region and in a low-energy collision cell. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1712–1716. [CrossRef]
15. Senko, M.W.; Beu, S.C.; McLafferty, F.W. High-resolution tandem mass spectrometry of carbonic anhydrase. *Anal. Chem.* **1994**, *66*, 415–418. [CrossRef] [PubMed]
16. Parcher, J.F.; Wang, M.; Chittiboyina, A.G.; Khan, I.A. In-source collision-induced dissociation (IS-CID): Applications, issues and structure elucidation with single-stage mass analyzers. *Drug Test. Anal.* **2018**, *10*, 28–36. [CrossRef] [PubMed]
17. Kelleher, N.L. Top-down proteomics. *Anal. Chem.* **2004**, *76*, 197A–203A. [CrossRef] [PubMed]
18. Loo, J.A.; Edmonds, C.G.; Smith, R.D. Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry. *Science* **1990**, *248*, 201–204. [CrossRef] [PubMed]
19. Loo, J.A.; Quinn, J.P.; Ryu, S.I.; Henry, K.D.; Senko, M.W.; McLafferty, F.W. High-resolution tandem mass spectrometry of large biomolecules. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 286–289. [CrossRef] [PubMed]
20. Beu, S.C.; Senko, M.W.; Quinn, J.P.; McLafferty, F.W. Improved fourier-transform ion-cyclotron-resonance mass spectrometry of large biomolecules. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 190–192. [CrossRef]
21. Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K.G.; Thomson, B.; Wilm, M.; Mann, M. Rapid 'de novo' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1015–1024. [CrossRef]
22. Morris, H.R.; Paxton, T.; Dell, A.; Langhorne, J.; Berg, M.; Bordoli, R.S.; Hoyes, J.; Bateman, R.H. High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 889–896. [CrossRef]
23. Vincent, D.; Elkins, A.; Condina, M.R.; Ezernieks, V.; Rochfort, S. Quantitation and Identification of Intact Major Milk Proteins for High-Throughput LC-ESI-Q-TOF MS Analyses. *PLoS ONE* **2016**, *11*, e0163471. [CrossRef] [PubMed]
24. Domon, B.; Aebersold, R. Mass spectrometry and protein analysis. *Science* **2006**, *312*, 212–217. [CrossRef] [PubMed]
25. Steen, H.; Kuster, B.; Mann, M. Quadrupole time-of-flight versus triple-quadrupole mass spectrometry for the determination of phosphopeptides by precursor ion scanning. *J. Mass Spectrom.* **2001**, *36*, 782–790. [CrossRef] [PubMed]
26. Nemeth-Cawley, J.F.; Rouse, J.C. Identification and sequencing analysis of intact proteins via collision-induced dissociation and quadrupole time-of-flight mass spectrometry. *J. Mass Spectrom.* **2002**, *37*, 270–282. [CrossRef] [PubMed]
27. Nemeth-Cawley, J.F.; Tangarone, B.S.; Rouse, J.C. "Top Down" characterization is a complementary technique to peptide sequencing for identifying protein species in complex mixtures. *J. Proteome Res.* **2003**, *2*, 495–505. [CrossRef] [PubMed]

28. Ginter, J.M.; Zhou, F.; Johnston, M.V. Generating protein sequence tags by combining cone and conventional collision induced dissociation in a quadrupole time-of-flight mass spectrometer. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1478–1486. [CrossRef] [PubMed]
29. Johnson, R.W.J.; Ahmed, T.F.; Miesbauer, L.J.; Edalji, R.; Smith, R.; Harlan, J.; Dorwin, S.; Walter, K.; Holzman, T. Protein fragmentation via liquid chromatography-quadrupole time-of-flight mass spectrometry: The use of limited sequence information in structural characterization. *Anal. Biochem.* **2005**, *341*, 22–32. [CrossRef] [PubMed]
30. Deepalakshmi, P.D. Top-down approach of completely sequencing a 4.9 kDa recombinant peptide using quadrupole time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 3747–3754. [CrossRef] [PubMed]
31. Armirotti, A.; Benatti, U.; Damonte, G. Top-down proteomics with a quadrupole time-of-flight mass spectrometer and collision-induced dissociation. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 661–666. [CrossRef] [PubMed]
32. Liu, J.; Jiang, J.; Wu, Z.; Xie, F. Antimicrobial peptides from the skin of the Asian frog, Odorrana jingdongensis: De novo sequencing and analysis of tandem mass spectrometry data. *J. Proteomics* **2012**, *75*, 5807–5821. [CrossRef] [PubMed]
33. Boeri Erba, E. Investigating macromolecular complexes using top-down mass spectrometry. *Proteomics* **2014**, *14*, 1259–1270. [CrossRef] [PubMed]
34. Raynes, J.K.; Vincent, D.; Zawadzki, J.L.; Savin, K.; Logan, A.; Mertens, D.; Williams, R.P.W. Investigation of age gelation UHT Milk. *Beverages* **2018**. submitted for publication.
35. Vincent, D.; Ezernieks, V.; Elkins, A.; Nguyen, N.; Moate, P.J.; Cocks, B.G.; Rochfort, S. Milk Bottom-Up Proteomics: Method Optimization. *Front. Genet.* **2015**, *6*, 360. [CrossRef] [PubMed]
36. O'Riordan, N.; Kane, M.; Joshi, L.; Hickey, R.M. Structural and functional characteristics of bovine milk protein glycosylation. *Glycobiology* **2014**, *24*, 220–236. [CrossRef] [PubMed]
37. Farrell, H.M., Jr.; Jimenez-Flores, R.; Bleck, G.T.; Brown, E.M.; Butler, J.E.; Creamer, L.K.; Hicks, C.L.; Hollar, C.M.; Ng-Kwai-Hang, K.F.; Swaisgood, H.E. Nomenclature of the proteins of cows' milk—Sixth revision. *J. Dairy Sci.* **2004**, *87*, 1641–1674. [CrossRef]

Sample Availability: Samples of the compounds are not available from the authors.

molecules

MDPI

Article

Comparative Proteomic Analysis of *Rana chensinensis* Oviduct

Hang Su [1], He Zhang [2,*], Xinghua Wei [3], Daian Pan [2], Li Jing [1], Daqing Zhao [4], Yu Zhao [4] and Bin Qi [5,*]

1 Practice Innovations Center, Changchun University of Chinese Medicine, Changchun 130117, China; suhang0720@live.cn (H.S.); prettygirl0122@163.com (L.J.)
2 School of Clinical Medicine, Changchun University of Chinese Medicine, Changchun 130117, China; 18744030771@163.com (D.P.)
3 Jilin Science Service Center, Changchun 130021, China; v_stars@163.com
4 Jilin Ginseng Academy, Changchun University of Chinese Medicine, Changchun 130117, China; cnzhaodaqing@126.com (D.Z.); yuzhao2016@163.com (Y.Z.)
5 College of Pharmacy, Changchun University of Chinese Medicine, Changchun 130117, China
* Correspondence: 13843148162@163.com (H.Z); qibin88@126.com (B.Q.)

Received: 8 March 2018; Accepted: 5 June 2018; Published: 8 June 2018

Abstract: As one of most important traditional Chinese medicine resources, the oviduct of female *Rana chensinensis* (Chinese brown frog) was widely used in the treatment of asthenia after sickness or delivery, deficiency in vigor, palpitation, and insomnia. Unlike other vertebrates, the oviduct of *Rana chensinensis* oviduct significantly expands during prehibernation, in contrast to the breeding period. To explain this phenomenon at the molecular level, the protein expression profiles of *Rana chensinensis* oviduct during the breeding period and prehibernation were observed using isobaric tags for relative and absolute quantitation (iTRAQ) technique. Then, all identified proteins were used to obtain gene ontology (GO) annotation. Ultimately, KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analysis was performed to predict the pathway on differentially expressed proteins (DEPs). A total of 4479 proteins were identified, and 312 of them presented different expression profiling between prehibernation and breeding period. Compared with prehibernation group, 86 proteins were upregulated, and 226 proteins were downregulated in breeding period. After KEGG enrichment analysis, 163 DEPs were involved in 6 pathways, which were lysosome, RNA transport, glycosaminoglycan degradation, extracellular matrix (ECM)–receptor interaction, metabolic pathways and focal adhesion. This is the first report on the protein profiling of *Rana chensinensis* oviduct during the breeding period and prehibernation. Results show that this distinctive physiological phenomenon of *Rana chensinensis* oviduct was mainly involved in ECM–receptor interaction, metabolic pathways, and focal adhesion.

Keywords: iTRAQ; proteomics; *Rana chensinensis* oviduct; regulation; differentially expressed proteins

1. Introduction

The oviduct is a reproductive organ of females, which plays several important roles in the events related to fertilization and embryo development. The oviduct is not only a passive channel for sperm and eggs transport, but is also a highly active secretory organ, such as estrous cycle and ovulation [1]. It provides the most efficient environment for the success of fertilization and early embryo development [2,3]. The component of oviduct of *Rana chensinensis* is rich in glands, including glycoprotein (such as mucins, collagen, enzymes, and hormones, etc.) and lipoprotein [4]. The dry oviduct of female *Rana chensinensis*, which is recorded in the Chinese Pharmacopoeia, possesses the function of improving immune system and lung function [5].

The hibernation for *Rana chensinensis* ranges from October to February next year, followed by the breeding period from February to June. After the breeding period, *Rana chensinensis* goes into the prehibernation period until October [6]. Distinguishing from the oviduct expansion during the breeding period in other species, a unique physiological phenomenon of *Rana chensinensis* is that its oviduct starts to expand after breeding, reaching a peak by October during the prehibernation [7]. For this reason, *Rana chensinensis* oviduct, as traditional Chinese medicine, was collected from frog in the autumn before hibernation. To ascertain the signaling pathways involved in the timing of oviduct expansion, we should clarify the changes of macromolecular components between prehibernation and breeding period. Proteomics is an efficient methodology to analyze protein expression, and it will disclose protein expression profiles in different physiological phases [8]. It is widely used to tackle biological problems, whereby the raw data obtained from proteomics is studied further by bioinformatic methodologies [8]. Isobaric tags for relative and absolute quantitation (iTRAQ) is an isobaric labeling method applied in quantitative proteomics by tandem mass spectrometry to identify the amount of proteins from different samples in a single experiment [9]. iTRAQ can separate and identify a variety of proteins, including membrane proteins, proteins of high molecular weight, insoluble proteins, acidic proteins, and alkaline proteins [10].

2. Results

2.1. Protein Identified through Itraq Technology

13216 unique peptides and 4479 proteins were identified using the iTRAQ technology (see Tables S1 and S2 in Supplementary Material). Among these identified proteins, 2422 were 0–20 kDa, 1758 were 20–60 kDa, and 299 were over 60 kDa (Figure 1A). Accordingly, we found that identified proteins were primarily below 20 kDa. To further prove the credibility of protein identification, peptide sequence coverage and the unique peptide numbers of proteins were also regarded as two important quality evaluation parameters. Figure 1B showed that peptide sequence coverage of these proteins was basically less than 30%. The number of unique peptides for identified proteins were mainly concentrated in 1 and 2, which makes up approximately 71% of the total unique peptides (Figure 1C).

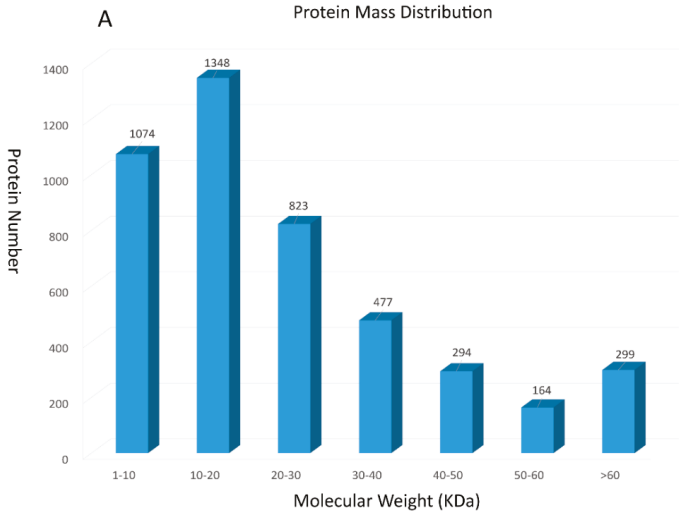


Figure 1. *Cont.*

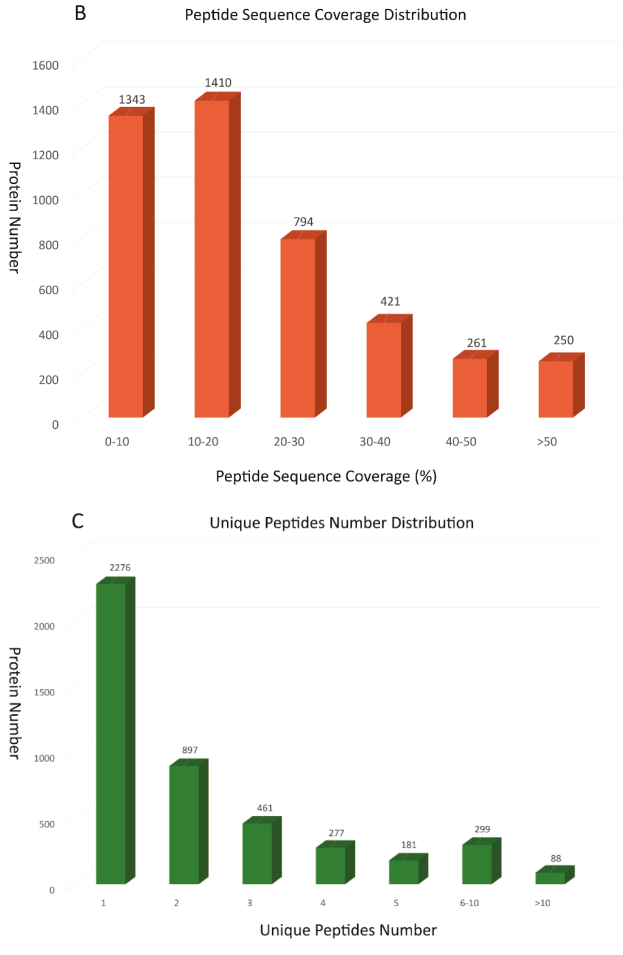


Figure 1. Summary for iTRAQ data: (**A**) The protein mass distribution shown as a histogram; (**B**) The peptide sequence coverage distribution shown as a histogram; (**C**) The number of unique peptides distribution shown as a histogram.

2.2. Differentially Expressed Proteins

Hereon, we defined proteins with expression level fold changes >2.0 or <0.5 with p-values < 0.05 as differentially expressed proteins (DEPs). Changes in the protein profile were analyzed, and 312 proteins exhibited a difference. Compared with the prehibernation group, 86 proteins were increased by more than 2-fold, and 226 proteins decreased to less than 0.5-fold during the breeding period. The detailed information of differential expressed proteins was shown in Table S3.

2.3. Gene Ontology Annotation and Classification Profiles

With annotating those 312 DEPs into Gene Ontology (GO) database, we found that 240 DEPs were classified into biological process, 245 DEPs were classified into molecular function, 244 DEPs were classified into cellular component (Table S4 in Supplementary Material). Further analysis of DEPs clustered into biological processes disclosed that 193 DEPs were related to cellular process, single-organism process (172 DEPs) second, and the next were metabolic process (142 DEPs), biological

regulation (117 DEPs), and regulation of biological process (108 DEPs) (Figure 2A). In the molecular function categories, DEPs were primarily associated with binding, indicating that proteins related in binding function will experience significant changes in the breeding period (Figure 2B). Cellular component ontology annotation showed that cell and cell parts possessed two-thirds of the whole DEPs, followed by organelles (Figure 2C).

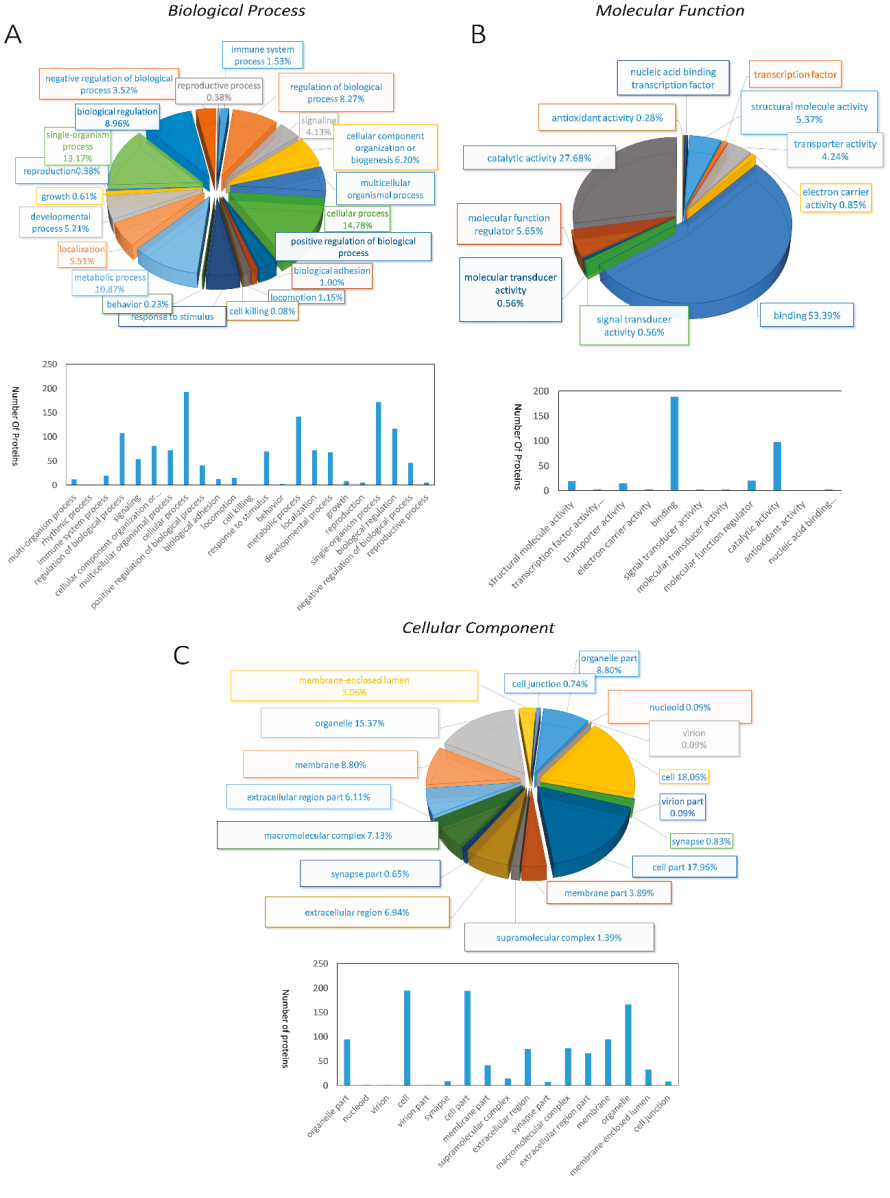


Figure 2. Pie diagrams and bar diagrams for gene ontology analysis of differentially expressed proteins between breeding period group and prehibernation group: (**A**) biological process, (**B**) molecular function, (**C**) cellular component.

2.4. KEGG Pathway Annotation and Reasonable Enrichment Analysis

By annotating the 312 DEPs into KEGG database, we found 177 pathways (Table S5 in Supplementary Material). The top 10 pathways were pathways in cancer, RNA transport, spliceosome, focal adhesion, PI3K–Akt signaling pathway, Huntington's disease, Alzheimer's disease, metabolism, extracellular matrix (ECM)–receptor interaction, and amoebiasis (Table 1).

Table 1. Top 10 pathway terms of differentially expressed proteins (DEPs) in KEGG annotation.

Pathway Terms [1]	DEPs with KEGG Annotation [2]	DEPs with KEGG Enrichment [3]
Pathways in cancer	9	0
RNA transport	8	4
Spliceosome	8	0
Focal adhesion	7	4
PI3K–Akt signaling	7	0
Huntington's disease	7	0
Alzheimer's disease	6	0
Metabolic pathways	24	11
ECM–receptor interaction	6	3
Amoebiasis	6	0

[1] Pathway terms in KEGG database; [2] The number of differentially expressed proteins annotated in KEGG database; [3] The number of differentially expressed proteins enriched in top 10 pathway terms.

KEGG can be used for systematic analysis of gene functions, which is linked with genomic information combined with information stored in the pathway database. The KEGG pathway is an aggregate of pathway maps showing the present knowledge on the molecular interaction networks. In our research, a database based on transcriptome data was used for identification of *Rana* oviduct proteins, which originated from different species. As a result, it is arduous for these DEPs to process KEGG pathway enrichment analysis, because the proteins were slightly different in various species. The distribution of differentially expressed proteins in various species is shown in Table 2. In the present study, the largest protein group (163 proteins, 52%) and the second largest one (61 proteins, 20%) of total DEPs was from *Xenopus tropicalis* and *Xenopus laevis*, respectively. To discover the possible signaling pathways underlying the oviduct of *Rana chensinensis*, the largest two groups of DEPs were used to process KEGG pathway analysis through KOBAS separately [11,12]. The result of KEGG enrichment was shown in Table S6. After enrichment analysis, 163 DEPs from *Xenopus tropicalis* were involved in 6 pathway terms, which were lysosome, RNA transport, glycosaminoglycan degradation, ECM–receptor interaction, metabolic pathways, and focal adhesion (Table 3). While 61 DEPs from *Xenopus laevis* were input into KEGG database for enrichment analysis, only one pathway term, pyruvate metabolism, was obtained (Table 3). Hence, KEGG enrichment for DEPs from *Xenopus tropicalis* and *Xenopus laevis* were selected to logically interpret physiological changes between prehibernation period and breeding period.

Table 2. The distribution of DEPs in different species.

Organism [1]	Number of Proteins [2]	Organism	Number of Proteins
Xenopus tropicalis	163	*Salmo salar*	2
Xenopus laevis	61	*Bos taurus*	2
Rana catesbeiana	13	*Oryctolagus cuniculus*	1
Taeniopygia guttata	11	*Mus musculus*	1
Gallus gallus	8	*Ovis aries*	1
Monodelphis domestica	7	*Zonotrichia albicollis*	1
Homo sapiens	5	*Pan troglodytes*	1
Rana esculenta	4	*Pyrrhocoris apterus*	1
Mus musculus	4	*Canis familiaris*	1

Table 2. *Cont.*

Organism [1]	Number of Proteins [2]	Organism	Number of Proteins
Ailuropoda melanoleuca	4	*Oryzias latipes*	1
Ornithorhynchus anatinus	3	*Equus caballus*	1
Danio rerio	3	*Trepomonas agilis*	1
Rattus norvegicus	3	*Anoplopoma fimbria*	1
Macaca mulatta	2	*Helicoverpa zea*	1
Bufo japonicus	2	*Sus scrofa*	1
Callithrix jacchus	2		

[1] Organism to which protein belongs; [2] The number of differentially expressed proteins originates from one species.

Table 3. KEGG enrichment analysis of DEPs.

Pathway Terms [1]	Input Number [2]	Background Number [3]	Gene Symbol of Input DEPs [4]	*p*-Value
Lysosome	5	123	acp2, lamp1, lgmn, hexb, LOC100496969	0.0049
RNA transport	4	140	thoc6, eif3a, eif4a2, eif4b	0.0292
Glycosaminoglycan degradation	2	18	LOC100496969, hexb	0.0292
ECM–receptor interaction	3	75	lama2, lamb2, col4a1	0.0292
Metabolic pathways	11	1163	idh3b, rpn2, coq3, pla2g6, lpin1, ak7, alas2, gamt, hexb, LOC100496969, acat2	0.0292
Pyruvate metabolism	3	43	acss2.2.L, pc.1.L, ldha.S	0.0338
Focal adhesion	4	191	lama2, lamb2, col4a1, actn1	0.0451

[1] Pathway terms in KEGG database; [2] The number of differentially expressed proteins enriched in one of pathway terms in KEGG database; [3] The total number of differentially expressed proteins in one of pathway terms in KEGG database; [4] The corresponding gene symbol of differentially expressed proteins enriched in one of pathway terms in KEGG database.

3. Discussion

3.1. Changes in Proteins Associated with Focal Adhesion and Extracellular Martix

There are three differential expressed proteins enriched in ECM (extracellular matrix)–receptor interaction pathway through KEGG enrichment analysis, collagen alpha-1 (IV) chain (COL4A1), laminin subunit alpha 2 (LAMA2), and laminin subunit beta 2 (LAMB2) (Table 3). These different kinds of ECM and ECM-associated proteins, which bind to cells through ECM receptor, play an important role in building the complex meshwork of extracellular matrices [13]. Therefore, the assembled manner of these macromolecules from ECM, including collagens, proteoglycans, laminins, and fibronectin, determine the structure and function of cell and tissue [14]. Furthermore, the resultant ECM will facilitate the biological function of a specific tissue. According to previous study, we know that the differential expression of extracellular matrix (ECM) proteins in tissue-specific cells will perform different biological processes and molecular functions [15]. Compared with the prehibernation group, COL4A1 protein in oviduct during the breeding period was upregulated with a 5-fold change. Type IV collagen, along with related other ECM proteins, is widely distributed in the basement membrane of various specific tissues [16]. Similarly, the membrane protein relevant to the morphology of oviduct was changed significantly across the estrous cycle through the bovine oviduct proteome [17]. The morphology of the oviduct muscle will be changed in *COL4A1* mutants, which then causes severe myopathy with centronuclear myofibers. The lack of biological function because of *COL4A1* mutants will lead to the gradual development of female sterility [18]. Accordingly, COL4A1 protein plays a vital role in maintaining the structure and function of the oviduct muscle. In our research, the expression of COL4A1 protein increases significantly during breeding period. Therefore, the upregulation of

COL4A1 protein in breeding period may be one of notable key factors for *Rana chensinensis* to allow the oviduct to maintain optimal morphology for ovulation.

Laminins are heterotrimeric molecules consisting of three subunit chains, named α, β, γ chain, and they significantly influence the biological processes related to cell adhesion, growth, morphology, differentiation, and migration [19,20]. There are another two laminin proteins (LAMA2, LAMB2) enriched in the ECM–receptor interaction pathway. Laminin, as a ubiquitous connective glycoprotein, is still also the predominant component of basement membrane in a specific tissue [21]. In a previous study, we know that laminin is the major component of the basement membrane in the oviduct of Japanese quail [22]. The oviduct is a key site for reproductive events that involve gamete maturation, fertilization, and early embryo development, processes that finally determine breeding success. The upregulation of LAMA2 and LAMB2, as important ECM components, could have a positive influence on the reproduction of *Rana chensinensis*. In this study, these two upregulated proteins during breeding period are supposed to be involved in preparing the oviduct for breeding time.

The second highly enriched pathway is focal adhesion, which involves four differentially expressed proteins. Focal adhesions are integrin-containing, multiplex protein structures that form mechanical connections between intracellular actin bundles and the extracellular matrix components in various cell types [23]. Besides three differentially expressed proteins from ECM–receptor interaction pathway, α-actinin (ACTN1) is enriched in the focal adhesion pathway. A previous research study has shown that mating will change the expression of α-actinin in *Drosophila* oviduct [24]. The oviduct muscles are visceral muscles that are supercontractile, meaning they have Z-bands with perforations that enlarge during contraction. The α-actinin is specific to the Z-band with perforation of supercontractile muscle, in which it crosslinks actin filaments from adjacent sarcomeres [25]. With mating of *Drosophila*, the abundance of the muscle protein α-actinin will increase markedly in the oviduct. The results in our research are consistent with this finding. By comparison with prehibernation group, the expression of α-actinin in *Rana chensinensis* oviduct is significantly upregulated during breeding time. Hence, we get the conclusion that upregulation of α-actinin (ACTN1) will probably play a vital role in maintaining reproductive function of the oviduct after hibernation.

3.2. Enzymes Involved in Energy Metabolism Pathway

Isocitrate dehydrogenase (NAD) subunit beta is an enzyme that, in humans, is encoded by the *IDH3B* gene. As an isocitrate dehydrogenase, IDH3 catalyzes the reversible oxidative decarboxylation of isocitrate to form α-ketoglutarate (α-KG) and CO_2 as part of the tricarboxylic acid (TCA) cycle in glucose metabolism [26]. This step also participates in the reduction of NAD+ to NADH, which is then applied to produce ATP through the electron transport chain. Notably, IDH3 acts as NAD-specific electron acceptor, in contrast to NADP-dependent IDH1 and IDH2 [27,28]. IDH3 activity is regulated by the energy demands of the cell [27,29]. When the energy metabolism of cells is under insufficiency, IDH3 is activated by ADP. Conversely, IDH3 is inhibited by ATP and NADH under sufficiency. According to Karl's research [30], ATP consuming processes during the breeding period could be compensated through the upregulation of ATP production. Our research has found that isocitrate dehydrogenase (IDH3B) was significantly upregulated in the breeding period, which is a method of upregulation of ATP production. The previous study also showed that the expression of isocitrate dehydrogenase (IDH) was upregulated in oviduct fluid from ewes during estrus [31]. The energy metabolism of *Rana chensinensis* oviduct, as an important site for fertilization, will involve the regulation of energy substrate and related enzymes during breeding period [32].

In this research, one another upregulated protein in energy metabolism is ubiquinone biosynthesis *O*-methyltransferase (COQ3) during the breeding period. This *O*-methyltransferase (COQ3) played a part in two steps of the reactions in the biosynthesis of ubiquinone (coenzyme Q). This enzyme methylated an early coenzyme Q intermediate, 3,4-dihydroxy-5-polyprenylbenzoic acid, as well as the final intermediate in the pathway, converting demethylated ubiquinone to coenzyme Q [33]. It is a component of the electron transport chain, and participates in aerobic cellular respiration,

which generates energy in the form of ATP [34]. The complex process of fertilization required energy, which is produced by mitochondria mostly via oxidative phosphorylation [34,35]. To compensate for excess energy consumption during the breeding period, the efficiency of electron transport chain was improved through the upregulation of *O*-methyltransferase.

Guanidinoacetate *N*-methyltransferase (GAMT) is an enzyme that converts guanidinoacetate to creatine, and is encoded by the gene *GAMT*. Creatine is involved in recycling of adenosine triphosphate (ATP), the energy currency of the cell, primarily in muscle and brain tissue. This is achieved by recycling adenosine diphosphate (ADP) to adenosine triphosphate ATP via donation of phosphate groups [36]. Maternal creatine homeostasis was influenced by the expression of creatine biosynthesis-related enzymes. In a previous study, the expression of GAMT protein in heart and brain tissue was decreased in pregnant spiny mouse [37]. The finding in our research showed that the expression of GAMT was also downregulated in breeding period. Without feeding during the hibernation and breeding periods, L-arginine and glycine, as staring material of creatine biosynthesis, were inevitably at low levels in *Rana chensinensis*. Hence, the expression of GAMT was down-regulated due to deficiency of starting material. Accordingly, this result indicated that creatine biosynthesis pathway was switched off.

Adenylate kinase 7 (AK7) is a member of the adenylate kinase family of enzymes, which catalyzes the reversible phosphorylation of adenine nucleotides. It has been proven that adenylate kinase (AK) isoforms have an essential role in ciliary function and energy homeostasis [38], and mutations of *AK7* in the mouse result in primary ciliary dyskinesia [39]. The tissue specificity of Ak7, which is abundant in trachea and oviduct, has been found by Fernandez-Gonzalez et al. [39]. The expression profile of genes related to the ciliary function of the oviduct have been analyzed in porcine oviduct transcriptome [40]. It has reported that differentially expressed genes (DEGs) of porcine oviduct were enriched in GO term "cell motion" in preovulatory phase [40]. Our research has shown that AK7 expression was upregulated in oviduct during the breeding period. It is known that eggs in the ovary of *Rana chensinensis* undergo the last stages of oogenesis, and are then released into the body cavity. Finally, they are driven into the oviduct by ciliary action. In order to maintain ciliary action, AK7 provided an efficient way to transport energy from ATP production sites to the ciliary of oviducts. Hence, upregulation of AK7 will help eggs of *Rana chensinensis* to be transferred into oviduct through ciliary movement.

3.3. Enzymes Changed in Pyruvate Metabolism Pathway

There are three differentially expressed proteins enriched in pyruvate metabolism pathway. In contrast to prehibernation period, the expression level of pyruvate carboxylase (PC.1) and lactate dehydrogenase A (LDHA.S) during breeding period were upregulated and acetyl-CoA synthase (ACSS2.2) was downregulated. The environmental hypoxia during hibernation period contributed to metabolic acidosis in amphibians, and increased lactate concentrations in amphibian tissues [41,42]. Hence, lactate levels during breeding time should be significantly higher in *Rana chensinensis* oviduct after long hibernation times, compared with prehibernation period. The upregulated expression of lactate dehydrogenase A (LDHA) in breeding time will catalyze the transformation from lactate to pyruvic acid when environmental stress in hibernation period is relieved [43]. Moreover, the upregulation of pyruvate carboxylase will increase the production of oxaloacetate (OAA), which can be converted to phosphoenolpyruvate in gluconeogenesis, or directly enter into the TCA cycle [44]. Without eating in hibernation and breeding period, the gluconeogenesis pathway will consume the accumulation of lactate in hibernation period to reduce metabolic acidosis and increase adipose [45,46]. Accordingly, we predict that the gluconeogenesis pathway is potentially at an active stage in breeding time to decrease lactate levels in oviduct tissue. Acetyl-CoA synthetase, which is downregulated in breeding time, plays key roles in the production of acetyl-CoA from free acetate for the synthesis of fatty acid [47]. The expression of acetyl-CoA synthetase drops down when animals are starved, which leads to a decline in fatty acid synthesis [48]. While animals was refed with high carbohydrate diet, the upregulated expression of acetyl-CoA synthetase will increase fatty acid synthesis [48].

Due to no eating during the whole breeding time, the deficiency of carbon source could lead to the downregulation of acetyl-CoA synthetase.

3.4. Down-Regulation of Lipid Metabolism during Breeding Period

Lipin-1, encoded by the *LPIN1* gene, possesses phosphatidate phosphatase activity [49]. In the mouse, the expression of lipin-1 is at high levels in adipose tissue and skeletal muscle, consistent with functions in lipid metabolism in these specific tissues. Actually, adipocytes in lipin-1-deficient mice fail to accumulate triacylglycerol (TAG), and do not develop mature adipocyte function [50]. In this study, the expression of lipin-1 was downregulated in the oviduct tissue during the breeding period. Accordingly, the accumulation of triacylglycerol significantly decreased. As a result, the morphology of oviduct appeared to be in a shriveled state, due to this reason. Another protein related with lipid metabolism, calcium-independent phospholipase A2 (PLA2G6), is also downregulated. PLA2 enzyme is involved in phospholipid metabolism, which is very important for many biological processes, including resembling cell membrane during cell cycle [51]. PLA2G6 helps to regulate the levels of phosphatidylcholine, which is a major membrane phospholipid [52]. Downregulation of PLA2G6 expression inhibited cell proliferation in culture, and tumorigenicity of ovarian cancer cell lines in nude mice [53]. Another reason for the shriveled state of the oviduct is that cell proliferation in *Rana chensinensis* oviduct was suppressed by the downregulation of PLA2G6 in the breeding period.

Acetyl-CoA acetyltransferase 2 (ACAT2), an enzyme converting cholesterol and fatty acid to cholesteryl esters, is involved in lipid metabolism [54]. The previous study shows that cholesterol and fatty acid stabilize ACAT2, which is ubiquitylated on cystine residue to degrade [55]. Lipids including cholesterol and fatty acid induce the generation of reactive oxygen species, which oxidize Cys277 of ACAT2, and subsequently prevent ACAT2 from degradation [55,56]. A low level of lipid induced ubiquitylation on Cys277 for degradation of ACAT2. Accordingly, the downregulation of ACAT2 protein in this study indicated that the concentration of sterol and fatty acid was at a relatively low level in the breeding period.

3.5. Expression of 5-Aminolevulinate Synthase (ALAS2) in Heme Biosynthesis

The expression of 5-aminolevulinate synthase (ALAS2) was upregulated during the breeding period. ALAS is the rate-limiting enzyme in protoporphyrin IX (PPIX) production, which is the final intermediate in the heme biosynthetic pathway. Transcription of ALAS2, unlike ALAS1, is regulated by erythroid-specific transcription factors, such as GATA1 [57]. Regarding the post-transcription of ALAS2, it is affected by the content of iron. In the absence of iron, the iron-free form of iron regulatory protein (IRP) binds to iron regulatory element (IRE), forming an IRE–IRP complex that prevents translation of ALAS2 [58]. Heme, also as a downregulation factor, inhibits the translation of ALAS2 and the import of ALAS2 precursor into mitochondria [59]. According to our results about the expression of ALAS2, we concluded that the content of heme in oviduct tissue is at low levels after hibernation. Otherwise, the expression of ALAS2 was inhibited by the adequate heme.

4. Materials and Methods

4.1. Animals and Treatment Procedure

In this study, we performed iTRAQ proteomic analysis to identify differentially expressed proteins from the oviduct of female *Rana chensinensis* between the breeding period and the prehibernation period. Sixty adult female *Rana chensinensis* were obtained in April (n = 30, the breeding period), and October (n = 30, the prehibernation period), 2017 from Jilin Province (125°16′57″ E~131°19′12″ E, 40°51′55″ N~44°38′54″ N), China. All animals were treated in strict accordance with the recommendations in the Guide for the International Cooperation Committee Animal Welfare (ICCAW). All experimental procedures were approved by the Committee on the Ethics of Animal Experiments of Changchun

University of Chinese Medicine. Frogs were anesthetized by diethyl ether. Each pair of oviducts was collected from *Rana chensinensis*.

4.2. Protein Extraction, Digestion, and iTRAQ Labeling

Each sample of 20 mg were frozen in liquid nitrogen and ground with a mortar and pestle. One milliliter of 10% TCA/acetone (v/v, 1:9) was added to the powder, and mixed by vortexing. The mixture was placed at −20 °C for 4 h, and then the precipitate was washed with acetone at 4 °C, until colorless. The resulting pellet was resolubilized in STD buffer (4% SDS, 100 mM DTT, 150 mM Tris-HCl, pH 8.0). The protein concentration was measured by the BCA Protein Assay kit (Beyotime, Hangzhou, China). Proteins were digested according to the FASP method [60]. The resulting mixture of each sample was labeled using iTRAQ reagent according to the manufacturer's instructions (Applied Biosystems, Branchburg, NJ, USA). The samples of *Rana dybowskii* oviduct in prehibernation period were labeled with iTRAQ tag 113 (A) and 114 (B), and samples from the breeding period were labeled with iTRAQ tag 115 (C) and 116 (D).

4.3. Peptide Fractionation with Strong Cation Exchange (SCX) Chromatography

The peptide mixture was constituted again and acidified with buffer A (10 mM KH_2PO_4 in 25% of ACN, pH 3.0), and loaded onto a polysulfethyl column (5 µm, 4.6 × 100 mm, 200 Å, PolyLC Inc., Columbia, MD, USA). The peptides were fractionated at a flow rate of 1 mL/min with a gradient of 0–10% buffer B (500 mM KCl, 10 mM KH_2PO_4 in 25% of ACN, pH 3.0) for 30 min, 10–60% buffer B during 30–50 min, 60–100% buffer B during 50–55 min, 100% buffer B during 55–60 min, and finally, buffer B was set to 0% after 60 min. The detector was set at 214 nm, and fractions were collected every 1 min. The collected fractions were desalted on C18 Cartridges (Empore™ SPE Cartridges C18, Sigma, St. Louis, MO, USA) and concentrated by vacuum centrifugation.

4.4. LC-MS/MS Analysis

Experiments were performed as described previously [61] using a Q Exactive mass spectrometer coupled to Easy nLC (Thermo Fisher Scientific, Waltham, MA, USA). Ten microliters of each fraction was injected for nano LC-MS/MS analysis. The peptide mixture (5 µg) was loaded onto a the C18-reversed phase column (3µm, 75 µm × 10 cm) in buffer A (0.1% formic acid) and separated with a linear gradient of buffer B (80% ACN and 0.1% formic acid) at a flow rate of 250 nL/min controlled by IntelliFlow technology over 140 min. MS data was acquired using a data-dependent top10 method dynamically. The duration of dynamic exclusion was 60 s. Survey scans were acquired at a resolution of 70,000 at m/z 200 and the resolution for HCD spectra was set to 17,500 at m/z 200. Normalized collision energy was 30 eV and the underfill ratio, which means the minimum percentage of the target value at maximum fill time, was set to 0.1%. The instrument was run with peptide recognition mode enabled.

4.5. Protein Identification and Data Analysis

In our previous study, the transcriptome of OR was sequenced, and the corresponding unigenes were generated [62]. In the present study, the amino acid sequences translated from the CDS of unigenes were used as the protein database. The raw files were analyzed using the Proteome Discoverer 1.3 software (Thermo Electron, San Jose, CA, USA). A search for the fragmentation spectra was performed using the MASCOT search engine [63]. The results were filtered based on a false discovery rate (FDR) of no more than 0.01. The protein identification was supported by at least one unique peptide. Isobaric Labeling Multiple File Distiller and Identified Protein iTRAQ Statistic Builder were applied to calculate the ratios of protein, in which REF was used as the reference, based on the weighted average of the intensities of report ions in each identified peptide [64]. The final ratios were then normalized with the median average protein ratio, assuming that most proteins remained unchanged in abundance. Protein ratios represent the median of the unique peptides of the protein [65].

For statistical analysis, two-way ANOVA was performed for each protein, and Student's *t*-test was used to evaluate the significant differences. The differentially expressed proteins ratio meets the fold change ($\geq$2.0 or $\leq$0.5, at $p < 0.05$).

4.6. Bioinformatics Analysis of Differentially Expressed Proteins

We carried out gene ontology (GO) annotation analysis on the differentially expressed proteins to catalog the molecular functions, biological processes, and cellular components [66]. Mapping the function of Blast2GO (Version 3.3.5) was applied to determine GO terms correlated with DEPs [67]. To further explore the interaction of differentially expressed protein physiological process in the body and discover relationships between DEPs, KEGG enrichment analysis was performed based on KOBAS server (Version 3.0) [11,12]. KEGG pathway enrichment analyses were applied based on the Fisher' exact test, considering the whole quantified protein annotations as background dataset. Only functional categories and pathways with *p*-values under a threshold of 0.05 were considered as significant.

Supplementary Materials: The following are available online.

Author Contributions: D.Z., B.Q. and H.Z. developed and framed the research questions. B.Q. and L.J. collected the samples of frog. H.S., H.Z. and X.W. analyzed proteomics data. D.P. and Y.Z. were involved in revising the manuscript. H.S. and H.Z. drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (Award No. 81503243).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, Y.; Weng, J.; Huang, S.; Shen, Y.; Sheng, X.; Han, Y.; Xu, M.; Weng, Q. Immunoreactivities of PPARγ2, leptin and leptin receptor in oviduct of Chinese brown frog during breeding period and pre-hibernation. *Eur. J. Histochem.* **2014**, *58*. [CrossRef] [PubMed]
2. Gabler, C.; Killian, G.J.; Einspanier, R. Differential expression of extracellular matrix components in the bovine oviduct during the oestrous cycle. *Reprod. Camb. Engl.* **2001**, *122*, 121–130. [CrossRef]
3. Mondéjar, I.; Acuña, O.; Izquierdo-Rico, M.; Coy, P.; Avilés, M. The Oviduct: Functional Genomic and Proteomic Approach: Transcriptome and Proteome of the Oviduct. *Reprod. Domest. Anim.* **2012**, *47*, 22–29. [CrossRef] [PubMed]
4. Weng, J.; Liu, Y.; Xu, Y.; Hu, R.; Zhang, H.; Sheng, X.; Watanabe, G.; Taya, K.; Weng, Q.; Xu, M. Expression of P450arom and Estrogen Receptor Alpha in the Oviduct of Chinese Brown Frog (*Rana dybowskii*) during Prehibernation. *Int. J. Endocrinol.* **2015**, *2015*, 1–9. [CrossRef] [PubMed]
5. The Pharmacopoeia Commission of PRC. *Pharmacopoeia of the People's Republic of China*; The Pharmacopoeia Commission of PRC: Beijing, China, 2000; pp. 255–256.
6. Shen, Y.; Liu, Y.; Ma, J.; Ma, X.; Tian, Y.; Zhang, H.; Li, L.; Xu, M.; Weng, Q.; Watanabe, G.; et al. Immunoreactivity of c-kit receptor protein during the prehibernation period in the oviduct of the Chinese brown frog, *Rana chensinensis*. *J. Vet. Med. Sci.* **2012**, *74*, 209–213. [CrossRef] [PubMed]
7. Huang, D.; Yang, L.; Wang, C.; Ma, S.; Cui, L.; Huang, S.; Sheng, X.; Weng, Q.; Xu, M. Immunostimulatory Activity of Protein Hydrolysate from Oviductus Ranae on Macrophage In Vitro. *Evid. Based Complement. Altern. Med.* **2014**, *2014*, 1–11. [CrossRef]
8. Wright, P.C.; Noirel, J.; Ow, S.-Y.; Fazeli, A. A review of current proteomics technologies with a survey on their widespread use in reproductive biology investigations. *Theriogenology* **2012**, *77*, 738–765. [CrossRef] [PubMed]
9. Zieske, L.R. A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. *J. Exp. Bot.* **2006**, *57*, 1501–1508. [CrossRef] [PubMed]
10. Wu, W.W.; Wang, G.; Baek, S.J.; Shen, R.-F. Comparative Study of Three Proteomic Quantitative Methods, DIGE, cICAT, and iTRAQ, Using 2D Gel- or LC−MALDI TOF/TOF. *J. Proteome Res.* **2006**, *5*, 651–658. [CrossRef] [PubMed]

11. Xie, C.; Mao, X.; Huang, J.; Ding, Y.; Wu, J.; Dong, S.; Kong, L.; Gao, G.; Li, C.-Y.; Wei, L. KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **2011**, *39*, W316–W322. [CrossRef] [PubMed]
12. Wu, J.; Mao, X.; Cai, T.; Luo, J.; Wei, L. KOBAS server: A web-based platform for automated annotation and pathway identification. *Nucleic Acids Res.* **2006**, *34*, W720–W724. [CrossRef] [PubMed]
13. Hynes, R.O.; Naba, A. Overview of the Matrisome—An Inventory of Extracellular Matrix Constituents and Functions. *Cold Spring Harb. Perspect. Biol.* **2012**, *4*, a004903. [CrossRef] [PubMed]
14. Rozario, T.; DeSimone, D.W. The extracellular matrix in development and morphogenesis: A dynamic view. *Dev. Biol.* **2010**, *341*, 126–140. [CrossRef] [PubMed]
15. Mouw, J.K.; Ou, G.; Weaver, V.M. Extracellular matrix assembly: A multiscale deconstruction. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 771–785. [CrossRef] [PubMed]
16. Kühn, K. Basement membrane (type IV) collagen. *Matrix Biol.* **1995**, *14*, 439–445. [CrossRef]
17. Lamy, J.; Labas, V.; Harichaux, G.; Tsikis, G.; Mermillod, P.; Saint-Dizier, M. Regulation of the bovine oviductal fluid proteome. *Reproduction* **2016**, *152*, 629–644. [CrossRef] [PubMed]
18. Kelemen-Valkony, I.; Kiss, M.; Csiha, J.; Kiss, A.; Bircher, U.; Szidonya, J.; Maróy, P.; Juhász, G.; Komonyi, O.; Csiszár, K.; et al. Drosophila basement membrane collagen col4a1 mutations cause severe myopathy. *Matrix Biol.* **2012**, *31*, 29–37. [CrossRef] [PubMed]
19. Tzu, J.; Marinkovich, M.P. Bridging structure with function: Structural, regulatory, and developmental role of laminins. *Int. J. Biochem. Cell Biol.* **2008**, *40*, 199–214. [CrossRef] [PubMed]
20. Koch, M.; Olson, P.F.; Albus, A.; Jin, W.; Hunter, D.D.; Brunken, W.J.; Burgeson, R.E.; Champliaud, M.-F. Characterization and Expression of the Laminin γ3 Chain: A Novel, Non-Basement Membrane–associated, Laminin Chain. *J. Cell Biol.* **1999**, *145*, 605–618. [CrossRef] [PubMed]
21. Saotome, K.; Isomura, T.; Seki, T.; Nakamura, Y.; Nakamura, M. Structural changes in gonadal basement membranes during sex differentiation in the frog Rana rugosa. *J. Exp. Zool. Part Ecol. Genet. Physiol.* **2010**, *313A*, 369–380. [CrossRef] [PubMed]
22. Madekurozwa, M.-C. Immunolocalization of Intermediate Filaments and Laminin in the Oviduct of the Immature and Mature Japanese Quail (*Coturnix coturnix japonica*). *Anat. Histol. Embryol.* **2014**, *43*, 210–220. [CrossRef] [PubMed]
23. Chen, C.S.; Alonso, J.L.; Ostuni, E.; Whitesides, G.M.; Ingber, D.E. Cell shape provides global control of focal adhesion assembly. *Biochem. Biophys. Res. Commun.* **2003**, *307*, 355–361. [CrossRef]
24. Kapelnikov, A.; Zelinger, E.; Gottlieb, Y.; Rhrissorrakrai, K.; Gunsalus, K.C.; Heifetz, Y. Mating induces an immune response and developmental switch in the Drosophila oviduct. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13912–13917. [CrossRef] [PubMed]
25. Vigoreaux, J.O.; Saide, J.D.; Pardue, M.L. Structurally differentDrosophila striated muscles utilize distinct variants of Z-band-associated proteins. *J. Muscle Res. Cell Motil.* **1991**, *12*, 340–354. [CrossRef] [PubMed]
26. Soundar, S.; Park, J.-H.; Huh, T.-L.; Colman, R.F. Evaluation by Mutagenesis of the Importance of 3 Arginines in α, β, and γ Subunits of Human NAD-dependent Isocitrate Dehydrogenase. *J. Biol. Chem.* **2003**, *278*, 52146–52153. [CrossRef] [PubMed]
27. Bzymek, K.P.; Colman, R.F. Role of α-Asp 181, β-Asp 192, and γ-Asp 190 in the Distinctive Subunits of Human NAD-Specific Isocitrate Dehydrogenase. *Biochemistry* **2007**, *46*, 5391–5397. [CrossRef] [PubMed]
28. Zeng, L.; Morinibu, A.; Kobayashi, M.; Zhu, Y.; Wang, X.; Goto, Y.; Yeom, C.J.; Zhao, T.; Hirota, K.; Shinomiya, K.; et al. Aberrant IDH3α expression promotes malignant tumor growth by inducing HIF-1-mediated metabolic reprogramming and angiogenesis. *Oncogene* **2015**, *34*, 4758–4766. [CrossRef] [PubMed]
29. Soundar, S.; O'Hagan, M.; Fomulu, K.S.; Colman, R.F. Identification of Mn^{2+}-binding Aspartates from α, β, and γ Subunits of Human NAD-dependent Isocitrate Dehydrogenase. *J. Biol. Chem.* **2006**, *281*, 21073–21081. [CrossRef] [PubMed]
30. Dumollard, R.; Campbell, K.; Halet, G.; Carroll, J.; Swann, K. Regulation of cytosolic and mitochondrial ATP levels in mouse eggs and zygotes. *Dev. Biol.* **2008**, *316*, 431–440. [CrossRef] [PubMed]
31. Soleilhavoup, C.; Riou, C.; Tsikis, G.; Labas, V.; Harichaux, G.; Kohnke, P.; Reynaud, K.; de Graaf, S.P.; Gerard, N.; Druart, X. Proteomes of the Female Genital Tract During the Oestrous Cycle. *Mol. Cell. Proteom.* **2016**, *15*, 93–108. [CrossRef] [PubMed]

32. Hugentobler, S.A.; Humpherson, P.G.; Leese, H.J.; Sreenan, J.M.; Morris, D.G. Energy substrates in bovine oviduct and uterine fluid and blood plasma during the oestrous cycle. *Mol. Reprod. Dev.* **2008**, *75*, 496–503. [CrossRef] [PubMed]
33. Clarke, C.F.; Williams, W.; Teruya, J.H. Ubiquinone biosynthesis in Saccharomyces cerevisiae. Isolation and sequence of COQ3, the 3,4-dihydroxy-5-hexaprenylbenzoate methyltransferase gene. *J. Biol. Chem.* **1991**, *266*, 16636–16644. [PubMed]
34. Dumollard, R.; Ward, Z.; Carroll, J.; Duchen, M.R. Regulation of redox metabolism in the mouse oocyte and embryo. *Dev. Camb. Engl.* **2007**, *134*, 455–465. [CrossRef] [PubMed]
35. Ben-Meir, A.; Burstein, E.; Borrego-Alvarez, A.; Chong, J.; Wong, E.; Yavorska, T.; Naranian, T.; Chi, M.; Wang, Y.; Bentov, Y.; et al. Coenzyme Q10 restores oocyte mitochondrial function and fertility during reproductive aging. *Aging Cell* **2015**, *14*, 887–895. [CrossRef] [PubMed]
36. Wyss, M.; Kaddurah-Daouk, R. Creatine and Creatinine Metabolism. *Physiol. Rev.* **2000**, *80*, 1107–1213. [CrossRef] [PubMed]
37. Ellery, S.J.; LaRosa, D.A.; Kett, M.M.; Della Gatta, P.A.; Snow, R.J.; Walker, D.W.; Dickinson, H. Maternal creatine homeostasis is altered during gestation in the spiny mouse: Is this a metabolic adaptation to pregnancy? *BMC Pregnancy Childbirth* **2015**, *15*. [CrossRef] [PubMed]
38. Ginger, M.L.; Ngazoa, E.S.; Pereira, C.A.; Pullen, T.J.; Kabiri, M.; Becker, K.; Gull, K.; Steverding, D. Intracellular Positioning of Isoforms Explains an Unusually Large Adenylate Kinase Gene Family in the Parasite Trypanosoma brucei. *J. Biol. Chem.* **2005**, *280*, 11781–11789. [CrossRef] [PubMed]
39. Fernandez-Gonzalez, A.; Kourembanas, S.; Wyatt, T.A.; Mitsialis, S.A. Mutation of Murine Adenylate Kinase 7 Underlies a Primary Ciliary Dyskinesia Phenotype. *Am. J. Respir. Cell Mol. Biol.* **2009**, *40*, 305–313. [CrossRef] [PubMed]
40. Acuña, O.S.; Avilés, M.; López-Úbeda, R.; Guillén-Martínez, A.; Soriano-Úbeda, C.; Torrecillas, A.; Coy, P.; Izquierdo-Rico, M.J. Differential gene expression in porcine oviduct during the oestrous cycle. *Reprod. Fertil. Dev.* **2017**, *29*, 2387. [CrossRef] [PubMed]
41. Warren, D.E. The role of mineralized tissue in the buffering of lactic acid during anoxia and exercise in the leopard frog Rana pipiens. *J. Exp. Biol.* **2005**, *208*, 1117–1124. [CrossRef] [PubMed]
42. Donohoe, P.H.; Boutilier, R.G. The use of extracellular lactate as an oxidative substrate in the oxygen-limited frog. *Respir. Physiol.* **1999**, *116*, 171–179. [CrossRef]
43. Abboud, J.; Storey, K.B. Novel control of lactate dehydrogenase from the freeze tolerant wood frog: Role of posttranslational modifications. *PeerJ* **2013**, *1*, e12. [CrossRef] [PubMed]
44. Kiss, A.J.; Muir, T.J.; Lee, R.E., Jr.; Costanzo, J.P. Seasonal Variation in the Hepatoproteome of the Dehydration- and Freeze-Tolerant Wood Frog, Rana sylvatica. *Int. J. Mol. Sci.* **2011**, *12*, 8406–8414. [CrossRef] [PubMed]
45. Costanzo, J.P.; do Amaral, M.C.F.; Rosendale, A.J.; Lee, R.E. Hibernation physiology, freezing adaptation and extreme freeze tolerance in a northern population of the wood frog. *J. Exp. Biol.* **2013**, *216*, 3461–3473. [CrossRef] [PubMed]
46. Kumashiro, N.; Beddow, S.A.; Vatner, D.F.; Majumdar, S.K.; Cantley, J.L.; Guebre-Egziabher, F.; Fat, I.; Guigni, B.; Jurczak, M.J.; Birkenfeld, A.L.; et al. Targeting Pyruvate Carboxylase Reduces Gluconeogenesis and Adiposity and Improves Insulin Resistance. *Diabetes* **2013**, *62*, 2183–2194. [CrossRef] [PubMed]
47. Ikeda, Y.; Yamamoto, J.; Okamura, M.; Fujino, T.; Takahashi, S.; Takeuchi, K.; Osborne, T.F.; Yamamoto, T.T.; Ito, S.; Sakai, J. Transcriptional Regulation of the Murine Acetyl-CoA Synthetase 1 Gene through Multiple Clustered Binding Sites for Sterol Regulatory Element-binding Proteins and a Single Neighboring Site for Sp1. *J. Biol. Chem.* **2001**, *276*, 34259–34269. [CrossRef] [PubMed]
48. Luong, A.; Hannah, V.C.; Brown, M.S.; Goldstein, J.L. Molecular Characterization of Human Acetyl-CoA Synthetase, an Enzyme Regulated by Sterol Regulatory Element-binding Proteins. *J. Biol. Chem.* **2000**, *275*, 26458–26466. [CrossRef] [PubMed]
49. Donkor, J.; Sariahmetoglu, M.; Dewald, J.; Brindley, D.N.; Reue, K. Three Mammalian Lipins Act as Phosphatidate Phosphatases with Distinct Tissue Expression Patterns. *J. Biol. Chem.* **2006**, *282*, 3450–3457. [CrossRef] [PubMed]
50. Phan, J.; Péterfy, M.; Reue, K. Lipin Expression Preceding Peroxisome Proliferator-activated Receptor-γ Is Critical for Adipogenesis in Vivo and in Vitro. *J. Biol. Chem.* **2004**, *279*, 29558–29564. [CrossRef] [PubMed]

51. Ramanadham, S.; Ali, T.; Ashley, J.W.; Bone, R.N.; Hancock, W.D.; Lei, X. Calcium-independent phospholipases A $_2$ and their roles in biological processes and diseases. *J. Lipid Res.* **2015**, *56*, 1643–1668. [CrossRef] [PubMed]
52. Barbour, S.E.; Kapur, A.; Deal, C.L. Regulation of phosphatidylcholine homeostasis by calcium-independent phospholipase A2. *Biochim. Biophys. Acta* **1999**, *1439*, 77–88. [CrossRef]
53. Song, Y.; Wilkins, P.; Hu, W.; Murthy, K.S.; Chen, J.; Lee, Z.; Oyesanya, R.; Wu, J.; Barbour, S.E.; Fang, X. Inhibition of calcium-independent phospholipase A_2 suppresses proliferation and tumorigenicity of ovarian carcinoma cells. *Biochem. J.* **2007**, *406*, 427–436. [CrossRef] [PubMed]
54. Chang, T.Y.; Chang, C.C.Y.; Cheng, D. Acyl-coenzyme A:cholesterol acyltransferase. *Annu. Rev. Biochem.* **1997**, *66*, 613–638. [CrossRef] [PubMed]
55. Wang, Y.-J.; Bian, Y.; Luo, J.; Lu, M.; Xiong, Y.; Guo, S.-Y.; Yin, H.-Y.; Lin, X.; Li, Q.; Chang, C.C.Y.; et al. Cholesterol and fatty acids regulate cysteine ubiquitylation of ACAT2 through competitive oxidation. *Nat. Cell Biol.* **2017**, *19*, 808–819. [CrossRef] [PubMed]
56. Boden, G. Obesity, insulin resistance and free fatty acids. *Curr. Opin. Endocrinol. Diabetes Obes.* **2011**, *18*, 139–143. [CrossRef] [PubMed]
57. Cable, E.E.; Miller, T.G.; Isom, H.C. Regulation of Heme Metabolism in Rat Hepatocytes and Hepatocyte Cell Lines: δ-Aminolevulinic Acid Synthase and Heme Oxygenase Are Regulated by Different Heme-Dependent Mechanisms. *Arch. Biochem. Biophys.* **2000**, *384*, 280–295. [CrossRef] [PubMed]
58. Wingert, R.A.; Galloway, J.L.; Barut, B.; Foott, H.; Fraenkel, P.; Axe, J.L.; Weber, G.J.; Dooley, K.; Davidson, A.J.; Schmidt, B.; et al. Deficiency of glutaredoxin 5 reveals Fe–S clusters are required for vertebrate haem synthesis. *Nature* **2005**, *436*, 1035–1039. [CrossRef] [PubMed]
59. Smith, S.J.; Cox, T.M. Translational control of erythroid delta-aminolevulinate synthase in immature human erythroid cells by heme. *Cell. Mol. Biol. (Noisy-Le-Grand)* **1997**, *43*, 103–114.
60. Wiśniewski, J.R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6*, 359–362. [CrossRef] [PubMed]
61. Zhao, Y.-L.; Zhou, Y.-H.; Chen, J.-Q.; Huang, Q.-Y.; Han, Q.; Liu, B.; Cheng, G.-D.; Li, Y.-H. Quantitative proteomic analysis of sub-MIC erythromycin inhibiting biofilm formation of S. suis in vitro. *J. Proteom.* **2015**, *116*, 1–14. [CrossRef] [PubMed]
62. Zhang, M.; Li, Y.; Yao, B.; Sun, M.; Wang, Z.; Zhao, Y. Transcriptome Sequencing and de novo Analysis for Oviductus Ranae of Rana chensinensis Using Illumina RNA-Seq Technology. *J. Genet. Genom.* **2013**, *40*, 137–140. [CrossRef] [PubMed]
63. Koenig, T.; Menze, B.H.; Kirchner, M.; Monigatti, F.; Parker, K.C.; Patterson, T.; Steen, J.J.; Hamprecht, F.A.; Steen, H. Robust Prediction of the MASCOT Score for an Improved Quality Assessment in Mass Spectrometric Proteomics. *J. Proteome Res.* **2008**, *7*, 3708–3717. [CrossRef] [PubMed]
64. Dong, M.; Gu, J.; Zhang, L.; Chen, P.; Liu, T.; Deng, J.; Lu, H.; Han, L.; Zhao, B. Comparative proteomics analysis of superior and inferior spikelets in hybrid rice during grain filling and response of inferior spikelets to drought stress using isobaric tags for relative and absolute quantification. *J. Proteom.* **2014**, *109*, 382–399. [CrossRef] [PubMed]
65. Yu, F.; Han, X.; Geng, C.; Zhao, Y.; Zhang, Z.; Qiu, F. Comparative proteomic analysis revealing the complex network associated with waterlogging stress in maize (*Zea mays* L.) seedling root cells. *Proteomics* **2015**, *15*, 135–147. [CrossRef] [PubMed]
66. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef] [PubMed]
67. Gotz, S.; Garcia-Gomez, J.M.; Terol, J.; Williams, T.D.; Nagaraj, S.H.; Nueda, M.J.; Robles, M.; Talon, M.; Dopazo, J.; Conesa, A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **2008**, *36*, 3420–3435. [CrossRef] [PubMed]

Sample Availability: Samples of *Rana chensinensis* oviduct are available from the authors.

molecules

MDPI

Article

Comparative Targeted Proteomics of the Central Metabolism and Photosystems in SigE Mutant Strains of *Synechocystis* sp. PCC 6803

Yuma Tokumaru [1], Kiyoka Uebayashi [1], Masakazu Toyoshima [1], Takashi Osanai [2], Fumio Matsuda [1] and Hiroshi Shimizu [1,*]

[1] Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan; yuma_tokumaru@ist.osaka-u.ac.jp (Y.T.); kiyoka_uebayashi@ist.osaka-u.ac.jp (K.U.); toyoshima104@ist.osaka-u.ac.jp (M.T.); fmatsuda@ist.osaka-u.ac.jp (F.M.)

[2] School of Agriculture, Meiji University, 1-1-1, Higashimita, Tama, Kawasaki, Kanagawa 214-8571, Japan; tosanai@meiji.ac.jp

* Correspondence: shimizu@ist.osaka-u.ac.jp; Tel.: +81-6-6879-7446

Academic Editor: Paolo Iadarola

Received: 9 April 2018; Accepted: 27 April 2018; Published: 1 May 2018

Abstract: A targeted proteome analysis was conducted to investigate the SigE dependent-regulation of central metabolism in *Synechocystis* sp. PCC 6803 by directly comparing the protein abundance profiles among the wild type, a *sigE* deletion mutant (Δ*sigE*), and a *sigE* over-expression (*sigE*ox) strains. Expression levels of 112 target proteins, including the central metabolism related-enzymes and the subunits of the photosystems, were determined by quantifying the tryptic peptides in the multiple reaction monitoring (MRM) mode of liquid-chromatography–triple quadrupole mass spectrometry (LC–MS/MS). Comparison with gene-expression data showed that although the abundance of Gnd protein was closely correlated with that of *gnd* mRNA, there were poor correlations for GdhA/*gdhA* and glycogen degradation-related genes such as GlgX/*glgX* and GlgP/*glgP* pairs. These results suggested that the regulation of protein translation and degradation played a role in regulating protein abundance. The protein abundance profile suggested that SigE overexpression reduced the proteins involved in photosynthesis and increased GdhA abundance, which is involved in the nitrogen assimilation pathway using NADPH. The results obtained in this study successfully demonstrated that targeted proteome analysis enables direct comparison of the abundance of central metabolism- and photosystem-related proteins.

Keywords: central carbon metabolism; targeted proteome analysis; NAPDH balance; *Synechocystis* sp. PCC 6803; *sigE*

1. Introduction

As a photosynthetic organism, cyanobacteria can be used as a host organism for the direct production of bioenergy and chemicals from solar energy and CO_2 [1–4]. A model cyanobacteria, *Synechocystis* sp. PCC 6803, can adapt metabolically to environmental perturbations by modulating protein expression levels [5,6]. A sigma factor, SigE, is a global regulator of central carbon metabolism, whose expression is induced under nitrogen-starved conditions in an NtcA-dependent manner [7–11]. The expression of SigE is also induced at the end of day time or 12 h after light exposure, suggesting that SigE is responsible for a metabolic regulation required to survive in dark conditions [12].

For a detailed investigation of SigE functions, *Synechocystis* sp. PCC 6803 strains lacking (Δ*sigE*) and over-expressing (*sigE*ox) the *sigE* gene have been constructed in previous studies [10,13].

Transcriptome analysis of the Δ*sigE* strain under normal photoautotrophic conditions showed that the expression levels of genes involved in the pentose phosphate pathway (*zwf*, *gnd*, *tal*, and *opcA*), glycolysis (*gap1*, *pfkA* (*sll1196*), and *pyk1*), and glycogen degradation (*glgX* (*slr0237*), *glgP* (*sll1356*)) were reduced by the loss of SigE function [10]. These sugar-metabolism genes that were down-regulated in the Δ*sigE* strain were also down-regulated in ΔsigCDE [14]. It was also reported that *sigE* over-expression in the *sigE*ox strain induced an increase in the expression of genes in the pentose phosphate pathway (*zwf*, *gnd*, *tal*, and *opcA*) and glycogen degradation pathways (*glgX* and *glgP*) [13]. These findings indicated that SigE is a global positive regulator of genes in the pentose phosphate and glycogen degradation pathways. Gene-expression analysis also suggested that the degradation of glycogen into glucose and the regeneration of NADPH by the oxidative pentose phosphate pathway were activated under dark conditions by the function of SigE. NADPH would be used for the maintenance of vital functions in dark conditions instead of being supplied from the photosystem.

A proteome-level analysis is required for a more detailed investigation of the metabolic regulation mechanism. Western blotting analysis of the *sigE*ox strain showed that the expression of enzyme proteins in the oxidative pentose phosphate pathway (Zwf, Gnd) and glycogen degradation pathways (GlgX and GlgP) increased and were consistent with that of gene-expression data [13,15]. A more comprehensive analysis is required because proteome analysis has disclosed the multiple post-transcriptional mechanisms that play important roles in the regulation of protein expression, protein–protein interactions, signaling, and enzymatic activity [16–18]. Thus, in this study, the SigE-dependent regulation of central metabolism in *Synechocystis* sp. PCC 6803 was confirmed by the targeted proteome analysis for the direct comparison of the protein abundance profiles among the wild-type (WT), Δ*sigE*, and *sigE*ox strains.

Shotgun proteomics has been employed for the discovery of novel regulation mechanisms in *Synechocystis* sp. PCC 6803 [19–23]. For instance, 45% of *Synechocystis* proteins were simultaneously identified and quantified by shotgun proteomics [20]. Data analysis showed that 155 proteins were differentially expressed across autotrophic, heterotrophic, photoheterotrophic and mixotrophic conditions [20]. On the other hand, targeted proteomics is a method for sensitive, precise quantification of expression levels of preselected proteins [24,25]. In targeted proteomics, a crude protein extract from cyanobacteria cells is digested by trypsin [26]. Expression levels of target proteins are determined by quantifying amounts of the preselected tryptic peptides by the selected or multiple reaction monitoring (MRM) mode of triple quadrupole mass spectrometry. The MRM assay method for cyanobacteria has been comprehensively developed for the central metabolism-related enzymes [27,28]. Targeted proteome analysis using the MRM assay method successfully quantified more proteins than shotgun proteome analysis in *Synechocystis* sp. PCC 6803 grown in iron-deficient conditions [29].

In this study, a targeted proteome analysis was conducted to measure directly the abundance of 144 target proteins in the central metabolism related-enzymes and the photosynthetic apparatus. The protein abundance data, including 112 proteins, were successfully obtained from the WT, Δ*sigE*, and *sigE*ox strains cultured under photoautotrophic conditions to understand the SigE-dependent regulation of central metabolism and photosynthesis at the protein layer. The results showed that one of the most important roles of SigE was as a positive regulator of oxidative pentose phosphate pathway (OxPPP) activity and NADPH reproduction, because the abundances of Gnd in OxPPP and NADPH/$NADP^+$ ratios were significantly changed in Δ*sigE* and *sigE*ox. The protein abundance profile also suggested that SigE overexpression increases GdhA abundance, which is involved in the nitrogen assimilation pathway using NADPH and downregulates the proteins involved in photosynthesis. Those results confirmed the SigE-dependent regulation of the C/N balance at the proteome level in *Synechocystis* sp. PCC 6803.

2. Results and Discussion

2.1. Targeted Proteome Analysis of Wild-Type, sigE Deleted and Overexpressed Strains of Synechocystis sp. PCC 6803

Three strains of *Synechocystis* sp. PCC 6803, WT, *sigE*-deleted (Δ*sigE*), and overexpressed (*sigE*ox) strains were cultivated in flasks under photoautotrophic conditions with continuous light at 40 μmol m^{-2} s^{-1}. Modified BG11 medium containing 5 mM NH_4Cl as a nitrogen source was employed throughout this study. The culture profile data (Figure 1) showed that cells actively grew until 48 h after the start of cultivation. Although the cell growth rate was reduced after 48 h, cell density was gradually increased and reached an OD_{730} = 2–3 (WT: 2.4, Δ*sigE*: 2.8, and *sigE*ox: 2.3) at 168 h. The growth curves of the three strains including WT, Δ*sigE*, and *sigE*ox were essentially similar to each other, as demonstrated in previous studies [10,13].

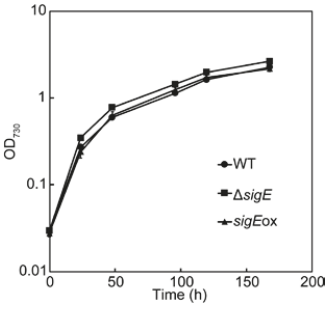


Figure 1. Cell growth curve of wild-type (WT), *sigE* deleted (Δ*sigE*) and overexpressed (*sigE*ox) strains of *Synechocystis* sp. PCC 6803. Flask-scale cultivations were performed in modified BG11 medium containing 5 mM NH_4Cl as a nitrogen source under photoautotrophic conditions with continuous light at 40 μmol m^{-2} s^{-1}. Means of triplicate cultivations were represented with standard deviations.

For the targeted proteome analysis, crude protein extracts were obtained from cells collected at mid-log phase (OD_{730} = 0.4–0.7). Following reductive alkylation and digestion by trypsin, the tryptic peptide samples were served for the nano liquid-chromatography–triple quadrupole mass spectrometry (LC–MS/MS) analysis. Here, we employed a series of MRM assay methods with 3065 channels to analyze 686 tryptic peptides derived from 144 target proteins including enzyme and subunit proteins responsible for central metabolism and photosynthetic apparatus (Table S1). In order to compare protein levels precisely, fully ^{15}N-labeled tryptic peptide samples were prepared from WT, Δ*sigE*, and *sigE*ox strains grown in modified BG11 medium containing $^{15}NH_4Cl$ as the nitrogen source. The fully ^{15}N-labeled tryptic peptide samples were used as internal standards. The tryptic peptides, whose signals were commonly observed among 3 strains in the targeted proteomics data with the largest signal-to-noise ratios, were employed for protein quantification (Table S2). The MRM assays in this study successfully determined the levels of tryptic peptides derived from the 112 proteins (Table S2, Figures S1–S5). Because the signal change of multiple peptides constituting one protein were not significantly different, one of these peptides was selected as a representative for the quantification of protein (Figures S1–S4).

Figures 2 and 3 show the heat map representations of the fold change in the relative abundance of central metabolic enzymes (Figure 2) and photosynthetic proteins (Figure 3) in Δ*sigE* (Δ*sigE*/WT) and *sigE*ox (*sigE*ox/WT) compared to those in WT. The enzyme abundance profiles were compared by volcano plots using thresholds with fold of change (FC) > 1.5, FC < 0.667 and p value < 0.05 (Figure 4). The comparison of the protein abundance levels of 112 target proteins between WT and Δ*sigE* showed that the abundances of four proteins decreased in Δ*sigE* including Fda, Gnd, TalB, and GltA (Figure 4A).

Although it was not statistically significant, the abundance of CbbA increased in Δ*sigE* (Figure 2). Although GltA is an enzyme for citrate synthase in the tricarboxylic acid (TCA) cycle, four proteins including CbbA, Fda, Gnd, and TalB were commonly responsible for the Calvin–Benson cycle and the OxPPP.

The protein abundance profile data showed that two enzymes in the pentose phosphate pathway (PPP), Gnd and TalB, were upregulated in *sigE*ox, whereas abundances of these enzymes were decreased in Δ*sigE* (Figure 2, Figure 4). These results suggested that the expression of Gnd and TalB are positively regulated by SigE. Furthermore, the relative abundances of Fda (Slr0943, fructose-1,6-bisphosphate aldolase, class I) levels decreased in both the Δ*sigE* and *sigE*ox strains while the expression of another aldolase, CbbA (Sll0018, fructose-1,6-bisphosphate aldolase class II), was negatively regulated by SigE (Figure 2, Figure 4). These results suggested that two aldolases, Fda and CbbA, had distinct roles in regulating the Calvin–Benson cycle. Indeed, it has been reported that sedoheptulose-1,7-bisphosphate (SBP) in addition to fructose-1,6-bisphosphate (FBP) could be a substrate of both Fda and CbbA. However, the SBP/FBP activity ratio of the CbbA was two times higher than that of the Fda [30]. The up- and down-regulation of proteins, however, did not affect the normal cell metabolic functions, as the three strains showed similar growth curves (Figure 1).

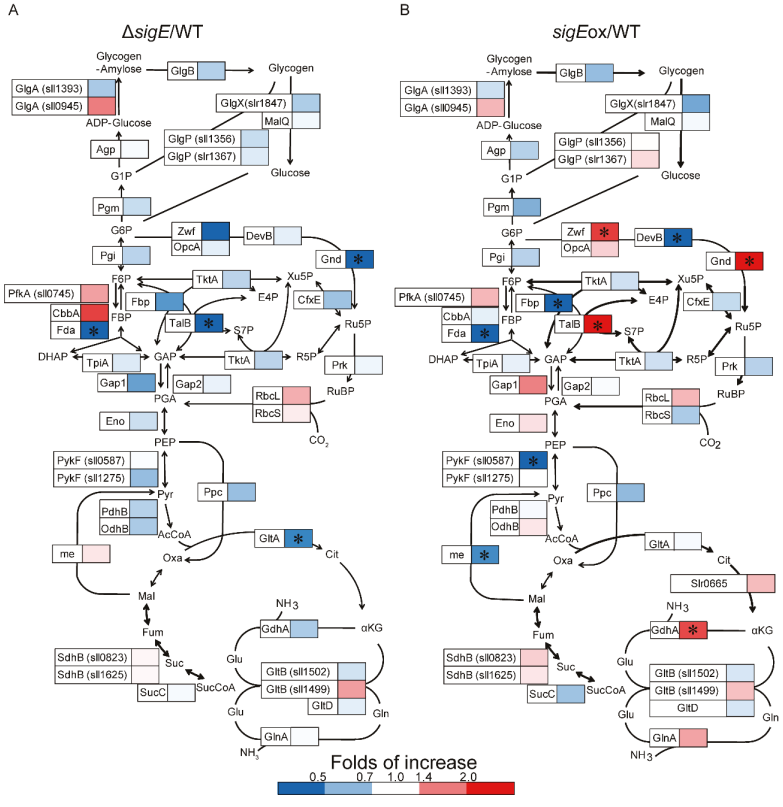


Figure 2. Effect of *sigE* deletion and overexpression on the expression of the central metabolism proteins of *Synechocystis* sp. PCC 6803. Relative expression levels [Δ*sigE*/WT (**A**) and *sigE*ox/WT (**B**)] of proteins determined by the targeted proteome analysis are represented by heat maps. Asterisks indicate a significant difference in terms of both two-sided Student's *t*-test (α 0.05) and the fold increase (>1.5 and <0.67).

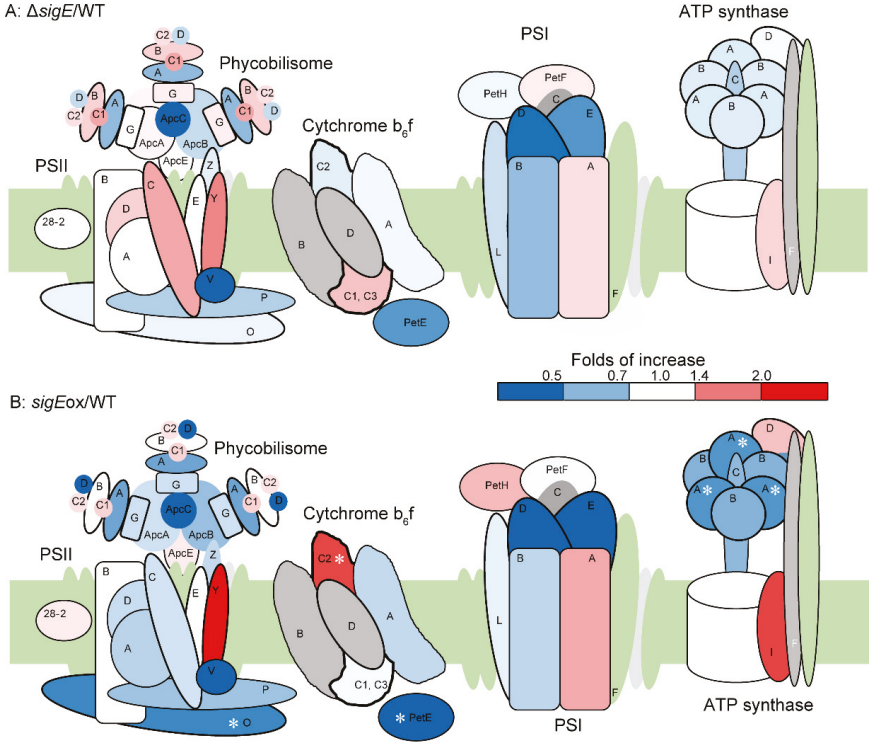


Figure 3. Effect of *sigE* deletion and overexpression on the expression of photosystem-related proteins in *Synechocystis* sp. PCC 6803. Relative expression levels [Δ*sigE*/WT (**A**) and *sigE*ox/WT (**B**)] of proteins determined by the targeted proteome analysis were represented by heat maps. Asterisks indicate a significant difference in terms of both two-sided Student's *t*-test (α 0.05) and the fold increase (>1.5 and <0.67).

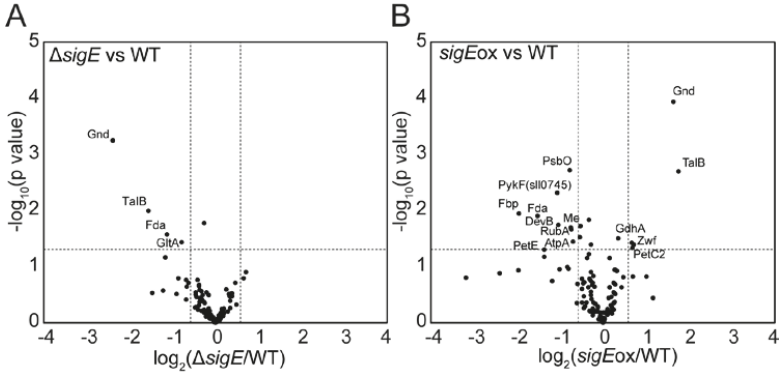


Figure 4. Volcano plot for finding proteins whose abundances were significantly changed in terms of both two-sided Student's *t*-test (α 0.05) and the fold increase (>1.5 and <0.67). (**A**) Δ*sigE* vs. WT. (**B**) *sigE*ox vs. WT. Values are means of three independent experiments.

2.2. Comparison with the Gene-Expression Data

The relative protein abundance levels determined by the targeted proteome analysis in this study were compared with the gene-expression data obtained by the transcriptome analysis performed in a previous study (Figure 5). As shown in the above section, the abundances of Gnd protein in Δ*sigE* and *sigE*ox strains were significantly lower and higher than that of WT, respectively. Essentially identical patterns were observed for *gnd* mRNA expression levels determined by the previous microarray analysis (Figure 5). The similarities between protein abundance and gene-expression patterns were also observed for Zwf/*zwf* and TalB/*talB* in the PPP, and CbbA/*cbbA*, and Gap1/*gap1* in glycolysis (Figure 5). For instance, protein abundances of GlnA and GdhA in the nitrogen assimilation pathway were different from that of the corresponding genes (Figure 5). Furthermore, the microarray analyses in the previous studies pointed out that the expression levels of glycogen degradation-related genes such as *glgX* (*slr0237*) and *glgP* (*sll1356*) were significantly changed by the regulation of *sigE*. However, this regulation was not observed in protein levels as the relative protein abundances of GlgP and GlgX were different from those of gene expressions (Figure 5). Although this discrepancy may be derived from differences in culture conditions (shaking flask in this study vs. bubbling flask in the previous study), results also suggested that the regulation of protein translation and degradation played a role in regulating protein abundance.

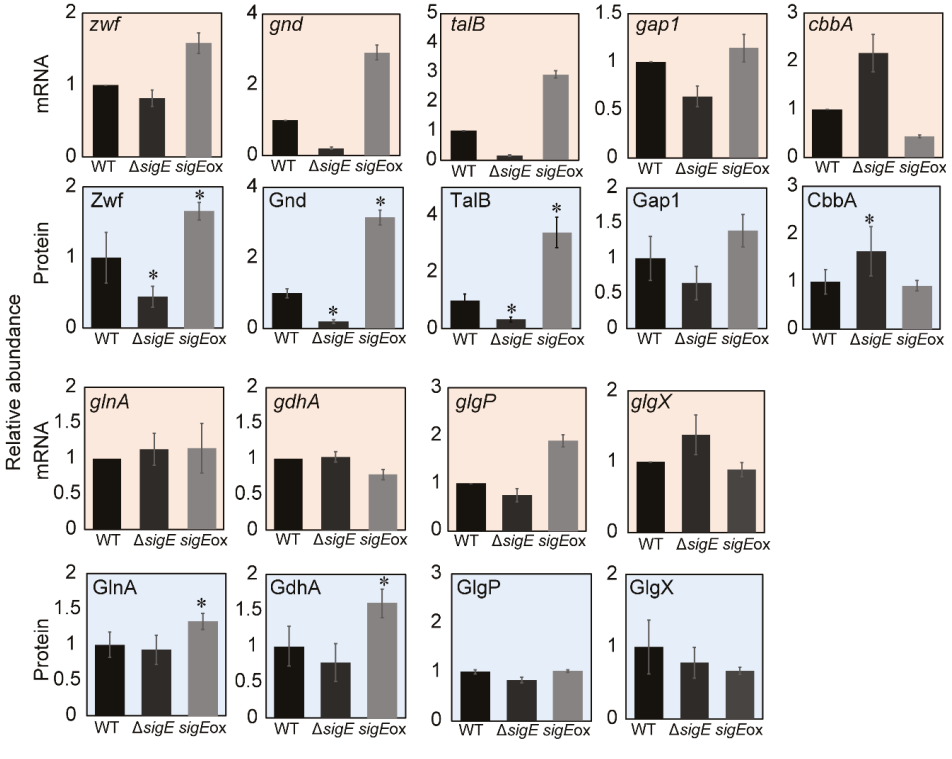


Figure 5. Comparison between gene expression and protein abundance in the wild-type (WT), Δ*sigE*, and *sigE*ox strains. The gene expression data obtained by the microarray analysis was obtained from a previous study [13]. Values are means ± standard deviation (SD) of three independent experiments. Asterisks indicate a significant difference assessed with two-sided Student's *t*-test with an α level of 0.05.

2.3. Effect of sigE Deletion and Overexpression on NADPH/NADP$^+$ Ratio

The targeted proteome analysis revealed that among the investigated proteins, the abundance of proteins in the PPP, such as Gnd and TalB, were most significantly and directly affected by SigE deletion and overexpression (Figure 2, Figure 4). The direct measurement of protein abundance confirmed that the activation and inactivation of the PPP via Gnd and TalB expression was a key mechanism in the SigE-dependent metabolic regulation in *Synechocystis* sp. PCC 6803.

Although TalB (transaldolase) is a part of the Calvin–Benson cycle for carbon fixation, Gnd (gluconate-6-phosphate dehydratase) is responsible for NADPH regeneration in the OxPPP. Therefore, the SigE-dependent regulation of Gnd level could affect the NADPH regeneration rate. The NADPH/NADP$^+$ assay shown in Figure 6A showed that the NADPH/NADP$^+$ ratio in *sigE*ox was 1.33 times larger than that of Δ*sigE*, and that the NADPH/NADP$^+$ ratios in Δ*sigE* and *sigE*ox tended to decrease and increase compared to that of wild type, respectively. These results revealed that the abundance of Gnd in the OxPPP was significantly under the control of SigE and that changes in Gnd abundance should affect the NADPH balance via the OxPPP. It has been demonstrated that the metabolic flux level in the OxPPP is up-regulated by treatment with the PSII inhibitor, DCMU [3-(3,4-dichlorophenyl)-1,1-dimethylurea]. In photoheterotrophic conditions, cyanobacteria utilize glucose or glycogen as carbon sources and survive by reproducing NADPH through the OxPPP [31,32].

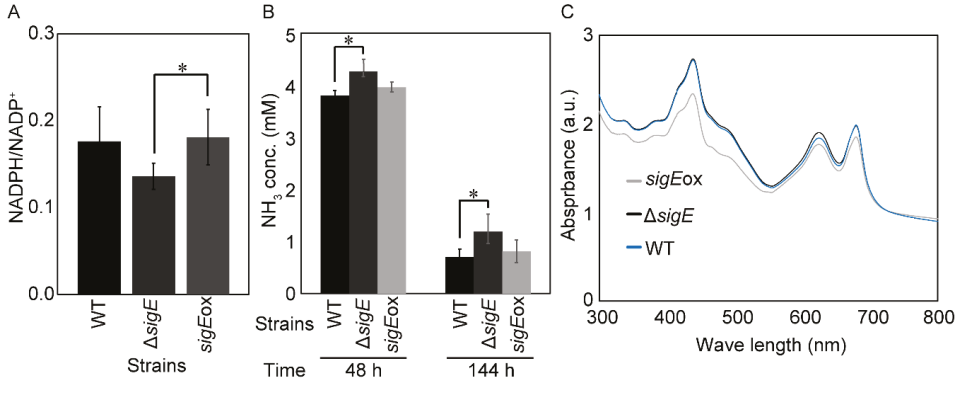


Figure 6. Effect of *sigE* deletion and overexpression on phenotypes of *Synechocystis* sp. PCC 6803. (**A**) NADPH/NADP$^+$ ratio. (**B**) NH_3 concentration in the medium. Values are means ± SD of three independent experiments. Differences were assessed with two-sided Student's *t*-tests with an alpha level of 0.05. Asterisks indicate a significant difference ($p < 0.05$). (**C**) UV-VIS spectra. Absorbance levels were normalized as ABS730 as 1.0.

2.4. SigE-Dependent Regulation of the Nitrogen Assimilation Pathway

In cyanobacteria, ammonium is incorporated into 2-oxoglutarate (2-OG) via the glutamine synthetase (GS, GlnA) and glutamate synthase (GOGAT, GltB) cycle, known as the GS–GOGAT pathway [33]. The GS-GOGAT cycle is regulated by the global nitrogen assimilation regulator NtcA. Alternatively, ammonium can be incorporated directly into glutamate by NADP-dependent glutamate dehydrogenase (GdhA), in a less efficient but less energy-consuming manner [34]. The targeted proteome analysis performed in this study showed that the abundance of GdhA was significantly changed in Δ*sigE* strains (Figure 2, Figure 4). The increase in GdhA abundances in *sigE*ox strains should associate with the regulation of the OxPPP as GdhA consumes NADPH for the nitrogen assimilation in *Synechocystis* sp. PCC 6803 [35]. However, increased expression of the *gdhA* gene was not found in the previous transcriptome analysis (Figure 5), suggesting that the regulation of GdhA abundance was sensitive to culture conditions and that protein translation and degradation play a role.

In order to confirm the effect of *sigE* deletion and over-expression on nitrogen assimilation, WT, *sigE*ox, and Δ*sigE* strains were cultivated again under the same conditions to determine concentrations of remaining ammonium in the medium (Figure 6B). The analysis showed that the ammonium concentration of Δ*sigE* at 48 h was 4.3 mM, which was 1.11 times larger than that of WT (Figure 6B). On the other hand, the concentration of remaining ammonium in the *sigE*ox strain was similar to that of WT. Similar results were also observed at 144 h (Figure 6B). Those results suggested that deleting SigE could reduce nitrogen assimilation, probably by decreasing the GS-GOGAT pathway genes such as GdhA abundance and NADPH supply from the oxidative pentose phosphate pathway.

2.5. Protein Abundance Profiles of the Photosynthetic Apparatus

It was expected that the assimilated nitrogen was mainly used for the biosynthesis of proteins such as the photosynthetic apparatus. The volcano plot of the targeted proteome data (Figure 4) showed that the abundance of PsbO, RubA, AtpA and PetE was significantly decreased in the *sigE*ox strains. PsbO is a manganese-stabilizing polypeptide in PSII that binds to a putative Mn-binding protein and keeps 2 of the 4 Mn-atoms. The absence of PsbO in *Synechocystis* affects the coordination of photosynthesis/respiration [36]. RubA is an iron-sulfur protein (rubredoxin) responsible for the assembly of PSI [37]. AtpA is the alpha subunit of ATP synthase. PetE is a plastocyanin that participates in electron transfer between P700 and the cytochrome b6-f complex in PSI [38]. The abundances of antenna proteins such as ApcA, ApcB and ApcC tended to decrease in the *sigE*ox strain although this effect was not statistically significant.

The comparison of ultraviolet–visible (UV–VIS) spectra among the WT, *sigE*ox, and Δ*sigE* strains revealed that the absorbance spectrum of Δ*sigE* was essentially similar to that of WT (Figure 6C). The result was consistent with the targeted proteome data as the abundance of the photosystem-related proteins was not significantly changed in Δ*sigE*. On the other hand, the absorbance of the carotenoid, the phycocyanin and the chlorophyll in the photosynthetic apparatus (about 550 nm, 630 nm and 680 nm, respectively) was reduced in *sigE*ox, suggesting that photosynthetic apparatus construction was disturbed by SigE overexpression. In particular, the significant decrease of the carotenoid is consistent with the previous studies in which photosynthesis activity was decreased under high light conditions and the activity of non-photochemical quenching (NPQ) was down-regulated in the *sigE*ox strain [15,39]. The decrease of the carotenoid may be related to the fact that metabolic flow was directed to synthesis of the polyhydroxybutyrate (PHB) [15]. However, the photosynthesis activity was maintained under usual light [39]. Accordingly, $NADPH/NADP^+$ ratio in *sigE*ox strain was not significantly different from WT strain (Figure 6A). In *sigE*ox strain, PPP and *gdhA* expression were much higher and photosystems expression were lower than in WT, but $NADPH/ NADP^+$ ratio, nitrogen assimilation and growth rate were not different. These results indicate that NADPH from PPP reduces the need from photosynthesis.

The down-regulation of photosynthetic apparatus could be explained in the context of the C/N balance [40]. As the photosynthetic apparatus are the most abundant protein complexes in cyanobacteria, it has been considered that antenna proteins could be a nitrogen sink in these species. For instance, phycobilisome and the associated linkers are degraded to supply nitrogen under conditions of N starvation [41]. It was reported that degradation of the nitrogen-rich phycobilisomes starts 12 h after nitrogen starvation [42], and also that reduced CO_2 fixation leads to the down-regulation of genes encoding proteins involved in nitrogen assimilation [43]. The targeted proteome analysis conducted in this study suggested that *sigE* globally regulated the metabolism to increase nitrogen assimilation in *sigE*ox. These results suggested that the overexpression of *sigE* mimicked the nitrogen-limiting condition in *Synechocystis* cell. This is consistent with in the previous studies, in which *sigE* plays a role in recovery from nitrogen deprivation [44–46]. Decreases in the antenna protein would be compensated by improved light reaction efficiency in the photosystem.

3. Materials and Methods

3.1. Bacterial Strains and Culture Conditions

Synechocystis sp. PCC 6803 GT strain, isolated by Williams [47], the Δ*sigE* strain [10] in which the *sigE* (*sll1689*) gene in the genome was disrupted, and the *sigE*ox strain [13] which constitutively expresses the gene were used in this study. These strains were grown in modified BG11 medium, containing 5 mM NH_4Cl as a nitrogen source. Cells were grown in 100 mL of medium in 500 mL Erlenmeyer flasks for batch culture under photoautotrophic conditions under continuous light (about 40 μmol m^{-2} s^{-1}) and 34 °C, and cells in the linear growth phase (OD_{730} = 0.4–0.7) [48] were collected for analysis. Kanamycin (10 μg mL^{-1}) was added in the precultures of Δ*sigE* and *sigE*ox.

3.2. Sample Preparation for Proteome Analysis

Total proteins were extracted as described in Picotti et al. [49]. Cell-culture medium containing cells in the logarithmic growth phase (100 mL, OD_{730} = 0.4–0.7) was collected by centrifugation (5000× *g*, 4 °C, 5 min). Pellets were resuspended in 1 mL of lysis buffer [50 mM HEPES, 15% Glycerol, 15 mM DTT (Dithiothreitol), 100 mM KCl, 5 mM EDTA (Ethylenediamine-*N*,*N*,*N*′,*N*′-tetraacetic acid, disodium salt, dihydrate), and one cOmplete protease inhibitors cocktail (Roche, Basel, Switzerland)]. The suspension was transferred to an Eppendorf tube containing zirconia beads (0.6 and 6 mm beads) and disrupted with a Beads Crusher μT-12 (TAITEC, Saitama, Japan) (3000 min^{-1}, 6 min). The resulting solution was centrifuged (15,000 rpm, 4 °C, 5 min), and the supernatant was transferred to a proteomics Eppendorf tube (protein low-adsorption tube) to obtain protein extraction samples. The protein concentration of the extracted sample was measured by the Bradford method, and the total protein amount was adjusted to 50 μg. Denatured buffer (500 mM Tris-HCl, 10 mM EDTA, 7 M Guanidine HCl) was added to the adjusted sample to a total volume of 220 μL.

3.3. Reduction and Alkylation/Methanol Chloroform Precipitation

One microliter of 50 mg mL^{-1} DTT was added and shaken at room temperature for 1 h using a tube mixer (CM-1000 Cute Mixer, EYELA, Tokyo, Japan), then 2.5 μL of 50 mg mL^{-1} iodoacetamide (IAA) was added and shaken for 1 h to reduce/alkylate the proteins. Next, the proteins were purified by methanol/chloroform precipitation. Proteins were purified as described in Wessel and Flügge [50]. Cold methanol (600 μL) was added to the sample solution and mixed by inversion, then 150 μL of cold chloroform was added and mixed by inversion. Cold milli-Q water (450 μL) was added and mixed by inversion, followed by centrifugation (15,000 rpm, 4 °C, 5 min). The upper layer was removed, and 450 mL of cold methanol was added and mixed gently by inversion. After centrifuging (15,000 rpm, 4 °C, 5 min) using a swing rotor, the supernatant was removed then additional centrifugation (15,000 rpm, 4 °C, 1 min) was performed to completely remove the supernatant.

3.4. Trypsin/LysC Digestion

Trypsin hydrolyzed the ester bonds on the carboxyl side of Arg and Lys and LysC on the carboxyl side of Lys. With trypsin alone, Lys which can not be leaved remains, but by combining LysC, the overall degradation efficiency and reproducibility are improved. Trypsin/LysC digestion was performed as described previously [26]. To the supernatant obtained from the above steps, 9 μL of 6 M urea was added and shaken for about 10 min at room temperature using a tube mixer. Thirty-six microliters of 0.1 M Tris-HCl (pH 8.5) was added and ultrasonic treatment and standing on ice were repeated twice for 30 s using an ultrasonic washer (Branson 2510, Danbury, CT, USA) to resuspend the protein precipitate. One microliter of 0.5 mg mL^{-1} LysC solution and 2.5 μL of 1% Protease Max solution ware added and mixed by tapping, then incubated at 25 °C for 3 h. One microliter of 0.5 mg mL^{-1} trypsin solution was added and mixed by tapping, then incubated at 37 °C for 16 h.

3.5. Desalting Samples

Sample desalination was performed as described previously [51–53]. Milli-Q water (7.5 µL) and 3 µL of 50% formic acid aqueous solution were added to the trypsin digestion product, and the mixture was stirred with a vortex mixer and centrifuged (15,000 rpm, 4 °C, 5 min). Fifty-eight microliters of the supernatant were obtained as a result. In order to carry out relative quantitative analysis, ^{15}N samples, obtained by culturing *Synechocystis* sp. PCC 6803 in BG11 medium whose nitrogen source was replaced with $^{15}NH_4Cl$, and ^{14}N samples were mixed so that the protein amount became 1:1 to prepare an analytical sample. The samples were diluted five times by Reagent A (5% acetonitrile, 0.1% formic acid). The diluted samples were desalted by handmade Stage-tip (3M Empore disk C18). The column was equilibrated with the same bed volume of Reagent B (80% acetonitrile, 0.1% formic acid). Then the column was washed with reagent A by centrifugation (4000× *g*, 10 min, RT). The samples corresponding to 5.8 µg were loaded onto the column, followed by centrifugation (4000× *g*, 10 min, RT). The column was washed by a bed volume of reagent A by centrifugation (4000× *g*, 10 min, RT) twice. The same volume of reagent B was added and centrifuged (4000× *g*, 10 min, RT), and the prepared sample was dried and solidified with a centrifugal concentrator, and then dissolved with 36 µL of 0.1% formic acid before nano LC–MS/MS analysis.

3.6. Design of Multiple Reaction Monitoring (MRM) Assay

First, the target 144 proteins related to the central metabolic pathway and photosynthetic apparatus were selected from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg/kegg2.html). The amino acid sequences of target proteins were obtained from Cyanobase (http://genome.microbedb.jp/CyanoBase). The MRM method used to quantify these 144 proteins was created by the open software Skyline version 2.6 [54]. Each protein was subjected to a tryptic peptide filter of 8 to 25 residues and five y-fragments (y1 to y5) were selected for each peptide. Samples of *Synechocystis* sp. PCC 6803 were analyzed once by nano LC–MS/MS (LCMS-8060, Shimadzu, Kyoto, Japan) by the provisional MRM method. From the results obtained by the analysis, peak picking was performed based on the shape, coelution, and intensity of the peak, and the best transitions up to 5 were selected, respectively. For proteins with no suitable tryptic peptides and transitions, transitions were quantified for all y fragments and b fragments, from which the tryptic peptides suitable for quantitation were selected again to create the final MRM method.

3.7. Nano Liquid-Chromatography–Triple Quadrupole Mass Spectrometry (LC–MS/MS) Analysis by MRM Assay

The trypsin-digested samples were analyzed by a quadrupole mass spectrometer (LCMS-8060, Shimadzu) as described previously [55]. Electrospray ionization (ESI) was performed, and sample separation was performed by nanoLC (LC-20ADnano, Shimadzu). The analytical conditions were as follows: high-performance liquid chromatography (HPLC) column, L-column ODS (pore size: 5 µm, 0.1 × 150 mm, CERI, Tokyo, Japan); trap column, L-column ODS (pore size: 5 µm, 0.3 × 5 mm, CERI); solvent system, water (0.1% formic acid) : acetonitrile (0.1% formic acid); gradient program, 10:90, *v*/*v* at 0 min, 10:90 at 10 min, 40:60 at 45 min, 95:5 at 55 min, and 90:10 at 65 min; and flow rate, 400 nL min^{-1}. Mass spectrometry was performed in MRM mode, ESI was 1.6 kV, capillary temperature was 150 °C, collision gas was 270 kPa, resolution of Q 1 and Q 3 was Low, dwell time was 1.0 ms, pause time was 1.0 ms, retention time window was 2 min. One tryptic peptide was selected for each protein and quantified by the peak area ratio of ^{14}N sample to ^{15}N sample.

3.8. Quantification of NH_3 in the Medium

The extracellular NH_3 assay was performed by F-kit (JK International) according to the manufacturer's protocol. Briefly, the medium was centrifuged (15,000 rpm, 4 °C, 1 min) and the supernatant was obtained, followed by heating at 80 °C for 20 min to inactivate remaining enzyme.

Reagent Mix (200 μL) was reacted with 6.7 μL of the sample and incubated. A_{340} was measured after 5 and 20 min.

3.9. Ultraviolet–Visible (UV–VIS) Spectrum

A DU 8000 was used to measure the UV–VIS spectrum. Absorption spectra of cell suspensions were measured according to the 'opal glass method,' with a translucent cuvette placed in front of the detector to minimize the effect of light scattering [56]. The results obtained were normalized to the absorbance at 730 nm of chlorophyll as 1.0.

4. Conclusions

Targeted proteome analysis was conducted in this study to directly compare the abundance of the central metabolism- and the photosystem-related proteins among WT, Δ*sigE*, and *sigE*ox strains of *Synechocystis* sp. PCC 6803. The analysis showed the SigE-dependent regulation of central metabolism at the protein abundance level. Among the investigated proteins, the most direct or tight regulation via gene expression was observed for the proteins in the pentose phosphate pathway such as TalB and Gnd. Further investigations of protein abundance and their modification states such as phosphorylation by proteomic analysis should uncover the detailed regulatory mechanism of central carbon metabolism in cyanobacteria.

Supplementary Materials: The following are available online, Figure S1: Selected reaction monitoring (SRM) spectra of the peptide EVTASLVGADAGK (ApcB), Figure S2: SRM spectra of the peptide SYFASGELR (ApcB), Figure S3: SRM spectra of the peptide AVLPQNLTQAQR (Gnd), Figure S4: SRM spectra of the peptide ELEPILTK (Gnd), Figure S5: SRM spectra of the peptide VPATIEEIAAR (Eno), Figure S6: Comparison with the relative abundances of mRNAs and proteins, Table S1: MRM assay method used in this study, Table S2: Targeted proteomics data; Table S3: All analyzed peptides.

Author Contributions: F.M. and H.S. designed the experiments. Y.T. and K.U. performed the experiments. Y.T., M.T. and F.M. wrote the main manuscript text. T.O., F.M. and H.S. commented and improved the manuscript. All authors have read and approved the manuscript.

Acknowledgments: We thank Yoshihiro Toya, Natsuki Hiasa, Atsumi Tomita (Osaka University), Ichiro Hirano, and Taito Ogura (Shimadzu Co.) for helpful comments on this manuscript and for technical support for the targeted proteome analysis. This work was supported in part by Grants-in-Aid for Scientific Research (grant no. 16H06552 and 16H06559).

Conflicts of Interest: The authors declare that they have no competing interests.

References

1. Dismukes, G.C.; Carrieri, D.; Bennette, N.; Ananyev, G.M.; Posewitz, M.C. Aquatic phototrophs: Efficient alternatives to land-based crops for biofuels. *Curr. Opin. Biotechnol.* **2008**, *19*, 235–240. [CrossRef] [PubMed]
2. Georgianna, D.R.; Mayfield, S.P. Exploiting diversity and synthetic biology for the production of algal biofuels. *Nature* **2012**, *488*, 329–335. [CrossRef] [PubMed]
3. Leite, G.B.; Abdelaziz, A.E.M.; Hallenbeck, P.C. Algal biofuels: Challenges and opportunities. *Bioresour. Technol.* **2013**, *145*, 134–141. [CrossRef] [PubMed]
4. Hondo, S.; Takahashi, M.; Osanai, T.; Matsuda, M.; Hasunuma, T.; Tazuke, A.; Nakahira, Y.; Chohnan, S.; Hasegawa, M.; Asayama, M. Genetic engineering and metabolite profiling for overproduction of polyhydroxybutyrate in cyanobacteria. *J. Biosci. Bioeng.* **2015**, *120*, 510–517. [CrossRef] [PubMed]
5. Fulda, S.; Mikkat, S.; Huang, F.; Huckauf, J.; Marin, K.; Norling, B.; Hagemann, M. Proteome analysis of salt stress response in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Proteomics* **2006**, *6*, 2733–2745. [CrossRef] [PubMed]
6. Kurian, D.; Phadwal, K.; Mäenpää, P. Proteomic characterization of acid stress response in *Synechocystis* sp. PCC 6803. *Proteomics* **2006**, *6*, 3614–3624. [CrossRef] [PubMed]
7. Joseph, A.; Aikawa, S.; Sasaki, K.; Teramura, H.; Hasunuma, T.; Matsuda, F.; Osanai, T.; Hirai, M.Y.; Kondo, A. Rre37 stimulates accumulation of 2-oxoglutarate and glycogen under nitrogen starvation in *Synechocystis* sp. PCC 6803. *FEBS Lett.* **2014**, *588*, 466–471. [CrossRef] [PubMed]

8. Osanai, T.; Azuma, M.; Tanaka, K. Sugar catabolism regulated by light- and nitrogen-status in the cyanobacterium *Synechocystis* sp. PCC 6803. *Photochem. Photobiol. Sci.* **2007**, *6*, 508–514. [CrossRef] [PubMed]
9. Osanai, T.; Imamura, S.; Asayama, M.; Shirai, M.; Suzuki, I.; Murata, N.; Tanaka, K. Nitrogen induction of sugar catabolic gene expression in *Synechocystis* sp. PCC 6803. *DNA Res.* **2006**, *13*, 185–195. [CrossRef] [PubMed]
10. Osanai, T.; Kanesaki, Y.; Nakano, T.; Takahashi, H.; Asayama, M.; Shirai, M.; Kanehisa, M.; Suzuki, I.; Murata, N.; Tanaka, K. Positive regulation of sugar catabolic pathways in the cyanobacterium *Synechocystis* sp. PCC 6803 by the group 2 σ factor SigE. *J. Biol. Chem.* **2005**, *280*, 30653–30659. [CrossRef] [PubMed]
11. Osanai, T.; Sato, S.; Tabata, S.; Tanaka, K. Identification of PamA as a PII-binding membrane protein important in nitrogen-related and sugar-catabolic gene expression in *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* **2005**, *280*, 34684–34690. [CrossRef] [PubMed]
12. Yoshimura, T.; Imamura, S.; Tanaka, K.; Shirai, M.; Asayama, M. Cooperation of group 2 σ factors, SigD and SigE for light-induced transcription in the cyanobacterium *Synechocystis* sp. PCC 6803. *FEBS Lett.* **2007**, *581*, 1495–1500. [CrossRef] [PubMed]
13. Osanai, T.; Oikawa, A.; Azuma, M.; Tanaka, K.; Saito, K.; Hirai, M.Y.; Ikeuchi, M. Genetic engineering of group 2 σ factor SigE widely activates expressions of sugar catabolic genes in *Synechocystis* species PCC 6803. *J. Biol. Chem.* **2011**, *286*, 30962–30971. [CrossRef] [PubMed]
14. Hakkila, K.; Antal, T.; Gunnelius, L.; Kurkela, J.; Matthijs, H.C.P.; Tyystjärvi, E.; Tyystjarvi, T. Group 2 sigma factor mutant ΔsigCDE of the cyanobacterium *Synechocystis* sp. PCC 6803 reveals functionality of both carotenoids and flavodiiron proteins in photoprotection of photosystem II. *Plant Cell Physiol.* **2013**, *54*, 1780–1790. [CrossRef] [PubMed]
15. Osanai, T.; Numata, K.; Oikawa, A.; Kuwahara, A.; Iijima, H.; Doi, Y.; Tanaka, K.; Saito, K.; Hirai, M.Y. Increased bioplastic production with an RNA polymerase sigma factor SigE during nitrogen starvation in *Synechocystis* sp. PCC 6803. *DNA Res.* **2013**, *20*, 525–535. [CrossRef] [PubMed]
16. Huokko, T.; Muth-Pawlak, D.; Battchikova, N.; Allahverdiyeva, Y.; Aro, E.-M. Role of type 2 NAD(P)H dehydrogenase NdbC in redox regulation of carbon allocation in *Synechocystis*. *Plant Physiol.* **2017**, *174*, 1863–1880. [CrossRef] [PubMed]
17. Veit, S.; Takeda, K.; Tsunoyama, Y.; Baymann, F.; Nevo, R.; Reich, Z.; Rögner, M.; Miki, K.; Rexroth, S. Structural and functional characterisation of the cyanobacterial PetC3 Rieske protein family. *Biochim. Biophys. Acta Bioenerg.* **2016**, *1857*, 1879–1891. [CrossRef] [PubMed]
18. Pade, N.; Mikkat, S.; Hagemann, M. Ethanol, glycogen and glucosylglycerol represent competing carbon pools in ethanol-producing cells of *Synechocystis* sp. PCC 6803 under high-salt conditions. *Microbiology (UK)* **2017**, *163*, 300–307. [CrossRef] [PubMed]
19. Borirak, O.; de Koning, L.J.; van der Woude, A.D.; Hoefsloot, H.C.J.; Dekker, H.L.; Roseboom, W.; de Koster, C.G.; Hellingwerf, K.J. Quantitative proteomics analysis of an ethanol- and a lactate-producing mutant strain of *Synechocystis* sp. PCC6803. *Biotechnol. Biofuels* **2015**, *8*, 111. [CrossRef] [PubMed]
20. Fang, L.; Ge, H.; Huang, X.; Liu, Y.; Lu, M.; Wang, J.; Chen, W.; Xu, W.; Wang, Y. Trophic mode-dependent proteomic analysis reveals functional significance of light-independent chlorophyll synthesis in *Synechocystis* sp. PCC 6803. *Mol. Plant* **2017**, *10*, 73–85. [CrossRef] [PubMed]
21. Gao, L.; Wang, J.; Ge, H.; Fang, L.; Zhang, Y.; Huang, X.; Wang, Y. Toward the complete proteome of *Synechocystis* sp. PCC 6803. *Photosynth. Res.* **2015**, *126*, 203–219. [CrossRef] [PubMed]
22. Plohnke, N.; Seidel, T.; Kahmann, U.; Rögner, M.; Schneider, D.; Rexroth, S. The proteome and lipidome of *Synechocystis* sp. PCC 6803 cells grown under light-activated heterotrophic conditions. *Mol. Cell. Proteom.* **2015**, *14*, 572–584. [CrossRef] [PubMed]
23. Ren, Q.; Shi, M.; Chen, L.; Wang, J.; Zhang, W. Integrated proteomic and metabolomic characterization of a novel two-component response regulator Slr1909 involved in acid tolerance in *Synechocystis* sp. PCC 6803. *J. Proteom.* **2014**, *109*, 76–89. [CrossRef] [PubMed]
24. Liebler, D.C.; Zimmerman, L.J. Targeted quantitation of proteins by mass spectrometry. *Biochemistry* **2013**, *52*, 3797–3806. [CrossRef] [PubMed]
25. Shi, T.; Song, E.; Nie, S.; Rodland, K.D.; Liu, T.; Qian, W.J.; Smith, R.D. Advances in targeted proteomics and applications to biomedical research. *Proteomics* **2016**, *16*, 2160–2182. [CrossRef] [PubMed]

26. Uchida, Y.; Tachikawa, M.; Obuchi, W.; Hoshi, Y.; Tomioka, Y.; Ohtsuki, S.; Terasaki, T. A study protocol for quantitative targeted absolute proteomics (QTAP) by LC-MS/MS: Application for inter-strain differences in protein expression levels of transporters, receptors, claudin-5, and marker proteins at the blood-brain barrier in ddY, FVB, and C57BL/6J mice. *Fluids Barriers CNS* **2013**, *10*, 21. [PubMed]
27. Vuorijoki, L.; Isojärvi, J.; Kallio, P.; Kouvonen, P.; Aro, E.M.; Corthals, G.L.; Jones, P.R.; Muth-Pawlak, D. Development of a quantitative SRM-based proteomics method to study iron metabolism of *Synechocystis* sp. PCC 6803. *J. Proteome Res.* **2016**, *15*, 266–279. [CrossRef] [PubMed]
28. Vuorijoki, L.; Kallio, P.; Aro, E.M. SRM dataset of the proteome of inactivated iron-sulfur cluster biogenesis regulator SufR in *Synechocystis* sp. PCC 6803. *Data BR* **2017**, *11*, 572–575. [CrossRef] [PubMed]
29. Vuorijoki, L.; Tiwari, A.; Kallio, P.; Aro, E.M. Inactivation of iron-sulfur cluster biogenesis regulator SufR in *Synechocystis* sp. PCC 6803 induces unique iron-dependent protein-level responses. *Biochim. Biophys. Acta* **2017**, *1861*, 1085–1098. [CrossRef] [PubMed]
30. Nakahara, K.; Yamamoto, H.; Miyake, C.; Yokota, A. Purification and characterization of class-I and class-II fructose-1,6-bisphosphate aldolases from the cyanobacterium *Synechocystis* sp. PCC6803. *Plant Cell Physiol.* **2003**, *44*, 326–333. [CrossRef] [PubMed]
31. Nakajima, T.; Kajihata, S.; Yoshikawa, K.; Matsuda, F.; Furusawa, C.; Hirasawa, T.; Shimizu, H. Integrated metabolic flux and omics analysis of *Synechocystis* sp. PCC 6803 under mixotrophic and photoheterotrophic conditions. *Plant Cell Physiol.* **2014**, *55*, 1606–1612. [CrossRef] [PubMed]
32. You, L.; He, L.; Tang, Y.J. Photoheterotrophic fluxome in *Synechocystis* sp. strain PCC 6803 and its implications for cyanobacterial bioenergetics. *J. Bacteriol.* **2015**, *197*, 943–950. [CrossRef] [PubMed]
33. Muro-Pastor, M.I.; Reyes, J.C.; Florencio, F.J. Ammonium assimilation in cyanobacteria. *Photosynth. Res.* **2005**, *83*, 135–150. [CrossRef] [PubMed]
34. Von Wobeser, E.A.; Ibelings, B.W.; Bok, J.; Krasikov, V.; Huisman, J.; Matthijs, H.C.P. Concerted changes in gene expression and cell physiology of the cyanobacterium *Synechocystis* sp. strain PCC 6803 during transitions between nitrogen and light-limited growth. *Plant Physiol.* **2011**, *155*, 1445–1457. [CrossRef] [PubMed]
35. Chávez, S.; Reyes, J.C.; Chauvat, F.; Florencio, F.J.; Candau, P. The NADP-glutamate dehydrogenase of the cyanobacterium *Synechocystis* 6803: Cloning, transcriptional analysis and disruption of the *gdhA* gene. *Plant Mol. Biol.* **1995**, *28*, 173–188. [CrossRef] [PubMed]
36. Schriek, S.; Aguirre-Von-Wobeser, E.; Nodop, A.; Becker, A.; Ibelings, B.W.; Bok, J.; Staiger, D.; Matthijs, H.C.P.; Pistorius, E.K.; Michel, K.P. Transcript profiling indicates that the absence of PsbO affects the coordination of C and N metabolism in *Synechocystis* sp. PCC 6803. *Physiol. Plant.* **2008**, *133*, 525–543. [CrossRef] [PubMed]
37. Shen, G.; Zhao, J.; Reimer, S.K.; Antonkine, M.L.; Cai, Q.; Weiland, S.M.; Golbeck, J.H.; Bryant, D.A. Assembly of photosystem I: I. Inactivation of the *rubA* gene encoding a membrane-associated rubredoxin in the cyanobacterium *Synechococcus* sp. PCC 7002 causes a loss of photosystem I activity. *J. Biol. Chem.* **2002**, *277*, 20343–20354. [CrossRef] [PubMed]
38. Wang, X.Q.; Jiang, H.B.; Zhang, R.; Qiu, B.S. Inactivation of the *petE* gene encoding plastocyanin causes different photosynthetic responses in cyanobacterium *Synechocystis* PCC 6803 under light-dark photoperiod and continuous light conditions. *FEMS Microbiol. Lett.* **2013**, *341*, 106–114. [CrossRef] [PubMed]
39. Osanai, T.; Kuwahara, A.; Iijima, H.; Toyooka, K.; Sato, M.; Tanaka, K.; Ikeuchi, M.; Saito, K.; Hirai, M.Y. Pleiotropic effect of *sigE* over-expression on cell morphology, photosynthesis and hydrogen production in *Synechocystis* sp. PCC 6803. *Plant J.* **2013**, *76*, 456–465. [CrossRef] [PubMed]
40. Coruzzi, G.M.; Zhou, L. Carbon and nitrogen sensing and signaling in plants: Emerging "matrix effects". *Curr. Opin. Plant Biol.* **2001**, *4*, 247–253. [CrossRef]
41. Richaud, C.; Zabulon, G.; Joder, A.; Thomas, J.C. Nitrogen or sulfur starvation differentially affects phycobilisome degradation and expression of the *nblA* gene in *Synechocystis* strain PCC 6803. *J. Bacteriol.* **2001**, *183*, 2989–2994. [CrossRef] [PubMed]
42. Krasikov, V.; Aguirre von Wobeser, E.; Dekker, H.L.; Huisman, J.; Matthijs, H.C.P. Time-series resolution of gradual nitrogen starvation and its impact on photosynthesis in the cyanobacterium *Synechocystis* PCC 6803. *Physiol. Plant.* **2012**, *145*, 426–439. [CrossRef] [PubMed]
43. Singh, A.K.; Elvitigala, T.; Bhattacharyya-Pakrasi, M.; Aurora, R.; Ghosh, B.; Pakrasi, H.B. Integration of carbon and nitrogen metabolism with energy production is crucial to light acclimation in the cyanobacterium *Synechocystis*. *Plant Physiol.* **2008**, *148*, 467–478. [CrossRef] [PubMed]

44. Antal, T.; Kurkela, J.; Parikainen, M.; Kårlund, A.; Hakkila, K.; Tyystjärvi, E.; Tyystjärvi, T. Roles of group 2 sigma factors in acclimation of the cyanobacterium *Synechocystis* sp. PCC 6803 to nitrogen deficiency. *Plant Cell Physiol.* **2016**, *57*, 1309–1318. [CrossRef] [PubMed]
45. Muro-Pastor, A.M.; Herrero, A.; Flores, E. Nitrogen-regulated group 2 sigma factor from *Synechocystis* sp. strain PCC 6803 involved in survival under nitrogen stress. *J. Bacteriol.* **2001**, *183*, 1090–1095. [CrossRef] [PubMed]
46. Heilmann, B.; Hakkila, K.; Georg, J.; Tyystjärvi, T.; Hess, W.R.; Axmann, I.M.; Dienst, D. 6S RNA plays a role in recovery from nitrogen depletion in *Synechocystis* sp. PCC 6803. *BMC Microbiol.* **2017**, *17*, 229. [CrossRef] [PubMed]
47. Williams, J.G.K. Construction of specific mutations in photosystem II photosynthetic reaction center by genetic engineering methods in *Synechocystis* 6803. *Methods Enzymol.* **1988**, *167*, 766–778.
48. Schuurmans, R.M.; Matthijs, J.C.P.; Hellingwerf, K.J. Transition from exponential to linear photoautotrophic growth changes the physiology of *Synechocystis* sp. PCC 6803. *Photosynth. Res.* **2017**, *132*, 69–82. [CrossRef] [PubMed]
49. Picotti, P.; Bodenmiller, B.; Mueller, L.N.; Domon, B.; Aebersold, R. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* **2009**, *138*, 795–806. [CrossRef] [PubMed]
50. Wessel, D.; Flügge, U.I. A Method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **1984**, *138*, 141–143. [CrossRef]
51. Rappsilber, J.; Mann, M.; Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2007**, *2*, 1896–1906. [CrossRef] [PubMed]
52. Ishihama, Y.; Rappsilber, J.; Mann, M. Modular stop and go extraction tips with stacked disks for parallel and multidimensional peptide fractionation in proteomics. *J. Proteome Res.* **2006**, *5*, 988–994. [CrossRef] [PubMed]
53. Rappsilber, J.; Ishihama, Y.; Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **2003**, *75*, 663–670. [CrossRef] [PubMed]
54. MacLean, B.; Tomazela, D.M.; Shulman, N.; Chambers, M.; Finney, G.L.; Frewen, B.; Kern, R.; Tabb, D.L.; Liebler, D.C.; MacCoss, M.J. Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26*, 966–968. [CrossRef] [PubMed]
55. Matsuda, F.; Tomita, A.; Shimizu, H. Prediction of hopeless peptides unlikely to be selected for targeted proteome analysis. *Mass Spectrom. (Tokyo)* **2017**, *6*, A0056. [CrossRef] [PubMed]
56. Toyoshima, M.; Mori, N.; Moriyama, T.; Misumi, O.; Sato, N. Analysisi of triacylglycerol accumulation under nitrogen deprivation in the red alga *Cyanidioschyzon merolae*. *Microbiology* **2016**, 803–812. [CrossRef] [PubMed]

Sample Availability: Samples of the compounds are available from the authors upon request.

molecules

MDPI

Article

Identification of *Ophiocordyceps sinensis* and Its Artificially Cultured *Ophiocordyceps* Mycelia by Ultra-Performance Liquid Chromatography/Orbitrap Fusion Mass Spectrometry and Chemometrics

Ping Zhang [1,2], Saina Li [3], Juan Li [4], Feng Wei [2], Xianlong Cheng [2], Guifeng Zhang [3], Shuangcheng Ma [2,*] and Bin Liu [1,*]

1 School of Chinese Material Medica, Beijing University of Chinese Medicine, No. 11Beisanhuan east Road, Beijing 100029, China;zping0227@sina.com

2 Research and Inspection Center of Traditional Chinese Medicine and Ethnomedicine, National Institutes for Food and Drug Control, State Food and Drug Administration, No. 2 TiantanXili, Beijing 100050, China; weifeng@nifdc.org.cn (F.W.); lncxl@sina.com (X.C.)

3 State Key Laboratory of Biochemical Engineering, Institute of Process Engineering, Chinese Academy of Sciences, No. 1 Second north Road of zhongguancun, Beijing 100190, China;seine2015@163.com (S.L.); gfzhang@ipe.ac.cn (G.Z.)

4 Institute of Microbiology, Chinese Academy of Sciences, No. 1 West Beichen Road, Beijing 100101, China; 18612037125@126.com

* Correspondence: masc@nifdc.org.cn (S.M.); liubinyn67@163.com (B.L.); Tel.: +86-10-135-0139-8172 (B.L.); Fax: +86-10-67023650 (S.M.)

Received: 25 March 2018; Accepted: 19 April 2018; Published: 26 April 2018

Abstract: Since the cost of *Ophiocordyceps sinensis*, an important fungal drug used in Chinese medicine, has increased dramatically, and the counterfeits may have adverse health effects, a rapid and precise marker using the peptide mass spectrometry identification system could significantly enhance the regulatory capacity. In this study, we determined the marker peptides in the digested mixtures of fungal proteins in wild *O. sinensis* fruiting bodies and various commercially available mycelium fermented powders using ultra-performance liquid chromatography/Orbitrap Fusion mass spectrometry coupled with chemometrics. The results indicated the following marker peptides: TLLEAIDSIEPPK (m/z 713.39) was identified in the wild *O. sinensis* fruiting body, AVLSDAITLVR (m/z 579.34) was detected in the fermented *O. sinensis* mycelium powder, FAELLEK (m/z 849.47) was found in the fermented *Ophiocordyceps* mycelium powder, LESVVTSFTK (m/z 555.80) was discovered in the artificial *Ophiocordyceps* mycelium powder, and VPSSAVLR (m/z 414.75) was observed in *O. mortierella* mycelium powder. In order to verify the specificity and applicability of the method, the five marker peptides were synthesized and tested on all samples. All in all, to the best of our knowledge, this is the first time that mass spectrometry has been employed to detect the marker peptides of *O.sinensis* and its related products.

Keywords: *Ophiocordyceps sinensis*; ultra-performance liquid chromatography/Orbitrap Fusion mass spectrometry; chemometrics; fungi marker peptide; quality control

1. Introduction

Chongcao (the sexual stage of the *Ophiocordyceps sinensis*) is an important traditional fungal drug that has been commonly used for hundreds of years as a tonic and/or drug. However, its safety was questioned, because the wild *Ophiocordyceps sinensis* was reported to contain a high amount of arsenic likely due to soil contamination [1]. Then *Ophiocordyceps sinensis* had the clinical

effect of tonifying the kidney and replenishing lung, stanching bleeding, and resolving phlegm. It could be used to treat 21 ailments and also be a potential adjuvant chemotherapeutic agent in non-small cell lung cancer, liver cancer, and breast therapy [2,3]. Despite the hazardous effects to the humanhealth, chongcao possesses manyanti-tumor and antioxidant activities, as well asthe capacity to modulate the immune system and treat fatigue, night sweating, hyposexuality, hyperglycemia, liver disease, and heart disease [4–9]. In recent years, due tothe limited natural resources and simultaneously increasing demand, thecost of *O. sinensis* has increased dramatically. In spite of the increasing price, its manufacture and sales were strictly regulated in 2016 by the China Food and Drug Administration (CFDA), because its natural fruiting bodies usually contain high amounts of arsenic, which is an environmental pollutant and could decrease neuronal migration, as well as cellular maturation, and it inhibits the proliferation of neural progenitor cells [1,10]. Considering the safe clinical use and the discrepancy between need and availability, other *Ophiocordyceps*-related fungi and the conidial forms of the artificially cultured *O. sinensis* fermentation mycelia have been used as substitutes in Chinese medicine and healthy food [11,12].

From the numerous species that have been reportedly isolated from *O. sinensis* [13,14], it is widely accepted among researchers that *Hirsutella sinensis* is a unique anamorph of *O. sinensis* [11,15–17], while other species such as *Paecilomyces hepialid*, *Gliocladium roseum*, and *Tolypocladium sinensis* represent endoparasitic fungi commonly found in natural *O. sinensis* [18,19]. Currently, four species isolated from *O. sinensis*, namely, fermented *O. sinensis* mycelium powder (*Hirsutella sinensis* species), fermented *Ophiocordyceps* mycelium powder (*Paecilomyces hepialid* species), artificial *Ophiocordyceps* mycelium powder (*Gliocladium roseum* species), and *O. mortierella* mycelium powder (*Tolypocladium sinensis* species), all of which are usually confused with the *O. sinensis* name, have been successfully cultured. Four more standardized mycelia fermentation products of *Ophiocordyceps* have been produced [20–24] and are widely employed as Chinese medical material in preparations in China [25].

Since different *Ophiocordyceps* species may have different health effects, authentication of *O. sinensis*-related products is essential in order to ensure safeuse and efficacy. Traditionally, *O. sinensis* is identified through morphological description, microscopic identification, or chemical composition assay [23,26–29]. However, since these methods lack objective standards or a specific component index, it is difficult to effectively identify and distinguish *O. sinensis* from various cultured *O.* mycelia [30–34]. Furthermore, although the polymerase chain reaction (PCR) has been successfully applied in the identification of *Ophiocordyceps sinensis* (*O.S.*) fruiting bodies [14,15,35], it cannot be used for cultured *Ophiocordyceps* mycelia, because the integrity of the DNA genome is compromised during the drying process [18,36,37].

Since species identification is an important and necessary procedure to control the quality and standardization of herbal medicines, it is crucial to develop methods to authenticate *O. sinensis* and the four aforementioned cultured *O.* mycelia. One approach to that could be the detection of the fungal proteins in the species. The fungal protein is a special protein, the amino acid sequence of which is different in the different fungi, thereby making it an important factor in the identification of species [38,39]. Moreover, these proteins are among the bioactive components in *O.S.* and, to the best of our knowledge, have rarely been studied and reported [40]. Biological mass spectrometry has been developed as an efficient method for proteomic analysis that exhibits superior mass accuracy and ultra-high resolution, and employs a segmented quadruple mass filter with improved selectivity and ion transmissibility [38,39,41–46]. However, for a specific genus, there is only limited data available to characterize the fungal protein. One example is the marker peptide in the digested mixture of *O. sinensis*-related products.

In this study, we first determinedthe digested mixture peptides by ultra-performance liquid chromatography/Orbitrap Fusion mass-spectrometry (UPLC/MS/MS). Then, the marker peptide ion was detected using chemometrics, and the marker peptide sequence was confirmed by comparisonwith the database for the identification of *O. sinensis* and relevant cultured *Ophiocordyceps* mycelia. The results demonstrated that this method could be used to authenticate not only wild *O. sinensis*

and its related cultured *Ophiocordyceps* mycelia powder but also the mixed commercial products. Moreover, the work presented herein is, to the best of our knowledge, the first extensive study on the authentication of *O. sinensis* and cultured *Ophiocordyceps* mycelia using mass spectrometry and chemometrics, thereby providing a powerful quality control tool.

2. Results

2.1. Size-Exclusion Chromatographic Analysis ofFungal Proteinsand Their Tryptic Digest Mixtures

Figure 1 shows the size-exclusion chromatograms (SEC) of fungal proteins in fermented *O. sinensis* mycelia and the sample aliquots withdrawn during the digestion process. It was found that the fungal proteins had a wide molecular weight range (Figure 1A). When the amount of trypsin was increased, the peak intensities arising from the digest mixture gradually increased, indicating that more peptides in the mixture were degraded (Figure 1B–D). Once the sample-to-trypsin ratio exceeded 100:10, no significant changes in the peak intensity in the elution profile were observed (Figure 1E). The molecular-weight ranges of the fungal proteins and the digest mixtures incubated at 37 °C for 18 h were assessed against a series of protein standards, namely immunoglobulin G (molecular weight (M_W = 150 kDa), bovine albumin (M_W = 68 kDa), globular actin (M_W = 42 kDa), trypsinogen (M_W = 24 kDa), lysozyme (M_W = 14 kDa), and bovine insulin (M_W = 6 kDa), which were analyzed by SEC under the same conditions. The results revealed that the molecular weights of the fungal proteins ranged from 42 to 14 kDa, whereas the molecular weights of the peptides in the digest mixture were <5 kDa, which is consistent with the theoretical molecular weight range of peptides resulting from the digestion of fungal proteins [47].

2.2. Multivariate Data Analysis

The total ion chromatograms (TICs) of the wild *O. sinensis* and four cultured *Ophiocordyceps* mycelia fungal proteins digested at 37 °C for 18 h, over the 350–1550 m/z scan range, are displayed in Figure 2. The results show that the marker peptides of each mycelium were concealed by a large number of tryptic peptides that were indistinguishable in the TICs at higher concentrations due to the homologies of the fungal proteins. As a consequence, the different types of mycelia were difficult to detect simply by visual inspection of their chromatograms, and further sample profiling of the digest mixtures was performed using multivariate statistical software tools. In this study, the three-dimensional (3-D) ultra-performance liquid chromatography/Orbitrap Fusion mass-spectrometry (UPLC/MS/MS) data were first converted into a 2-D matrix containing exact-mass/retention-time (EMRT) pairs using the Progenesis QI for Proteomics, which is the application manager for Progenesis QI™. The data set was visualized using unsupervised principal component analysis (PCA) in order to check for outliers and classification trends among the mycelia. Preliminary PCA was performed on all observations using pareto-scaled variables. The final PCA score plot revealed that five different types of mycelia clusters formed, all of which lay inside the Hotelling T2 (0.95) ellipse (Figure 3a). In the PCA scores plot, the fermented *Ophiocordyceps* mycelia powder and *O.* mortierella mycelia powder lay close to each other but were located much further away from the wild O. sinensis, the fermented *O.S*. mycelia powder, and the artificial *Ophiocordyceps* mycelia powder.

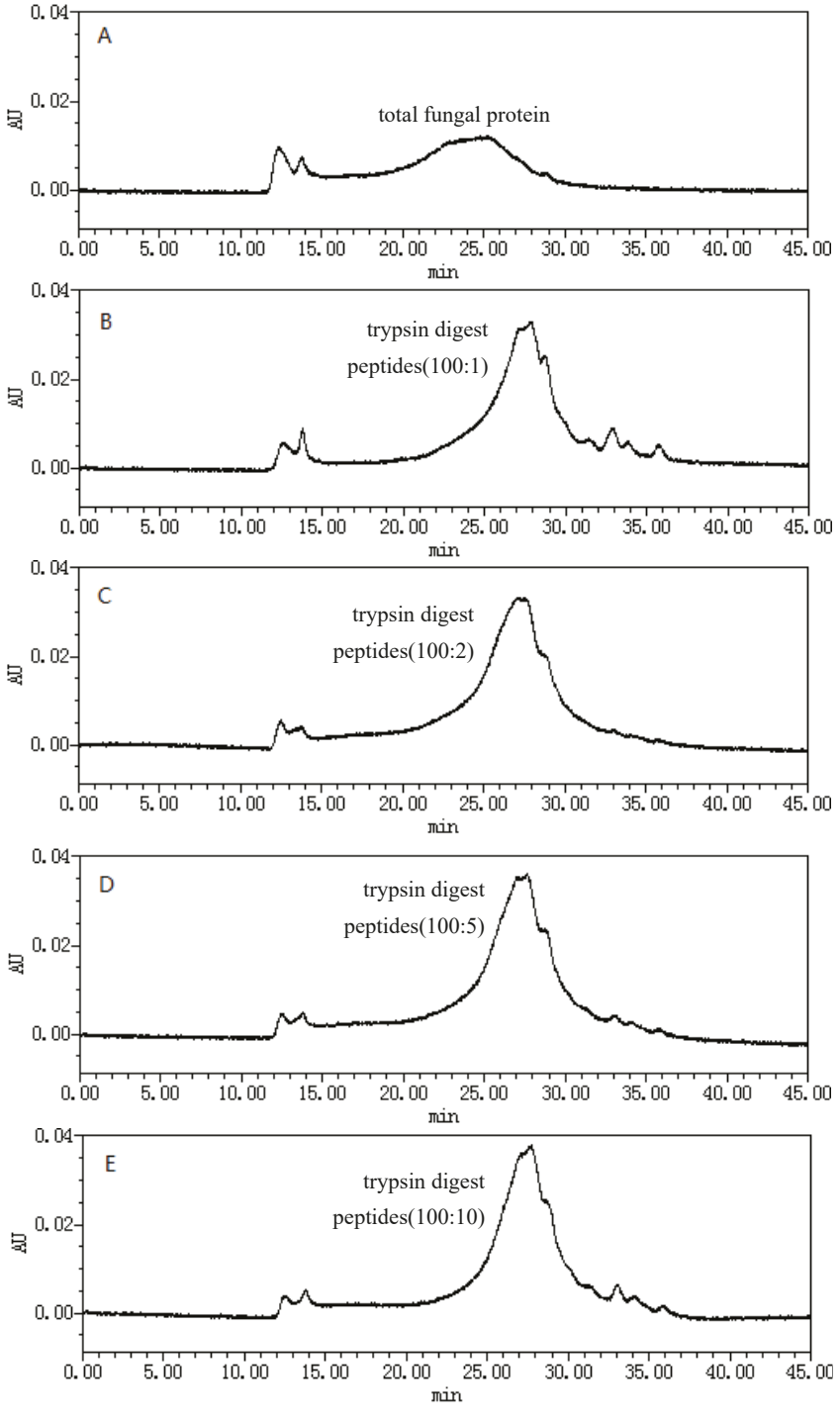


Figure 1. The gelfiltration chromatograms of (**A**) fungal protein from fermented *O. sinensis* mycelia powder and (**B**–**E**) therespective digest peptidesincubated at 37 °C for 18 h with sample-to-trypsin ratios of (**B**) 100:1; (**C**) 100:2; (**D**) 100:5; and (**E**) 100:10.

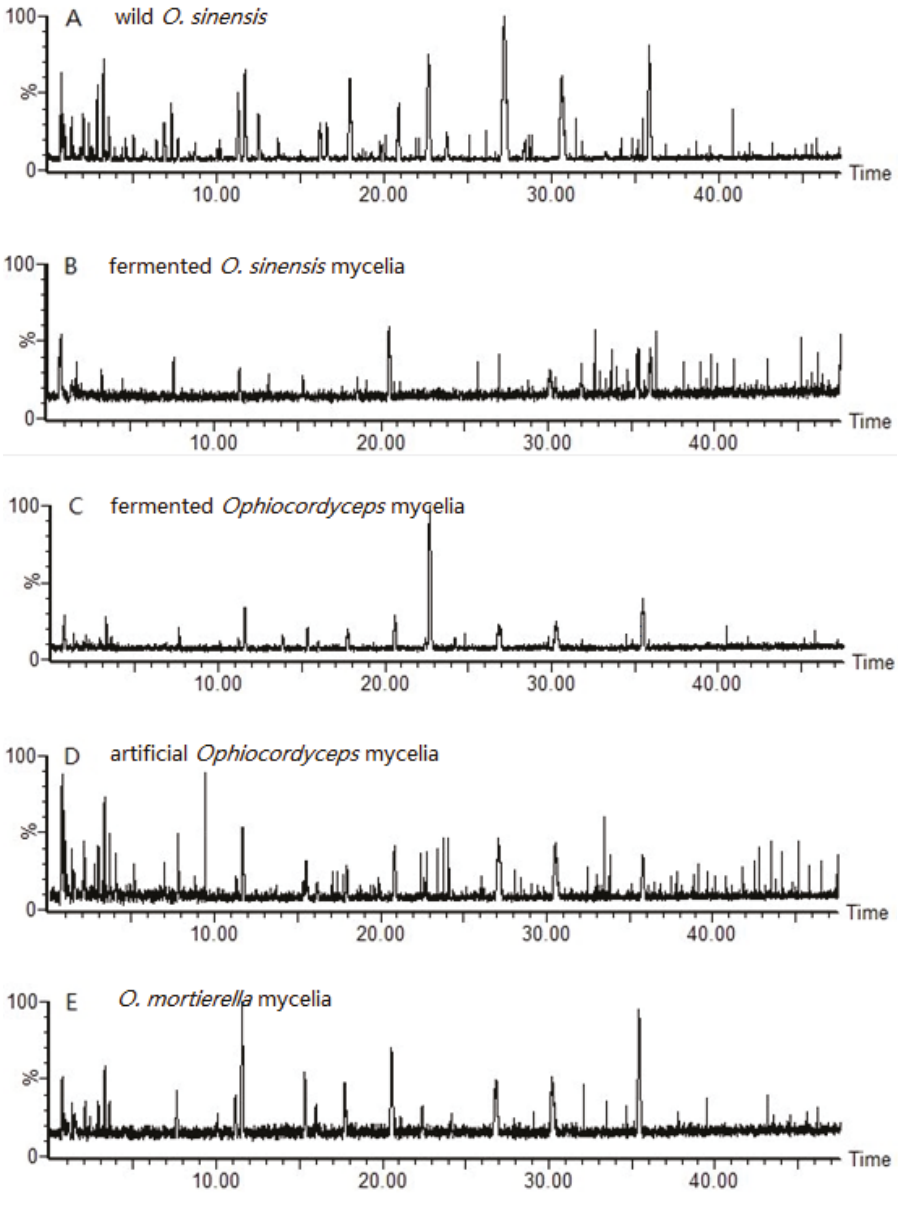


Figure 2. The positive-ion base-peak-intensity chromatograms of the digest peptides of: (**A**) wild *O. sinensis* fruiting body; (**B**) fermented *O. sinensis* mycelia powder; (**C**) fermented *Ophiocordyceps* mycelia powder; (**D**) artificial *Ophiocordyceps* mycelia powder; and (**E**) *O. mortierella* mycelia powder.

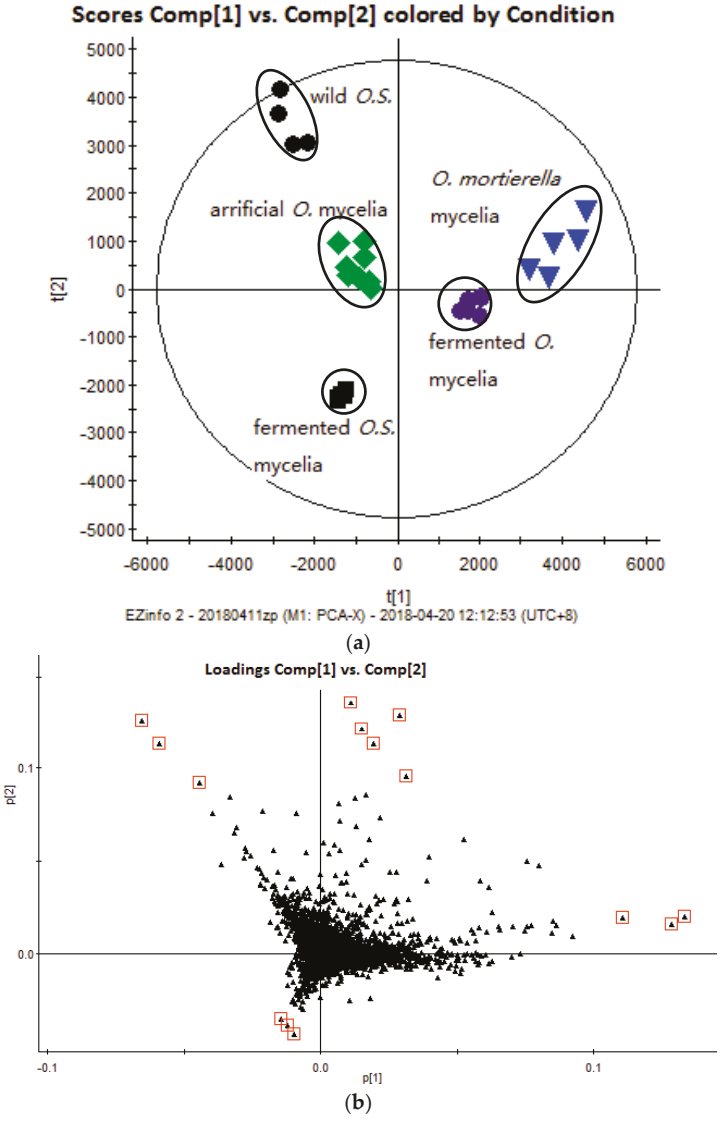


Figure 3. (**a**) The principal component analysis score plot of (●): wild *O. sinensis* fruiting body, (■): fermented *O.S.* mycelia powder, (◆): artificial *Ophiocordyceps* mycelia powder, (●): fermented *Ophiocordyceps* mycelia powder, (▼): *O. mortierella* mycelia powder, and (**b**) the loading plot of wild *O. sinensis* and four cultured *Ophiocordyceps* mycelia.

2.3. Identification of Marker Peptides in Digested Mixtures

The results from this study demonstrate that it is possible to isolate and identify marker peptides that play important roles in the authentication of various *Ophiocordyceps* mycelia. The loading plot from the PCA, based on the UPLC/MS data, is shown in Figure 3b. The ions that correspond to the EMRT pairs of 41.17–713.3967, 35.03–579.3498, 22.68–849.4741, 26.05–555.8072, and 13.60–414.7517 were chosen as marker peptides for each sample. The amino acid sequence of

the marker peptides was determined using Mascot v2.5.1. Mascot was calibrated by searching the data on *O. sinensis*, *Hirsutella sinensis*, *Paecilomyces hepialid*, *Gliocladium roseum*, and *Tolypocladium sinensis*, which was obtained from the Universal Protein (UniProt) database. The Mascot search was performed with a fragment ion mass tolerance of 0.60 Da and a parent ion tolerance of 15.0 ppm. Carbamidomethylated iodoacetamide (IAM) was specified in Mascot as a fixed modification, while oxidized dithiothreitol (DTT) was specified as a variable modification.

The results of the study showed that the ion with m/z = 713.3967 was only found in the spectrum of the digested mixture of the wild *O.S.* fruiting body (Figure 4A). Moreover, extracted-ion mass spectrometry (MS^E) revealed that this ion was doubly charged and that the MS/MS spectrum indicated that it corresponded to the TLLEAIDSIEPPK amino acid sequence. The partial LEAIDSIEPPK amino acid sequence was derived from the observed single charged y(11) m/z 1211.6518 ion, while the partial TL amino acid sequence was confirmed by the b(2) m/z 215.1390 ion. In contrast, the ion with m/z = 579.3498 was only detected in the spectrum of the digested mixture of the fermented *O.S*. mycelia powder (Figure 4B). Similarly, the respective MS^E spectrum revealed that this ion was doubly charged, while the MS/MS spectrum indicated that it corresponded to the AVLSDAITLV amino acid sequence. The partial AVL amino acid sequence was derived from the observed doubly charged $b^{++}(3)$ m/z 142.6021 ion, while the partial SDAITLVR amino acid sequence was confirmed by the y(8) m/z 874.4993 ion. In addition, the ion with m/z =849.4741 was only found in the spectrum of the digested mixture of the fermented *Ophiocordyceps* mycelia powder (Figure 4C). The MS^E spectrum revealed that this ion was single charged, and that its MS/MS spectrum indicated that it corresponded to the FAELLEK amino acid sequence. The partial FAE amino acid sequence was derived from the observed single charged b(3) m/z 348.1554 ion, while the partial LLEK amino acid sequence was confirmed by the y(4) m/z 502.3235 ion. Moreover, the ion with m/z = 555.8072 was only detected in the spectrum of the digested mixture of the artificial *Ophiocordyceps* mycelia powder (Figure 4D). The MS^E spectrum revealed that this ion was doubly charged, and that its MS/MS spectrum indicated that it corresponded to the LESVVTSFTK amino acid sequence. The partial LESV amino acid sequence was derived from the observed doubly charged $b^{++}(4)$ m/z 215.1208 ion, while the partial VTSFTK amino acid sequence was confirmed by the y(6) m/z 682.3770 ion. Lastly, the ion with m/z = 414.7517 was only found in the spectrum of the digested mixture of the *O. mortierella* mycelia powder (Figure 4E). The corresponding MS^E spectrum revealed that this ion was doubly charged, and that its MS/MS spectrum indicated that it corresponded to the VPSSAVLR amino acid sequence. The partial VPS amino acid sequence was derived from the observed doubly charged $b^{++}(3)$ m/z 142.5839 ion, while the partial SAVLR amino acid sequence was confirmed by the y(5) m/z 545.3406 ion.

The amino acid sequences of the marker peptides of each *Ophiocordyceps* species were aligned using the Basic Local Alignment Search Tool (BLAST) of the UniProt database in order to identify the original protein types. The results were as follows: translation elongation factor 1-α (gi:A4U9H1), belonging to *Ophiocordyceps brunneipunctata* (Table 1) and recognized by Mascot matching as a precursor of the tryptic peptide TLLEAIDSIEPPK (m/z 713.3967), was chosen as a marker of the wild *O. sinensis* fruiting body; linoleate diol synthase (gi: T5AC53), belonging to *Hirsutella sinensis* and recognized by Mascot matching as a precursor of tryptic peptide AVLSDAITLVR (m/z 579.3498), was chosen as a marker for unambiguous identification of the fermented *O. sinensis* mycelia powder; the adenosine triphosphate (ATP) synthase subunit α (gi: A0A0B7JUZ6), belonging to *Bionectria ochroleuca* (*Gliocladium roseum*) and recognized by Mascot matching as a precursor of tryptic peptide LESVVTSFTK (m/z 555.8072), was chosen as a marker of the artificial *Ophiocordyceps* mycelia powder(*Gliocladium roseum* species). However, the two ions at m/z 849.4741 and 414.7517 were not assigned to any protein by the Mascot matching. The selected ion monitoring chromatograms of the marker peptides and the corresponding spectra (MS^E) are shown in Figure 4. In most cases, the amino acid sequences were recognized to belong to specific proteins of the analyzed species, while in other cases, the peptide was not assigned to any protein, with the engine indicating only partial matching (in brackets, Table 1).

Table 1. The multianalyte results of the marker peptides from *Ophiocordyceps sinensis* and the four cultured *O.* mycelia.

		Marker				Mascot Matching		
Item	No.of Ions	Precursor Ion, (m/z)	Charge	Fragment Ion, (m/z)	Time, (min)	Peptide Match	Peptide Fragmentation	Protein Match
Ophiocordyceps sinensis	1	713.39	2		41.1	TLLEAIDSIEPPK		gi:A4U9H1 (*Ophiocordyceps brunneipunctata*)
				969.52			y	
				898.48			y	
				785.4			y	
				215.13			b	
				452.25			y^0	
				881.46			y *	
	2	851.74	3		46.2	SVEMHHEQLTEGLPGDNVGFNVK		gi: 0A060IK44 (*Ophiocordyceps sinensis*)
				1046.52			y	
				949.47			y	
				497.7			b++	
				596.25			b *++	
				777.42			y	
				1208.54			b	
fermented *O. sinensis* mycelia	3	579.34	2		35	AVLSDAITLVR		gi: T5AC53 (*Hirsutella sinensis*)
				987.58			y	
				874.49			y	
				787.46			y	
				672.44			y	
				488.31			y	
				142.6			b++	
	4	670.3	2		14.5	NAGSGCPTYTVGR		gi: T5AC53 *(Hirsutella sinensis)*
				1154.52			y	
				1010.47			y	
				397.21			y++	
				370.13			b *	
				387.16			b	
	5	614.85	2		54.7	MVEVLGIIQAR		gi: T5AC53 *(Hirsutella sinensis)*
				657.4			y	
				770.48			y	
				998.59			y	
				231.11			b	
				360.15			b	

Table 1. *Cont.*

	Marker				Mascot Matching		
artificial *Ophiocordyceps* mycelia	6	555.8	2	26	LESVVTSFTK		gi: 0A0B7JUZ6 (*Gliocladium roseum*)
			868.47			y	
			682.37			y	
			243.13			b	
			215.12			b++	
			482.75			b++	
	7	637.34	2	44.9	HALVIYDDLSK		gi: 0A0B7JUZ6 *(Gliocladium roseum)*
			952.49			y	
			740.34			y	
			577.28			y	
			209.1			b	
			534.33			b	
fermented *Ophiocordyceps* mycelia	8	849.47	1	22.6	(FAELLEK)		Unassigned
			502.32			y	
			389.23			y	
			348.15			b	
	9	799.44	2	58	YLEIIKETSNFIK		gi: A0A172PXZ9 (*Ophiostoma pseudocatenulatum*)
			966.52			y	
			596.85			y++	
			406.19			b	
			990.55			b	
			487.26			b *++	
O. mortierella mycelia	10	414.75	2	13.6	(VPSSAVLR)		Unassigned
			458.3			y	
			370.24			y *	
			387.27			y	
			142.58			b++	

Note: y *(m/z = 881.46) ion was dehydroxylated ion of y (m/z = 898.48); b *++ (m/z = 596.25) ion was dehydroxylated ion of b (m/z = 1208.54) with two charges, b * (m/z = 370.13) ion was dehydroxylated ion of b (m/z = 387.16), b *++ (m/z = 487.26) ion was dehydroxylated ion of b (m/z = 990.55) with two charges, and y * (m/z = 370.24) ion was dehydroxylated ion of y (m/z = 387.27).

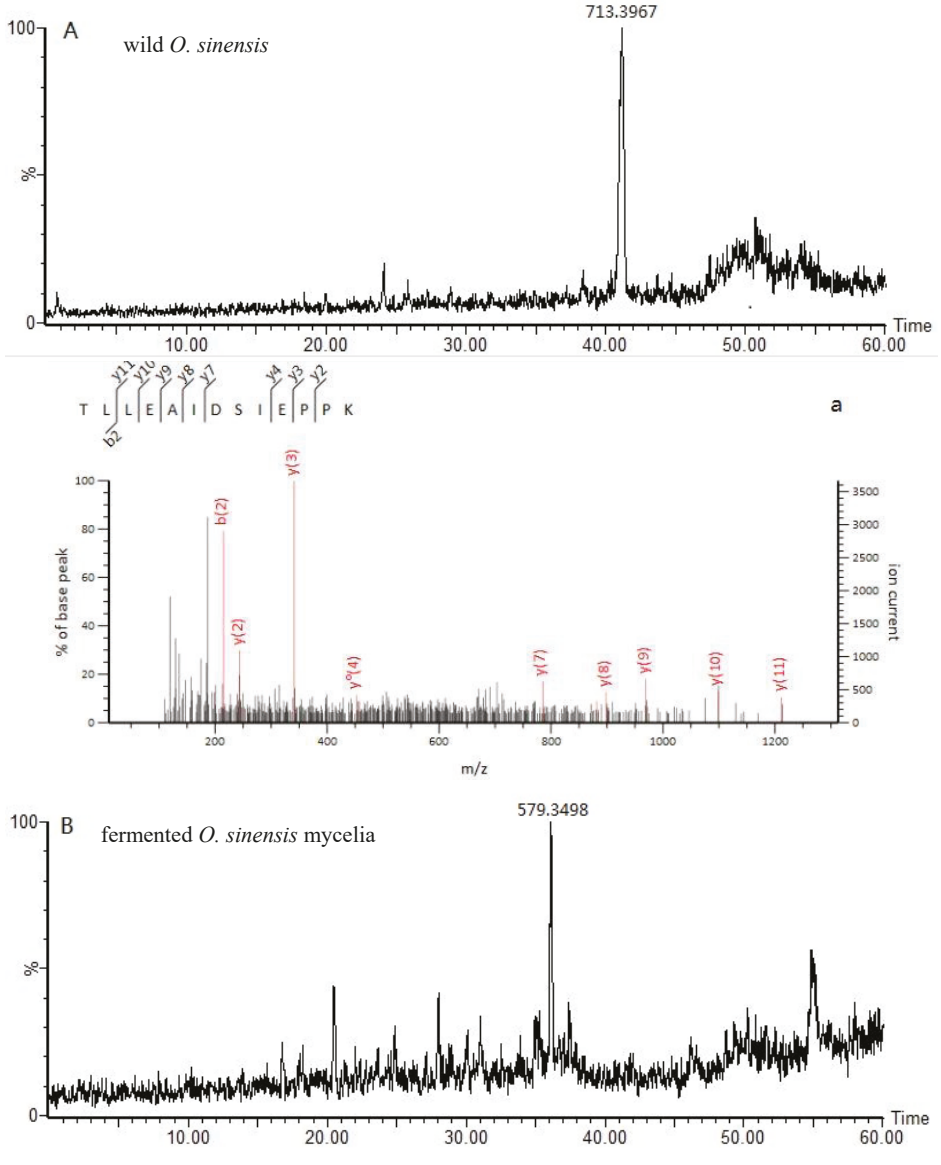


Figure 4. *Cont.*

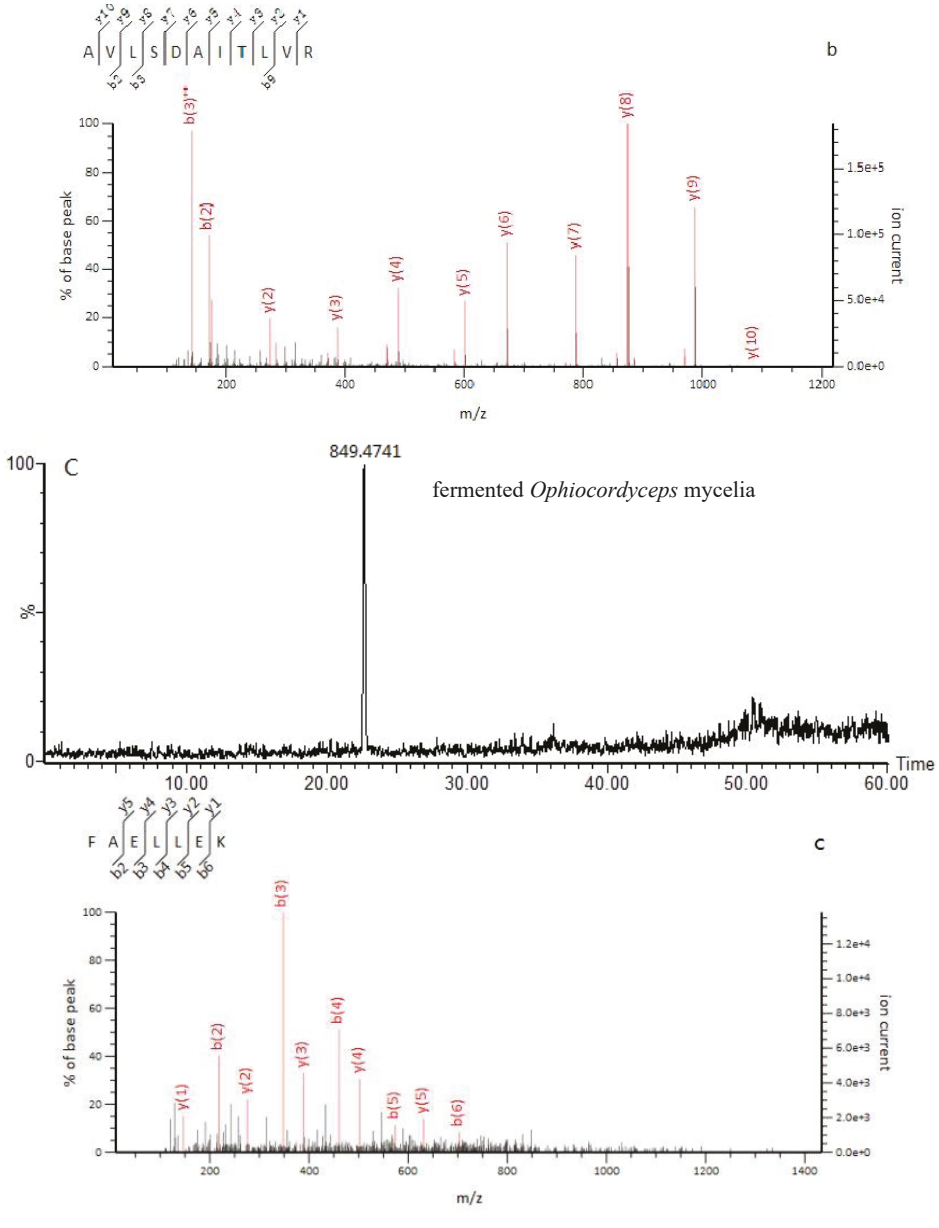


Figure 4. *Cont.*

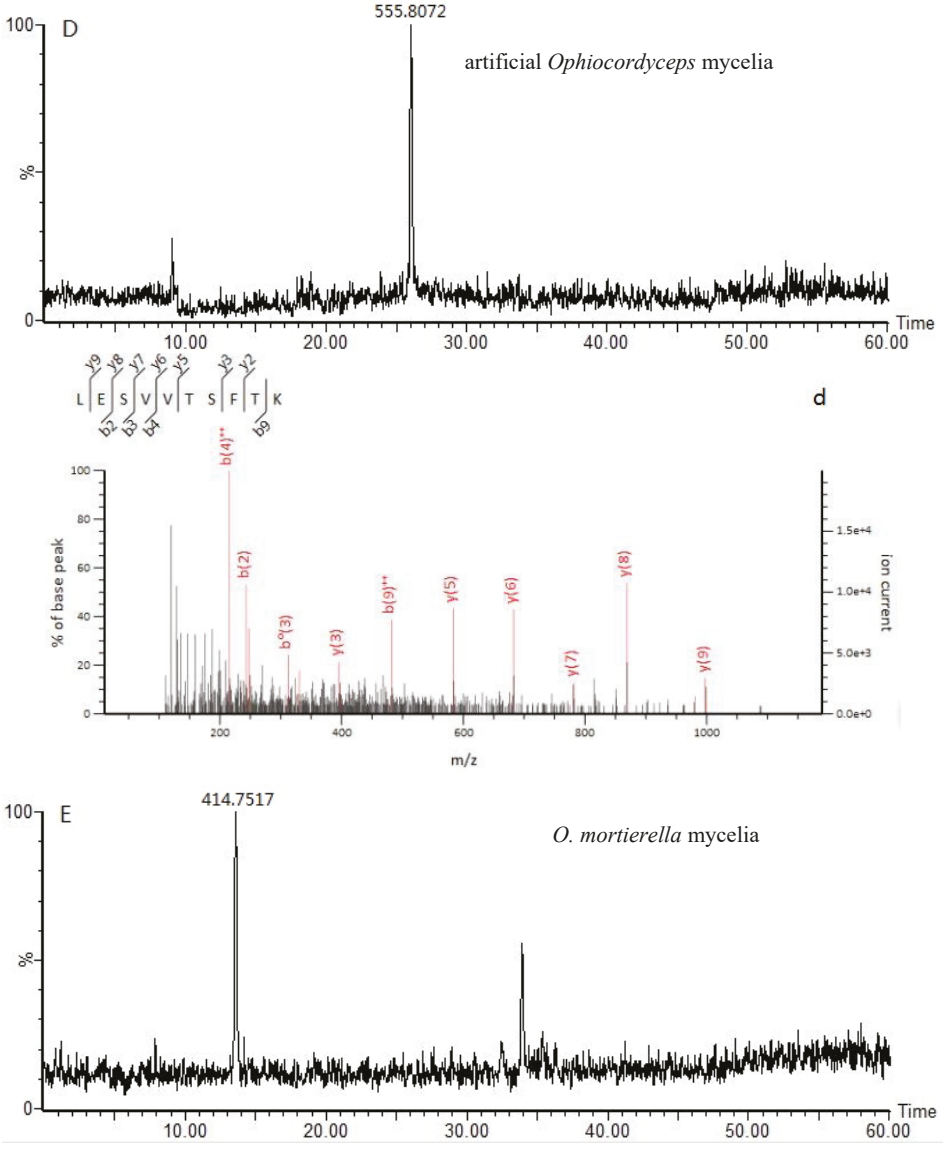


Figure 4. *Cont.*

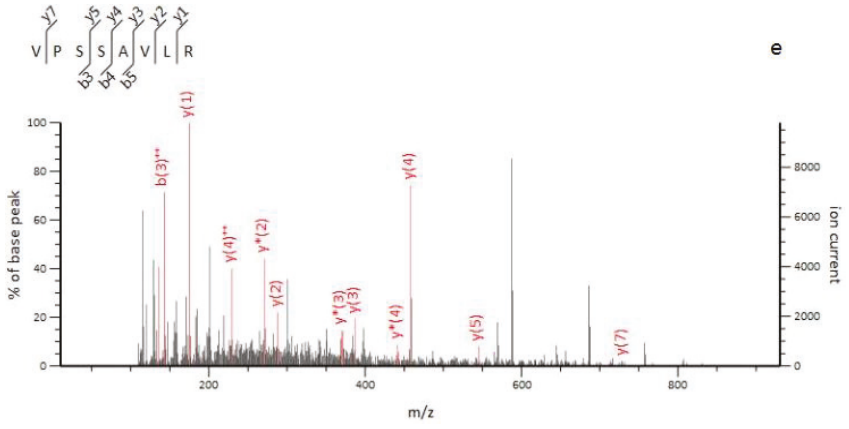


Figure 4. The selected ion-monitoring chromatograms of marker peptides in (**A**) wild *O. sinensis* fruiting body, *m*/*z* 713.39, a doubly charged TLLEAIDSIEPPK fragment ion; (**B**) fermented *O. sinensis* mycelia powder, *m*/*z* 579.34, a doubly charged AVLSDAITLVR fragment ion; (**C**) fermented *Ophiocordyceps* mycelia powder, *m*/*z* 849.47, a singly charged FAELLEK fragment ion; (**D**) artificial *Ophiocordyceps* mycelia powder, *m*/*z* 555.80, a doubly charged LESVVTSFTK fragment ion; and (**E**) *O. mortierella* mycelia powder, *m*/*z* 414.75, a doubly charged VPSSAVLR fragment ion. Correspondingly, the fragment ion mass spectrogram of marker peptides in (**a**) wild *O. sinensis* fruiting body; (**b**) fermented *O. sinensis* mycelia powder; (**c**) fermented *Ophiocordyceps* mycelia powder; (**d**) artificial *Ophiocordyceps* mycelia powder; and (**e**) *O. mortierella* mycelia powder.

3. Discussion

Due to its apparent beneficial clinical and health effects, *O. sinensis* has been employed in China as a highly valued traditional Chinese medicine for centuries. Recently, it has become increasingly popular and is now widely used, especially by elderly and unhealthy people in China and abroad, as a dietary supplement or natural remedy [25,48]. However, the market price for *O. sinensis* has increased remarkably due to insufficient resources and growing demand. Moreover, other *Ophiocordyceps*-related fungi and the conidial form of the artificially cultured *O. sinensis* fermentation mycelia have also been used as substitutes in Chinese medicine and healthy foods [11,12], thereby causing confusion in the market management and challenging the safe use of *O. sinensis*. Therefore, it is crucial to develop a reliable and practical method to differentiate *O. sinensis* from its substitutes, especially the cultured mycelia.

In this study, the marker peptides in the digest mixtures of fungal proteins were determined by UPLC/MS/MS coupled with chemometrics using wild *O.S.* fruiting bodies and several commercially available mycelium fermented powders. Moreover, the marker peptides were detected, and the amino acid sequences of the marker peptides were identified. The obtained results showed that the marker peptides could provide accurate species identification for the *Ophiocordyceps* samples by biological mass spectrometry. To the best of our knowledge, the first extensive study on the authentication of *O. sinensis* and revelent-cultured *Ophiocordyceps* mycelia by biological mass spectrometry combined with chemometrics, thereby provided a powerful quality control tool.

Previous studies report different macroscopic and microscopic methods that can be used to identify *O. sinensis*-related products [23,26–29]. Nevertheless, the probability of the accurate identification of the species level was not the same for the different species [32,49–52]. Most studies focused on morphological characterizations, microscopy studies, determination of the chemical composition, and PCR amplifications. However, none of them analyzed the specific fungal protein or, more specifically, the marker peptide that identifies the species level of *O. sinensis*.

The fungal protein is a special protein that is differentiated by the type of fungus [40]. One efficient method for proteomic analysis is biological mass spectrometry [39,53,54]. In recent years, many studies have focused on the fungal proteins. Two dimensional electrophoresis (2-DE) and sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was used to examine and identify *O. sinensis* [40,55]. Then, five proteins were identified using MALDI-TOF-TOF/MS. Based on the proteomic profile of *O. sinensis*, 2-DE identification pattern was provided, and this approach was a foundation for intensive study of *O. sinensis* proteins. Another isobaric tag for relative and absolute quantification (i TRAQ)-coupled two-dimensional liquid chromatography tandem mass spectrometry(2D LC-MS/MS) proteomics approach was used to analyze the protein profiles of samples of the larva and various development stages of Chinese *Cordyceps*. This bioinformatics analysis revealed that i TRAQ-coupled 2D LC-MS/MS was a unique method for identifying protein groups of Chinese *Cordyceps* at different development stages [56]. None of these methods can effectively identify and distinguish *Ophiocordyceps sinensis* and its revelent-fermented *Ophiocordyceps* mycelia. In this study, we employed this approach to identify the marker peptides of the specific fungal proteins in wild *O. sinensis* and four revelent-fermented *Ophiocordyceps* mycelia powders. Then, as a result, TLLEAIDSIEPPK (m/z 713.39) was detected in wild *O. sinensis* fruiting bodies, which was matched to protein of translation elongation factor 1-α, belonging to *Ophiocordyceps brunneipunctata*. AVLSDAITLVR (m/z 579.34) was discovered in the fermented *O. sinensis* mycelium powder, which was matched to protein of linoleate diol synthase, belonging to *Hirsutella sinensis.* FAELLEK (m/z 849.47) was found in fermented *O.* mycelium powder, which was not matched to any protein. LESVVTSFTK (m/z 555.80) was identified in artificial *O.* mycelium powder, which was matched to protein of ATP synthase subunit α, belonging to *Gliocladium roseum*. VPSSAVLR (m/z 414.75) was detected in *O. mortierella* mycelium powder, which was not matched to any protein. All in all, three marker peptides were matched to the corresponding species; two marker peptides were not matched to corresponding species, but they were specific peptides.

In order to verify their specificity, all marker peptides were synthesized and tested on the samples. The results from the measurements of the aforementioned five samples revealed the following information on the marker peptides: TLLEAIDSIEPPK was only present in wild *O. sinensis*, AVLSDAITLVR was only detected in fermented *O. sinensis* mycelium powder, FAELLEK was only observed in fermented *Ophiocordyceps* mycelium powder, LESVVTSFTK was only present in artificial *O.* mycelium powder, and VPSSAVLR was only found in *O. mortierella* mycelium powder. Previous studies revealed the proteins from 26 different producing areas were obviously different in the numbers and abundance of protein spots of protein profiles, and this showed certain association with producing areas [40]. Another report revealed five proteins of *O. sinensis* were identified in 2-DE, but the specific protein was not reported [55]. Isobaric tags for relative and absolute quantification (i TRAQ)-coupled two-dimensional liquid chromatography tandem mass spectrometry (2D LC-MS/MS) proteomics approach was used to analyze the protein profiles of samples of the larva and various development stages of Chinese *Cordyceps*. The results indicated that protein composition of mummified larva, sclerotium, and stroma were significantly different from commercial *cordyceps* [56]. These were the results of studying the producing area and various development stages of *O. sinensis*. Few study results of the differential proteins of *O. sinensis* and various cultured *Ophiocordyceps* mycelia were reported. In our study, the specific marker peptides were found by chemomatrics first and identified the sequence using MASCOT. Thus, we could examine the marker peptides to identify the *O. sinensis* and revelent-fermented *Ophiocordyceps* mycelia products.

The method developed in this study could be applied not only to qualitatively identify the *O. sinensis*-related species, but also to quantitatively determine the contents of the marker peptides to control the qualityof the *Ophiocordyceps* related products. Moreover, the biological mass spectrometry method is also suitable for the identification of Chinese medicinal materials derived from animals, especially processed animal medicinal materials such as *Cicadae periostracum*, processed *Manis squama*, etc. However, because Chinese herbal medicine contains many complex ingredients, the extraction

of high-purity and high-quality protein components is a key requirement of this method and its application to traditional Chinese medicine (TCM).

4. Materials and Methods

4.1. Materials and Reagents

Polyacrylamide and Coomassie Brilliant Blue were purchased from Sigma-Aldrich (St. Louis, MO, USA). Acetic acid was purchased from Beijing Chemical Reagent Co. (Beijing, China), while dithiothreitol (DTT), iodoacetamide (IAM), and trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (St. Louis, MO, USA). The reagents used were either of chemical or analytical reagent grade. Ammonium hydrogen carbonate (analytical reagent grade) was purchased from Beijing Chemical Reagent Co. (Beijing, China); formic acid was obtained from Sigma-Aldrich (St. Louis, MO, USA); and acetonitrile (HPLC grade) was purchased from Fisher Scientific (Fair Lawn, NJ, USA). The ultra-high-purity water was prepared using a Milli-Q water purification system (Millipore Corporation, Bedford, MA, USA). Trypsin (sequencing grade) was obtained from Pierce (Thermo Scientific, Waltham, MA, USA). The syringe filters (diameter: 0.22 μm) were purchased from Millipore (Billerica, MA, USA). Four wild *O.S.* samples were collected from the Tibetan province in China. The sources were identified by Chief Pharmacist Zhang nan-ping of National Institutes for Food and Drug Control (NIFDC) in China. Five fermented *O. sinensis* mycelia samples were provided by Hangzhou Sino-US Huadong Pharmaceutical Co., Ltd. (Hangzhou, China), and six fermented *Ophiocordyceps* mycelia samples were obtained from Jiangxi Jiminkexin Pharmaceutical Co. Ltd. (Nanchang, China). Eight artificial *Ophiocordyceps* mycelia samples were provided by Hebei Chang Tian Pharmaceutical Co., Ltd. (Shijiazhuang, China), and five *O. mortierella* mycelia samples were obtained from Zhejiang Changxing Pharmaceutical Co. Ltd. (Hangzhou, China) (Table 2). These strains of four fermented *Ophiocordyceps* mycelia were identified by Institute of Microbiology, Chinese Academy of Sciences in China.

Table 2. The *Ophiocordyceps*-related samples' information included in this study.

Sample Status	Claimed Names [a]	No. of Samples	Locations
Wild fruiting body	*Ophiocordyceps sinensis*	4	Tibet
Cultured mycelium powder [b]	Fermented *O.S.* mycelia	5	Zhejiang
	Fermented *Ophiocordyceps* mycelia	6	Jiangxi
	Artificial *Ophiocordyceps* mycelia	8	Hebei
	O. mortierella mycelia	5	Zhejiang

[a] Sample names when they were collected; [b] cultured mycelium powder was collected from the manufacturing enterprise of each sample.

4.2. Extraction of Ophiocordyceps Fungal Proteins

First, deionized water (0.5 mL) was added to a collected sample (10 mg) in a 1.5-mL microfuge tube. The sample was mixed, centrifuged for 5 min at 20,000× *g*, and the supernatant was completely removed. Subsequently, a lysis buffer (30 μL) and silica powder (Φ 0.5 mm) were added to the tube, and its contents were ground repeatedly with a plastic pestle for 2 min using twisting motions. Next, lysis buffer (150 μL) was added, and the sample was ground again for 30 s with the same pestle. The tube was centrifuged for 5 min at 14,000× *g*, the supernatant was collected in a new tube, and another aliquot of a lysis buffer (150 μL) was added to the prime tube. The sample was ground again for 30 s with the same pestle, and the tube was centrifuged for 5 min at 14,000× *g*. The supernatant was collected and transferred to a Millipore 3K ultrafiltration spin column placed in a 2-mL collection tube. The spin column was centrifuged for 25 min at 12,000× *g*, after which the filtrate was discarded and NH_4HCO_3 (300 μL, 0.05 mol/L, pH 8.0) was added. The spin column was centrifuged for 25 min at 12,000× *g*, and the filtrate was discarded again. This step was repeated twice

before the spin column was inverted into a new2-mL collection tube and centrifuged for 5 min at 14,000× *g*. Finally, another aliquot of NH_4HCO_3 (300 μL, 0.05 mol/L, pH 8.0) was added, and the sample was diluted to a concentration of 1 mg/mL [57–59].

4.3. Purification and Tryptic Digestion

Each protein solution (100 μL) was purified by polyacrylamide gel electrophoresis with a 16% polyacrylamide concentrate gel at 100 V for 10 min. The protein gel was stained with Coomassie Brilliant Blue for 2 h and then decolorized with acetic acid for 2 h. The blue bands were cut into small pieces and rinsed three times with water. Subsequently, DTT (10 mM) was added to the gel pieces at 56 °C, and the mixture was incubated for 45 min. After this, the solution was removed, and iodoacetamide (55 mM) was added to the gel. The reaction was left to proceed for 30 min at room temperature in the dark. Next, the NH_4HCO_3 (0.05 M)/ACN (1:1, *v*/*v*) solution (20 mL) was used to decolorize the gel, and then, the decolorizing agent was added twice every 30 min. The gel pieces were dehydrated rapidly with acetonitrile and vacuum-dried for 30 min. A trypsin/0.05 M NH_4HCO_3solution (1:20, *v*/*v*) was employed to digest the proteins in the gel. Each mixture was incubated at 37 °C for 18 h, and then eachsolution was transferred into a new 2-mL tube. The gel pieces were incubated at 37 °C overnight after adding ACN/H_2O (1:1, *v*/*v*, containing 5% TFA). Subsequently, the solution was collected and dried using a vacuum centrifugal concentrator. The residue was dissolved in aqueous formic acid (0.1%) and analyzed by UPLC/Orbitrap Fusion MS/MS [60–62].

4.4. Size Exclusion Chromatography of the Digest Mixture

The molecular weight ranges of the digest mixtures were determined on a TSK G2000SWL column (7.8 mm id × 300 mm length; particle size, 5 μm) (Tosoh Bioscience, Tokyo, Japan) using a Waters 2695–2998 liquid chromatography system (Waters Instruments Co., Rochester, MN, USA). The mobile phase was comprised of a phosphate buffer (0.02 mol/L) containing sodium sulfate (0.1 mol/L). The flow rate was set to 0.5 mL/min. A 10-μL aliquot of the sample was withdrawn from the digest mixture and injected directly onto the column. The UV detection was recorded at 220 nm. The size exclusion chromatography results are displayed in Figure 1.

4.5. Chromatographic Separation and Mass Spectrometry

The liquid chromatography (LC) separation was conducted using a Thermo Scientific™ EASY-nLC™ 1000 HPLC system (Thermo Fisher Scientific Inc., Waltham, MA, USA). The mobile phases were composed of (A)water (with 0.1% of formic acid) and (B)acetonitrile (with 0.1% of formic acid). The peptides were loaded directly onto a homemade C18 column (75 μm id × 15 cm, 3 μm, 120 Å). The analytical mobile phase gradient was 2–6% B from 0–1 min, 6–25% B from 1–46 min, 25–35% B from 46–61 min, 35–80% B from 61–62 min, and finally 80% B for an additional 8 min. The flow rate was set to 300 nL/min for these analytical gradients. The column and autosampler were maintained at temperatures of 40 and 8 °C, respectively. The injection volume was 5 μL [63,64].

All the separated peptide fractions were analyzed using a Thermo Orbitrap Fusion™ (Thermo Fisher Scientific, Waltham, MA, USA) mass spectrometer. The data were acquired at a resolution of 120,000 (@ 200 *m*/*z*) for full MS scans, followed by a high-energy-collision dissociation (HCD) fragmentation and detection of the fragment ions in the ion trap. The MS parameters were set as follows. Full Scan for MS: resolution (@ *m*/*z* 200) 120,000; scan range (*m*/*z*): 350–1550; max injection (ms): 50; and automatic gain control (AGC) target: 2.00×10^5. Data-dependent MS/MS: Fragmentation HCD; NCE (%): 35; detector type: Orbitrap; AGC target: 5.00×10^4; max injection (ms): 35; and dynamic exclusion (s): 60. All acquisitions and data analyses were controlled using the Progenesis QI for proteomics v3.0 (QIP) (Waters, Great Bookham, UK) and Mascot v2.5.1 (Matrix Science, London, UK) software.

LC/MS peptides mass spectrogram fingerprint method was validated under the regulation of Chinese Pharmacopoeia Commission. Seven different ions (RT 4.29 min, *m*/*z* 330.1976; RT 7.47 min,

m/*z* 508.2743; RT 10.52 min, *m*/*z* 615.3333; RT 14.47 min, *m*/*z* 577.2941; RT 20.56 min, *m*/*z* 218.2128; RT 26.49 min, *m*/*z* 802.4413; and RT 31.95 min, *m*/*z* 246.2455) were selected for repeatability, precision, and stability, because the detected ion intensities were generated using the RT and *m*/*z* data pairs in LC/MS peptides mass spectrogram. In six mass spectrograms of the same sample solution, the RSD values of retention time and exact mass of seven ions were less than 1.0%. This suggested that the precision of method was better. In six mass spectrograms of six sample solutions prepared from the same sample, the RSD values of retention time and exact mass of seven ions were also less than 1.0%, and it revealed the repeatability of the method was better. The RSD values of retention time and exact mass of seven ions detected in 0, 2, 4, 6, 8, and 10 h were less than 1.2%, which showed the sample solution was quite stable within 10 h.

4.6. Multivariate Data Analyses

Progenesis QIP was used to analyze the raw data. The following parameters were employed: retention time range: 1.0–61.0 min; detected mass range: 100–2000 Da; mass tolerance: 0.05 Da; noise elimination level: 6.00; intensity threshold: 100 counts; mass window: 0.05 amu; and retention time (RT) window: 0.2 min. No specific mass or adduct was excluded. A list of the detected peak intensities was generated using the RT and *m*/*z* data pairs. Ions in different samples were considered to be identical when they demonstrated identical RT (tolerance of 0.2 min) and *m*/*z* values (tolerance of 0.05 Da). The ion intensities for each detected peak were normalized against the sum of the peak intensities within that sample using Progenesis QIP. The resulting three-dimensional data comprising of the peak number (RT-*m*/*z* pair), sample name, and ion intensity were analyzed via unsupervised PCA. All variables were pareto-scaled prior to analysis. The scheme of the developed method was shown in Figure 5.

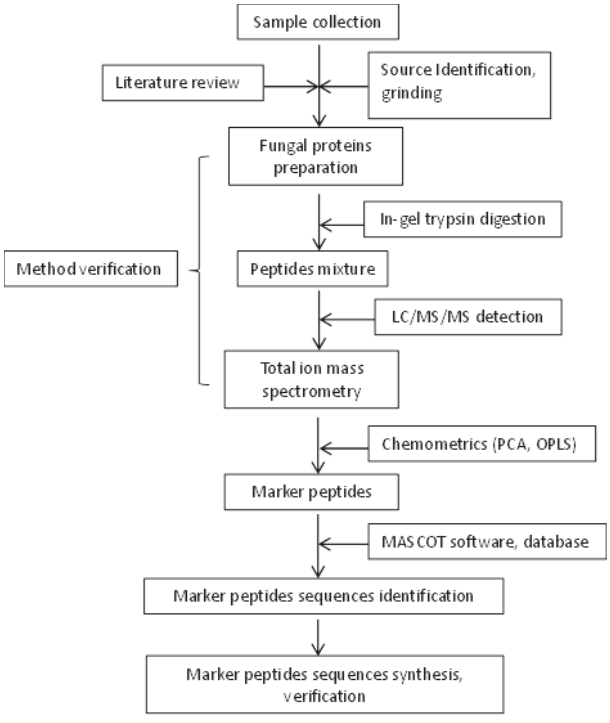


Figure 5. Scheme of the developed method.

5. Conclusions

In this work, we established an efficient method that employs UPLC/Orbitrap-Fusion-MS/MS coupled with chemometrics for the identification of wild *O. sinensis* and four cultured *O.* mycelia products by identifying the marker peptide. This approach allowed for the profiling of the details of each sample so that the different marker peptides could be detected. Hence, the marker peptides could be used as powerful indexes for the identification of these mycelia and to distinguish the different mycelia in mixtures. The present approach provided a foundation for detecting the ion pairs, which came from parent ion and fragment ion of marker peptides using the MRM mode by LC/MS/MS and for developing the sensitive, stable, rapid quality control standard of the valuable Chinese medicine *O. sinensis* and its various cultured mycelia products.

Acknowledgments: This work was partly supported by grants from the project "The Platform of Quality Safety Inspection and Risk Control Technology Research of Traditional Chinese Medicine" (Grant No. 2014ZX09304-307-02) from the Important Program of the Ministry of Science and Technology of the People's Republic of China. Additionally, we sincerely thank engineer Haiwei Xi from Waters Technology Co., Ltd. (Shanghai, China) and Professor Yuanming Luo from the Institute of Microbiology, Chinese Academy of Sciences for their expert technical assistance.

Author Contributions: Ping Zhang carried out the experimental work and wrote the paper. Saina Li and Juan Li acquired and analyzed the data. Feng Wei and Xianlong Cheng collected the test samples, and Guifeng Zhang interpreted the data. Shuangcheng Ma and Bin Liu conceived and designed the work and provided guidance for the implementation of the experimental scheme.

Conflicts of Interest: The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Zhou, L.; Hao, Q.X.; Wang, S.; Kang, C.Z.; Yang, W.Z.; Guo, L.P. Study on distribution of five heavy metal elements in different parts of *Cordyceps sinensis* by microwave digestion ICP-MS. *China J. Chin. Mater. Med.* **2017**, *42*, 2934–2938. [CrossRef]
2. Lee, E.J.; Jang, K.H.; Im, S.Y.; Farooq, M.; Farhoudi, R.; Lee, D.J. Physico-chemical properties and cytotoxic potential of *Cordyceps sinensis* metabolites. *Nat. Prod. Res.* **2015**, *29*, 455–459. [CrossRef] [PubMed]
3. Panda, A.K.; Swain, K.C. Traditional uses and medicinal potential of *Cordyceps sinensis* of Sikkim. *J. Ayurveda Integr. Med.* **2011**, *2*, 9–13. [CrossRef] [PubMed]
4. Bao, Z.D.; Wu, Z.G.; Zheng, F. Amelioration of aminoglycoside nephrotoxicity by *Cordyceps sinensis* in old patients. *Chin. J. Integr. Tradit. West. Med.* **1994**, *14*, 271–273.
5. Bao, T.T.; Wang, G.F.; Yang, J.L. Pharmacological actions of *Cordyceps sinensis*. *Chin. J. Mod. Dev. Tradit. Med.* **1988**, *8*, 325–326, 352–354.
6. Buenz, E.J.; Bauer, B.A.; Osmundson, T.W.; Motley, T.J. The traditional Chinese medicine *Cordyceps sinensis* and its effects on apoptotic homeostasis. *J. Ethnopharmacol.* **2005**, *96*, 19–29. [CrossRef] [PubMed]
7. Buenz, E.J.; Weaver, J.G.; Bauer, B.A.; Chalpin, S.D.; Badley, A.D. *Cordyceps sinensis* extracts do not prevent Fas-receptor and hydrogen peroxide-induced T-cell apoptosis. *J. Ethnopharmacol.* **2004**, *90*, 57–62. [CrossRef] [PubMed]
8. Zhao, X.; Li, L. *Cordyceps sinensis* in protection of the kidney from cyclosporine A nephrotoxicity. *Chin. Med. J.* **1993**, *73*, 410–412, 447.
9. Illana Esteban, C. *Cordyceps sinensis*, a fungi used in the Chinese traditional medicine. *Rev. Iberoam. Micol.* **2007**, *24*, 259–262. [CrossRef]
10. Jahan-Abad, A.J.; Morteza-Zadeh, P.; Negah, S.S.; Gorji, A. Curcumin attenuates harmful effects of arsenic on neural stem/progenitor cells. *Avicenna J. Phytomed.* **2017**, *7*, 376–388.
11. Cao, L.; Ye, Y.; Han, R. Fruiting body production of the medicinal Chinese caterpillar mushroom, *Ophiocordyceps sinensis* (Ascomycetes), in artificial medium. *Int. J. Med. Mushrooms* **2015**, *17*, 1107–1112. [CrossRef] [PubMed]
12. Zhou, X.W.; Li, L.J.; Tian, E.W. Advances in research of the artificial cultivation of *Ophiocordyceps sinensis* in China. *Crit. Rev. Biotechnol.* **2014**, *34*, 233–243. [CrossRef] [PubMed]

13. Yao, Y.S.; Zhu, J.S. Indiscriminate use of Latin name for natural *Cordyceps sinensis* insect-fungi complex and multiple *Ophiocordyceps sinensis* fungi. *China J. Chin. Mater. Med.* **2016**, *41*, 1361–1366. [CrossRef]
14. Zhu, J.S.; Halpern, G.M.; Jones, K. The scientific rediscovery of an ancient Chinese herbal medicine: *Cordyceps sinensis*: Part I. *J. Altern. Complement. Med.* **1998**, *4*, 289–303. [CrossRef] [PubMed]
15. Chen, Y.Q.; Wang, N.; Qu, L.; Li, T.; Zhang, W. Determination of the anamorph of *Cordyceps sinensis* inferred from the analysis of the ribosomal DNA internal transcribed spacers and 5.8S rDNA. *Biochem. Syst. Ecol.* **2001**, *29*, 597–607. [CrossRef]
16. Liu, Z.Q.; Lin, S.; Baker, P.J.; Wu, L.F.; Wang, X.R.; Wu, H.; Xu, F.; Wang, H.Y.; Brathwaite, M.E.; Zheng, Y.G. Transcriptome sequencing and analysis of the entomopathogenic fungus *Hirsutella sinensis* isolated from *Ophiocordycepssinensis*. *BMC Genom.* **2015**, *16*, 106. [CrossRef]
17. Chen, S.J.; Yin, D.H.; Li, L.; Zhou, X.L.; Za, X. Studies on anamorph of *Cordyceps sinensis* (Berk) from Naqu Tibet. *China J. Chin. Mater. Med.* **2001**, *26*, 453–454.
18. Yang, J.L.; Xiao, W.; He, H.X.; Zhu, H.X.; Wang, S.F.; Cheng, K.D.; Zhu, P. Molecular phylogenetic analysis of *Paecilomyces hepiali* and *Cordyceps sinensis*. *Acta Pharm. Sin.* **2008**, *43*, 421–426. [CrossRef] [PubMed]
19. Leung, P.H.; Zhang, Q.X.; Wu, J.Y. Mycelium cultivation, chemical composition and antitumour activity of a *Tolypocladium* sp. fungus isolated from wild *Cordyceps sinensis*. *J. Appl. Microbiol.* **2006**, *101*, 275–283. [CrossRef] [PubMed]
20. Balon, T.W.; Jasman, A.P.; Zhu, J.S. A fermentation product of *Cordyceps sinensis* increases whole-body insulin sensitivity in rats. *J. Altern. Complement. Med.* **2002**, *8*, 315–323. [CrossRef] [PubMed]
21. Zan, K.; Huang, L.L.; Guo, L.N.; Liu, J.; Zheng, J.; Ma, S.C.; Qian, Z.M.; Li, W.J. Comparative study on specific chromatograms and main nucleosides of cultivated and wild *Cordyceps sinensis*. *China J. Chin. Mater. Med.* **2017**, *42*, 3957–3962. [CrossRef]
22. Xie, C.; Xu, N.; Shao, Y.; He, Y. Using FT-NIR spectroscopy technique to determine arginine content in fermented *Cordyceps sinensis* mycelium. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2015**, *149*, 971–977. [CrossRef] [PubMed]
23. Li, S.P.; Yang, F.Q.; Tsim, K.W. Quality control of *Cordyceps sinensis*, a valued traditional Chinese medicine. *J. Pharm. Biomed. Anal.* **2006**, *41*, 1571–1584. [CrossRef] [PubMed]
24. Ko, Y.F.; Liau, J.C.; Lee, C.S.; Chiu, C.Y.; Martel, J.; Lin, C.S.; Tseng, S.F.; Ojcius, D.M.; Lu, C.C.; Lai, H.C.; et al. Isolation, culture and characterization of *Hirsutella sinensis* mycelium from caterpillar fungus fruiting body. *PLoS ONE* **2017**, *12*, e0168734. [CrossRef] [PubMed]
25. Yue, K.; Ye, M.; Lin, X.; Zhou, Z. The artificial cultivation of medicinal caterpillar fungus, *Ophiocordyceps sinensis* (Ascomycetes): A review. *Int. J. Med. Mushrooms* **2013**, *15*, 425–434. [CrossRef] [PubMed]
26. Guo, L.X.; Xu, X.M.; Hong, Y.H.; Li, Y.; Wang, J.H. Stable carbon isotope composition of the lipids in natural *Ophiocordyceps sinensis* from major habitats in China and its substitutes. *Molecules* **2017**, *22*, 1567. [CrossRef] [PubMed]
27. Du, C.; Zhou, J.; Liu, J. Identification of Chinese medicinal fungus *Cordyceps sinensis* by depth-profiling mid-infrared photoacoustic spectroscopy. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2017**, *173*, 489–494. [CrossRef] [PubMed]
28. Au, D.; Wang, L.; Yang, D.; Mok, D.K.; Chan, A.S.; Xu, H. Application of microscopy in authentication of valuable Chinese medicine I—*Cordyceps sinensis*, its counterfeits, and related products. *Microsc. Res. Tech.* **2012**, *75*, 54–64. [CrossRef] [PubMed]
29. Teng, W.Z.; Song, J.; Meng, F.X.; Meng, Q.F.; Lu, J.H.; Hu, S.; Teng, L.R.; Wang, D.; Xie, J. Study on the detection of active ingredient contents of *Paecilomyces hepiali* mycelium via near infrared spectroscopy. *Spectrosc. Spectr. Anal.* **2014**, *34*, 2645–2651.
30. Hou, X.R.; Luan, L.J.; Cheng, Y.Y. Quantitative analysis of the nucleosides in *Cordyceps sinensis* with capillary zone electrophoresis. *China J. Chin. Mater. Med.* **2005**, *30*, 447–449.
31. Huang, L.F.; Guo, F.Q.; Liang, Y.Z.; Chen, B.M. Determination of adenosine and cordycepin in *Cordyceps sinensis* and *C. militarris* with HPLC-ESI-MS. *China J. Chin. Mater. Med.* **2004**, *29*, 762–764.
32. Li, S.P.; Li, P.; Dong, T.T.; Tsim, K.W. Determination of nucleosides in natural *Cordyceps sinensis* and cultured *Cordyceps* mycelia by capillary electrophoresis. *Electrophoresis* **2001**, *22*, 144–150. [CrossRef]
33. Xie, J.W.; Huang, L.F.; Hu, W.; He, Y.B.; Wong, K.P. Analysis of the main nucleosides in *Cordyceps sinensis* by LC/ESI-MS. *Molecules* **2010**, *15*, 305–314. [CrossRef] [PubMed]

34. Guo, F.Q.; Li, A.; Huang, L.F.; Liang, Y.Z.; Chen, B.M. Identification and determination of nucleosides in *Cordyceps sinensis* and its substitutes by high performance liquid chromatography with mass spectrometric detection. *J. Pharm. Biomed. Anal.* **2006**, *40*, 623–630. [CrossRef] [PubMed]
35. Zhang, Y.; Xu, L.; Zhang, S.; Liu, X.; An, Z.; Wang, M.; Guo, Y. Genetic diversity of *Ophiocordyceps sinensis*, a medicinal fungus endemic to the Tibetan Plateau: Implications for its evolution and conservation. *BMC Evol. Biol.* **2009**, *9*, 290. [CrossRef] [PubMed]
36. Lam, K.Y.C.; Chan, G.K.L.; Xin, G.Z.; Xu, H.; Ku, C.F.; Chen, J.P.; Yao, P.; Lin, H.Q.; Dong, T.T.X.; Tsim, K.W.K. Authentication of *Cordyceps sinensis* by DNA analyses: Comparison of ITS sequence analysis and RAPD-derived molecular markers. *Molecules* **2015**, *20*, 22454–22462. [CrossRef] [PubMed]
37. Liang, H.H.; Cheng, Z.; Yang, X.L.; Li, S.; Ding, Z.Q.; Zhou, T.S.; Zhang, W.J.; Chen, J.K. Genetic diversity and structure of *Cordyceps sinensis* populations from extensive geographical regions in China as revealed by inter-simple sequence repeat markers. *J. Microbiol.* **2008**, *46*, 549–556. [CrossRef] [PubMed]
38. Schiavone, N.M.; Sarver, S.A.; Sun, L.; Wojcik, R.; Dovichi, N.J. High speed capillary zone electrophoresis-mass spectrometry via an electrokinetically pumped sheath flow interface for rapid analysis of amino acids and a protein digest. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2015**, *991*, 53–58. [CrossRef] [PubMed]
39. Lee, S.H.; Matsushima, K.; Miyamoto, K.; Oe, T. UV irradiation-induced methionine oxidation in human skin keratins: Mass spectrometry-based non-invasive proteomic analysis. *J. Proteom.* **2016**, *133*, 54–65. [CrossRef] [PubMed]
40. Li, C.H.; Zuo, H.L.; Zhang, Q.; Wang, F.Q.; Hu, Y.J.; Qian, Z.M.; Li, W.J.; Xia, Z.N.; Yang, F.Q. Analysis of soluble proteins in natural *Cordyceps sinensis* from different producing areas by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and two-dimensional electrophoresis. *Pharmacogn. Res.* **2017**, *9*, 34–38. [CrossRef]
41. Hu, S.; Qiu, N.; Liu, Y.; Zhao, H.; Gao, D.; Song, R.; Ma, M. Identification and comparative proteomic study of quail and duck egg white protein using 2-dimensional gel electrophoresis and matrix-assisted laser desorption/ionization time-of-flight tandem mass spectrometry analysis. *Poult. Sci.* **2016**, *95*, 1137–1144. [CrossRef] [PubMed]
42. Zhang, H.; Gan, J.; Shu, Y.Z.; Humphreys, W.G. High-resolution mass spectrometry-based background subtraction for identifying protein modifications in a complex biological system: Detection of acetaminophen-bound microsomal proteins including argininosuccinate synthetase. *Chem. Res. Toxicol.* **2015**, *28*, 775–781. [CrossRef] [PubMed]
43. Legg, K.M.; Powell, R.; Reisdorph, N.; Reisdorph, R.; Danielson, P.B. Verification of protein biomarker specificity for the identification of biological stains by quadrupole time-of-flight mass spectrometry. *Electrophoresis* **2017**, *38*, 833–845. [CrossRef] [PubMed]
44. Jin, L.; Wang, D.; Gooden, D.M.; Ball, C.H.; Fitzgerald, M.C. Targeted mass spectrometry-based approach for protein-ligand binding analyses in complex biological mixtures using a phenacyl bromide modification strategy. *Anal. Chem.* **2016**, *88*, 10987–10993. [CrossRef] [PubMed]
45. Jiang, L.; Liu, Q.; Hu, D.; Xu, H.; Wang, H. Identification of a new protein from silkworm pupas by biological mass spectrometry. In Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Shanghai, China, 1–4 September 2005; Volume 6, pp. 5709–5711.
46. Magni, F.; Van Der Burgt, Y.E.M.; Chinello, C.; Mainini, V.; Gianazza, E.; Squeo, V.; Deelder, A.M.; Kienle, M.G. Biomarkers discovery by peptide and protein profiling in biological fluids based on functionalized magnetic beads purification and mass spectrometry. *Blood Transfus.* **2010**, *8*, s92–s97. [CrossRef] [PubMed]
47. Bai, K.C.; Sheu, F. A novel protein from edible fungi *Cordyceps militaris* that induces apoptosis. *J. Food Drug Anal.* **2018**, *26*, 21–30. [CrossRef] [PubMed]
48. Zhou, X.; Gong, Z.; Su, Y.; Lin, J.; Tang, K. *Cordyceps* fungi: Natural products, pharmacological functions and developmental products. *J. Pharm. Pharmacol.* **2009**, *61*, 279–291. [CrossRef] [PubMed]
49. Feng, K.; Wang, S.; Hu, D.J.; Yang, F.Q.; Wang, H.X.; Li, S.P. Random amplified polymorphic DNA (RAPD) analysis and the nucleosides assessment of fungal strains isolated from natural *Cordyceps sinensis*. *J. Pharm. Biomed. Anal.* **2009**, *50*, 522–526. [CrossRef] [PubMed]
50. Liu, Y.; Wang, J.; Wang, W.; Zhang, H.; Zhang, X.; Han, C. The chemical constituents and pharmacological actions of *Cordyceps sinensis*. *Evid. Based Complement. Altern. Med.* **2015**, *2015*. [CrossRef]

51. Hu, H.; Xiao, L.; Zheng, B.; Wei, X.; Ellis, A.; Liu, Y.M. Identification of chemical markers in *Cordyceps sinensis* by HPLC-MS/MS. *Anal. Bioanal. Chem.* **2015**, *407*, 8059–8066. [CrossRef] [PubMed]
52. Bedlovičová, Z.; Ungvarská, L.M.; Salayová, A.; Harvanová, J.; Očenáš, P. Determination of biologically active compounds in the fungi of the genus *Cordyceps sinensis* by HPLC and NMR. *CeskaSlov. Farm.* **2015**, *64*, 202–205.
53. Zhang, L.; Ge, Y.; Li, J.; Hao, J.; Wang, H.; He, J.; Gao, X.M.; Chang, Y.X. Simultaneous determination of columbianetin-β-d-glucopyranoside and columbianetin in a biological sample by high-performance liquid chromatography with fluorescence detection and identification of other columbianetin-β-d-glucopyranoside metabolites by ultra high-performance liquid chromatography coupled with quadrupole-time of flight mass spectrometry. *J. Pharm. Biomed. Anal.* **2018**, *153*, 221–231. [CrossRef] [PubMed]
54. Woolman, M.; Gribble, A.; Bluemke, E.; Zou, J.; Ventura, M.; Bernards, N.; Wu, M.; Ginsberg, H.; Das, S.; Vitkin, A.; et al. Optimized Mass Spectrometry Analysis Workflow with Polarimetric Guidance for ex vivo and in situ Sampling of Biological Tissues. *Sci. Rep.* **2017**, *7*, 468. [CrossRef] [PubMed]
55. Zhang, S.; Lai, X.; Wu, C.; Wang, S.; Chen, X.; Huang, J.; Yang, G. Application of Differential Proteomic Analysis to Authenticate *Ophiocordyceps sinensis*. *Curr. Microbiol.* **2016**, *72*, 337–343. [CrossRef] [PubMed]
56. Zhang, H.X.; Qian, Z.M.; Yao, S.U.; Liu, X.Z.; Wen, J.L.; Dong, C.H. Comparative analyses of proteomic profile at different development stages of Chinese cordyceps by iTRAQ-coupled 2D LC-MSMS. *Microbiology* **2012**, *39*, 853–864.
57. Ashoub, A.; Berberich, T.; Beckhaus, T.; Bruggemann, W. A competent extraction method of plant proteins for 2-D gel electrophoresis. *Electrophoresis* **2011**, *32*, 2975–2978. [CrossRef] [PubMed]
58. Salem, M.; Bernach, M.; Bajdzeienko, K.; Giavalisco, P. A Simple Fractionated Extraction Method for the Comprehensive Analysis of Metabolites, Lipids, and Proteins from a Single Sample. *J. Vis. Exp.* **2017**, *124*. [CrossRef] [PubMed]
59. Tan, N.J.; Daim, L.D.; Jamil, A.A.; Mohtarrudin, N.; Thilakavathy, K. An effective placental cotyledons proteins extraction method for 2D gel electrophoresis. *Electrophoresis* **2017**, *38*, 633–644. [CrossRef] [PubMed]
60. Huynh, M.L.; Russell, P.; Walsh, B. Tryptic digestion of in-gel proteins for mass spectrometry analysis. *Methods Mol. Biol.* **2009**, *519*, 507–513. [CrossRef] [PubMed]
61. Albright, J.C.; Dassenko, D.J.; Mohamed, E.A.; Beussman, D.J. Identifying gel-separated proteins using in-gel digestion, mass spectrometry, and database searching: Consider the chemistry. *Biochem. Mol. Biol. Educ.* **2009**, *37*, 49–55. [CrossRef] [PubMed]
62. Lill, J.R.; Nesatyy, V.J. Microwave-assisted protein staining, destaining, and in-gel/in-solution digestion of proteins. *Methods Mol. Biol.* **2012**, *869*, 521–532. [CrossRef] [PubMed]
63. Guo, L.; Wang, Q.; Weng, L.; Hauser, L.A.; Strawser, C.J.; Rocha, A.G.; Dancis, A.; Mesaros, C.; Lynch, D.R.; Blair, I.A. Liquid Chromatography-High Resolution Mass Spectrometry Analysis of Platelet Frataxin as a Protein Biomarker for the Rare Disease Friedreich's Ataxia. *Anal Chem.* **2018**, *90*, 2216–2223. [CrossRef] [PubMed]
64. Randall, D.R.; Park, P.S.; Chau, J.K. Identification of altered protein abundances in cholesteatoma matrix via mass spectrometry-based proteomic analysis. *J. Otolaryngol. Head Neck Surg.* **2015**, *44*, 50. [CrossRef] [PubMed]

Sample Availability: Samples of the compounds marker peptides of TLLEAIDSIEPPK, AVLSDAITLVR, FAELLEK, and LESVVTSFTK are available from the authors.

molecules

MDPI

Review

Trends in the Design of New Isobaric Labeling Reagents for Quantitative Proteomics

Remigiusz Bąchor *, Mateusz Waliczek, Piotr Stefanowicz and Zbigniew Szewczuk *

Faculty of Chemistry, University of Wrocław, F. Joliot-Curie 14, 50-383 Wrocław, Poland; mateusz.waliczek@chem.uni.wroc.pl (M.W.); piotr.stefanowicz@chem.uni.wroc.pl (P.S.)

* Correspondence: remigiusz.bachor@chem.uni.wroc.pl (R.B.); zbigniew.szewczuk@chem.uni.wroc.pl (Z.S.); Tel.: +48-71-375-7218 (R.B.); +48-71-375-7212 (Z.S.); Fax: +48-71-328-2348 (R.B.); +48-71-328-2348 (Z.S.)

Academic Editor: Paolo Iadarola

Received: 29 December 2018; Accepted: 13 February 2019; Published: 15 February 2019

Abstract: Modern mass spectrometry is one of the most frequently used methods of quantitative proteomics, enabling determination of the amount of peptides in a sample. Although mass spectrometry is not inherently a quantitative method due to differences in the ionization efficiency of various analytes, the application of isotope-coded labeling allows relative quantification of proteins and proteins. Over the past decade, a new method for derivatization of tryptic peptides using isobaric labels has been proposed. The labels consist of reporter and balanced groups. They have the same molecular weights and chemical properties, but differ in the distribution of stable heavy isotopes. These tags are designed in such a way that during high energy collision induced dissociation (CID) by tandem mass spectrometry, the isobaric tag is fragmented in the specific linker region, yielding reporter ions with different masses. The mass shifts among the reporter groups are compensated by the balancing groups so that the overall mass is the same for all forms of the reagent. Samples of peptides are labeled with the isobaric mass tags in parallel and combined for analysis. Quantification of individual peptides is achieved by comparing the intensity of reporter ions in the tandem mass (MS/MS) spectra. Isobaric markers have found a wide range of potential applications in proteomics. However, the currently available isobaric labeling reagents have some drawbacks, such as high cost of production, insufficient selectivity of the derivatization, and relatively limited enhancement of sensitivity of the analysis. Therefore, efforts have been devoted to the development of new isobaric markers with increased usability. The search for new isobaric markers is focused on developing a more selective method of introducing a tag into a peptide molecule, increasing the multiplexicity of markers, lowering the cost of synthesis, and increasing the sensitivity of measurement by using ionization tags containing quaternary ammonium salts. Here, the trends in the design of new isobaric labeling reagents for quantitative proteomics isobaric derivatization strategies in proteomics are reviewed, with a particular emphasis on isobaric ionization tags. The presented review focused on different types of isobaric reagents used in quantitative proteomics, their chemistry, and advantages offer by their application.

Keywords: isobaric labeling reagents; quantitative proteomics; ESI-MS; ionization enhancers; quaternary ammonium salts

1. Introduction

Mass spectrometry (MS) has become a powerful tool for the analysis of complex protein mixtures. This method combined with various separation techniques offers extremely high resolution and sensitivity; however, obtaining quantitative results is often problematic. Nowadays the problem of quantification is solved in two ways: using label free methods based on spectral count or chromatographic peak area/height, or by methods using stable isotopes. The literature data suggest

that the label-free method is more efficient in terms of number of identified and quantified proteins; however, isotopic methods are characterized by better reproducibility and precision [1]. Another advantage of the methods based on isotopic labeling is multiplexing, allowing parallel analysis of several samples, which saves analysis time and increases the overall reliability of the method.

The incorporation of isotopic label may be performed on various stages of analytical procedure. Currently used methods allow the labeling of intact proteins or the peptides obtained by the enzymatic hydrolysis of the analyzed sample. The labeling of whole proteins is often based on biological systems including cell cultures (SILAC—stable isotope labeling by amino acids in cell culture) [2], or whole organisms (SILAM—stable isotope labeling in mammals [3], HILEP—hydroponic isotope labeling of entire plants [4]). The recombinant proteins or constructs composed of many sequences corresponding to peptides present in the analyzed sample (QconCAT—quantitative concatenation) may be used as isotopically labeled standards [5]. The labeling of whole proteins has many advantages, since standard and sample may be combined before analytical separation, e.g., by electrophoresis, and this procedure is not compromised for the incomplete tryptic hydrolysis.

The most frequently used method uses chemical modification of peptides obtained by the enzymatic hydrolysis of protein sample. The specially designed reagents target amino, sulfhydryl and carboxyl groups of peptides. These labeling reagents incorporate stable heavy isotopes like ^{13}C, ^{15}N, ^{18}O and ^{2}H. The incorporation of isotopic label should result in formation of two isotopologues of peptide, which have the same chemical and chromatographic properties, but differ in molecular masses and can be distinguished in the MS spectrum, allowing direct comparison of the abundance of peptide in labeled and unlabeled sample. It is worth noting that the deuterium isotope is easily available and relatively inexpensive, but may affect a retention time of isotopologues during chromatographic separation of labeled peptides [6].

The isotopic dilution method is efficient in the analysis of samples of moderate complexity. However, the proteomic samples often contain thousands of components. In such situations, the presence of compounds with similar molecular masses and retention times may cause severe quantification errors resulting from the overlapping of the matrix component peaks and peaks of analyte or isotopically labeled standards. Moreover, the system based on the mass differences caused by isotopic substitution is limited to a binary (2-plex), ternary (3-plex) or quaternary (4-plex) set of tags. Therefore, the comparison of samples corresponding to multiple states cannot be achieved in one experiment [7]. The problem of multiplicity in peptide analysis was solved by designing the sets of isobaric reagents with different distribution of heavy isotopes. The products obtained by chemical modification of peptides with these labels are isobaric, with the same chemical and physical properties, and therefore are chromatographically indistinguishable. The relative abundancies of two isotopic variants of peptide is revealed in MS/MS experiment, when fragmentation produces reporter ions with different molecular masses. At the same time, the peptide fragment ions are used for identification of peptide sequences characteristic for the analyzed protein. The structure of isobaric reagents is in general composed of three groups. First of all, they contain a reporter group, a moiety which is easily released from the labeled peptide during collision induced dissociation (CID). The reporter groups may have different molecular masses because of isotopic substitution. This difference allows distinguishing and quantification of peptides from various samples. Another component of the isobaric reagent is a balance group. The balance group is also isotopically substituted assuring that the total mass of the reporter ion and the balance group is the same in all versions of the reagent. The isobaric reagents contain also a reactive group —the moiety responsible for selective reaction of reagent with selected functional groups in a peptide. The most common reactive group is the succinimidyl ester, which reacts with moderate selectivity with amino groups in peptides forming stable amide bond [8]. The active esters react also with phenolic group of tyrosine, imidazole in histidine side chain as well as SH group of cysteine. There are also reports on applying other reactive groups like pentafluorophenyl esters [9], dimetoxytriazyne esters [10], groups selective for thiol moiety, e.g., iodoacetamide [11]

and groups targeting carbonyl moiety in carbonylated proteins, e.g., substituted hydrazines [12] and hydroxylamines [13].

The first example of isobaric tag, *Tandem Mass Tag* (TMT, duplex), was demonstrated by Thomson [14] who has shown that in this strategy, a very high signal-to-noise ratio is achieved compared to MS1 mode. The structure of TMT reagent was a relatively complex and involved deuterium substitution, which may lead to the chromatographic separation of labeled peptides and quantification errors. Another example of isobaric tag, isobaric tags for relative and absolute quantification (iTRAQ), was demonstrated by Ross et al. who tested this reagent for comparison of protein expression in isogenic yeast strains [15]. The isobaric iTRAQ in version developed by Ross was available as a tetraplex. Over the past decade, chemical labeling with isobaric tandem mass tags, such as iTRAQ and TMT reagents, has been employed in a wide range of different clinically orientated serum and plasma proteomics studies [16].

Currently, labeling with isobaric tags is a routine method in quantitative proteomics studies, including comparative proteomics and absolute protein quantification. However, there are some limitations of this group of reagents.

The recent papers demonstrate many examples of the modified isobaric tags. The main directions in development of a new classes of isobaric labeling reagents for quantitative proteomics include:

- Increase in sensitivity of detection of the labeled peptides. This goal may be accomplished by incorporating the permanent electric charge to the labeled molecule, which facilitates electrospray ionization (ESI).
- Increase in the multiplexing capacity of isobaric markers. This modification allows to compare bigger number of proteomic samples in a single experiment.
- Modification of reactive group in isobaric tags in order to increase the selectivity and yield of derivatization.
- Simplification of the structure of isobaric tag in order to reduce prices of this group of reagents.

These trends in a development of isobaric tags will be discussed herein.

The mass spectrometry-based quantitative proteomics can be applied as an absolute or relative method. The first of them allows the quantification of changes in protein expression, whereas the second one is commonly use to establish the protein downregulation in a relation to the control sample. Therefore, in the relative type of analysis, the introduction of a chemically equivalent differential isotopic mass tag is required for comparative quantitation of proteins in different samples. The role of the applied tag consists of the protein or peptide mass change without peptide or protein molecules; this may be achieved by metabolic or in vivo labeling, involving the incorporation of stable isotopes during protein biosynthesis, enzymatic reaction or chemical modification [17]. Among all the developed and applied methods for quantitative proteomic analysis, chemical labeling with isobaric tags is now one of the most popular solutions in proteomics. Isobaric labeling-based quantification has many advantages compared to other stable isotope labeling techniques, one of which is the ability to perform high-throughput quantification due to sample multiplexing. Over the years, several different chemical labeling methods have been proposed for isobaric peptide or protein formation. Usually the modification is performed before protein digestion, which requires complete denaturation to overcome errors in quantitation. One of the oldest and cost-efficient methods of isotopically peptide labeling is H3/D3 acetylation and H5/D5 propionylation. This solution is similar to the as isotope-coded protein label (ICPL) technique discussed below, except that the proteins are labeled with H6- or D6-acetic anhydride or H10- or D10-propionic anhydride [17]. In 2002, Zhang and co-workers [18] applied H5/D5 propionylation as a method for phosphopeptide quantification. They used a gentle chemical labeling of N-termini of all peptides with the H5/D5 propionyl group and measurement of abundance ratio of the isotopically labeled/non-labeled pairs using mass spectrometry. The method, although easy to perform and cost-efficient, requires protein denaturation prior to labeling, whereas

for complex samples, the presence of the overlapping isotopes clusters may require a high-resolution mass instrument like FT-ICR mass spectrometer (Fourier-Transform Ion Cyclotron Resonance).

2. ICPL

In 2000, James and co-workers developed a method based on isotopic labeling of all free amino groups in the protein called isotope-coded protein label (ICPL) [19]. The D4 or H4 forms of nicotinyl-*N*-hydroxysuccinimide (Nic-NHS) were used for *N*-terminal nicotinylation of peptides for their quantitative analysis, resulting in a mass difference of 4 Da per label (Figure 1). Lottspeich and co-workers developed H4 or D4-labeled *N*-nicotinoyloxy-succinimide to label free amino groups in peptides for their relative quantitation [20]. The ICPL reagent is now marketed by Bruker Daltonics (Bremen, Germany). In December 2008, Bruker introduced a 4-plex version of this reagent, the Serva ICPL 4plex Kit. This means that there are actually 4 closely related labels: (1) all natural isotopes, (2) X = D, shift of 4 Da relative to all natural isotopes version, (3) X = H, 6 × ^{13}C, shift of 6 Da relative to natural version, (4) X = D, 6 × ^{13}C, shift of 10 Da relative to all natural isotopes.

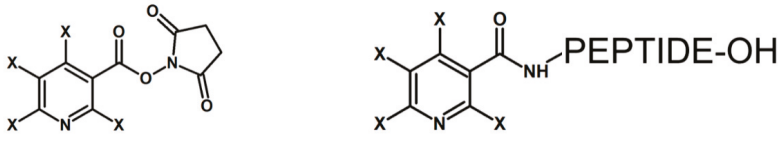


Figure 1. The D4/H4 forms of nicotinyl-*N*-hydroxysuccinimide (Nic-NHS) developed by James et al. [19] and *N*-terminally labeled peptide, which can be applied for quantitation. X-H or D atom.

3. ICAT

Gygi and co-workers introduced in 1999 one of the earliest chemical reagents for quantitative proteomics called Isotope Coded Affinity Tag (ICAT) [21]. The proposed solution of labeling involved iodoacetamide as a thiol-reactive group, biotin, no deuterium atoms in the light form and 8 deuterium atoms in a heavy form of the tag (Figure 2). The biotin-avidin pair was used for selective affinity purification of thiopeptides.

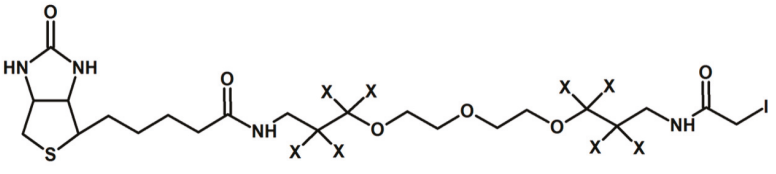


Figure 2. Isotope Coded Affinity Tag proposed by Gygi et al. [21]. X depicts H or D atoms in the molecule.

A new version of the ICAT reagent, where nine ^{13}C atoms were introduced instead of 8 deuteriums and biotin moiety was acid-cleavable, was proposed in 2003 and called cICAT [22,23]. The new form of ICAT reagent eliminates the isotope effect of deuterium during chromatographic separation and removes the potential confusion between a double ICAT label and oxidation, both resulting in a +16 Da mass shift. It was observed that the heavy form is retained stronger than the light form [24]. The use of the ^{13}C stable label removed this retention time shift. The known advantage of ICAT reagent is the reduction in sample complexity, because the label specifically targets cysteines, a relatively rare amino acid making up only 1.42% of all amino acids [25]. However, this simultaneously reduces the reliability of the quantification, as the experiment is based on a limited number of peptides per protein. It also makes it impossible to detect changes in the ~10–13% of proteins that do not contain cysteine residues [26,27]. Additional limitation of ICAT is that there are only two label forms available, which could result in multiple experiments if more than two versions need to be compared, and

would increase the cost accordingly. The need for comparisons of larger numbers of treatments led to the development of the 2- or 4-plex ICPL, the 4- or 8-plex iTRAQ, and the 2- or 6-plex Tandem Mass Tag (TMT) labeling techniques, which can compare up to four, eight or six samples in a single analysis, respectively. Unlike the previously described labeling techniques, which use the parent ion peak heights or peak areas from the MS spectra, the iTRAQ labels are currently the only tagging technologies commercially available where quantitation is carried out in the MS/MS mode.

4. mTRAQ

Mass differential tags for relative and absolute quantification (mTRAQ) are amine-reactive, non-isobaric peptide-labeling reagents introduced by Applied Biosystems, providing the absolute quantification by multiple reaction monitoring (MRM) [28]. The mTRAQ contains a *N*-hydroxysuccinimide ester group enabling primary amine labeling. Specifically, there are two versions of the mTRAQ reagent available: the "heavy" version, which is identical to the iTRAQ 117 tag, and the "light" version that is devoid of any intentional isotopic enrichment. Consequently, labeling of a peptide with the light version adds 140 Da, whereas 144 Da is gained by labeling with the heavy version. This mass difference allows unique MRM transition to be generated for any given peptide labeled with mTRAQ tags. In turn, this implies that, in a sample mixture where peptides originating from different sources are tagged separately with heavy and light labels, the two versions can be monitored independently and distinctly. The mTRAQ comes in three forms, and is not an isobaric tag like iTRAQ [28]. The mTRAQ was designed to be used after the biomarker discovery stage, during the validation stage of a project, as an alternative to having to synthesize deuterated standard peptides for MRM-based quantitation [17]. The precursor-ion mass shifts of these ^{13}C and ^{15}N-labeled reagents are 0, 4 and 8 Da for arginine and 0, 8 and 16 Da for lysine *C*-terminal tryptic peptides. The mTRAQ technique has already been successfully applied by the Siu laboratory to the quantitation of endogenous levels of a potential cancer biomarker in endometrial tissues [29]. Interestingly, in this study, the expression ratios found by mTRAQ were significantly higher than those found in the iTRAQ-based discovery phase. This provides more evidence of the 'compression' of iTRAQ-determined expression ratios.

5. iTRAQ and TMTs

In 2004, Ross and co-workers described the isobaric tags for relative and absolute quantification (iTRAQ) [15], which were commercialized by Applied Biosystems. The iTRAQ reagent contains three regions, namely a mass reporter group, a mass balancer group and a protein reactive group *N*-hydroxysuccinimide ester (NHS) that introduces a highly basic group at lysine side chains and at peptide *N*-termini (Figure 3). The first type of iTRAQ reagent, a 4-plex has the variable mass between 114–117 Da within the reporter group.

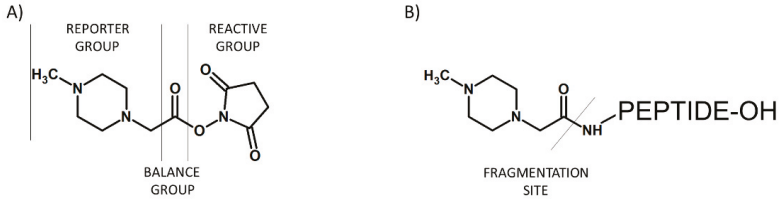


Figure 3. Schematic presentation of iTRAQ structure (**A**) and peptide labeled by iTRAQ (**B**).

The modification of 4-plex iTRAQ based on the introduction of larger balancing group resulted in the formation of 8-plex iTRAQ reagent with masses between 113–121 Da in the reporter group. In a MS scan mode, the iTRAQ labeled peptides are characterized by a single peak due to the isobaric masses. The signal intensity summing from all forms of a given labeled peptide in both MS and MS/MS modes

increases the sensitivity of detection of analyzed compounds. Additionally, the presence of a basic group having higher affinity to the H^+ cation under standard ionization conditions increases ionization efficiency. The MS/MS analysis of iTRAQ labeled peptides causes the release of the reporter group in a form of singly charged ions at m/z 114–117 (for 4-plex) or m/z 113, 114, 115, 116, 117, 118, 119 and 121 (for 8-plex). The m/z 120 iTRAQ 8-plex was omitted to avoid overlapping with the phenylalanine immonium ion (m/z 120.08). The relative quantitation of peptides in the sample is based on the analysis of intensities of these reporter ions in the MS/MS mode. Based on the presence of such reporter ions, the iTRAQ labeling reagent can be applied in the analysis of four different samples using 4-plex kit, or up to eight samples using the 8-plex kit. Both 4-plex and 8-plex are commonly used in the quantitative proteomic analysis of peptides; however, some differences in identification rates have been observed. In 2010, Pichler and co-workers revealed that 8-plex allowed to identify the lowest number of peptides in comparison to 4-plex labeling using iTRAQ or 6-plex TMT [30]. It was concluded that the bulkier structure of the 8-plex label may restrict peptide identification. This results were questioned in 2012 by Pottiez et al., as they achieved more accurate quantitation using 8-plex iTRAQ over 4-plex iTRAQ without sacrificing protein identification rates [31]. However, the observed differences may result from the different instruments and search algorithms used in both experiments [32].

6. TMT

The tandem mass tags (TMTs) are chemical labeling reagents commonly used for mass spectrometry-based quantification and identification of peptides and proteins. The idea of TMTs is based on a similar principle in comparison to iTRAQ, with up to six possible labels. The TMT reagent, proposed by Thompson et al. in 2003, is composed of a mass normalization group (balance group) that balances mass differences from individual reporter ions to ensure the same overall mass of the reagents, a reporter group that provides the abundance of a peptide upon MS/MS in individual samples being mixed, and an amino group reactive functionality making the modification of primary amines possible (Figure 4) [14].

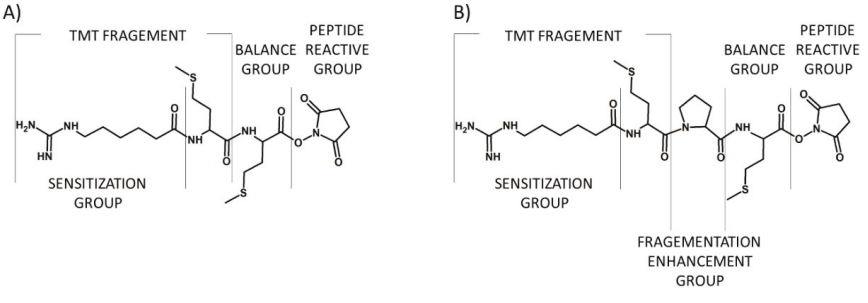


Figure 4. First (**A**) and second (**B**) generation of tandem mass tags developed by Thomson and co-workers [14].

Several TMT reagents are commercially available as TMTzero, TMT duplex, TMT 6-plex, and TMT 10-plex. They have the same chemical structure but contain different numbers and combinations of ^{13}C and ^{15}N isotopes in the reporter group (Figure 5). The chemical structure of the TMT tag enables the introduction of five heavy isotopes (^{13}C or ^{15}N) in the reporter group and five heavy isotopes (^{13}C or ^{15}N) in the balancer group to provide six isobaric tags. Each of the six tags of TMT 6-plex has a specific reporter ion that appears at m/z 126, 127, 128, 129, 130, and 131.

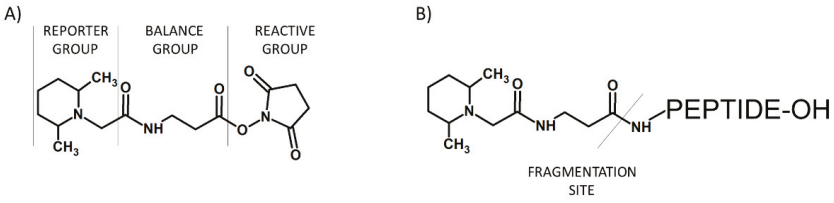


Figure 5. The structure of commercially available TMT (**A**) and the peptide modified by tandem mass tag (**B**) [17].

TMT 10-plex is an expansion of TMT 6-plex generated by combining current TMT 6-plex reagents with four isotope variants of the tag with 6.32 mDa mass differences between ^{15}N and ^{13}C isotopes [32]. Even though the mass difference between these reporter ion isotopologues is incredibly small, current generation high resolution and high mass accuracy analyzers can resolve these ions. The numbers of identified peptides and proteins in shotgun proteomics experiments have been compared for the three commercially available isobaric mass tags: iTRAQ 4-plex, TMT 6-plex, and iTRAQ 8-plex [30,32]. Even though the number of identified proteins and peptides was highest with iTRAQ 4-plex, followed by TMT 6-plex, and lowest with iTRAQ 8-plex, the precision at the level of peptide−spectrum matches and protein level dynamic range was similar.

Recently, to increase the multiplexing capacity of quantitative strategies, isobaric labeling and mass difference labeling have been combined [33]. One of the proposed strategy includes triplex labeling with TMT 6-plex, which achieved 18-plex quantitation and with the addition of medium and heavy sets of 6-plex TMT, 54-plex quantitation was demonstrated. Reagents have been produced to permit 2, 4, 6, 8, 10, and 12-plex comparisons. With the use of isotopologues, where the 6.32 mDa mass difference between positional variations of ^{13}C and ^{15}N substitutions are combined (as are currently used in the 10 and 12-plex reagents), multiplexing up to 18-plex has been suggested as a future possibility [34].

Previously, Sonnet and co-workers devised a system of protein quantification using complement reporter ions (TMTc), which nearly eliminates interference [35]. Additional complement reporter ions are formed as a result of the intact peptide remaining fused to the mass balancing region of the TMT tag. The TMTc ions encode different experimental conditions in the same way the low m/z reporter ions do, with the added benefit that the TMTc ions' mass is different for each peptide. The TMTc quantification does not need the additional gas-phase isolation step of the slow MS3 scan; it holds potential to be significantly more sensitive and is compatible with comparatively simple instruments like iontrap Orbitraps, quadrupole Orbitraps, and Q-TOFs.

However, the commercially available reagents for isobaric peptides labeling (TMT and iTRAQ) have some drawbacks, such as high cost in experiments, especially in quantitation for the modified peptides, and inconvenient handling for variable sizes of samples. Recently, Ren and co-workers [36] developed a set of 10-plex isobaric tags (IBT) with high stability and low cost, which is an improved version of 6-plex DiART. IBT 10-plex reagents were applied for dynamic monitoring of the quantitative responses of stimulated phosphoproteome.

7. Advantages and Disadvantages

The iTRAQ and other isobaric-tag labeling are performed on digested protein; therefore, every peptide formed during proteolysis should be labeled. The detection of multiple peptides originating from one protein may be achieved providing multiple quantitation per protein. This feature makes possible the identification and quantification of low-abundance proteins in complex samples. The ability of analysing of up to eight samples in parallel suggests that iTRAQ and similar mass balanced labeling tags offer the most promise for quantitative biomarker discovery [37]. However, the disadvantages of the described labeling methods include possible variability in labeling efficiencies,

difficulties in protein digestion, and price of the reagents. The protocol of sample preparation involves several steps to achieve reproducible and reliable results. The increasing multiplexicity of labeling reagents also entails a necessity of application of high resolution mass spectrometer to avoid erroneous analysis.

8. DiLeu

The attractive alternative for iTRAQ and TMT based on the *N*,*N*-dimethyl leucine (DiLeu) was proposed in 2010 by Xiang and co-workers (Figure 6) [10]. DiLeu provides relative quantification of peptides and amine-containing metabolites [38]. The amino-reactive group of this reagent targets the *N*-terminus and ε-amino group of the lysine side chain of a peptide. During the MS/MS experiment, this isobaric-labeling reagent produces reporter ions at m/z 115, 116, 117, and 118. Other characteristic features of DiLeu reagent are modest mass of the tag (mass shift of 145.1 Da), higher intensity of reporter ions compared to iTRAQ and TMT, and enhanced collision-induced MS^n fragmentation of labeled peptides at reduced collision energies.

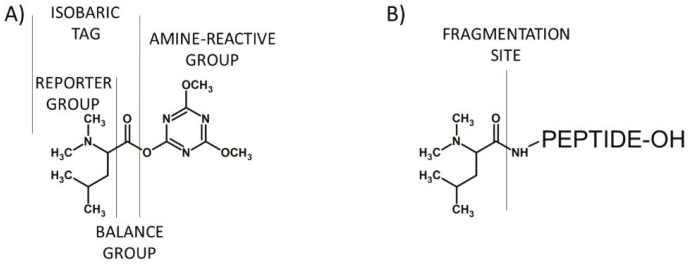


Figure 6. *N*,*N*-dimethyl leucine (DiLeu) tag (**A**) developed by Xiang and co-workers [10] and DiLeu modified peptide (**B**).

In 2015, inventors of the DiLeu isobaric labeling reagent developed the synthesis of a 12-plex set of DiLeu reagents, which was made possible by exploiting subtle mass differences imparted by mass defects between $^{12}C/^{13}C$, $^{14}N/^{15}N$, and $^{1}H/^{2}H$ [39]. This approach enabled a 3-fold increase in multiplexing capacity, keeping the simplicity of synthesis characteristic for the 4-plex set of reagents. However, the small mass difference of ~6 mDa separating the 115, 116, 117, and 118 variants of the 12-plex DiLeu reporters can be baseline-resolved for accurate quantitation at an MS^n resolving power of 30 k (at 400 m/z), which is achievable on the Orbitrap, FTICR, and some QTOF instruments. Acquiring at a resolving power of 60 k baseline resolves isotopic peaks and allows more accurate isotopic interference correction at full multiplexing capacity. Reduced multiplexing configurations allow for highly accurate 9-plex and 7-plex quantitation at resolving powers of 30 k and 15 k, respectively. Additionally, Frost and co-workers [40] presented a cost-effective chemical labeling approach that couples duplex stable isotope dimethyl labeling with 12-plex DiLeu isobaric tags in a combined precursor isotopic labeling and isobaric tagging (cPILOT) strategy. It was found that this approach is compatible with a wide variety of biological samples and permits 24-plex quantification in a single LC-MS/MS experiment.

9. DiART

A good alternative to iTRAQ and TMT for isobaric tagging in quantitative proteomic analysis is Deuterium isobaric Amine Reactive Tag (DiART) [41,42]. DiART reagents consist of an amine-reactive site (NHS ester) for coupling, a balancer, and a reporter (*N*,*N*′-dimethylleucine) with a m/z range of 114–119 (Figure 7). This labeling reagent uses the same labeling protocol as TMT and iTRAQ.

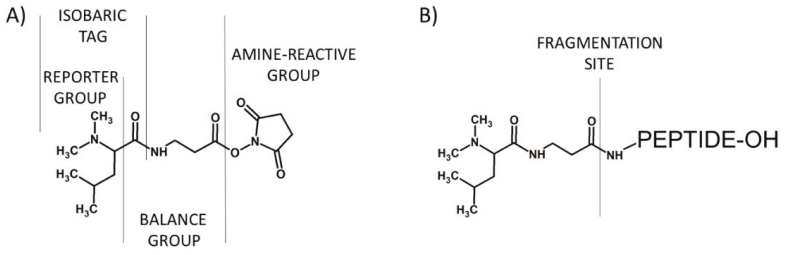


Figure 7. Schematic presentation of Deuterium isobaric Amine Reactive Tag (DiART) [32] (**A**) and DiART modified peptide (**B**).

The application of DiART allows labeling and MS analysis up to six samples. This reagent is characterized by high isotope purity; therefore, the isotopic impurities' correction characteristic for iTRAQ, TMT, or DiLeu labeling is not required during data analysis of DiART-labeled samples. DiART tags allow generation of high-intensity reporter ions compared to those with iTRAQ. In comparison to iTRAQ or TMT reagents, the reporter ions of DiART may easily undergo fragmentation; therefore, a lower collision energy during MS/MS experiment is required.

Protein Post-Translational Modifications

Co- and post-translational modifications pose a particular challenge to proteome analysis; however, they are frequently transient, substoichiometric, or chemically unstable (or all of the three), and may negatively affect analysis by mass spectrometry, for example by interfering with ionization or fragmentation [43]. Isobaric labeling-based quantification was successfully applied in the analysis of post-translational modifications of some protein, including carbonylated residues in protein. The iTRAQ hydrazide (iTRAQH) was proposed by Palmese et al. as a novel reagent providing the selective labeling of carbonyl groups in proteins and their relative quantification (Figure 8) [12]. The reaction between iTRAQH and carbonylated peptide involves the hydrazine moiety. The iTRAQH reporter ions in the low m/z region of the MS/MS spectrum provide relative abundance information on the carbonylated proteins in the analyzed samples.

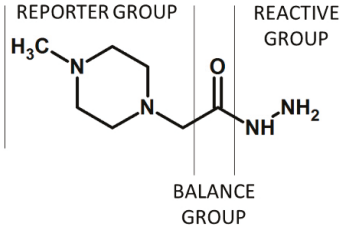


Figure 8. iTRAQ hydrazide (iTRAQH) proposed by Palmese et al. [12].

The proposed solution facilitates the analysis of carbonylated peptides by elimination of some sample preparation steps, including enrichment of modified peptides prior to LC-MS/MS analysis. Additional isotopically-labeled carbonyl-reactive tandem mass tags (glyco-TMTs) have been developed and successfully applied for quantification of *N*-linked glycans [13]. These tags are TMT derivatives, containing hydrazide or aminoxy carbonyl-reactive groups (Figure 9).

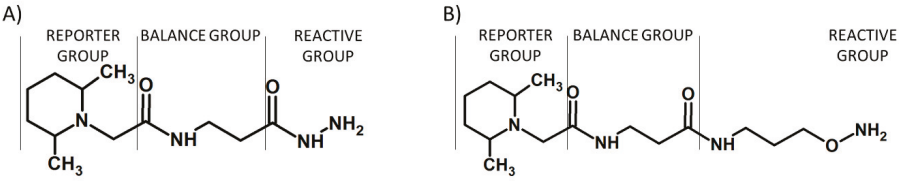


Figure 9. Schematic presentation of the glyco-TMT reagents containing hydrazide (**A**) or aminoxy carbonyl-reactive groups (**B**) [13].

The proposed glyco-TMT tags can be applied in the quantification of a broad range of biologically important molecules containing carbonyl groups, including oxidized proteins, carbohydrates, and steroids. The aminoxy-functionalized glyco-TMTs are commercially available from Thermo Scientific (Rockford, IL, USA) as aminoxyTMT zero and aminoxyTMT sixplex reagents.

Cysteine, a proteinogenic amino acid containing thiol group, plays a crucial role in protein structure and function. Moreover, the cysteine sulfhydryl groups in proteins are potential sites of reversible oxidative modification because of the unique redox chemistry of this amino acid [11]. It was found that 91% of the proteins contain at least one cysteine residue, which is present in more than 24% of the predicted tryptic peptides [44]. That makes the cysteine-containing peptides an attractive target for proteomic analysis.

In 2012, Tambor et al. [45] reported a novel method of cysteine-containing peptide quantification based on amine-reactive iTRAQ labeling with enrichment of thiol-containing peptides in an approach named cysTRAQ. The proposed methodology allows deeper sampling of the cysteinyl proteome in a manner similar to the ICAT reagents [32].

In 2014, Pan et al. developed a selective reagent for isobaric labeling of cysteine-containing peptides based on TMT, containing the iodoacetyl reactive group (Figure 10) [46]. iodoTMT zero and iodoTMT sixplex are commercially available from Thermo Scientific (Rockford, IL, USA). The proposed isobaric iodoTMT 6-plex reagent within a set has the same nominal mass and consists of a thiol reactive iodoacetyl functional group for covalent and irreversible labeling of cysteine, a balancer, and a reporter group. The iodoTMT sixplex has been commonly used to specifically detect and quantify protein S-nitrosylation [47]. The proposed iodoTMT reagent facilitates the procedure of proteomic analysis of cysteine-containing peptides by reducing the necessity of their enrichment.

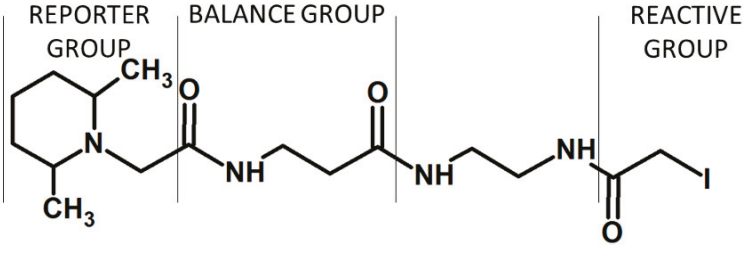


Figure 10. The iodoTMT reagent for cystein-containing peptide labeling and quantification proposed by Pan and co-workers [46].

The isobaric labeling-based proteomic has many advantages in comparison to other stable isotope labeling techniques, one of which is the ability to perform high-throughput quantification due to sample multiplexing. The ability to combine and analyze several samples within one experiment eliminates the need to compare multiple LC-MS/MS data sets, thereby reducing overall analytical time and run-to-run variation. The isobaric mass tags also offer advantages in the MS/MS mode—there is no interference with peptide fragmentation, as the peptide length distribution profile and amino acid content of the isobarically derivatized peptides are similar to those obtained using other MS-based approaches [48].

In fact, isobaric tags have been reported to improve the efficiency of MS/MS fragmentation and result in increased signal intensities of native peptides in samples of human parotid saliva that, in general, lack the uniform architecture of tryptic cleavage products, e.g., a basic C-terminal amino acid residue [49]. Isobaric labeling-based methods in proteomics have been successfully applied in the analysis of endogenous levels of a potential cancer biomarker in endometrial tissues [29], quantification of human plasma samples [39], and other organisms [17,32]. However, the structure of commercially available isobaric-labeling reagents does not provide a high-proton affinity group and partly introduces hydrophobicity, which may lead to higher ionization efficiency of modified peptides. Additionally, the synthesis of ^{13}C, ^{15}N and ^{18}O isotopically labeled reagents is relatively expensive. Access to deuterated compounds could be significantly more efficient and cost-effective by exchange of hydrogen by deuterium in the target molecule than by de novo synthesis. However, the deuterated compounds are not considered as good internal standards due to the possible chromatographic deuterium effect, which may affect the isotopologues' co-elution during LC-MS. The synthesis of isotopically labeled tags could be performed using commercially available precursors; however, it frequently involves long synthetic routes and high costs of starting materials. Therefore, labeling by direct hydrogen/deuterium exchange may be faster and a more cost-effective method for preparation of isotopically labeled reagents. Recently, we observed an unusual hydrogen-deuterium exchange (HDX) at the α-C of acetylated *N*-methylglycine (sarcosine) residue in peptides, which occurs in 1% *N*,*N*,*N*-triethylamine (TEA)/D_2O mixture (pD = 12.3) at room temperature [50]. Additionally, the back exchange (DHX) of introduced deuterons was not observed at neutral pH. The results were confirmed by ESI-MS and NMR data. We also found that such exchange does not occur in the case of *N*-methylalanine residue in model peptides [51]. The method was successfully applied for deuterium labeling of denatonium benzoate (Bitrex) via HDX reaction at the α-carbon situated between carbonyl and the quaternary ammonium group [52]. The proposed strategy is rapid, cost-efficient, and does not require special derivatization reagents or further purification. The LC-MS analysis of denatonium cation isotopologues revealed that the introduced deuterons do not undergo back-exchange under acidic conditions and the co-elution of deuterated and non-deuterated forms was observed. We also developed a method of preparation of deuterium-labeled cyclosporine A standards via H/D exchange of their α-carbon hydrogen atoms occurring in D_2O under basic conditions [53]. We proved that the preparation of deuterated standards of several different compounds may be obtained by simple H/D exchange, where the cost of the reaction depends only on the price of D_2O. However, to provide the multiplexity required in the design of isobaric reagents for quantitative proteomics, the introduction of other isotopes is required, which is discussed below.

10. Sensitivity Problem

Sensitivity of detection is one of the most important issues in mass spectrometry, which in combination with liquid chromatography systems allows to analyze enormously complex mixtures of peptides. Although this technique has been considered as the most versatile and sensitive for analysis of peptides, there are many problems and limitations in analysis of trace amounts of peptides. This issue results from insufficient ionization efficiency of several peptides, which in turn causes the limited detection of these compounds. Moreover, the insufficient intensity of fragmentation peaks may result in ambiguous sequencing, which impedes the successful interpretation of obtained data.

The aforementioned techniques based on the formation of isobaric peptides as well as the multiplexion of samples do not solve the problem of analysis of trace amounts of peptides. Although the multiplexion has been very useful and enables analysis of up to 12 samples simultaneously (12-plex), it may result in decrease of sensitivity. According to literature data and protocols of producers, 25–50 μg of tagged protein has usually been used for the iTRAQ experiment. However, there are some losses during sample cleanup; therefore, using 200–500 μg of proteins is recommended [54,55]. Thus, a great effort is made on improving the sensitivity of the analysis.

One of the common practices to increase sensitivity in mass spectrometry is based on the introduction of permanent charge to the molecule of interest, thereby further ionization (i.e., protonation) is not required for the detection. This approach includes formation of quaternary ammonium salts (QAS) [56–58] or phosphonium salts (QPS) [59]. The commonly used example of derivatization was coupling of a peptide with *N*,*N*,*N*-trimethylammonium acetyl (TMAA, betaine). Although the introduction of fixed charge resulted in observation of N-terminal fragment principally, the QAS modified peptides did not exhibit signal enhancement [60]. The comparison of signal intensity for derivatized and unmodified compounds showed the higher peaks for unmodified peptides. One of the possible reasons is improvement of the hydrophilic character of peptide due to the presence of fixed charge. Thus, there was a need to change the TMAA for more hydrophobic group, which in turn allowed to overcome this issue. It was also described that the *N*,*N*,*N*-trialkilammonium modified peptides undergo Hofmann elimination [61]. This phenomenon has been regarded as problematic, since it complicates the analysis of obtained data. Moreover, the induction of peptide degradation via intramolecular cyclization was observed. Setner and co-workers solved this problem developing QAS-peptide conjugates based on bicyclic 1-azabicyclo[2.2.2]octane (ABCO) and 1,4-diazabicyklo[2.2.2]octane (DABCO) [62,63]. These structures are cyclic, thereby not susceptible to Hofmann elimination. The sequencing of peptides containing the above-mentioned modifications is facilitated, due to the presence of mainly *a* and *b* ion type series. One of the most interesting features of ABCO and DABCO modified peptides is high tolerance for salt concentration, which means that peptides can be detected even in the presence of 10 mM NaCl. This may be of interest during the analysis of trace amounts of peptides, especially if there are concerns about the loss of the sample during desalting on solid support. Bąchor and co-workers studied the limit of detection for DABCO modified peptides and showed the possibility of detection of 15 attomole sample using nano-LC-ESI-MRM [64]. Moreover, these results were used to design a novel quaternary ammonium-based isobaric tag for relative and absolute quantification of peptides (QAS-iTRAQ 2-plex) (Figure 11) [9].

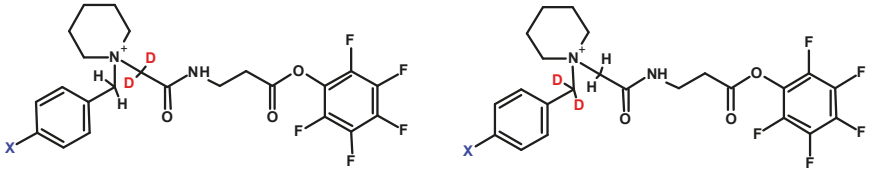


Figure 11. The structures of QAS-iTRAQ 2-plex (X = Br, I) [9].

One form of tag is synthesized using commercially available 4-bromobenzyl[d_2] bromide, the other is a result of facile H/D exchange in the presence of 1% triethylamine, which results in the introduction of two deuterium atoms in the balance group [65]. This new tag liberates stable benzylic-type cationic reporter upon collisional activation. Thus, this approach introduces an ionization tag, which enables isotopic labeling as well as results in increased ionization efficiency. Although this approach is cost-effective and allows sensitive analysis due to the presence of an ionization enhancer, there are still problems related to limited selectivity because of the active ester reactivity.

We recently proposed another class of ionization tags based on 2,4,6-trisubstituted pyridinium scaffold [66]. The reaction relies on a simple conversion of ε-amino groups of lysine to 2,4,6-trisubstituted pyridinium derivative using pyrylium salt; hence, the highly improved sensitivity of peptide detection by mass spectrometry. Pyrylium salts have been known for their reactivity towards sterically unhindered primary amines leading to the formation of pyridinium salt. This chemical property makes them especially selective toward the amine group of the lysine side chain, even in the presence of other nucleophiles such as thiol group. The scheme of reaction is presented in Figure 12.

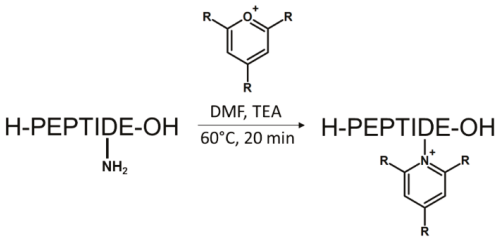


Figure 12. Reaction of peptide with 2,4,6-trisubtituted pyrylium salt (R*-alkyl or aryl).

It is worth noting that derivatization by pyrylium salt does not require the application of an active ester, e.g., succinimidyl group like in the case of other commonly used derivatizing agents. Thus, there is no risk of a side reaction with the sulfhydryl group of cysteine or the imidazole group of histidine. However, either desalting or lyophilization of the tryptically digested sample is required due to the side reaction with ammonium ions originating from ammonium bicarbonate buffer. Therefore, the use of triethylammonium bicarbonate should be considered. The study on detection limits revealed the possibility of detection of 2,4,6-triphenylpyridinium modified peptide at an attomole level in multiple reaction monitoring mode (Figure 13). The MS/MS fragmentation spectra formed upon collisional activation are characterized mainly by *y* type ions series, due to the presence of permanent charge at *C*-terminus. Thus, the sequencing is facilitated as well as becomes more predictable.

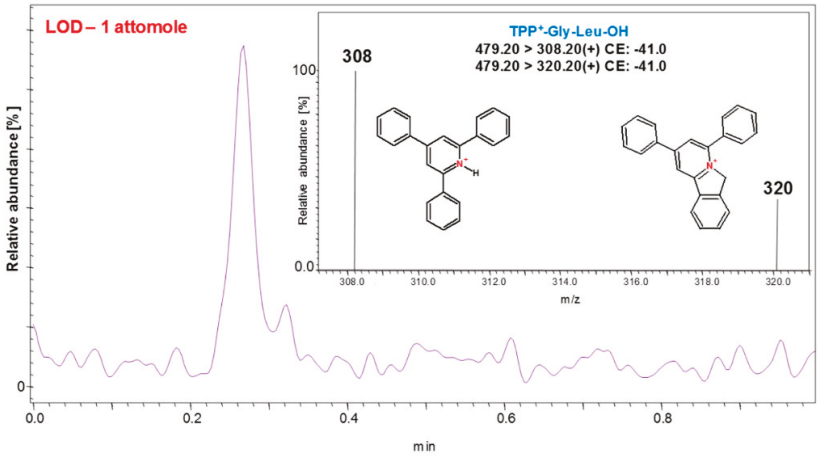


Figure 13. ESI-MS-MRM chromatogram obtained for H-Gly-Leu-OH derivatized with 2,4,6-triphenylpyrylium salt and registered for 1 attomole [66].

The great advantage of pyrylium salts' application has been their commercial availability as well as cost-effectiveness. Furthermore, the compounds can be easily synthesized with a good yield by a one-pot cyclization of two equivalents of benzalacetophenone and one equivalent of benzaldehyde [67–69]. This fact prompted Waliczek and co-workers to prepare isotopically labeled 2,4,6-triphenylpyrylium (tetra-2,3,5,6-$^{13}C_4$) and then apply it to the formation of isobaric peptides (2-plex) [70]. This novel approach (Figure 14) exploited the combination of $^{16}O/^{18}O$ labeling, commonly used in proteomics and formation of a pyridinium-based ionization enhancer [71,72].

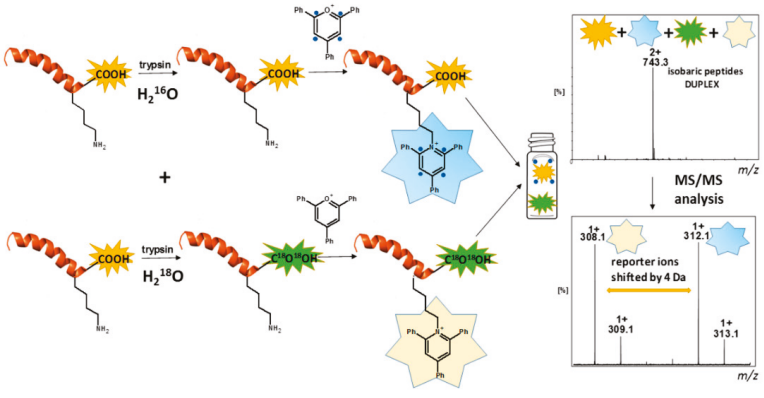


Figure 14. Scheme presenting the formation of a pyridinium-based duplex (the blue circle indicates the ^{13}C labeling) [70].

The enzymatic tryptic digestion and simultaneous $^{18}O_2$-labeling has become a useful strategy in comparative proteomics and allows for quantitative comparison of two samples during a single LC-MS run. This technique has been featured by easy adaptation to clinical samples, cost-effectiveness, and simplicity of performance. Despite many advantages, there are drawbacks, including peaks overlapping and back-exchange [73]. The first one occurs particularly in multiple charged peptides due to decrease of m/z difference between the labeled and unlabeled peptide. This may overcome the problem of partial overlapping of the natural isotope peaks of unlabeled and labeled compounds. The scheme of this approach is presented in Figure 14. The two samples are subjected to enzymatic $^{16}O/^{18}O$ exchange (trypsin) of the carboxyl group oxygen atoms in the C-terminal lysine residue, while the second one remains unlabeled. The mass difference between these two forms is 4 Da. Then, the $^{18}O_2$-labeled first sample is derivatized by 2,4,6-triphenylpyrylium salt, while the unlabeled sample is treated with 2,4,6-triphenylpyrylium salts containing four ^{13}C atoms. Finally, the obtained samples are pooled and analyzed by LC-MRM MS. The previous study revealed the presence of an abundant protonated pyridinium cation in the fragmentation spectrum, thereby playing a role of the balancing group (m/z 308.14 and 312.15) in this method. Thus, a comparison of these two signal allows relative quantitation. It should be noted here that the ^{18}O labeling approach possesses some limitations because of incomplete $^{16}O/^{18}O$ exchange in case of certain tryptic peptides. Therefore, in rare cases, the proposed method may not be efficient enough and their exclusion from further analysis should be considered. Moreover, this method enables the simultaneous analysis of only two samples and at the current stage there is no possibility to expand it to e.g., 4-plex or 8-plex. However, it is worth noting that a combination of $^{16}O/^{18}O$ labeling with derivatization by pyridinium salt solves some problems regarding this approach. First of all, the relative quantitation of peptides relies on comparison of singly charged and is shifted by 4 Da reporter ions, formed as a result of fragmentation of isobaric peptides. Thus, no risk of peak overlapping occurs. Secondly, the pyridinium modified peptides are no longer recognized by trypsin, which in turn prevents the back-exchange. Thirdly, the presence of ionization tag in peptides increases ionization efficiency and allows the sensitive analysis of tryptic digests.

Summarizing the presented data, there is still a need for multiplexing of samples, thereby enabling the simultaneous analysis of many samples. However, due to the frequently occurring decrease in sensitivity with increased multiplexity, there is also a great need for application of ionization tags. Thus, a design of the modern isobaric tags should in our opinion involve the ionization enhancer, thus improving the sensitivity of analysis.

11. Conclusions

The presented manuscript provides an overview of the current role of isobaric markers in MS-based proteomics. In recent years, state-of-the-art applications in this field have expanded enormously, proving its validity and utility. Currently, the application of labeling strategies dominates in the field of quantitative proteomic and post-translational modifications' profiling. Due to its success, chemical isobaric labeling methods play a crucial role in proteomics and, supported by recent advances in mass spectrometric instrumentation and bioinformatics tools, may help solve several problems related to the early diagnosis, prognosis, and monitoring of some diseases. It can be concluded that there is not one perfect and universal method for protein quantification; however, several usable methods have been developed. Additionally, the continuous development of mass spectrometers and new multiplex labeling techniques may significantly improve quantitative proteomic analysis. Although the possibilities of wide use of isobaric tracers in quantitative studies of proteins are beyond doubt, their application in screening medical tests is limited due to the high costs of isobaric reagents and too small increase in the sensitivity of measurements. Therefore, new methods of inexpensive labeling of peptide biomarkers are being intensively sought, which will include not only an isobaric labeling function, but also ionization markers, thus significantly reducing the amount of biological material needed for quantitative analysis. It can be expected that in the future the new isobaric markers will be developed with elevated usable features, including the lowest possible price for production of a set of isobaric reagents; high stability of the reagent allowing its long-term storage at room temperature; high reactivity and specificity in reacting with a specific chemical group of biomarkers; ability to smoothly release certain reporter ions from the labeled biomarker during CID, which will provide reliable quantitative results; high multiplexicity of markers, which will ensure the possibility of employing of a large variety and complex distribution of heavy atoms in the reporter and balancing groups; and high sensitivity of the analysis, by using stable ionization markers in the ionizing marker reporter group, such as cyclic quaternary ammonium salts.

Author Contributions: Writing—original draft preparation, R.B. and M.W., P.S., Z.S.; writing—review and editing, R.B.; conceptualization and project administration, Z.S. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding: This work was supported by Grant No. UMO-2016/23/B/ST4/01036 from the National Science Centre, Poland.

Conflicts of Interest: The authors have declared that no competing interests exist.

References

1. Collier, T.S.; Sarkar, P.; Franck, W.L.; Rao, B.M.; Dean, R.A.; Muddiman, D.C. Direct comparison of stable isotope labeling by amino acids in cell culture and spectral counting for quantitative proteomics. *Anal. Chem.* **2010**, *82*, 8696–8702. [CrossRef] [PubMed]
2. Ong, S.E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D.B.; Steen, H.; Pandey, A.; Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002**, *1*, 376–386. [CrossRef] [PubMed]
3. Wu, C.C.; MacCoss, M.J.; Howell, K.E.; Matthews, D.E.; Yates, J.R., III. Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal. Chem.* **2004**, *76*, 4951–4959. [CrossRef] [PubMed]
4. Bindschedler, L.V.; Palmblad, M.; Cramer, R. Hydroponic isotope labelling of entire plants (HILEP) for quantitative plant proteomics; an oxidative stress case study. *Phytochemistry* **2008**, *69*, 1962–1972. [CrossRef] [PubMed]
5. Pratt, J.M.; Simpson, D.M.; Doherty, M.K.; Rivers, J.; Gaskell, S.J.; Beynon, R.J. Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat. Protoc.* **2006**, *1*, 1029–1043. [CrossRef]
6. Zhang, R.; Regnier, F.E. Minimizing resolution of isotopically coded peptides in comparative proteomics. *J. Proteome Res.* **2002**, *1*, 139–147. [CrossRef]

7. Chahrour, O.; Cobice, D.; Malone, J. Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. *J. Pharm. Biomed. Anal.* **2015**, *10*, 2–20. [CrossRef]
8. Wiktorowicz, J.E.; English, R.D.; Wu, Z.; Kurosky, A. Model studies on itraq modification of peptides: Sequence-dependent reaction specificity. *J. Proteome Res.* **2012**, *11*, 1512–1520. [CrossRef]
9. Setner, B.; Stefanowicz, P.; Szewczuk, Z. Quaternary ammonium isobaric tag for a relative and absolute quantification of peptides. *J. Mass Spectrom.* **2018**, *53*, 115–123. [CrossRef]
10. Xiang, F.; Ye, H.; Chen, R.; Fu, Q.; Li, L. *N*,*N*-Dimethyl leucines as novel isobaric tandem mass tags for quantitative proteomics and peptidomics. *Anal. Chem.* **2010**, *82*, 2817–2825. [CrossRef]
11. Thomas, J.A.; Mallis, R.J. Aging and oxidation of reactive protein sulfhydryls. *Exp. Gerontol.* **2001**, *36*, 1519–1526. [CrossRef]
12. Palmese, A.; De Rosa, C.; Chiappetta, G.; Marino, G.; Amoresano, A. Novel method to investigate protein carbonylation by iTRAQ strategy. *Anal. Bioanal. Chem.* **2012**, *404*, 1631–1635. [CrossRef] [PubMed]
13. Hahne, H.; Neubert, P.; Kuhn, K.; Etienne, C.; Bomgarden, R.; Rogers, J.C.; Kuster, B. Carbonyl-reactive tandem mass tags for the proteome-wide quantification of *N*-linked glycans. *Anal. Chem.* **2012**, *84*, 3716–3724. [CrossRef] [PubMed]
14. Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A.K.; Hamon, C. Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **2003**, *75*, 1895–1904. [CrossRef] [PubMed]
15. Ross, P.L.; Huang, Y.N.; Marchese, J.N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **2004**, *3*, 1154–1169. [CrossRef] [PubMed]
16. Moulder, R.; Bhosale, S.D.; Goodlett, D.R.; Lahesmaa, R. Analysis of the plasma proteome using iTRAQ and TMT-based Isobaric labeling. *Mass Spectrom. Rev.* **2018**, *37*, 583–606. [CrossRef] [PubMed]
17. Elliott, M.H.; Smith, D.S.; Parker, C.E.; Borchers, C. Current trends in quantitative proteomics. *J. Mass Spectrom.* **2009**, *44*, 1637–1660. [CrossRef]
18. Zhang, X.; Jin, Q.K.; Carr, S.A.; Annan, R.S. N-Terminal peptide labeling strategy for incorporation of isotopic tags: A method for the determination of site-specific absolute phosphorylation stoichiometry. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 2325–2332. [CrossRef]
19. Munchbach, M.; Quadroni, M.; Miotto, G.; James, P. Quantitation and facilitated de novo sequencing of proteins by isotopic *N*-terminal labeling of peptides with a fragmentation-directing moiety. *Anal. Chem.* **2000**, *72*, 4047–4057. [CrossRef]
20. Schmidt, A.; Kellermann, J.; Lottspeich, F. A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **2005**, *5*, 4–15. [CrossRef]
21. Gygi, S.P.; Rist, B.; Gerber, S.A.; Turecek, F.; Gelb, M.H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999**, *17*, 994–999. [CrossRef]
22. Li, J.; Steen, H.; Gygi, S.P. Protein profiling with cleavable isotopecoded affinity tag (cICAT) reagents: The yeast salinity stress response. *Mol. Cell. Proteomics* **2003**, *2*, 1198–1204. [CrossRef] [PubMed]
23. Hansen, K.C.; Schmitt-Ulms, G.; Chalkley, R.J.; Hirsch, J.; Baldwin, M.A.; Burlingame, A.L. Mass spectrometric analysis of protein mixtures at low levels using cleavable 13C-isotopecoded affinity tag and multidimensional chromatography. *Mol. Cell. Proteomics* **2003**, *2*, 299–314. [CrossRef]
24. Zhang, R.; Sioma, C.S.; Thompson, R.A.; Xiong, L.; Regnier, F.E. Controlling deuterium isotope effects in comparative proteomics. *Anal. Chem.* **2002**, *74*, 3662–3669. [CrossRef] [PubMed]
25. Gevaert, K.; Impens, F.; Ghesquiere, B.; Van Damme, P.; Lambrechts, A.; Vandekerckhove, J. Stable isotopic labeling in proteomics. *Proteomics* **2008**, *8*, 4873–4885. [CrossRef] [PubMed]
26. Leitner, A.; Lindner, W. Current chemical tagging strategies for proteome analysis by mass spectrometry. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2004**, *25*, 1–26. [CrossRef] [PubMed]
27. Vaughn, C.P.; Crockett, D.K.; Lim, M.S.; Elenitoba-Johnson, K.S.J. Analytical characteristics of cleavable isotope-coded affinity tag- LC-tandem mass spectrometry for quantitative proteomic studies. *J. Mol. Diagn.* **2006**, *8*, 513–520. [CrossRef]
28. Kang, U.-B.; Yeom, J.; Kim, H.; Lee, C. Quantitative analysis of mTRAQ-Labeled proteome using full MS scans. *J. Proteome Res.* **2010**, *9*, 3750–3758. [CrossRef] [PubMed]

29. De Souza, L.V.; Taylor, A.M.; Li, W.; Minkoff, M.S.; Romaschin, A.D.; Colgan, T.J.; Siu, K.W. Multiple reaction monitoring of mTRAQ-labeled peptides enables absolute quantification of endogenous levels of a potential cancer marker in cancerous and normal endometrial tissues. *J. Proteome Res.* **2008**, *7*, 3525–3534. [CrossRef]
30. Pichler, P.; Kocher, T.; Holzmann, J.; Mazanek, M.; Taus, T.; Ammerer, G.; Mechtler, K. Peptide labeling with isobaric tags yields higher identification rates using iTRAQ 4-plex compared toTMT 6-plex and iTRAQ 8-plex on LTQ Orbitrap. *Anal. Chem.* **2010**, *82*, 6549–6558. [CrossRef]
31. Pottiez, G.; Wiederin, J.; Fox, H.S.; Ciborowski, P. Comparison of 4-plex to 8-plex iTRAQ Quantitative Measurements of Proteins in Human Plasma Samples. *J. Proteome Res.* **2012**, *11*, 3774–3781. [CrossRef] [PubMed]
32. Rauniyar, N.; Yates, J.R., III. Isobaric labeling-based relative quantification in shotgun proteomics. *J. Proteomics* **2012**, *75*, 857–867. [CrossRef] [PubMed]
33. Everley, R.A.; Kunz, R.C.; McAllister, F.E.; Gygi, S.P. Increasing throughput in targeted proteomics assays: 54-plex quantitation in a single mass spectrometry run. *Anal. Chem.* **2013**, *85*, 5340–5346. [CrossRef] [PubMed]
34. McAlister, G.C.; Huttlin, E.L.; Haas, W.; Ting, L.; Jedrychowski, M.P.; Rogers, J.C.; Kuhn, K.; Pike, I.; Grothe, R.A.; Blethrow, J.D.; et al. Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Anal. Chem.* **2012**, *84*, 7469–7478. [CrossRef] [PubMed]
35. Sonnett, M.; Yeung, E.; Wühr, M. Accurate, sensitive, and precise multiplexed proteomics using the complement reporter ion cluster. *Anal. Chem.* **2018**, *90*, 5032–5039. [CrossRef] [PubMed]
36. Ren, Y.; He, Y.; Lin, Z.; Zi, J.; Yang, H.; Zhang, S.; Lou, X.; Wang, Q.; Li, S.; Liu, S. Reagents for Isobaric Labeling Peptides in Quantitative Proteomics. *Anal. Chem.* **2018**, *90*, 12366–12371. [CrossRef] [PubMed]
37. Simpson, K.L.; Whetton, A.D.; Dive, C. Quantitative mass spectrometry-based techniques for clinical use: Biomarker identification and quantification. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2009**, *877*, 1240–1249. [CrossRef]
38. Arul, A.B.; Robinson, R.A.S. Sample Multiplexing Strategies in Quantitative Proteomics. *Anal. Chem.* **2019**, *91*, 178–189. [CrossRef]
39. Frost, D.C.; Greer, T.; Li, L. High-resolution enabled 12-plex DiLeu isobaric tags for quantitative proteomics. *Anal. Chem.* **2015**, *87*, 1646–1654. [CrossRef]
40. Frost, D.C.; Rust, C.J.; Robinson, R.A.S.; Li, L. Increased *N*,*N*-Dimethyl Leucine Isobaric Tag Multiplexing by a Combined Precursor Isotopic Labeling and Isobaric Tagging Approach. *Anal. Chem.* **2018**, *90*, 10664–10669. [CrossRef]
41. Chen, Z.; Wang, Q.; Lin, L.; Tang, Q.; Edwards, J.L.; Li, S.; Liu, S. Comparative evaluation of two isobaric labeling tags, DiART and iTRAQ. *Anal. Chem.* **2012**, *84*, 2908–2915. [CrossRef]
42. Zhang, J.; Wang, Y.; Li, S. Deuterium isobaric amine-reactive tags for quantitative proteomics. *Anal. Chem.* **2010**, *82*, 7588–7595. [CrossRef]
43. Leitner, A. A review of the role of chemical modification methods in contemporary mass spectrometry-based proteomics research. *Anal. Chim. Acta* **2018**, *1000*, 2–19. [CrossRef]
44. Lin, D.; Li, J.; Slebos, R.J.; Liebler, D.C. Cysteinyl peptide capture for shotgun proteomics: Global assessment of chemoselective fractionation. *J. Proteome Res.* **2010**, *9*, 5461–5472. [CrossRef]
45. Tambor, V.; Hunter, C.L.; Seymour, S.L.; Kacerovsky, M.; Stulik, J.; Lenco, J. CysTRAQ—A combination of iTRAQ and enrichment of cysteinyl peptides for uncovering and quantifying hidden proteomes. *J. Proteomics* **2012**, *75*, 857–867. [CrossRef]
46. Pan, K.T.; Chen, Y.Y.; Pu, T.H.; Chao, Y.S.; Yang, C.Y.; Bomgarden, R.D.; Rogers, J.C.; Meng, T.C.; Khoo, K.H. Mass spectrometry-based quantitative proteomics for dissecting multiplexed redox cysteine modifications in nitric oxide-protected cardiomyocyte under hypoxia. *Antioxid. Redox Signal.* **2014**, *20*, 1365–1381. [CrossRef]
47. Qu, Z.; Meng, F.; Bomgarden, R.D.; Viner, R.I.; Li, J.; Rogers, J.C.; Cheng, J.; Greenlief, C.M.; Cui, J.; Lubahn, D.B.; et al. Proteomic quantification and site-mapping of S-nitrosylated proteins using isobaric iodoTMT reagents. *J. Proteome Res.* **2014**, *13*, 3200–3211. [CrossRef] [PubMed]
48. Aggarwal, K.; Choe, L.H.; Lee, K.H. Shotgun proteomics using the iTRAQ isobaric tags. *Brief. Funct. Genomics Proteomics* **2006**, *5*, 112–120. [CrossRef]
49. Hardt, M.; Witkowska, H.E.; Webb, S.; Thomas, L.R.; Dixon, S.E.; Hall, S.C.; Fisher, S.J. Assessing the effects of diurnal variation on the composition of human parotid saliva: Quantitative analysis of native peptides using iTRAQ reagents. *Anal. Chem.* **2005**, *77*, 4947–4954. [CrossRef] [PubMed]

50. Bąchor, R.; Setner, B.; Kluczyk, A.; Stefanowicz, P.; Szewczuk, Z. The unusual hydrogen-deuterium exchange of α-carbon protons in N-substituted glycine containing peptides. *J. Mass Spectrom.* **2014**, *49*, 43–49. [CrossRef] [PubMed]
51. Bąchor, R.; Rudowska, M.; Kluczyk, A.; Stefanowicz, P.; Szewczuk, Z. Hydrogen-deuterium exchange of α-carbon protons and fragmentation pathways in N-methylated glycine and alanine-containing peptides derivatized by quaternary ammonium salts. *J. Mass Spectrom.* **2014**, *49*, 529–536. [CrossRef] [PubMed]
52. Bąchor, R.; Kluczyk, A.; Stefanowicz, P.; Szewczuk, Z. Facile synthesis of deuterium-labeled denatonium cation and its application in the quantitative analysis of Bitrex by liquid chromatography-mass spectrometry. *Anal. Bioanal. Chem.* **2015**, *407*, 6557–6561. [CrossRef] [PubMed]
53. Bąchor, R.; Kluczyk, A.; Stefanowicz, P.; Szewczuk, Z. Preparation of novel deuterated cyclosporin A standards for quantitative LC-MS analysis. *J. Mass Spectrom.* **2017**, *52*, 817–822. [CrossRef] [PubMed]
54. Vetter, D.E.; Basappa, J.; Turcan, S. Multiplexed isobaric tagging protocols for quantitative mass spectrometry approaches to auditory research. *Methods Mol. Biol.* **2009**, *493*, 345–366. [PubMed]
55. Protein Quantitation (iTRAQ). Available online: https://www.proteomics.com.au/analytical-services/itraq/ (accessed on 13 December 2018).
56. Stults, J.T.; Lai, J.; McCune, S.; Wetzel, R. Simplification of high-energy collision spectra of peptides by amino-terminal derivatization. *Anal. Chem.* **1993**, *65*, 1703–1708. [CrossRef] [PubMed]
57. Vath, J.E.; Biemann, K. Microderivatization of peptides by placing a fixed positive charge at the *N*-terminus to modify high-energy collision fragmentation. *Int. J. Mass Spectrom. Ion Processes.* **1990**, *100*, 287–299. [CrossRef]
58. Mirzaei, H.; Regnier, F. Enhancing electrospray ionization efficiency of peptides by derivatization. *Anal. Chem.* **2006**, *78*, 4175–4183. [CrossRef]
59. Liao, P.C.; Huang, Z.H.; Allison, J. Charge remote fragmentation of peptides following attachment of a fixed positive charge. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 501–509. [CrossRef]
60. Blagojevic, V.; Zhidkov, N.; Tharmaratnam, S.; Pham, V.T.; Kaplan, H.; Bohme, D.K. Peptide quantitation with methyl iodide isotopic tags and mass spectrometry. *Analyst* **2010**, *135*, 1456–1460. [CrossRef]
61. Cydzik, M.; Rudowska, M.; Stefanowicz, P.; Szewczuk, Z. The competition of charge remote and charge directed fragmentation mechanisms in quaternary ammonium salt derivatized peptides—An isotopic exchange study. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 2103–2107. [CrossRef]
62. Setner, B.; Rudowska, M.; Kluczyk, A.; Stefanowicz, P.; Szewczuk, Z. The 5-azoniaspiro[4.4]nonyl group for improved MS peptide analysis: A novel non-fragmenting ionization tag for mass spectrometric sensitive sequencing of peptides. *Anal. Chim. Acta* **2017**, *986*, 71–81. [CrossRef]
63. Cydzik, M.; Rudowska, M.; Stefanowicz, P.; Szewczuk, Z. Derivatization of peptides as quaternary ammonium salts for sensitive detection by ESI–MS. *J. Pept. Sci.* **2011**, *17*, 445–453. [CrossRef]
64. Bąchor, R.; Mielczarek, P.; Rudowska, M.; Silberring, J.; Szewczul, Z. Sensitive detection of charge peptides at the attomole level using nano-LC-ESI-MRM analysis. *Int. J. Mass Spectrom.* **2014**, *362*, 32–38.
65. Rudowska, M.; Wojewska, D.; Kluczyk, A.; Bąchor, R.; Stefanowicz, P.; Szewczuk, Z. The hydrogen-deuterium exchange at α-carbon atom in *N*,*N*,*N*-trialkylglycine residue: ESI-MS studies. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1024–1028. [CrossRef] [PubMed]
66. Waliczek, M.; Kijewska, M.; Rudowska, M.; Setner, B.; Stefanowicz, P.; Szewczuk, Z. Peptides labeled with pyridinium salts for sensitive detection and sequencing by electrospray tandem mass spectrometry. *Sci. Rep.* **2016**, *6*. [CrossRef]
67. Balaban, A.T.; Boulton, A.J. 2,4,6-Trimethyl-pyrylium tetrafluoroborate. *Org. Synth.* **1973**, *5*, 1112–1113.
68. Balaban, A.T.; Boulton, A.J. 2,4,6-Trimethyl-pyrylium trifluoromethanesulfonate. *Org. Synth.* **1973**, *5*, 1114–1116.
69. Dimroth, K.; Reichardt, C.; Vogel, K. 2,4,6-triphenylpyrylium tetrafluoroborate. *Org. Synth.* **1973**, *5*, 1135.
70. Waliczek, M.; Bąchor, R.; Kijewska, M.; Gąszczyk, D.; Panek-Laszczyńska, K.; Dąbrowska, K.; Witkiewicz, W.; Marek-Bukowiec, K.; Tracz, J.; et al. Isobaric duplex based on a combination of $^{16}O/^{18}O$ enzymatic exchange and labeling with pyrylium salts. *Anal. Chim. Acta* **2019**, *1048*, 96–104. [CrossRef]
71. Yao, X.; Afonso, C.; Fenselau, C. Dissection of proteolytic ^{18}O labeling: Endoprotease-catalyzed ^{16}O-to-^{18}O exchange of truncated peptide substrates. *J. Proteome Res.* **2003**, *2*, 147–152. [CrossRef]
72. Rao, K.C.; Carruth, R.T.; Miyagi, M. Proteolytic ^{18}O labeling by peptidyl-Lys metalloendopeptidase for comparative proteomics. *J. Proteome Res.* **2005**, *4*, 507–514. [CrossRef] [PubMed]

73. Mori, M.; Abe, K.; Yamaguchi, H.; Goto, J.; Shimada, M.; Mano, N. Production of ^{18}O-single labeled peptide fragments during trypsin digestion of proteins for quantitaive proteomcis using nanoLC-ESI-MS/MS. *J. Proteome Res.* **2010**, *9*, 3741–3749. [CrossRef] [PubMed]

molecules

MDPI

Review

Elucidation of the Mechanism of Action for Metal Based Anticancer Drugs by Mass Spectrometry-Based Quantitative Proteomics

Shuailong Jia [1], Runjing Wang [1], Kui Wu [2], Hongliang Jiang [1] and Zhifeng Du [1,*]

[1] Tongji School of Pharmacy, Huazhong University of Science and Technology, Wuhan 430030, China; jsl1993521@163.com (S.J.); wangrjcn@foxmail.com (R.W.); jianghongliang@hust.edu.cn (H.J.)
[2] School of Chemistry and Chemical Engineering, Wuhan University of Science and Technology, Wuhan 430081, China; wukui@wust.edu.cn
* Correspondence: duzhifeng@hust.edu.cn

Academic Editor: Paolo Iadarola
Received: 14 January 2019; Accepted: 2 February 2019; Published: 6 February 2019

Abstract: The discovery of the anticancer activity of cisplatin and its clinical application has opened a new field for studying metal-coordinated anticancer drugs. Metal-based anticancer drugs, such as cisplatin, can be transported to cells after entering into the human body and form metal–DNA or metal–protein adducts. Then, responding proteins will recognize adducts and form stable complexes. The proteins that were binding with metal-based anticancer drugs were relevant to their mechanism of action. Herein, investigation of the recognition between metal-based anticancer drugs and its binding partners will further our understanding about the pharmacology of cytotoxic anticancer drugs and help optimize the structure of anticancer drugs. The "soft" ionization mass spectrometric methods have many advantages such as high sensitivity and low sample consumption, which are suitable for the analyses of complex biological samples. Thus, MS has become a powerful tool for the identification of proteins binding or responding to metal-based anticancer drugs. In this review, we focused on the mass spectrometry-based quantitative strategy for the identification of proteins specifically responding or binding to metal-based anticancer drugs, ultimately elucidating their mechanism of action.

Keywords: anticancer drugs; metal complexes; mechanism; mass spectrometry; quantitative proteomics

1. Introduction to Metal-Based Anticancer Drugs

As early as the 1960s, Rosenerg first discovered that platinum complexes have an inhibitory effect on tumor cell growth, and used platinum complexes to treat tumors [1]. In the past 30 years, cisplatin, carboplatin, nedaplatin, oxaliplatin, and lobaplatin (Figure 1) have been successfully developed and used for the clinical treatment of cancer. In particular, cisplatin, oxaliplatin, and carboplatin are used worldwide as anticancer drugs. To date, all of the clinically used platinum drugs contain a single Pt^{II} center with two exchangeable ligands in cis geometry. The interaction of these drugs with cellular biomolecules such as sulfur-containing glutathione and metallothionein can deactivate them before reaching their pharmacological target, DNA [2,3]. After platinum drugs enter into the body, the intracellular Cl^- concentration is low, and the drug is easily hydrolyzed to form an active molecule. The active hydrolysate has a positive charge, and is electrostatically attracted by DNA, which is a negatively charged genetic material located in the nucleus [4]. Binding to DNA forms DNA intra-strand cross-linking, inter-strand cross-linking, and DNA protein cross-linking. The formation of DNA cross-linking affects DNA strand synthesis, replication, and ultimately leads to cell death [5–9].

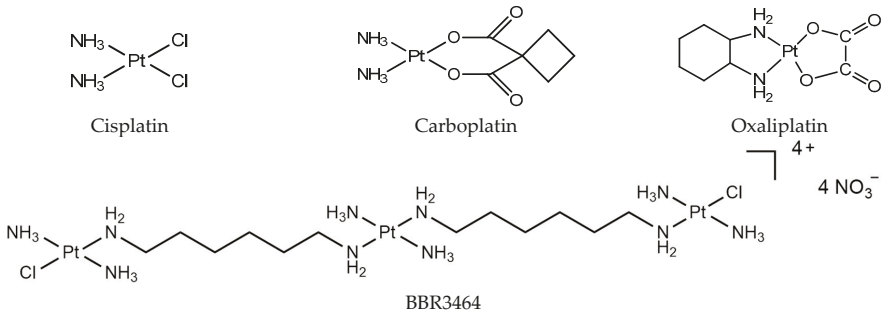


Figure 1. Platinum chemotherapeutic drugs: cisplatin, carboplatin, oxaliplatin, and BBR3464.

The study of the mechanism of classical platinum drugs has helped chemists develop new platinum drugs. Polynuclear platinum complexes (PPCs) are a new class of platinum anticancer complexes, which are structurally different from cisplatin, and exhibit a different mode of DNA binding, such as the phosphate clamp DNA binding mode of substitution-inert PPCs [10]. Especially, conformational changes induced by long-range inter-strand and intra-strand cross-links are distinctly different from those induced by mononuclear platinum complexes [11]. The prototype of this class, BBR3464 [{*trans*-PtCl(NH_3)$_2$}$_2${μ-*trans*-Pt(NH_3)$_2$($H_2N(CH_2)_6NH_2$)}]$^{4+}$ (Figure 1), is the only platinum compound without two exchangeable ligands in cis, and has reached Phase II clinical trials [12]. It is cytotoxic in cisplatin-resistant cell lines, and shows high efficacy in p53 mutant tumor cells [13]. BBR3464 can be deactivated in human plasma [14,15].

In addition to PPCs, there are also other *trans*-platinum complexes. Early studies have suggested that transplatin is inactive, but recent studies have found that some *trans*-platinum complexes have good in vitro and in vivo anti-tumor activity. The anti-tumor mechanism of this complex is still unclear. Generally, although DNA has long been believed to be the major target of platinum anticancer drugs, several proteins/enzymes have recently been proposed to be involved in the action of platinum complexes [16].

The discovery of the anticancer activity of cisplatin and its clinical application has triggered the study of metal-coordinated anticancer drugs; however, the serious side effects and intrinsic or acquired drug resistance of cisplatin largely limited its further clinical application. This has led medicinal chemists to explore other metal-based anticancer candidates, for example Ti, Os, and Ir complexes, to circumvent the problems associated with cisplatin administration [17–20]. A number of non-platinum complexes have entered clinical trials. Ruthenium compounds are regarded as promising alternatives to anticancer platinum drugs based on several advantages, for example, ruthenium compounds have lower toxicity and less drug resistance [21,22]. The most important developments comprise the clinically tested Ru^{III} compounds indazolium *trans*-[tetrachloridobis(1*H*-indazole)ruthenate(III)] (KP1019), imidazolium *trans*-[tetrachlorido(DMSO)(1*H*-imidazole)ruthenate(III)] (NAMI-A), and Ru^{II} compounds RAPTA (Figure 2). The structure of KP1019 was slightly modified due to solubility reasons, and was renamed as NKP-1339 (also IT-139). NKP-1339 is currently in clinical trials, and obtained orphan status from the Food and Drug Administration (FDA) in 2017. While the clinical trial for NAMI-A has been abandoned. Ru^{III} has a wide range of coordination numbers and geometries, as well as accessible redox states, which offer the medicinal chemists a wide spectrum of reactivities that can be exploited. RAPTA complexes are a promising class of organometallic Ru^{II} compounds that inhibit processes related to metastasis in vitro and exhibit pronounced antimetastatic activity in vivo, but only low antiproliferative activity [23,24]. Moreover, it has been shown that RAPTA compounds preferentially bind to proteins, even in the presence of DNA [25,26].

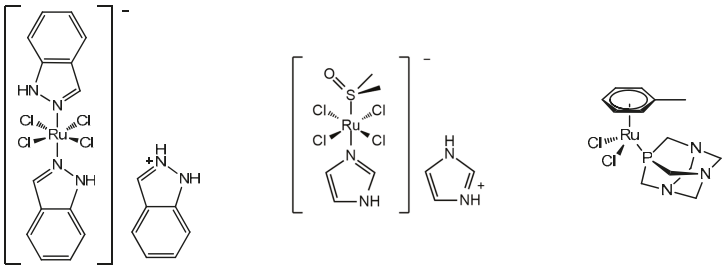


Figure 2. Chemical structure of ruthenium anticancer drugs: Ru^{III} compounds indazolium *trans*-[tetrachloridobis(1*H*-indazole)ruthenate(III)] (KP1019), imidazolium *trans*-[tetrachlorido(DMSO)(1*H*-imidazole)ruthenate(III)] (NAMI-A), RAPTA-T.

Except platinum and ruthenium, arsenic and gold complexes are also used as anticancer agents. As^{III} and As^{V} are the main oxidation states of arsenic. The cytotoxic activity of arsenic compounds in the trivalent state is strongly associated with the enhanced production of reactive oxygen species (ROS) [27,28]. The biological activity of pentavalent arsenic is mainly based on substitution for phosphate (e.g., in ATP) [29]. Amongst the various forms of arsenicals, the greatest clinical success has been the one of arsenic trioxide (ATO; As_2O_3, Trisenox®) (Figure 3) in the treatment of hematological cancers, especially acute promyelocytic leukemia (APL) [30,31]. Arsenic trioxide forms inorganic $As(OH)_3$ in an aqueous environment, and this form can be transported intracellularly via aquaglyceroporin channels due to its similarity to glycerol [32]. In addition to arsenic trioxide, other organic and inorganic arsenic agents are undergoing clinical trials for hematological malignancies, such as *S*-dimethylarsino-glutathione (Figure 3) [33].

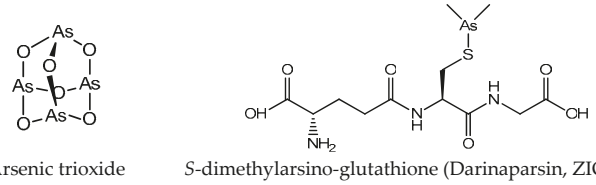


Figure 3. Chemical structure of arsenic anticancer drugs.

In the early days, gold compounds were mainly used for the treatment of rheumatoid arthritis (RA). Auranofin (Ridaura) was approved by the FDA for the treatment of RA in 1985 (Figure 4) [34]. Au compounds can be considered as prodrugs, and require activation (achieved by ligand exchange reactions) before they can develop their full pharmacological potential [29]. Several Au compounds are found to resolve the resistance of platinum compounds confirming different mechanisms of action [35]. It seems that the main targets of Au compounds are proteins rather than DNA, since scientists found the selenoprotein thioredoxin reductase plays a crucial role in the biological actions of gold compounds and acts as a major and general receptor for gold compounds, which can interact with specific thiol-containing and seleno-containing peptide moieties [36,37].

In addition to the above-mentioned metal drugs, many other metal complexes can be potentially used for the treatment of cancer, such as vanadium (V), rhodium (Rh), zinc (Zn), and cobalt (Co) [26]. Generally, metal-based anticancer drugs have high affinity for sulfur-containing biomolecules, such as proteins with Cys and Met residue. Thus, protein may play an important role in the mechanisms of those drugs such as drug resistance, toxicity, and metabolism. A comprehensive investigation of the interaction between metal-based anticancer drugs and their binding proteins or cell will further our understanding about the pharmacology of cytotoxic anticancer drugs from a molecular level.

Mass spectrometry-based quantitative strategy has been widely used for the identification of proteins specifically responding or binding to metal-based anticancer drugs.

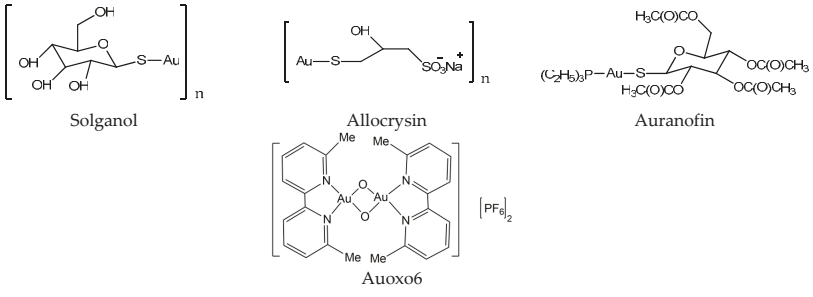


Figure 4. Chemical structure of gold anticancer drugs.

2. Quantitative Proteomics: General Remarks

Proteomics is the large-scale study of proteins, with a particular emphasis on their structures and functions. Proteins form a crucial part of the living organisms, as they are the main components of the metabolic and signaling pathways of cells while playing very important structural roles. Mass spectrometry is widely used in proteomic research because of its various advantages, including high sensitivity, high throughput, and good compatibility. It is even more powerful when combined with other separation techniques such as capillary electrophoresis (CE) and liquid chromatography (LC). Among the mass spectrometric ion source, ESI and MALDI are the most commonly used. Briefly, ESI generates multiply charged ions for biomolecules, while MALDI yields mostly singly charged pseudomolecular ions of analytes. ESI-MS can be easily hyphenated with separation techniques, whereas MALDI-TOF-MS cannot be directly combined to chromatographic methods; thus, MALDI-TOF-MS is usually accompanied with two-dimensional (2D) gel electrophoresis in proteomics, while ESI-MS is often coupled with nano-LC in proteomics to achieve higher sensitivity. Consequently, both ion sources are utilized for proteomic research.

Most of the mass spectrometers that are available offer one or more fragmentation methods that are used to provide information about the structure and composition of the ion of interest. Collision-induced dissociation (CID) is the most frequently employed fragmentation technique in current MS-based proteomics. In CID, selected peptides are subjected to collisions with inert gas molecules such as helium and nitrogen [38]. HCD is another fragmentation method, which is available for the LTQ Orbitrap [39]. In HCD, ions are fragmented in a collision cell rather than an ion trap, and then transferred back through the C-trap for analysis in the Orbitrap [40]. Besides CID and HCD, electron-based approaches such as electron capture dissociation (ECD) and electron transfer dissociation (ETD) are also used in MS-based proteomics as fragmentation techniques. The mechanism of electron-driven fragmentation techniques is fundamentally different from CID [41].

There are two general strategies applied in proteomic, which are called "bottom–up" and "top–down". Top–down describes the fragmentation of intact proteins without enzyme digestion. This approach is limited to small and pure proteins, while bottom–up involves enzymatic digestion of the analytes prior to fragmentation, and requires more sample preparation before data acquisition. A variety of chromatographic separation and enrichment methods could be used to achieve better peptide coverage and more identification. Owing to the introduction of enzymatic digestion, this strategy is well suited for investigating lager proteins and protein mixtures. For the quantitative proteomics research of biological systems, the "bottom–up" approach is more widely used.

Current methods for protein quantitation can be classified into two main categories: stable isotope labeling and label-free quantitation. Stable isotope labeling can introduce isotopically labeled forms into target components for relative quantification. The light labeled and heavy labeled proteins

are mixed proportionally before LC/MS analysis, and the peak area ratio of ion pairs generated by isotopically labeled analytes is used to quantify the components in the sample [42]. The workflow is shown in Figure 5. ICAT, TMT, iTRAQ, and SILAC belong to this category, and they are commonly used in quantitative proteomics research. A comparison of those methods is shown in Table 1.

Table 1. Comparison of commonly used methods in quantitative proteomics.

Quantitation Methods	Advantages	Disadvantages	Metal Drugs Investigated by Each Method
ICAT	Procedure is easy.	Only two samples can be labeled, which is only applicable to proteins containing cysteine.	Cisplatin [43,44]
TMT	Quantification on multiple sets of protein samples.	Expensive	Cisplatin [45]
iTRAQ	Quantification on multiple sets of protein samples.	Expensive	Cisplatin [44]
SILAC	Applicable to cultured cell.	It cannot be applied to samples such as tissues and body fluids. Expensive, time-consuming, and complicated.	Cisplatin [46] Gold (III) porphyrins [35]
LFQP	Straightforward and cost-effective.	It requires more rigorous analytical measurements and statistical validation than isotope-coded measurements.	Cisplatin [47] [Pd(sac)(terpy)](sac) [48] Plecstatin [49] RAPTA agent [50]
2-DE MS	It resolves thousands of intact protein species in a single run.	Time-consuming and labor-intense.	Cisplatin [51] Auranofin [52] Auoxo6 [52] Gold (III) NHC complexes [53]

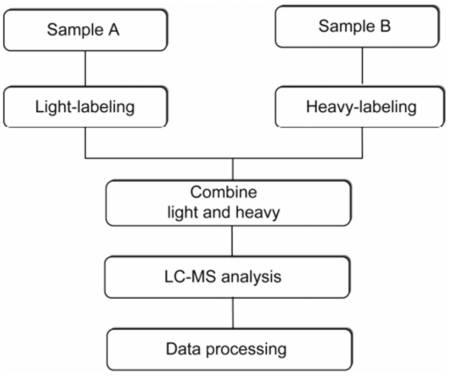


Figure 5. The workflow of stable isotope labeling-based quantitative proteomics. This figure is adapted from reference [44].

The ICAT reagent mainly consists of three parts: the first part is the affinity tag composed of biotin, which is used to separate ICAT-labeled peptide. The second part is the linker, and it is used to introduce the stable isotope. The third part is the reactive group that specifically binds to the thiol group of a cysteine residue in a peptide. The ICAT reagent exists in two forms—light (contains no deuterium) and heavy (contains eight deuterium atoms) form—and the mass difference among them is exactly eight Da (Figure 6). After the isotopic label is introduced into the peptide, the response intensity of two labeled forms for the same peptide is compared by MS. ICAT can analyze most proteins in cells, tissues, body fluids, etc., with good compatibility, but it should be noted that this method is only applicable to proteins containing cysteine [43,54–56].

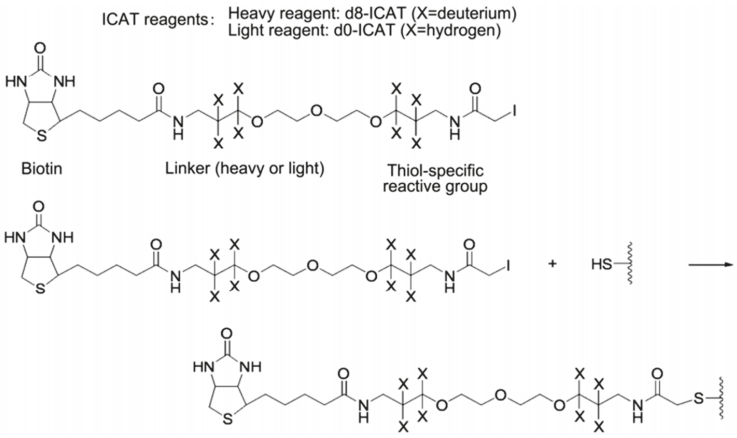


Figure 6. Structure and reaction mechanism of the ICAT reagent. This figure is adapted from reference [57].

In order to solve the shortcomings of the ICAT reagent, Thompson et al. synthesized TMT (tandem mass tags) reagent. The chemical structure of the commercial TMT agent is shown in Figure 7. It consists of a mass reporter region, a cleavable linker region, a mass normalization region, and a reactive group. The reactive group can specifically bind to the $-NH_2$ group of the peptide. When utilized for the relative quantification of multiple sets of protein samples, it enables the isotopically-labeled forms of target molecules to have identical chromatographic behavior and primary MS characteristic. The different label forms of target peptides are dissociated in the cleavable linker region and form different reporter ions for mass spectrometry (MS/MS) detection. Therefore, the relative content of target protein in different samples can be determined after comparing the intensity of reported ions.

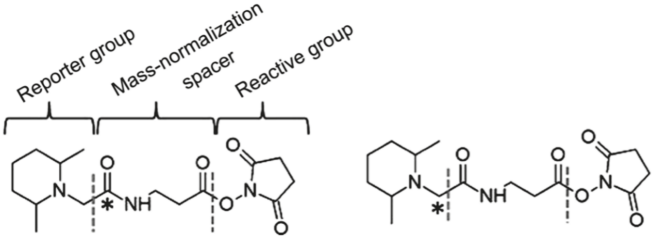


Figure 7. The chemical structure of two-plex tandem mass tag (TMT) reagents. This figure is adapted from reference [58].

iTRAQ is based on the same labeling strategy as TMT. The reagent consists of three parts (Figure 8), and the four available tags have identical overall mass. For four-plex iTRAQ reagents, the reporter group (green, N-methylpiperazine) mass is 114, 115, 116, and 117, respectively. The balance group (blue, carbonyl group) has a mass of 31 Da, 30 Da, 29 Da, and 28 Da accordingly. The reactive group (red, NHS ester) selectively reacts with the amino group of the peptide (the *N*-terminus and amino groups of the side chain). After LC-MS analysis, proteins are quantified by the relative intensity of the reporter group in MS/MS spectra. The iTRAQ technique has been widely used for the quantitative study of proteins due to its high accuracy and precision. What's more, it is able to simultaneously and quantitatively analyze up to eight samples [59,60]. However, iTRAQ reagents are expensive, and can significantly increase the cost of research.

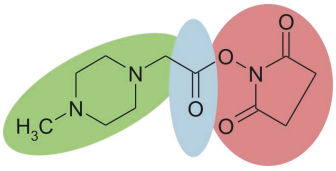


Figure 8. Chemical structure of the iTRAQ™ reagent. This figure is adapted from reference [61].

The stable isotope labeling methods mentioned above are all in vitro methods, and the typical example of the in vivo method for the relative quantitation of proteins is the stable isotope labeling by amino acids in cell culture (SILAC) [62]. The basic principle of this technique (Figure 9) is to add light and heavy isotopically-labeled essential amino acids (usually lysine and arginine) to the cell culture medium. After five to six doubling cycles, the amino acids of the newly synthesized protein in the cell are almost completely labeled with stable isotopes. Therefore, the accurate quantification of proteins can be achieved based on the peak intensity or area ratio of the two isotopically labeled peptides in the mixed sample. SILAC technology is able to mix samples at the protein level, which can effectively avoid the quantitative error caused by subsequent enzymatic digestion. It has high labeling efficiency and high quantitative accuracy. However, the presence of isotope-labeled arginine metabolism and the proline formation lead to low labeling efficiency and decreased quantification accuracy [63]. At the same time, this technique is only applicable to the cultured cell, and cannot be applied to samples such as tissues and body fluids, which are commonly used in medical research.

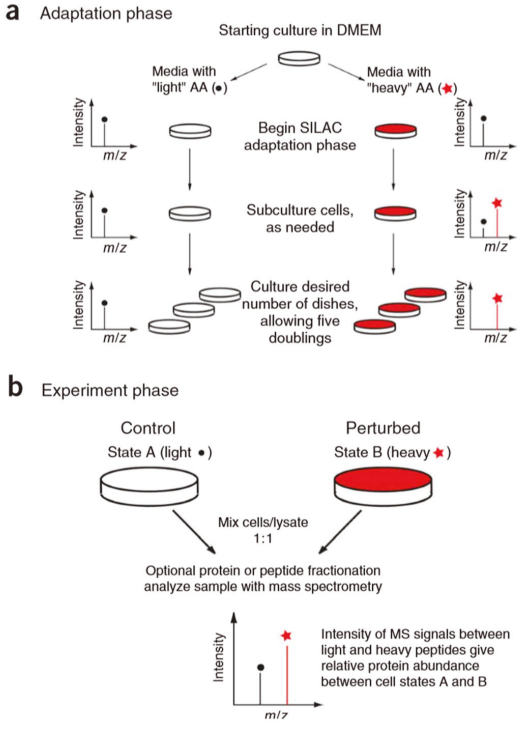


Figure 9. Overview of stable isotope labeling by amino acids in cell culture (SILAC) protocol. The SILAC experiment consists of two distinct phases: an adaptation (**a**) and an experimental (**b**) phase. This figure is adapted from reference [64].

Label-free quantitative proteomics (LFQP) is based on spectral counting and peak intensity for comparative analysis of the abundance of proteins. It provides a straightforward option for the large-scale analysis of biological samples. In contrast to label-based methods, LFQP mainly quantifies the number of identified peptides and the area of primary mass spectral peak by calculating the peptide fragment matching [65,66]. LFQP is cost-effective, and does not require expensive isotope reagents. In addition, LFQP is not time-consuming compared with some label-based methods that require tedious labeling steps [67,68]. Due to the aforementioned reasons and the increase of analytical capabilities of LC-MS/MS instrumentation in terms of resolution, accuracy, and sensitivity, LFQP has gained more acceptance in biomedical research. The characteristic peptides of each protein can be used for quantification [69], which can effectively improve the accuracy of non-labeled quantification. However, the label-free approach requires more rigorous analytical measurements and statistical validation compared with isotope-coded measurements.

As we can see, both labeling and label-free approaches have their own advantages and limitations (Table 1). More and more methods are developed for quantitative proteomic analysis in recent years. Those are the foundation for studying the mechanism of metal-based anticancer drugs.

3. Application of Quantitative Proteomics for Elucidation of the Mechanism of Metal-Based Anticancer Drugs

3.1. Proteomic Analysis of Cellular Responses to Metal-Based Anticancer Drugs

It's very important to monitor, in real time, the proteomic responses of cells to cytotoxic metallodrugs, as such responses might provide valuable information on the mechanism of action of the drug itself, and highlight which metabolic or signaling pathways of the cell are primarily affected and/or activated. If the damage is too intense to be repaired, specific biochemical pathways will be triggered, ultimately leading to cell apoptosis [70]. A number of studies utilizing this kind of strategy have appeared in recent literatures.

Cho et al. exploited the proteomic method based on a label-free quantitation strategy to study the cisplatin-induced hepatotoxicity [71]. Results showed that 76 proteins were up-regulated, and 19 proteins were down-regulated. The up-regulated proteins in the cisplatin-treated group include FBP1 (fructose 1,6-bisphosphatase 1), FASN (fatty acid synthase), CAT (catalase), PRDX1 (peroxiredoxin-1), HSPD1 (60-kDa heat shock protein), MDH2 (malate dehydrogenase 2), and ARG1 (arginase 1). Down-regulated proteins in the cisplatin-treated group include TPM1 (tropomyosin 1), TPM3 (tropomyosin 3), and CTSB (cathepsin B), which were further confirmed by Western blot analysis. Subsequent pathways analysis revealed that drug metabolism, fatty acid metabolism, glycolysis/TCA cycle, urea cycle, and inflammation metabolism were involved in cisplatin-induced hepatotoxicity.

Cytotoxic gold-based complexes have different modes of action from cisplatin. It was found that auranofin and Auoxo6 are more active than cisplatin against both the cisplatin-sensitive A2780/S and resistant A2780/R cell lines. Guidi et al. reported the use of 2-DE and MS analysis to find the molecular mechanisms through which auranofin and Auoxo6 caused their biological effects [52]. 2-DE combined with MALDI-TOF analyses showed that 10 proteins were down-regulated, and one protein was up-regulated in A2780/R cells treated with auranofin versus controls. For Auoxo6, 12 proteins were down-regulated and four proteins were up-regulated. After investigation of the altered proteins, they proposed that auranofin mostly acts by altering the amount of proteasome proteins, while Auoxo6 mainly modifies proteins related to mRNA splicing, trafficking, and stability. Interestingly, thioredoxin-like protein 1, which is involved in oxidative stress defense, was greatly reduced after treatment with both gold compounds.

The characteristics of high anticancer efficiency of cisplatin and the clinical inactivity of its trans isomer (transplatin) have been considered a paradigm for the classical structure-activity relationships of platinum drugs. However, some new analogues of transplatin exhibit activity in cisplatin-resistant tumor cells, but the mechanism behind this activity is unknown. An MS-based proteomic strategy combined with functionalized gold nanoparticles as affinity probes was developed to study the

cellular proteins responding to damaged DNA by cisplatin and *trans*-PtTz (Figure 10) [72]. To exclude nonspecific binding proteins to the platinated DNA, the negative probe with native double-stranded DNA on gold nanoparticles was utilized as a control. As a result, the well-known protein HMGB1 was identified as a binding partner of platinated DNA by cisplatin, and the nuclear protein positive cofactor PC4 was found to bind specifically with cross-linked DNA by *trans*-PtTz, which will shed light on the mechanism of this active transplatinum complex.

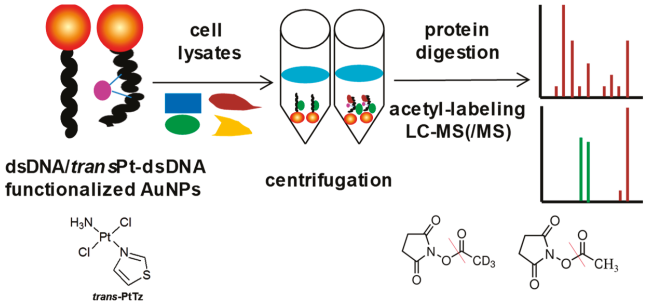


Figure 10. Overview of the MS-based proteomic strategy to study the cellular proteins responding to damaged DNA by platinum drugs. This figure is adapted from reference [72].

Palladium (Pd)-based compounds have gained the interest of many researchers, as they exhibit similar coordination chemistry and better solubility compared with platinum drugs [73]. [Pd(sac)(terpy)](sac)·$4H_2O$ (sac = saccharinate, and terpy = 2,2′:6′,2″-terpyridine), a palladium-based compound, was reported to be more potential than cisplatin in breast cancer cells. To figure out the mechanism of action, Adiguzel et al. performed nanoLC-MS/MS analysis to investigate the global proteomic changes after treatment of this Pd(II) complex [48]. Eventually, 681 proteins were identified, among which 335 protein groups were quantified through the label-free quantitative method. Furthermore, 30 differentially expressed proteins were identified between drug-treated cells and untreated cells. These proteins were involved in the regulation of apoptosis, proliferation, protein degradation, and DNA repair, etc. Ingenuity pathway analysis revealed that the involved pathway appeared as the protein ubiquitination; therefore, apoptosis is the mechanism of cell death in response to Pd(II) treatment, as a significant increase in ubiquitination was identified. Finally, they proposed that the mode for the cytotoxic action of the Pd(II) complex was that cells initially attempt to repair Pd(II)-induced damage, yet prolonged damage or exposure of higher doses lead to protein degradation and apoptosis in cancer cells.

3.2. Analysis of the Molecular Basis of Platinum Resistance through Comparative Proteomic Analysis of Pt Sensitive versus Pt-Resistant Cell Lines

There are lots of studies about the molecular basis of platinum drugs resistance. Indeed, acquired resistance is often the reason for treatment failure, and therefore attracts great interest. Understanding the molecular mechanism of resistance might help to circumvent it. Thus, a conspicuous number of proteomic studies were specifically devoted to this issue.

A quantitative proteomic screening was performed to identify the proteins that are differentially expressed in drug-resistant cell lines through which the mechanisms involved in cisplatin resistance may be found out [46]. The SILAC approach with nano-LC-MS was employed in this experiment. As a result, a total of 856 proteins, among which 374 proteins were differentially expressed between the cisplatin-resistant cell (HeLa/CDDP) and sensitive cells (HeLa) were identified. The expression of a few key proteins, including CD44, DDB-1, DJ-1, and XRCC5, were confirmed by Western blotting, which was highly consistent with the proteomic analysis. A further protein interaction network based on the differentially expressed proteins was constructed, and finally, the biological pathways, including

carbohydrates, energy-producing, regulation of apoptosis, and protein folding were involved in the drug resistance.

A number of studies have found that mitochondria was correlated with the cisplatin resistance. Mitochondrial DNA and membrane proteins were reported as preferential targets of cisplatin. It was also shown by several groups that mitochondria impairment appeared to play an important role in the platinum resistance of ovarian cancer cells. Using 2D DIGE integrated with MALDI-TOF-MS, Dai et al. investigated the mitochondrial proteins difference between platinum-sensitive human ovarian cancer cell lines (SKOV3 and A2780) with that of four platinum-resistant sublines (SKOV3/CDDP, SKOV3/CBP, A2780/CDDP, and A2780/CBP) [51]. Through a 2D DIGE experiment, 236 spots were identified, among which 128 spots were down-regulated in platinum-resistant cells, and 108 spots were up-regulated in these resistant cells. Eleven spots that had more than threefold changes in platinum-resistant cells compared with platinum-sensitive cell lines were analyzed and verified by MALDI-TOF-MS. Five of the proteins, ETF, PRDX3, PHB, ATP-a, and ALDH, were identified. ATP-a, PHB, and PRDX3 have been validated as mitochondrial proteins of ovarian cancer cells, and they were further confirmed through immunoblotting. The expressions of ATP-a, PHB, and PRDX3 were further validated in the clinical ovarian cancer sections; as a result, a significant difference existed in PHB expression between the sensitive group and the resistant group, demonstrating that PHB might be a correlative candidate protein for platinum resistance in the mitochondria of ovarian cancer cells.

Using the ICAT approach, Stewart et al. profiled the nuclear, cytosolic, and microsomal fractions of IGOV-1 (cisplatin-sensitive) and IGOV-1/CP (cisplatin-resistant) ovarian cancer cell lines [43]. A total of 1117 proteins were identified and quantified, among which 121 proteins were expressed differentially in cisplatin-resistant cell lines compared with the sensitive ones. Sixty-three proteins were overexpressed in cisplatin-sensitive cells, and 58 proteins had low expression in these cells. Among the 63 overexpressed proteins, several proteins were overexpressed at least fivefold in resistant cells, including cell recognition molecule CASPR3 (13.3-fold), S100 protein family members (8.7-fold), junction adhesion molecule Claudin 4 (7.2-fold), and CDC42-binding protein kinase (5.4-fold). Other proteins exhibited low expression for at least fivefold in resistant cells, including hepatocyte growth factor inhibitor 1B (13.3-fold) and programmed cell death 6-interacting protein (12.7-fold). They also compared the expression of mRNA with that of protein in a subset of 92 highly differentially expressed proteins, and the expression level of 37 proteins are in the same direction with that of mRNA, and 55 are discordant, possibly reflecting the post-transcriptional control of protein expression. By Gene Ontology (GO) analysis, many processes, including RNA splicing, processing, and DNA replication were increased in cisplatin-resistant cells. The increased activities of these biological processes may lead to faster repairs of cisplatin-induced DNA damages, thus resulting in a resistant phenotype. They also found that three pathways (glycolysis, the interleukin signaling pathway, and the PI 3-kinase pathway) were significantly up-regulated in cisplatin-sensitive cells, which were involved in cell apoptosis.

Cisplatin-based chemotherapy is currently used for bladder cancer (BC), but it lacks efficiency on patients who have acquired or developed resistance. In order to reveal the molecular mechanisms underlying this resistance, Jung et al. carried out a multidimensional proteomic analysis on cisplatin-sensitive (T24S) and resistant (T24R) T24 human BC cell lines [45]. It was reported that the aberrant expression or mutations of the EGFR family are related to the carcinogenesis of bladder cancer (BC); therefore, the temporal changes in protein abundance and phosphorylation in T24S and T24R cells after EGF stimulation were also investigated. Sixplex TMT reagents were used to label peptide samples. Consequently, the global proteome profiles in both T24R and T24S cells changed slightly. Whereas, phosphoproteome in T24S cells changed more than T24R cells, which revealed that T24S cells were impacted more greatly than T24R cells by EGF stimulation. The analysis of altered proteins revealed associations of cisplatin resistance with DNA damage, repair, and cell cycle regulation, which is consistent with previous reports [74]. Several key regulators linked to

cisplatin resistance were confirmed. These results are promising, but need more clinical specimens for phosphorproteomic analysis.

Neuroblastoma is a challenging childhood malignancy with a very high percentage of patients relapsing after the acquisition of drug resistance. In order to investigate the molecular pathways involved in the drug resistance of neuroblastoma, Piskareva et al. characterized three cisplatin-sensitive/resistant cell line pairs using the label-free quantitative method [47]. As a result, 46/72, 68/43, and 34/63 proteins were found to be up-regulated and down-regulated for the three cell line pairs, respectively. Differentially-expressed proteins for each individual cell line pair were used to analyze the molecular and cellular functions that were involved through ingenuity pathway analysis. Known mutual interactions among differentially expressed proteins for each cell line pair were used to construct protein networks. Consequently, four proteins were in common across these networks, including betatubulin (TUBB), beta-actin (ACTB), vimentin (VIM), and 78 kDa glucose-regulated protein (HSPA5). Pathways analysis suggested that the epithelial-to-mesenchymal transition is a feature during the development of drug resistance in neuroblastoma.

3.3. Mass Spectrometry-Based Quantitative Proteomics for Identification of Target Proteins for Metal-Based Anticancer Drugs

Target identification utilizing photoaffinity groups together with clickable moieties recently emerged as a useful strategy to identify macromolecular binding partners for small organic molecules, but it was seldom used for the target identification of anticancer metal complexes. Fung et al. developed a chemical probe for the target identification of $[Au^{III}(C\hat{}N\hat{}C)(NHC)]OTf$ by introducing a small photoaffinity diazirine group and a clickable alkyne moiety on NHC [53]. HeLa cells were treated with the probe, and then irradiated followed by a click reaction; eventually, six photoaffinity-labeled proteins were identified by gel electrophoresis accompanied by MALDI-TOF-MS, and these proteins were also found in the protein samples from NCI-H460 and HCT116 cells treated with the same probe. The six proteins were identified as mitochondrial heat shock protein 60 (HSP60), vimentin (VIM), nucleoside diphosphate kinase A (NDKA), nucleophosmin (NPM), nuclease-sensitive element binding protein (Y box binding protein, YB-1), and peroxiredoxin1 (PRDX1) by MALDI-TOF-MS/MS analysis. Except for YB-1, other proteins were verified by HPLC-LTQ-Orbitrap MS analysis with high confidence. Besides, gold(III) meso-tetraphenylporphyrin is notable for its high stability in biological environments and potent in vitro and in vivo anticancer activities. Extensive chemical biology approaches, including photoaffinity labeling, click chemistry, chemical proteomics, and the SILAC technique were used to find the protein target of gold(III) porphyrins [35]. Compelling evidence revealed that heat-shock protein 60 (Hsp60), a mitochondrial chaperone and potential anticancer target, is a direct target of gold(III) porphyrins in vitro and in cells. Structure-activity studies with a series of non-porphyrin gold(III) complexes and other metalloporphyrins revealed that Hsp60 inhibition is specifically dependent on both the gold(III) ion and the porphyrin ligand.

Organometallic anticancer agents often require ligand exchange for their anticancer activity, which is generally believed to possess low selectivity for potential cellular targets. However, Meier et al. found an unexpected target selectivity of a ruthenium(arene) pyridinecarbothioamide (plecstatin) (Figure 11). They utilized a label-free quantitative method to seek out the potential target of plecstatin. To address nonspecific interactions, a competition experiment was conducted by pretreatment with drug [49]. As a result, roughly 400 proteins were identified, among which only outer dense fiber protein 2 (ODF2, 210-fold) and plectin (PLEC, 160-fold) were considered as potential targets with high enrichment factors. The latter one is a scaffold protein and cytolinker with pronounced effects on the organization of non-mitotic microtubules, which was considered as an unexpected target for plecstatin. Moreover, non-mitotic microtubules are an underappreciated drug target, and their disturbance by plectin-targeting agents affects the motility of cancer cells, which may develop a promising anticancer strategy.

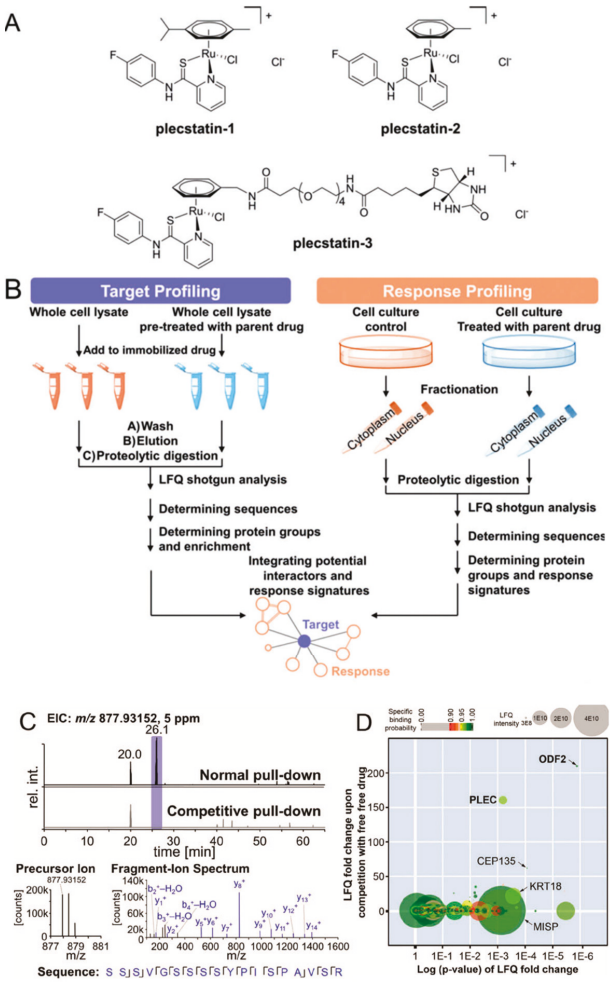


Figure 11. Potential protein targets of ruthenium(arene) pyridinecarbothioamides (plecstatins) were obtained by integrated target-response profiling. (**A**) Chemical structures. (**B**) Schematic representation of the integrated target-response profiling workflow. (**C**) Selected chromatograms and mass spectra of a plectin-specific tryptic peptide. (**D**) Plot of the target profiling experiment. PLEC: plectin; ODF2: outer dense fiber protein 2. This figure is adapted from reference [49].

RAPTA compounds have attracted researchers' attention, as their modes of action are substantially different from commonly used platinum-based chemotherapeutics. In this case, drug pull-down combined with affinity chromatography, mass spectrometry-based proteomics, and bioinformatics were utilized to figure out the target proteins of antimetastatic agent RAPTA in ovarian cancer lysate [50]. A competitive experiment using an acetylated RAPTA analogue was carried out to remove the non-specific binding proteins. Pull-down experiments without competitive experiments resulted in the identification of a total of 184 proteins. After comparing data with the competitive experiment, the number of high-affinity proteins decreased to 29, which can be classified into four types: extracellular proteins, cell cycle-regulating proteins, histone-related proteins, and ribosomal proteins. Among the 29 proteins, 15 proteins were found to be cancer-related. What's more, the identified

proteins, including the cytokines midkine, pleiotrophin, fibroblast growth factor-binding protein 3, guanine nucleotide-binding protein-like 3, and FAM32A were consistent with the hypothesis that the antiproliferative activity of RAPTA compounds is due to the induction of a G2/M arrest and histone proteins [75].

4. Conclusion Remarks

In recent years, metal-based anticancer drugs have played an important role in the clinical chemotherapy of cancer, especially platinum-based anticancer drugs represented by cisplatin. The adverse effects and acquired resistance of platinum-based anticancer drugs triggered researchers to exploit novel anticancer metallodrugs and figure out the underlying mechanism of anticancer activity and drug resistance.

Initial studies identified DNA as the primary target of platinum-based drugs, but current research has revealed that metal-based drugs can also bind with proteins, which was correlated with their activity, drug resistance, toxicity, and metabolism. Proteomics is the large-scale study of proteins, and mass spectrometry has been widely used in proteomic research because of its high sensitivity, high throughput, and good compatibility. Current methods for protein quantitation include stable isotope labeling and label-free quantitation, which have been widely used for the studies of target proteins and cellular response proteins for metal-based anticancer drugs. The altered expression of proteins found by quantitative proteomics revealed that several metabolic and signaling pathways, for example protein ubiquitination, were involved in the mechanism of metal drugs. Cellular protein targets identified through quantitative proteomics pinpointed the molecular pharmacology of metal-based anticancer drugs, such as Hsp60 for gold (III) porphyrins. Therefore, mass spectrometry-based quantitative proteomics becomes a powerful tool to elucidate the mechanism of action for metal-based anticancer drugs, which will facilitate the design of more efficient anticancer drugs.

Author Contributions: Z.D. came with the idea and edited the whole text, R.W. wrote Sections 1 and 2, S.J. wrote Sections 3 and 4. K.W. and H.J. revised the manuscript.

Funding: This work was supported by National Natural Science Foundation of China (no. 81874309).

Conflicts of Interest: The authors declare that there are no conflicts of interest.

Abbreviations

ESI	electron spray ionization
MALDI	matrix assisted laser desorption ionization
TOF	time of flight
HCD	higher energy collisional dissociation
ICAT	isotope-coded affinity tag
ITRAQ	isobaric tag for relative and absolute quantitation
TMT	tandem mass tags
SILAC	stable isotope labeling by amino acids in cell culture
LFQP	label-free quantitative proteomics
2-DE	two-dimensional (2D) gel electrophoresis
HMGB1	high mobility group box 1 protein
PC4	positive cofactor 4
DDB-1	DNA damage-binding protein 1
DJ-1	Protein/nucleic acid deglycase DJ-1
XRCC5	X-ray repair cross-complementing protein 5
2-D DIGE	Two-dimensional fluorescence difference gel electrophoresis
ALDH	aldehyde dehydrogenase
ETF	electron transfer flavoprotein
PHB	prohibitin
PRDX3	peroxiredoxin III
EGFR	epidermal growth factor receptor

References

1. Rosenberg, B.; VanCamp, L.; Trosko, J.E.; Mansour, V.H. Platinum compounds: A new class of potent antitumour agents. *Nature* **1969**, *222*, 385–386. [CrossRef] [PubMed]
2. Wang, X.; Guo, Z. The Role of Sulfur in Platinum Anticancer Chemotherapy. *Anticancer Agents Med. Chem.* **2007**, *7*, 19–34. [CrossRef]
3. Meijer, C.; Mulder, N.H.; Timmer-Bosscha, H.; Sluiter, W.J.; Meersma, G.J.; de Vries, E.G. Relationship of cellular glutathione to the cytotoxicity and resistance of seven platinum compounds. *Cancer Res.* **1992**, *52*, 6885–6889. [PubMed]
4. Hall, M.D.; Okabe, M.; Shen, D.-W.; Liang, X.-J.; Gottesman, M.M. The role of cellular accumulation in determining sensitivity to platinum-based chemotherapy. *Annu. Rev. Pharmacol. Toxicol.* **2008**, *48*, 495–535. [CrossRef] [PubMed]
5. Zwelling, L.A.; Anderson, T.; Kohn, K.W. DNA-protein and DNA interstrand cross-linking by cis- and trans-platinum(II) diamminedichloride in L1210 mouse leukemia cells and relation to cytotoxicity. *Cancer Res.* **1979**, *39*, 365–369. [PubMed]
6. Jamieson, E.R.; Lippard, S.J. Structure, recognition, and processing of cisplatin-DNA adducts. *Chem. Rev.* **1999**, *99*, 2467–2498. [CrossRef]
7. Jung, Y.; Lippard, S.J. Direct cellular responses to platinum-induced DNA damage. *Chem. Rev.* **2007**, *107*, 1387–1407. [CrossRef]
8. Desoize, B.; Madoulet, C. Particular aspects of platinum compounds used at present in cancer treatment. *Crit. Rev. Oncol. Hemat.* **2002**, *42*, 317–325. [CrossRef]
9. Kartalou, M.; Essigmann, J.M. Recognition of cisplatin adducts by cellular proteins. *Mutat. Res. Fundam. Mol. Mech. Mutag.* **2001**, *478*, 1–21. [CrossRef]
10. Farrell, N.P. Multi-platinum anti-cancer agents. Substitution-inert compounds for tumor selectivity and new targets. *Chem. Soc. Rev.* **2015**, *44*, 8773–8785. [CrossRef]
11. Mangrum, J.B.; Farrell, N.P. Excursions in polynuclear platinum DNA binding. *Chem. Commun.* **2010**, *46*, 6640–6650. [CrossRef] [PubMed]
12. Farrell, N.P. Progress in platinum-derived drug development. *Drugs Future* **2012**, *37*, 795–806. [CrossRef]
13. Manzotti, C.; Pratesi, G.; Menta, E.; Di Domenico, R.; Cavalletti, E.; Fiebig, H.H.; Kelland, L.R.; Farrell, N.; Polizzi, D.; Supino, R.; et al. BBR 3464: A novel triplatinum complex, exhibiting a preclinical profile of antitumor efficacy different from cisplatin. *Clin. Cancer Res.* **2000**, *6*, 2626–2634. [PubMed]
14. Oehlsen, M.E.; Qu, Y.; Farrell, N. Reaction of polynuclear platinum antitumor compounds with reduced glutathione studied by multinuclear (H-1, H-1-N-15 gradient heteronuclear single-quantum coherence, and Pt-195) NMR spectroscopy. *Inorg. Chem.* **2003**, *42*, 5498–5506. [CrossRef] [PubMed]
15. Vacchina, V.; Torti, L.; Allievi, C.; Lobinski, R. Sensitive species-specific monitoring of a new triplatinum anti-cancer drug and its potential related compounds in spiked human plasma by cation-exchange HPLC-ICP-MS. *J. Anal. At. Spectrom.* **2003**, *18*, 884–890. [CrossRef]
16. Fang, T.; Cao, K.; Cheng, L.; Zhao, L.; Liu, Y. Protein interaction in the mechanism of platinum anticancer drugs. *Sci. Sin. Chim.* **2017**, *47*, 200–219. [CrossRef]
17. Cini, M.; Bradshaw, T.D.; Woodward, S. Using titanium complexes to defeat cancer: The view from the shoulders of titans. *Chem. Soc. Rev.* **2017**, *46*, 1040–1051. [CrossRef]
18. Hanif, M.; Babak, M.V.; Hartinger, C.G. Development of anticancer agents: Wizardry with osmium. *Drug Discov. Today* **2014**, *19*, 1640–1648. [CrossRef]
19. Hanif, M.; Hartinger, C.G. Anticancer metallodrugs: Where is the next cisplatin? *Future Med. Chem.* **2018**, *10*, 615–617. [CrossRef]
20. Liu, Z.; Sadler, P.J. Organoiridium Complexes: Anticancer Agents and Catalysts. *Acc. Chem. Res.* **2014**, *47*, 1174–1185. [CrossRef]
21. Bergamo, A.; Sava, G. Ruthenium anticancer compounds: Myths and realities of the emerging metal-based drugs. *Dalton Trans.* **2011**, *40*, 7817–7823. [CrossRef] [PubMed]
22. Pal, M.; Nandi, U.; Mukherjee, D. Detailed account on activation mechanisms of ruthenium coordination complexes and their role as antineoplastic agents. *Eur. J. Med. Chem.* **2018**, *150*, 419–445. [CrossRef] [PubMed]
23. Bergamo, A.; Masi, A.; Dyson, P.J.; Sava, G. Modulation of the metastatic progression of breast cancer with an organometallic ruthenium compound. *Int. J. Oncol.* **2008**, *33*, 1281–1289. [CrossRef] [PubMed]

24. Scolaro, C.; Bergamo, A.; Brescacin, L.; Delfino, R.; Cocchietto, M.; Laurenczy, G.; Geldbach, T.J.; Sava, G.; Dyson, P.J. In vitro and in vivo evaluation of ruthenium(II)-arene PTA complexes. *J. Med. Chem.* **2005**, *48*, 4161–4171. [CrossRef] [PubMed]
25. Wu, B.; Ong, M.S.; Groessl, M.; Adhireksan, Z.; Hartinger, C.G.; Dyson, P.J.; Davey, C.A. A Ruthenium Antimetastasis Agent Forms Specific Histone Protein Adducts in the Nucleosome Core. *Chem. Eur. J.* **2011**, *17*, 3562–3566. [CrossRef]
26. Adhireksan, Z.; Davey, G.E.; Campomanes, P.; Groessl, M.; Clavel, C.M.; Yu, H.; Nazarov, A.A.; Yeo, C.H.F.; Ang, W.H.; Droege, P.; et al. Ligand substitutions between ruthenium-cymene compounds can control protein versus DNA targeting and anticancer activity. *Nat. Commun.* **2014**, *5*, 3462. [CrossRef]
27. Jing, Y.K.; Dai, J.; Chalmers-Redman, R.M.E.; Tatton, W.G.; Waxman, S. Arsenic trioxide selectively induces acute promyelocytic leukemia cell apoptosis via a hydrogen peroxide-dependent pathway. *Blood* **1999**, *94*, 2102–2111.
28. Zheng, Y.H.; Yamaguchi, H.; Tian, C.J.; Lee, M.W.; Tang, H.; Wang, H.G.; Chen, Q. Arsenic trioxide (As_2O_3) induces apoptosis through activation of Bax in hematopoietic cells. *Oncogene* **2005**, *24*, 3339–3347. [CrossRef]
29. Englinger, B.; Pirker, C.; Heffeter, P.; Terenzi, A.; Kowol, C.R.; Keppler, B.K.; Berger, W. Metal Drugs and the Anticancer Immune Response. *Chem. Rev.* **2018**. [CrossRef]
30. Liu, J.X.; Zhou, G.B.; Chen, S.J.; Chen, Z. Arsenic compounds: Revived ancient remedies in the fight against human malignancies. *Curr. Opin. Chem. Biol.* **2012**, *16*, 92–98. [CrossRef]
31. Wang, Y.; Wang, H.; Li, H.; Sun, H. Metallomic and metalloproteomic strategies in elucidating the molecular mechanisms of metallodrugs. *Dalton Trans.* **2015**, *44*, 437–447. [CrossRef] [PubMed]
32. Chau, D.; Ng, K.; Chan, T.S.-Y.; Cheng, Y.-Y.; Fong, B.; Tam, S.; Kwong, Y.-L.; Tse, E. Azacytidine sensitizes acute myeloid leukemia cells to arsenic trioxide by up-regulating the arsenic transporter aquaglyceroporin 9. *J. Hematol. Oncol.* **2015**, *8*, 46. [CrossRef] [PubMed]
33. Torka, P.; Al Ustwani, O.; Wetzler, M.; Wang, E.S.; Griffiths, E.A. Swallowing a bitter pill-oral arsenic trioxide for acute promyelocytic leukemia. *Blood Rev.* **2016**, *30*, 201–211. [CrossRef]
34. Eisler, R. Chrysotherapy: A synoptic review. *Inflamm. Res.* **2003**, *52*, 487–501. [CrossRef] [PubMed]
35. Hu, D.; Liu, Y.; Lai, Y.-T.; Tong, K.-C.; Fung, Y.-M.; Lok, C.-N.; Che, C.-M. Anticancer Gold(III) Porphyrins Target Mitochondrial Chaperone Hsp60. *Angew. Chem. Int. Ed.* **2016**, *55*, 1387–1391. [CrossRef] [PubMed]
36. Saggioro, D.; Rigobello, M.P.; Paloschi, L.; Folda, A.; Moggach, S.A.; Parsons, S.; Ronconi, L.; Fregona, D.; Bindoli, A. Gold(III)—Dithiocarbamato complexes induce cancer cell death triggered by thioredoxin redox system inhibition and activation of ERK pathway. *Chem. Biol.* **2007**, *14*, 1128–1139. [CrossRef] [PubMed]
37. Nobili, S.; Mini, E.; Landini, I.; Gabbiani, C.; Casini, A.; Messori, L. Gold Compounds as Anticancer Agents: Chemistry, Cellular Pharmacology, and Preclinical Studies. *Med. Res. Rev.* **2010**, *30*, 550–580. [CrossRef]
38. Mcluckey, S.A. Principles of collisional activation in analytical mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **1992**, *3*, 599–614. [CrossRef]
39. Olsen, J.V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* **2007**, *4*, 709. [CrossRef]
40. Jedrychowski, M.P.; Huttlin, E.L.; Haas, W.; Sowa, M.E.; Rad, R.; Gygi, S.P. Evaluation of HCD- and CID-type Fragmentation Within Their Respective Detection Platforms for Murine Phosphoproteomics. *Mol. Cell. Proteomics* **2011**, 10. [CrossRef]
41. Zubarev, R.A.; Zubarev, A.R.; Savitski, M.M. Electron capture/transfer versus collisionally activated/induced dissociations: Solo or duet? *J. Am. Soc. Mass. Spectrom.* **2008**, *19*, 753–761. [CrossRef] [PubMed]
42. Toyo'oka, T. LC-MS determination of bioactive molecules based upon stable isotope-coded derivatization method. *J. Pharm. Biomed. Anal.* **2012**, *69*, 174–184. [CrossRef] [PubMed]
43. Stewart, J.J.; White, J.T.; Yan, X.W.; Collins, S.; Drescher, C.W.; Urban, N.D.; Hood, L.; Lin, B.Y. Proteins associated with cisplatin resistance in ovarian cancer cells identified by quantitative proteomic technology and integrated with mRNA expression levels (vol 5, pg 433, 2006). *Mol. Cell. Proteomics* **2006**, *5*, 433–443. [CrossRef] [PubMed]
44. Wang, Y.; Zhang, L.; Zheng, X.; Zhong, W.; Tian, X.; Yin, B.; Tian, K.; Zhang, W. Long non-coding RNA LINC00161 sensitises osteosarcoma cells to cisplatin-induced apoptosis by regulating the miR-645-IFIT2 axis. *Cancer Lett.* **2016**, *382*, 137–146. [CrossRef] [PubMed]

45. Jung, J.H.; You, S.; Oh, J.W.; Yoon, J.; Yeon, A.; Shahid, M.; Cho, E.; Sairam, V.; Park, T.D.; Kim, K.P.; et al. Integrated proteomic and phosphoproteomic analyses of cisplatin-sensitive and resistant bladder cancer cells reveal CDK2 network as a key therapeutic target. *Cancer Lett.* **2018**, *437*, 1–12. [CrossRef] [PubMed]
46. Chavez, J.D.; Hoopmann, M.R.; Weisbrod, C.R.; Takara, K.; Bruce, J.E. Quantitative Proteomic and Interaction Network Analysis of Cisplatin Resistance in HeLa Cells. *PLoS ONE* **2011**, *6*, 19892. [CrossRef] [PubMed]
47. Piskareva, O.; Harvey, H.; Nolan, J.; Conlon, R.; Alcock, L.; Buckley, P.; Dowling, P.; O'Sullivan, F.; Bray, I.; Stallings, R.L. The development of cisplatin resistance in neuroblastoma is accompanied by epithelial to mesenchymal transition in vitro. *Cancer Lett.* **2015**, *364*, 142–155. [CrossRef]
48. Adiguzel, Z.; Baykal, A.T.; Kacar, O.; Yilmaz, V.T.; Ulukaya, E.; Acilan, C. Biochemical and Proteomic Analysis of a Potential Anticancer Agent: Palladium(II) Saccharinate Complex of Terpyridine Acting through Double Strand Break Formation. *J. Proteome Res.* **2014**, *13*, 5240–5249. [CrossRef]
49. Meier, S.M.; Kreutz, D.; Winter, L.; Klose, M.H.M.; Cseh, K.; Weiss, T.; Bileck, A.; Alte, B.; Mader, J.C.; Jana, S.; et al. An Organoruthenium Anticancer Agent Shows Unexpected Target Selectivity for Plectin. *Angew. Chem. Int. Ed.* **2017**, *56*, 8267–8271. [CrossRef]
50. Babak, M.V.; Meier, S.M.; Huber, K.V.M.; Reynisson, J.; Legin, A.A.; Jakupec, M.A.; Roller, A.; Stukalov, A.; Gridling, M.; Bennett, K.L.; et al. Target profiling of an antimetastatic RAPTA agent by chemical proteomics: Relevance to the mode of action. *Chem. Sci.* **2015**, *6*, 2449–2456. [CrossRef]
51. Dai, Z.; Yin, J.; He, H.; Li, W.; Hou, C.; Qian, X.; Mao, N.; Pan, L. Mitochondrial comparative proteomics of human ovarian cancer cells and their platinum-resistant sublines. *Proteomics* **2010**, *10*, 3789–3799. [CrossRef] [PubMed]
52. Guidi, F.; Landini, I.; Puglia, M.; Magherini, F.; Gabbiani, C.; Cinellu, M.A.; Nobili, S.; Fiaschi, T.; Bini, L.; Mini, E.; et al. Proteomic analysis of ovarian cancer cell responses to cytotoxic gold compounds. *Metallomics* **2012**, *4*, 307–314. [CrossRef]
53. Fung, S.K.; Zou, T.; Cao, B.; Lee, P.-Y.; Fung, Y.M.E.; Hu, D.; Lok, C.-N.; Che, C.-M. Cyclometalated Gold(III) Complexes Containing N-Heterocyclic Carbene Ligands Engage Multiple Anti-Cancer Molecular Targets. *Angew. Chem. Int. Ed.* **2017**, *56*, 3892–3896. [CrossRef]
54. Han, D.K.; Eng, J.; Zhou, H.L.; Aebersold, R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* **2001**, *19*, 946–951. [CrossRef] [PubMed]
55. Li, C.; Hong, Y.; Tan, Y.X.; Zhou, H.; Ai, J.H.; Li, S.J.; Zhang, L.; Xia, Q.C.; Wu, J.R.; Wang, H.Y.; et al. Accurate qualitative and quantitative proteomic analysis of clinical hepatocellular carcinoma using laser capture microdissection coupled with isotope-coded affinity tag and two-dimensional liquid chromatography mass spectrometry. *Mol. Cell. Proteomics* **2004**, *3*, 399–409. [CrossRef] [PubMed]
56. Chen, R.; Sheng, P.; Yi, E.C.; Donohoe, S.; Bronner, M.P.; Potter, J.D.; Goodlett, D.R.; Aebersold, R.; Brentnall, T.A. Quantitative proteomic profiling of pancreatic cancer juice. *Proteomics* **2006**, *6*, 3871–3879. [CrossRef] [PubMed]
57. Gygi, S.P.; Rist, B.; Gerber, S.A.; Turecek, F.; Gelb, M.H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999**, *17*, 994. [CrossRef]
58. Dayon, L.; Sanchez, J.-C. *Relative Protein Quantification by MS/MS Using the Tandem Mass Tag Technology, in Quantitative Methods in Proteomics*; Marcus, K., Ed.; Humana Press: Totowa, NJ, USA, 2012; pp. 115–127.
59. Pierce, A.; Unwin, R.D.; Evans, C.A.; Griffiths, S.; Carney, L.; Zhang, L.; Jaworska, E.; Lee, C.-F.; Blinco, D.; Okoniewski, M.J.; et al. Eight-channel iTRAQ enables comparison of the activity of six leukemogenic tyrosine kinases. *Mol. Cell. Proteomics* **2008**, *7*, 853–863. [CrossRef]
60. Casado-Vela, J.; Jose Martinez-Esteso, M.; Rodriguez, E.; Borras, E.; Elortza, F.; Bru-Martinez, R. iTRAQ-based quantitative analysis of protein mixtures with large fold change and dynamic range. *Proteomics* **2010**, *10*, 343–347. [CrossRef]
61. Boehm, A.M.; Pütz, S.; Altenhöfer, D.; Sickmann, A.; Falk, M. Precise protein quantification based on peptide quantification using iTRAQ™. *BMC Bioinf.* **2007**, *8*, 1–18. [CrossRef]
62. Ong, S.E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D.B.; Steen, H.; Pandey, A.; Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002**, *1*, 376–386. [CrossRef] [PubMed]

63. Bicho, C.C.; Alves, F.D.; Chen, Z.A.; Rappsilber, J.; Sawin, K.E. A Genetic Engineering Solution to the "Arginine Conversion Problem" in Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC). *Mol. Cell. Proteomics* **2010**, *9*, 1567–1577. [CrossRef] [PubMed]
64. Ong, S.-E.; Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat. Protoc.* **2007**, *1*, 2650. [CrossRef] [PubMed]
65. Kito, K.; Ito, T. Mass spectrometry-based approaches toward absolute quantitative proteomics. *Curr. Genomics* **2008**, *9*, 263–274. [CrossRef] [PubMed]
66. Kalra, H.; Adda, C.G.; Liem, M.; Ang, C.-S.; Mechler, A.; Simpson, R.J.; Hulett, M.D.; Mathivanan, S. Comparative proteomics evaluation of plasma exosome isolation techniques and assessment of the stability of exosomes in normal human blood plasma. *Proteomics* **2013**, *13*, 3354–3364. [CrossRef] [PubMed]
67. Abdallah, C.; Dumas-Gaudot, E.; Renaut, J.; Sergeant, K. Gel-based and gel-free quantitative proteomics approaches at a glance. *Int. J. Plant Genomics* **2012**, *2012*, 494572. [CrossRef] [PubMed]
68. Anand, S.; Samuel, M.; Ang, C.-S.; Keerthikumar, S.; Mathivanan, S. Label-Based and Label-Free Strategies for Protein Quantitation. *Methods Mol. Biol.* **2017**, *1549*, 31–43.
69. Zhang, Y.; Wen, Z.; Washburn, M.P.; Florens, L. Improving Label-Free Quantitative Proteomics Strategies by Distributing Shared Peptides and Stabilizing Variance. *Anal. Chem.* **2015**, *87*, 4749–4756. [CrossRef]
70. Chiara, G.; Francesca, M.; Alessandra, M.; Luigi, M. Proteomic and Metallomic Strategies for Understanding the Mode of Action of Anticancer Metallodrugs. *Anticancer Agents Med. Chem.* **2010**, *10*, 324–337.
71. Cho, Y.-E.; Singh, T.S.K.; Lee, H.-C.; Moon, P.-G.; Lee, J.-E.; Lee, M.-H.; Choi, E.-C.; Chen, Y.-J.; Kim, S.-H.; Baek, M.-C. In-depth Identification of Pathways Related to Cisplatin-induced Hepatotoxicity through an Integrative Method Based on an Informatics-assisted Label-free Protein Quantitation and Microarray Gene Expression Approach. *Mol. Cell. Proteomics* **2012**, *11*, M111.010884.
72. Du, Z.; Luo, Q.; Yang, L.; Bing, T.; Li, X.; Guo, W.; Wu, K.; Zhao, Y.; Xiong, S.; Shangguan, D.; et al. Mass Spectrometric Proteomics Reveals that Nuclear Protein Positive Cofactor PC4 Selectively Binds to Cross-Linked DNA by a trans-Platinum Anticancer Complex. *J. Am. Chem. Soc.* **2014**, *136*, 2948–2951. [CrossRef]
73. Tušek-Božić, L.; Furlani, A.; Scarcia, V.; De Clercq, E.; Balzarini, J. Spectroscopic and biological properties of palladium(II) complexes of ethyl 2-quinolylmethylphosphonate. *J. Inorg. Biochem.* **1998**, *72*, 201–210. [CrossRef]
74. Kim, W.T.; Kim, J.; Yan, C.; Jeong, P.; Choi, S.Y.; Lee, O.J.; Chae, Y.B.; Yun, S.J.; Lee, S.C.; Kim, W.J. S100A9 and EGFR gene signatures predict disease progression in muscle invasive bladder cancer patients after chemotherapy. *Ann. Oncol.* **2014**, *25*, 974–979. [CrossRef] [PubMed]
75. Chatterjee, S.; Kundu, S.; Bhattacharyya, A.; Hartinger, C.G.; Dyson, P.J. The ruthenium(II)-arene compound RAPTA-C induces apoptosis in EAC cells through mitochondrial and p53-JNK pathways. *J. Biol. Inorg. Chem.* **2008**, *13*, 1149–1155. [CrossRef] [PubMed]

molecules

MDPI

Review

Pseudotrypsin: A Little-Known Trypsin Proteoform

Zdeněk Perutka and Marek Šebela *

Department of Protein Biochemistry and Proteomics, Centre of the Region Haná for Biotechnological and Agricultural Research, Faculty of Science, Palacký University, Šlechtitelů 27, 783 71 Olomouc, Czech Republic; Perutka.Zdenek@seznam.cz

* Correspondence: marek.sebela@upol.cz; Tel.: +420-585-634-927

Academic Editor: Paolo Iadarola

Received: 11 September 2018; Accepted: 9 October 2018; Published: 14 October 2018

Abstract: Trypsin is the protease of choice for protein sample digestion in proteomics. The most typical active forms are the single-chain β-trypsin and the two-chain α-trypsin, which is produced by a limited autolysis of β-trypsin. An additional intra-chain split leads to pseudotrypsin (ψ-trypsin) with three chains interconnected by disulfide bonds, which can be isolated from the autolyzate by ion-exchange chromatography. Based on experimental data with artificial substrates, peptides, and protein standards, ψ-trypsin shows altered kinetic properties, thermodynamic stability and cleavage site preference (and partly also cleavage specificity) compared to the above-mentioned proteoforms. In our laboratory, we have analyzed the performance of bovine ψ-trypsin in the digestion of protein samples with a different complexity. It cleaves predominantly at the characteristic trypsin cleavage sites. However, in a comparison with common tryptic digestion, non-specific cleavages occur more frequently (mostly after the aromatic residues of Tyr and Phe) and more missed cleavages are generated. Because of the preferential cleavages after the basic residues and more developed side specificity, which is not expected to occur for the major trypsin forms (but may appear anyway because of their autolysis), ψ-trypsin produces valuable information, which is complementary in part to data based on a strictly specific trypsin digestion and thus can be unnoticed following common proteomics protocols.

Keywords: autolysis; chain; cleavage; digestion; peptide; proteoform; pseudotrypsin; specificity; trypsin

1. Cleavage Specificity of Trypsin

Trypsin, a serine protease, is commonly used as an important enzymatic reagent in biochemistry and biology. It is almost indispensable especially for the digestion of protein samples to peptides in bottom-up proteomics [1]. Apart from this application, trypsin is a tool in working with cell cultures. During trypsinization, surface adhesion proteins are degraded, which allows adherent cells to be detached from each other and the walls of plastic containers or plates in which they are being cultured. In industry, interestingly, trypsin is applied to hydrolyze allergenic proteins for the production of hypoallergenic milk [2]. In proteomics, sample digests are typically analyzed for protein identification by nanoflow liquid chromatography coupled to tandem mass spectrometry (nLC-MS/MS). MS-based data on peptides are searched against amino acid sequence databases, which benefits from the relative stringent cleavage specificity of trypsin as the search algorithms incorporate the cleavage rule as a filtering criterion. According to a study with complex samples from 2004, the enzyme cleaves peptide bonds in proteins at pH 8.3 exclusively at the carboxyl end of arginine and lysine residues [3]. This is in agreement with the canonical trypsin cleavage rule postulated a long time ago [4] even though it was built on results obtained at that time only with a limited amount of substrates [5]. Thus, it has long been accepted that trypsin does not cleave before proline and its activity is suppressed either if a cysteine appears next to Arg/Lys or the basic residue is N- or C-terminally adjacent to an acidic

residue. This old rule was questioned when a large data set of 14.5 million MS/MS spectra of peptides from *Shewanella oneidensis* was processed to statistically evaluate the cleavage sites [5]. Interestingly, numerous cleavages before proline were found. Their number was even higher than that referring to the cleavages before cysteine.

An average length of tryptic peptides is 14 amino acids. This number has been deduced from an in silico digestion of human proteins in the UniProt database [1]. Because of this reasonable size as well as the presence of a positive charge at the C-terminal Arg or Lys, which enhances the ionization process in the positive ionization mode, tryptic peptides are highly amenable to mass spectrometric measurements. In fact, there are at least two defined positive charges in tryptic peptides (at both N- and C-termini), which is favorable for a good fragmentation in MS/MS analyses [1,2].

2. Nonspecific and Missed Cleavage Sites

In addition to the cleavages after Arg or Lys, proteomics studies have often reported the formation of semitryptic and nonspecific peptides during the digestion process involving trypsin [1,6]. The semitryptic cleavage assumes that one of the cleavage sites is tryptic, but the other site may be at any residue. A minor chymotrypsin or chymotrypsin-like activity yields nonspecific cleavages C-terminal to phenylalanine, tyrosine, tryptophan, or leucine residues [7]. This can result either from the presence of a chymotrypsin contamination, which is variable in trypsin preparations supplied by different vendors [1], or pseudotrypsin (ψ-trypsin), a product of trypsin autolysis, which possesses such an activity in addition to the characteristic trypsin properties [8]. The presence of nonspecific peptides in tryptic digests (excluding C-terminal peptides) is also elucidated by a secondary non-enzymatic cleavage between Asp and Pro residues yielding peptides with an N-terminal proline. This is because of the lability of the respective bond, which is easily hydrolyzed in solution as well as broken in the gas phase [3].

When the protein substrate is not cleaved to a completion, missed cleavages occur, which make the assignment of experimental data to amino acid sequence databases less specific and straightforward. Missed cleavage sites were investigated for example by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) using human cell line or *Mycobacterium* proteins separated by two-dimensional gel electrophoresis and digested in-gel by a porcine trypsin [9]. The analysis showed that about 90 % of the detected peptides with missed cleavage sites could be attributed to the following sequence motifs: (1) Arg or Lys with a neighboring proline at the C-terminal side, (2) two successive basic residues (Arg-Arg, Arg-Lys, Lys-Arg, Lys-Lys), and (3) Arg or Lys with an aspartic acid or glutamic acid residue at either N-terminal or C-terminal side. When processing peptide MS- and MS/MS-based data by database searches (during which theoretical peptide sequences are generated according to the selected input cleavage rules), the user may, among others, adjust a maximum number of missed cleavage sites. Usually, a setting of 0, 1, or 2 missed cleavages for peptides is recommended. The presence of missed cleavages in tryptic peptides represents a challenge in quantitative MS-based proteomics, which uses peptides as surrogates for their parent proteins [10]. Under optimal conditions, peptides should be stoichiometric with the parent protein to enable accurate quantitation. However, if a protein is digested into multiple overlapping peptides, the specific signal is attenuated and in consequence the quantification becomes underestimated.

There are also side reactions of trypsin known, for example its transpeptidase activity, which yields the addition of a single amino acid (Arg or Lys) to peptides in the reaction mixture. There are also dipeptides added or two longer peptides are combined together. The additions have been described for both N- and C-termini of peptides [7]. However, the incidence of the modified peptides is much lower compared to their unmodified counterparts.

3. Trypsin Forms

Wilhelm Kühne, who discovered trypsin and coined its name in 1876, noticed that the enzyme was produced as an inactive zymogen (trypsinogen) in pancreatic cells [11]. In the 1930s, Moses Kunitz and

John H. Northrop elaborated procedures to isolate both trypsinogen and trypsin in crystalline forms, which largely contributed to the development of enzymology [12]. Trypsinogen is stable at acidic pH (2–4). In a neutral or alkaline solution, it is activated by a limited proteolysis catalyzed by either duodenal enteropeptidase (enterokinase) or trypsin itself [4]. Thus, in the latter case, an autocatalytic process occurs. The complete amino acid sequence of bovine trypsinogen was deduced from peptide sequencing experiments by several independent groups in the 1960s [13,14] with later corrections at ambiguous positions (some Asn/Asp and Gln/Glu were not distinguished in early studies because of limitations of the sequencing methods used at that time). In addition to the dominant cationic trypsinogen, a minor anionic trypsinogen is produced in bovine pancreas [15]. The UniProt accession numbers are P00760 (TRY1_BOVIN) and Q29463 (TRY2_BOVIN), respectively; the corresponding mature trypsin sequences share 72% identity.

The cationic trypsinogen sequence spans the length of 229 amino acids (Figure 1). Altogether, there are six disulfide bonds in the molecule connecting the following cysteine residues: 13 and 143, 31 and 47, 115 and 216, 122 and 189, 154 and 168 plus 179 and 203 [14]. By the removal of the N-terminal hexapeptide VDDDDK, the single chain β-trypsin is formed as the predominant product of trypsinogen activation. Trypsin autolysis with a cleavage between Lys-131 and Ser-132 (numbered according to the trypsinogen convention) produces α-trypsin, which has two chains connected by disulfide bonds) [4]. A further cleavage at the bond Lys-176−Asp-177 yields ψ-trypsin with three discrete chains [16]. Another known active trypsin forms include for example the two chain δ-trypsin [17–19] and γ-trypsin [20] with chain splits at the bonds Arg-105−Val-106 and Lys-155−Ser-156, respectively. On the contrary, a cleavage between Lys-49 and Ser-50 has been shown to inactivate the enzyme [18]. More degraded autolysis products are considered inactive [4]. Bovine trypsin is characterized by a molar absorption coefficient ε_{280} of 40,000 mol·L^{-1}·cm^{-1} [21] resulting from the presence of 4 Trp, 10 Tyr, and 12 Cys residues (oxidized to disulfides in the protein) in the sequence. A certain degree of variability in these numbers can be found for trypsin enzymes from different mammalian or other sources of origin [4]. All mentioned variants are often called "trypsin isoforms" in the literature but this is not satisfactory. According to the International Union of Pure and Applied Chemistry (IUPAC), the term isoform refers only to genetic differences and not to a variation at the protein level. Hence, the real isoforms are e.g., the cationic and anionic trypsin. To solve this terminological confusion, the term "proteoform" has recently been introduced, which covers all molecular forms encoded by a single gene, including changes due to genetic variations, alternatively spliced transcripts and posttranslational modifications [22].

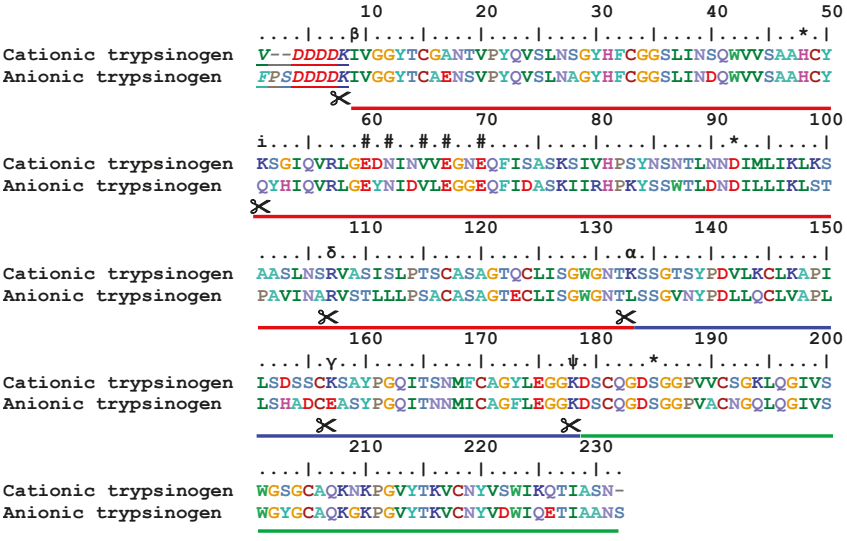


Figure 1. An alignment of the cationic and anionic bovine trypsinogen sequences. Italic letters and a thin underlining at the beginning of the sequences highlight the activation peptides, which are cleaved off during the trypsinogen conversion to mature trypsin. In the ruler line, asterisks (*) indicate active-site residues (the charge relay system, also catalytic triad) and hash (#) symbols mark calcium-binding residues. Greek alphabet letters indicate the cleavage sites in the active single-chain β-trypsin (223 residues), which yield the other trypsin proteoforms (α, γ, δ, and ψ) upon autolysis. The small-case letter 'i' denotes the cleavage site leading to an inactive trypsin variant. All above mentioned cleavage sites are additionally highlighted by a symbol of scissors. Thick red, blue, and green lines indicate individual ψ-trypsin chains. The alignment was made using BioEdit 7.2.5 with trypsinogen sequences obtained from the UniProt database. The accession numbers are P00760 (TRY1_BOVIN, cationic) and Q29463 (TRY2_BOVIN, anionic). The numbering refers to anionic trypsinogen. For cationic trypsinogen, it is shifted towards the canonical numbering by two units up because of the alignment.

4. Pseudotrypsin Purification from the Autolyzate

In the original report from 1969, ψ-trypsin was purified from an autolyzate of bovine trypsin using isocratic ion-exchange chromatography [16]. The autodigestion proceeded at pH 8.0 and 25 °C in the presence of calcium ions for up to 6.5 h. Then N^a-*p*-tosyl-L-lysine-chloromethyl ketone (TLCK), which is an irreversible trypsin inhibitor, was added at pH 7.0 to abolish any detectable activity. After a removal of low-molecular weight compounds from the protein by gel permeation chromatography (1 mM HCl as a mobile phase) and the subsequent lyophilization, the TLCK-treated autolyzate was separated on an Sulfoethyl (SE)-Sephadex C-50 column (1 × 48 cm) in 100 mM Tris-HCl, pH 7.1, containing 20 mM $CaCl_2$. This procedure had previously been developed for a reliable resolving α- and β-trypsin [23]. In this arrangement, ψ-trypsin was eluted prior to the elution of α- and finally β-trypsin. It was found not to contain an alkylation resulting from the TLCK treatment, which was elucidated by its decreased activity towards trypsin substrates [16]. In a similar way, ψ-trypsin was purified on SE-Sephadex C-50 in several other studies. Some differences appeared in the column length, e.g., 1.5 × 150 cm [24], or the authors optimized the composition of the mobile phase by varying sample loading, pH, flow rate and concentration of NaCl [25].

In our laboratory, we used a HEMA-BIO 1000 SB column (0.75 × 25 cm) in a medium-pressure protein liquid chromatography to separate trypsin autolyzate components [8]. The flow rate was adjusted to 2 mL·min^{-1} and the whole time window to resolve isocratically ψ-, α-, and β-trypsin (in

the given order of elution times) at pH 7.1 was 45 min long. We have recently replaced the HEMA-BIO 1000 SB column by a Uno S12 column (15 × 68 mm), which allowed reducing the separation time, at the expense of resolution, but still yielded a pure ψ-trypsin (Figure 2). In this case, however, the use of a gradient elution was necessary—the buffer B contained 1 M NaCl (Perutka et al., unpublished results). In the 1970s, a French group, which studied kinetic properties of ψ-trypsin at that time, developed a convenient purification method based on affinity chromatography [26,27]. The yield was 15–20 % of the initial fully active trypsin. To prepare the affinity column, a trypsin inhibitor from egg-white chicken ovomucoid was attached to aminoethyl-cellulose using glutaraldehyde as a coupling reagent. The equilibration buffer for separation runs was 0.1 M Tris-HCl, containing 50 mM $CaCl_2$, pH 7.1. ψ-Trypsin appeared already in the flow-through fraction (just after the void volume). Thus, it was necessary to include a size-exclusion chromatographic step prior to the affinity separation in order to remove low-molecular-weight contaminants such as peptides. In contrast, the proteoforms α and β remained bound and could be eluted (as an unresolved mixture) by applying an acidic elution buffer of pH 2.3 [27]. This low affinity of ψ-trypsin is in agreement with results on the formation of its complex with pancreatic trypsin inhibitor, where the corresponding dissociation constant was increased by five orders of magnitude compared to that for a mixture of α- and β-trypsin [28].

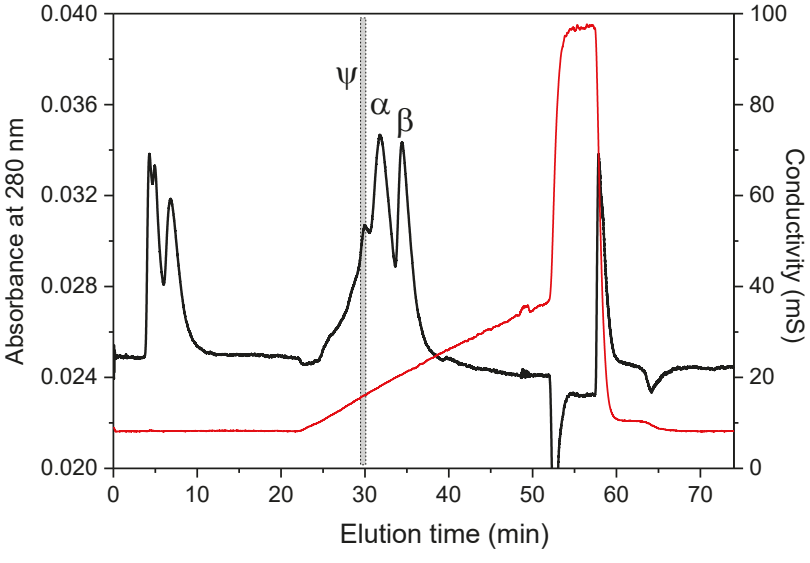


Figure 2. Ion-exchange chromatography of trypsin autolyzate. The depicted separation was performed on a Uno S12 column (15 × 68 mm; Bio-Rad, Hercules, CA, USA) using a linear gradient of NaCl (0–1 M) in 100 mM Tris-HCl, pH 7.1, containing 10 mM $CaCl_2$. The gray box indicates a time window, in which the respective ψ-trypsin fraction was collected. The black line shows absorbance at 280 nm, the red line refers to conductivity of the eluate.

5. Molecular Properties and Structure of Trypsin and Pseudotrypsin

Accurate experimental molecular mass values of bovine ψ-trypsin were first determined by electrospray ionization (ESI)-MS [29,30]. For β-trypsin, α-trypsin, and ψ-trypsin, the following average numbers (relative monoisotopic molecular masses) were obtained in the given order: 23296, 23310, and 23325 [29] or 23294, 23312 and 23328 [30]. These numbers are in accordance with sequence-based calculated values of 23293, 23311, and 23329, respectively [30], which reflect a relative mass difference of 18 resulting from each consecutive autolytic cleavage. For ψ-trypsin, the most recent data are 23332 ± 4 (MALDI-MS) and 23330 ± 0.1 (ESI-MS) [8]. Thus, based on an accurate molecular mass

determination, the purity of ψ-trypsin preparations can easily be evaluated, also to rule out the presence of a chymotrypsin contamination coming from the original trypsin material.

Isoelectric points of proteins can be estimated from their amino acid sequences for example using the software tool ProtParam (https://web.expasy.org/cgi-bin/protparam/). A theoretical p*I* value of the cationic bovine trypsin (Uniprot accession number P00760, positions 24–246) is 8.69. Similarly, for porcine trypsin (Uniprot accession number P00761, positions 9–231), the result is 8.26. Interestingly, higher experimental values of 10.0/10.5 [31,32] were published for the bovine enzyme and 10.2/10.8 for the porcine enzyme [33,34]. For ψ-trypsin, no experimental data on p*I* are available to our knowledge, but similar values could be expected. The anionic bovine trypsin (Uniprot accession number Q29463, positions 24–247) has a theoretical p*I* value of 4.90, which agrees well with an experimental result [35].

Chymotrypsin and trypsin were among first proteins with experimentally determined spatial structures. The crystal structure of bovine chymotrypsin appeared already in 1967 [36] and was refined later on [37]. At that time, a similarity in the three-dimensional folding of trypsin and chymotrypsin could be assumed because of the amino acid sequence homology and matching positions of disulfide bonds. The crystal structures of bovine β-trypsin and trypsinogen were solved in the 1970s [38–40]. Figure 3 shows a view of the trypsin structure (PDB accession code 1AQ7 [41]) visualized using PyMol 1.3. Trypsin is a globular protein. Its overall fold comprises two six-stranded Greek-key β-barrels [42]. The active site with the catalytic triad of amino acids is located between the two barrels. His-57 and Asp-102 belong to the N-terminal barrel whereas Ser-195 originates from the C-terminal barrel (this numbering is according to the chymotrypsinogen convention, see in Walsh and Neurath) [13]. Helices represent only minor secondary structure components, for example at the C-terminus. The enzyme contains a calcium ion, which is important for activity. Its coordination chemistry involves several residues from the calcium-binding loop [38]. The calcium ion interacts with the side-chain oxygens (in the trypsinogen numbering) of Glu-58, Glu-65 (this one via a coordinated water molecule), and Glu-68 plus the carbonyl oxygens of Asn-60 and Val-63. No crystal structure of ψ-trypsin has been solved up to now. As this proteoform contains two chain splits (between: 1. Lys-131 and Ser-132, 2. Lys 176 and Asp-177; according to the trypsinogen numbering), the whole molecule is loosened somehow in comparison to that of β-trypsin. The bond Lys-176−Asp-177 is located close to the anionic binding site (i.e., specificity site: Asp-177, which is the position 189 when expressed in the chymotrypsinogen numbering). Upon the autolytic splitting, the binding site arrangement is disconnected. In consequence, the affinity for polypeptide trypsin substrates is lowered and the cleavage specificity broadened [8,24]. This structural alteration does not prevent from binding of a pancreatic trypsin inhibitor; only the dissociation constant of the enzyme-inhibitor complex is increased [24,28]. ψ-Trypsin still keeps a certain level of specificity, which is based on hydrophobic interactions, as confirmed using synthetic ester substrates [27].

After trypsinogen activation, the new N-terminal residue (Ile-16, in chymotrypsinogen numbering) inserts into a cleft, where it establishes an ion pair (via α-amino group) with Asp-194 next to the catalytic serine. This results in a conformational rearrangement. The amino group of Gly-193 moves into a position, which completes the oxyanion hole at the active site [2]. The hole is formed by the trypsin amide hydrogens of Gly-193 and Ser-195 in favor of stabilization of the developing negative charge on the carbonyl oxygen atom of the cleaved substrates.

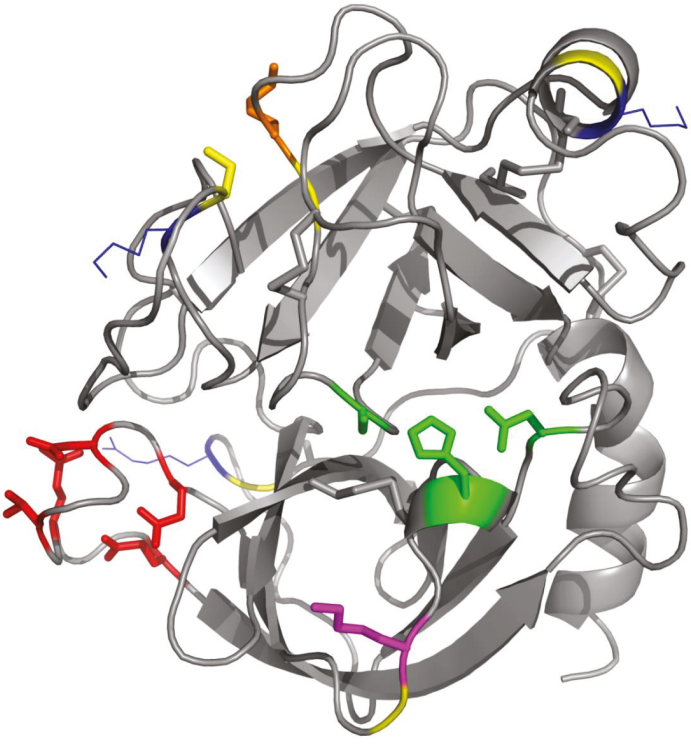


Figure 3. The crystal structure of bovine β-trypsin. This overall view was made in PyMOL 1.3 using a pdb formatted file (accession number 1AQ7 [41]) downloaded from the PDB database (http://www.rcsb.org/pdb). Trypsin fold comprises two β-barrels. The active site with the catalytic triad of amino acids is located between them (in the center of this figure) and highlighted in green. Red color on the bottom left shows the calcium-binding residues. The KD bond, which is cleaved to generate ψ-trypsin is expressed as an orange side chain (Lys-176) at a yellow backbone segment (Asp-177). Blue-line side chains at other yellow segments indicate the presence of the basic residues, where the β-trypsin polypeptide chain is cleaved to produce α-trypsin (Lys-131, above the calcium site and not far away from Lys-176 in this projection), γ-trypsin (Lys-155, top right, in the helix), and δ-trypsin (Arg-105, behind the calcium site). Finally, magenta color at the bottom indicates Lys-49, which is disconnected from Ser-50 to yield the inactive autolytic form.

6. Pseudotrypsin Activity with Artificial Substrates (Enzyme Kinetics)

Compared to α-trypsin and β-trypsin, the overall structural change resulting from the additional intrachain split in ψ-trypsin yields differences in the activity and specificity. This was evaluated already in the 1960s and 1970s by measuring kinetic parameters for low-molecular-weight substrates (Table 1). Smith and Shaw [16] recognized during the chromatographic purification that ψ-trypsin did not show any measurable activity with N^{α}-benzoyl-D,L-arginine-4-nitroanilide (Bz-Arg-pNA) as a substrate under experimental conditions optimized for α-trypsin (i.e., no amidase activity). Nevertheless, they could demonstrate a hydrolytic activity of ψ-trypsin by detecting a stoichiometric incorporation of [^{14}C] diisopropyl fluorophosphate at the active site, which is typical for serine proteases or esterases in general. However, the rate of ^{14}C incorporation was very slow, which indicated a decreased reactivity of ψ-trypsin compared to α-trypsin. Similarly, a slow conversion of the active site titrant *p*-nitrophenyl-*p*′-guanidinobenzoate (NPGB) was observed. Comparative kinetic data measured with several artificial ester substrates are shown in Table 1. For example, the affinity of ψ-trypsin for the

cationic N^{α}-benzoyl-L-arginine ethyl ester (Bz-Arg-OEt) represented by the determined Michaelis constant value is lower by three orders of magnitude than that of α-trypsin [16]. Such a big difference in affinity was observed also for carboxybenzyl-L-lysine esters or N^{α}-*p*-tosyl-L-arginine methyl ester [27]. Not only the affinity, but also the activity (catalytic constant) is often largely different. A typical chymotrypsin substrate N^{α}-acetyl-L-tyrosine ethyl ester (Ac-Tyr-OEt) is hydrolyzed by α-trypsin almost 3000 times faster than by ψ-trypsin (Table 1). At the same time, the respective K_m values are comparable as the binding of this neutral substrate is not affected so much by the disconnection of Asp-177 (the specificity site) in ψ-trypsin [16]. As a consequence of the reduced affinity and activity towards cationic substrates, the efficiency constant values k_{cat}/K_m are decreased by many orders of magnitude (up to 6) for ψ-trypsin [16,27].

Table 1. Kinetic data for bovine α-trypsin and ψ-trypsin with artificial substrates. This table is adapted from Smith and Shaw [16] and completed by results with non-fractionated trypsin [43]. All experiments were performed at pH 8.0 and 25 °C.

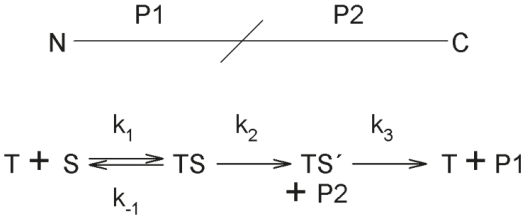
Enzyme	Substrate [a]	k_{cat} (s^{-1})	K_m (mol·L^{-1})	K_i (mol·L^{-1})
α-Trypsin	Bz-Arg-OEt (0.05 M $CaCl_2$)	24	2.5×10^{-6}	-
ψ-Trypsin	Bz-Arg-OEt (0.001 M $CaCl_2$)	0.19	1.1×10^{-2}	-
ψ-Trypsin	Bz-Arg-OEt (0.05 M $CaCl_2$, benzamidine)	-	-	3.7×10^{-2}
Trypsin (non-fractionated)	Bz-Arg-OEt (0.025 M $CaCl_2$)	14.6	4.3×10^{-6}	-
α-Trypsin	Ac-Tyr-OEt (0,05 M $CaCl_2$, 10% 2-propanol)	57	6.2×10^{-2}	-
ψ-Trypsin	Ac-Tyr-OEt (0.001 M $CaCl_2$, 10% 2-propanol)	0.02	3.7×10^{-2}	-
Trypsin (non-fractionated)	Ac-Tyr-OEt (0.05 M $CaCl_2$, 5% dioxane)	14.5	4.2×10^{-2}	-
Trypsin (non-fractionated)	Bz-Arg-pNA (benzamidine) [b]	-	-	1.8×10^{-5}

[a] Further information on the composition of the reaction mixture is provided in parentheses; [b] This reaction was performed at pH 8.15 and 15 °C.

Benzamidine has been shown a potent competitive inhibitor of trypsin. It is approximately of the same size as the side chain of lysine and arginine and contains both a positively charged group and hydrophobic moiety in its structure. The K_i value of benzamidine for the reaction of trypsin with Bz-Arg-pNA is 1.8×10^{-5} mol·L^{-1} [4]. In contrast, the binding of the inhibitor to ψ-trypsin is characterized by an increased K_i value 3.7×10^{-2} mol·L^{-1}, which is in agreement with the K_m value for the neutral (and thus non-specific) substrate Ac-Tyr-OEt (Table 1). The irreversible trypsin inhibitor TLCK does not inactivate ψ-trypsin at all [16]. Further evidence of the reduced ψ-trypsin affinity to positively charged ligands was observed with a basic pancreatic trypsin inhibitor. The second-order rate constant of the association is decreased by a factor of 16 when ψ-trypsin is used instead of α- and β-trypsin and the dissociation constant is increased by a factor of 1.5×10^5 from 6×10^{-14} mol·L^{-1} (a quasi-irreversible binding in the case of trypsin) to 9×10^{-9} mol·L^{-1}. This value is similar to that of chymotrypsin, which also associates with this inhibitor but lacks the trypsin specificity site [28]. When the disulfide bond Cys-179—Cys-203 in trypsin is selectively reduced and the emerged cysteines subsequently carboxymethylated, the resulting enzyme derivative still binds the inhibitor efficiently. Interestingly, the observed kinetic parameters of the association and dissociation correspond to those determined for ψ-trypsin [28].

The reaction mechanism of trypsin is illustrated in Scheme 1 (based on that in [27]) Acylation and deacylation rate constants (k_2 and k_3, respectively) were measured with the active site titrant NPGB and pure bovine α-, β-, and ψ-trypsin preparations [27]. Whereas the acylation rate was found to be 1000 slower for ψ-trypsin than for the other proteoforms at optimum pH, the deacylation rates were rather comparable. A similar difference in the acylation rate was observed with *p*-acetoxyphenylguanidine *p*-toluenesulfonate. This compound belongs to "inverse substrates" as it contains the specific cationic center within the leaving group instead of the acyl moiety [44]. Measurements with different

carboxybenzyl-L-lysine esters (cationic i.e., specific trypsin substrates) indicated that acylation is largely the rate-limiting step for ψ-trypsin contrary to deacetylation as it appears in the case of α-/β-trypsin [27,45]. For neutral (nonspecific) substrates, acylation is rate limiting also for the major proteoforms. The k_{cat} values for ψ-trypsin were found sensitive to the substrate leaving group in contrast to those observed for α-trypsin, which shows that the catalysis by this form is not completely nonspecific and hydrophobic binding subsites are preserved.



Scheme 1. The reaction mechanism of trypsin. A protein substrate (S) with the indicated N- and C-termini and the cleavage site is cleaved by trypsin (T) into two large peptides or polypeptides P1 and P2. TS stands for a Michaelis trypsin-substrate complex, TS′ represent an acyl-enzyme intermediate. The reaction rate constants are as follows: k_1—enzyme-substrate association rate constant; k_{-1}—enzyme-substrate dissociation rate constant; k_2—enzyme-substrate acylation rate constant; k_3—enzyme-substrate deacylation rate constant; $k_{cat} = k_2k_3/(k_2 + k_3)$. This scheme has been adapted from the reference [27].

7. Cleavage Specificity towards Peptides and Proteins

Both major forms of trypsin (α and β) prefer Arg-X sites to Lys-X sites during the hydrolysis of synthetic as well as polypeptide substrates at an optimal pH of 8–9. The preference ratio is even more pronounced at higher pH values (>pH 10) because the ε-amino group of lysine is largely discharged under these conditions and becomes less attractive [4]. The cleavage specificity of ψ-trypsin towards peptide bonds was first analyzed with two peptide substrates: a mixture of bovine and porcine glucagon and a heptapeptide fragment of the B chain of insulin with a sequence of GFFYTPK [24]. The heptapeptide was selected because of the content of aromatic residues to evaluate whether ψ-trypsin shows a preference similar to that of chymotrypsin. The glucagon sequence contains three canonical trypsin cleavage sites (Lys-12, Arg-17, and Arg-18). In a parallel work, the same group demonstrated that pure α- and β-trypsin preparations did not show any effect on the insulin-derived heptapeptide and carboxy sites of aromatic amino acid residues in glucagon [46]. Conversely to the effect of α-chymotrypsin used as a 1% model contamination in a commercial trypsin preparation (control digest), no effect of ψ-trypsin on the heptapeptide was observed. The digestion of glucagon was performed in 0.1 ammonium carbonate containing 20 mM $CaCl_2$ at pH 8.0 and ψ-trypsin generated the same fragments as it had been observed in the parallel study with α- and β-trypsin. However, the yield of these fragments was lower than those produced by the major forms, indicating its decreased affinity to polypeptides. However, ψ-trypsin showed an additional ability to cleave bonds adjacent to the aromatic amino acids Phe and Trp [24].

Dyčka et al. [8] performed overnight in-gel digestions of six standard proteins with monomer molecular masses of 12–95 kDa (cytochrome *c*, lysozyme, myoglobin, glucose oxidase, serum albumin, and glycogen phosphorylase) using ψ-trypsin, non-fractionated trypsin (treated by *N*-*p*-tosyl-L-phenylalanine chloromethyl ketone—TPCK—to inactivate a possible chymotrypsin contamination) and chymotrypsin to obtain a more complex view of the cleavage specificity. The numbers of sequence-matched peptides with the respective sequence coverage values from nLC-MALDI-MS/MS were similar for ψ-trypsin and trypsin. A majority of the registered ψ-trypsin cleavage sites (77 %) were produced by its action upon the C-termini of Arg and Lys residues (in the case of trypsin, it was 86 %). Additional cleavages appeared particularly after Phe and Tyr residues,

which confirmed the previous data obtained with glucagon [24]. Interestingly, ψ-trypsin provided 1.5-fold higher number of peptides containing missed cleavage (Arg and Lys) sites, which probably results from the lower substrate binding ability of this proteoform [8].

8. The Use of Pseudotrypsin for Protein Identification in Proteomics

At the present time, ψ-trypsin is not commonly used in biochemistry and proteomics as it is not commercially available and its purification takes a few days with a low yield of pure protein at the end. From 200 mg of bovine trypsin as a starting material [8], only milligrams of the final product could be obtained. The applicability of ψ-trypsin has a potential in protein identification experiments involving ESI- or MALDI-MS/MS as it generates more peptides with missed cleavage sites than native trypsin and also nonspecific peptides terminated mainly by Phe and Tyr residues (resembling partially the chymotrypsin mode of action). In consequence, higher sequence coverage values and increased number of matched peptides can be achieved. Interestingly, pseudotrypsin has been hypothesized to cause the cleavage of a Cys–Gly bond in NDRG1 protein [47]. Another possibility of application resides in studying posttranslational modifications. Such a modification may under a common routine interfere with trypsin digestion, for example when a phosphorylation occurs close to an arginine or lysine, and would then require selecting of another protease [48]. On the other hand, for its broader cleavage specificity, ψ-trypsin is not recommendable for mass spectrometry-based protein quantification experiments [49] because of the possible distribution of the same predicted canonical cleavage site in more peptides (multiple cleavage products due to missed cleavages or additional cleavage sites).

A comparative in-gel digestion of a gel fraction of rat urine proteins resulted in 22 identifications after the use of ψ-trypsin. The same number was reached with a commercial non-fractionated trypsin, but only 17 were identified in both cases. Hence a simple combination of the two digestions provided about 20 % more identification [8]. The total number of the matched peptides was 233 and 199, respectively. The numbers of the cleaved Arg and Lys sites were comparable (around 130); in the case of Phe and Tyr site cleavages, two times more peptides were produced by ψ-trypsin. Recently, this proteoform was applied to analyze nuclear proteins after a DNase treatment of barley nuclei and sodium dodecyl sulfate polyacrylamide gel electrophoresis of the released protein material (Perutka et al., unpublished results). Peptides from the digests were analyzed by MALDI and ESI MS/MS and the results compared with parallel tryptic digestions. The identified nonspecific peptides in the ψ-tryptic digests represented 15–20 % of the total peptide number (compared to 7 % for a standard trypsin). In agreement with previous reports [8,24], peptides with C-terminal Tyr and Phe residues (and also Leu) were found in a significant percentage representation but were only minor to those resulting from the characteristic trypsin cleavage after the basic Arg and Lys residues (Figure 4). The higher number of Arg-ending over Lys-ending peptides in the MALDI-TOF/TOF MS/MS results (but not in the ESI-based results; not shown) probably reflected the fact that Arg-peptides provide more intense signals in MALDI-TOF MS [50], and thus they are preferably selected for the subsequent data-dependent fragmentation. Database searches allowed identifying novel proteins which had not previously been recognized based on standard tryptic peptides and deposited in the database UNcleProt [51]. They accounted for around 10 % of all identifications. The digestion performance of ψ-trypsin was compared relatively to that of a commercial trypsin using MALDI-TOF MS-based quantification of bovine serum albumin peptides (Perutka et al., unpublished results). Tryptic digestion was performed in a buffer made of $H_2{}^{18}O$ [52]. During proteolysis, labeled standards were generated by incorporating ^{18}O isotope into the carbonyl group of the nascent peptides. The ψ-tryptic digest was made in a buffered $H_2{}^{16}O$ and then mixed equivolumetrically with the labeled standards. The ratios of non-labeled versus labeled peptides were calculated from the areas of isotopically resolved peaks in MALDI-TOF MS spectra. As a result, the observed overall overnight digestion performance of ψ-trypsin was found lower by around 20 %.

P1 amino acid	G	A	P	V	L	I	M	F	W	Y	S	T	C	N	Q	K	H	R	D	E
side chain	Nonpolar							Aromatic			Polar					Basic			Acidic	
Ψ-trypsin	2.0%	0.9%	0.3%	0.5%	2.2%	0.1%	0.9%	2.1%	0.4%	4.2%	0.7%	0.6%	0.0%	1.1%	0.4%	32.7%	0.4%	49.2%	1.2%	0.1%
α/β trypsin	0.8%	0.2%	0.0%	0.2%	1.1%	0.0%	0.0%	2.4%	0.0%	2.1%	0.0%	0.1%	0.0%	0.7%	0.0%	31.3%	0.5%	60.5%	0.1%	0.0%

Figure 4. C-terminal amino acids (P1 cleavage site) of peptides in ψ-tryptic and tryptic digests. This table summarizes results obtained after in-gel digestions of barley nuclear proteins. The percentage values were averaged from data for almost 40 distinct protein fraction samples. MALDI TOF/TOF MS/MS allowed to identify 1199 and 1238 peptides, respectively. PEAKS Studio 8 software (Bioinformatics Solutions, Waterloo, ON, Canada) was used to process the MGF-formatted files by searching against a *Hordeum vulgare* protein sequence database downloaded from the National Center for Biotechnology Information, USA, at 1% peptide false discovery rate (Perutka, Z. et al.; unpublished results). Peptide sequences containing C-termini of the identified proteins were not included in the calculation.

9. Concluding Remarks

Early studies identified ψ-trypsin as a proteoform resulting from trypsin autolysis. Amino acid analyses of its polypeptide chains revealed the existence of an additional split between Lys-176 and Asp-177 compared to the primary structure of α-trypsin. The enzyme can be obtained from a trypsin autolyzate by ion exchange chromatography. Purification protocols that are available utilize one of the characteristic features of ψ-trypsin: in contrast to α- or β-trypsin, it is not modified by the trypsin inactivator TLCK and shows a minimum retention on cation exchangers such as Sulfoethyl-Sephadex.

Enzyme kinetics studies with synthetic amino acid esters demonstrated a largely decreased affinity and activity towards cationic substrates. This has been elucidated by the presence of the characteristic chain split, which disconnects the specificity site (Asp-177) from the active site (Ser-183, His-46, Asp-90; all indicated according to the trypsinogen numbering convention). The reaction mechanism of ψ-trypsin with specific substrates differs from that of α- or β-trypsin in the rate limiting step. No crystal structure of ψ-trypsin has been solved yet, but the anticipated existence of the modified ('neutral') active site and the ability to hydrolyze tyrosine-derived ester substrates lead to cleavages characteristic of chymotrypsin. So far, only a few studies have focused on the digestion of peptides and proteins by ψ-trypsin. These studies have confirmed its preferential action on Arg and Lys residues but also at Phe and Tyr residues (which is typical for chymotrypsin) as minor cleavage sites. The produced peptides contain frequent missed cleavages because of the absent specificity and reduced affinity towards the cationic sites. However, overnight digestions of protein samples provide enough peptides for identification experiments involving nLC-MALDI or nLC-ESI MS/MS. Subsequently, combining data from tryptic and ψ-tryptic digestions has been shown to be advantageous in order to increase the number of matched peptides and sequence coverage values. Furthermore, peptides with missed cleavage sites could be beneficial for studying posttranslational modifications of proteins.

ψ-Trypsin should not be used as an equivalent substitute of trypsin or chymotrypsin and, in fact, there is no need to do it. However, as it makes preferential cleavages after basic Arg and Lys residues and has a more developed side specificity for aromatic and Leu residues, which is not expected to occur for pure major trypsin forms α and β (but may appear anyway because of their autolysis), it produces valuable complementary information. The unavailability of any commercial material represents a big obstacle for the application of ψ-trypsin in common proteomics research. On the other hand, a single-step affinity chromatographic method has already been introduced [27], which could make the preparation of the enzyme easier (after a revision and transformation with the use of modern chromatographic materials).

Author Contributions: Marek Šebela came with the idea, wrote the Sections 1–5 and finally edited the whole text, Zdeněk Perutka wrote the Sections 6–9 and provided some of his unpublished results.

Funding: This work was supported by grant no. LO1204 (National Program of Sustainability I) from the Ministry of Education, Youth and Sports, Czech Republic.

Conflicts of Interest: The authors declare that there is no conflict of interests.

Abbreviations

Ac-Tyr-OEt	N^{α}-acetyl-L-tyrosine ethyl ester
Bz-Arg-OEt	N^{α}-benzoyl-L-arginine ethyl ester
Bz-Arg-pNA	N^{α}-benzoyl-D,L-arginine-4-nitroanilide
ESI	electrospray ionization
IUPAC	International Union of Pure and Applied Chemistry
MALDI	matrix-assisted laser desorption/ionization
MS	mass spectrometry
MS/MS	tandem mass spectrometry
NPGB	*p*-nitrophenyl-*p*′-guanidinobenzoate
nLC	nanoflow liquid chromatography
PDB	Protein Data Bank
SE-Sephadex	Sulfoethyl-Sephadex
TLCK	N^{α}-*p*-tosyl-L-lysine-chloromethyl ketone
TOF	time-of-flight
TPCK	*N*-*p*-tosyl-L-phenylalanine chloromethyl ketone

References

1. Burkhart, J.M.; Schumbrutzki, C.; Wortelkamp, S.; Sickman, A.; Zahedi, R.P. Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. *J. Proteom.* **2012**, *75*, 1454–1462. [CrossRef] [PubMed]
2. Vandermarliere, E.; Mueller, M.; Martens, L. Getting intimate with trypsin. The leading protease in proteomics. *Mass Spectrom. Rev.* **2013**, *32*, 453–465. [CrossRef] [PubMed]
3. Olsen, J.V.; Ong, S.E.; Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteom.* **2004**, *3*, 608–614. [CrossRef] [PubMed]
4. Keil, B. Trypsin. *Enzymes* **1971**, *3*, 249–275. [CrossRef]
5. Rodriguez, J.; Gupta, N.; Smith, R.D.; Pevzner, P.A. Does trypsin cut before proline? *J. Proteom. Res.* **2008**, *7*, 300–305. [CrossRef] [PubMed]
6. Picotti, P.; Aebersold, R.; Domon, B. The implications of proteolytic background for shotgun proteomics. *Mol. Cell. Proteom.* **2007**, *6*, 1589–1598. [CrossRef] [PubMed]
7. Schaefer, H.; Chamrad, D.C.; Marcus, K.; Reidegeld, K.A.; Blüggel, M.; Meyer, H.E. Tryptic transpeptidation products observed in proteome analysis by liquid chromatography-tandem mass spectrometry. *Proteomics* **2005**, *5*, 846–852. [CrossRef] [PubMed]
8. Dyčka, F.; Franc, V.; Fryčák, P.; Raus, M.; Řehulka, P.; Lenobel, R.; Allmaier, G.; Marchetti-Deschmann, M.; Šebela, M. Evaluation of pseudotrypsin cleavage specificity towards proteins by MALDI-TOF mass spectrometry. *Protein Pept. Lett.* **2015**, *22*, 1123–1132. [CrossRef] [PubMed]
9. Thiede, B.; Lamer, S.; Mattow, J.; Siejak, F.; Dimmler, C.; Rudel, T.; Jungblut, P.R. Analysis of missed cleavage sites, tryptophan oxidation and N-terminal pyroglutamylation after in-gel tryptic digestion. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 496–502. [CrossRef]
10. Lawless, C.; Hubbard, S.J. Prediction of missed proteolytic cleavages for the selection of surrogate peptides for quantitative proteomics. *OMICS* **2012**, *16*, 449–456. [CrossRef] [PubMed]
11. Kühne, W. *Über das Trypsin (Enzym des Pankreas). Verhandlungen des Naturhistorisch-medizinischen Vereins zu Heidelberg*; Carl Winter's Universitätsbuchhandlung: Heidelberg, Germany, 1877; Volume 1, pp. 194–198, reprinted in *FEBS Lett.* **1976**, *62*, E8–E12. [CrossRef]
12. Kunitz, M.; Northrop, J.H. Isolation from beef pancreas of crystalline trypsinogen, trypsin, a trypsin inhibitor, and an inhibitor-trypsin compound. *J. Gen. Physiol.* **1936**, *19*, 991–1007. [CrossRef] [PubMed]

13. Walsh, K.A.; Neurath, H. Trypsinogen and chymotrypsinogen as homologous proteins. *Proc. Natl. Acad. Sci. USA* **1964**, *52*, 884–889. [CrossRef] [PubMed]
14. Mikeš, O.; Holeyšovský, V.; Tomášek, V.; Šorm, F. Covalent structure of bovine trypsinogen. The position of remaining amides. *Biochem. Biophys. Res. Commun.* **1966**, *24*, 346–352. [CrossRef]
15. Puigserver, A.; Desnuelle, P. Identification of an anionic trypsinogen in bovine pancreas. *Biochim. Biophys. Acta* **1971**, *236*, 499–502. [CrossRef]
16. Smith, R.L.; Shaw, E. Pseudotrypsin. A modified bovine trypsin produced by limited autodigestion. *J. Biol. Chem.* **1969**, *244*, 4704–4712. [PubMed]
17. Maroux, S.; Rovery, M.; Desnuelle, P. An autolyzed and still active form of bovine trypsin. *Biochim. Biophys. Acta* **1967**, *140*, 377–380. [CrossRef]
18. Maroux, S.; Desnuelle, P. On some autolyzed derivatives of bovine trypsin. *Biochim. Biophys. Acta* **1969**, *181*, 59–72. [CrossRef]
19. Kumazaki, T.; Ishi, S. Characterization of active derivatives produced by acetamidination and selective autolysis of bovine trypsin. *J. Biochem.* **1979**, *85*, 581–590. [CrossRef] [PubMed]
20. Lacerda, C.D.; Teixeira, A.E.; de Oliveira, J.S.; Silva, S.F.; Vasconcelos, A.V.B.; Gouveia, D.G.; da Silva, A.R.; Santoro, M.M.; dos Mares-Guia, M.L.; Santos, A.M.C. Gamma trypsin: Purification and physicochemical characterization of a novel bovine trypsin isoform. *Int. J. Biol. Macromol.* **2014**, *70*, 179–186. [CrossRef] [PubMed]
21. Walsh, K.A.; Wilcox, P.E. Serine proteases. *Methods Enzymol.* **1970**, *19*, 31–41. [CrossRef]
22. Smith, L.M.; Kelleher, N.L. The Consortium for Top Down Proteomics. Proteoform: A single term describing protein complexity. *Nat. Methods* **2013**, *10*, 186–187. [CrossRef] [PubMed]
23. Schroeder, D.D.; Shaw, E. Chromatography of trypsin and its derivatives. Characterization of a new active form of bovine trypsin. *J. Biol. Chem.* **1968**, *243*, 2943–2949. [PubMed]
24. Keil-Dlouhá, V.V.; Zylber, N.; Imhoff, J.M.; Tong, N.T.; Keil, B. Proteolytic activity of pseudotrypsin. *FEBS Lett.* **1971**, *16*, 291–295. [CrossRef]
25. Santos, A.M.C.; Oliveira, J.S.D.; Bittar, E.R.; Silva, A.L.D.; Guia, M.L.D.M.; Bemquerer, M.P.; Santoro, M.M. Improved purification process of β-and α-trypsin isoforms by ion-exchange chromatography. *Braz. Arch. Biol. Technol.* **2008**, *51*, 511–521. [CrossRef]
26. Foucault, G.; Kellershohn, N.; Seydoux, F.; Yon, J.; Parquet, C.; Arrio, B. Comparative study of some conformational properties of α, β and ψ trypsins. *Biochimie* **1974**, *56*, 1343–1350. [CrossRef]
27. Foucault, G.; Seydoux, F.; Yon, J. Comparative kinetic properties of α, β and ψ forms of trypsin. *Eur. J. Biochem.* **1974**, *47*, 295–302. [CrossRef] [PubMed]
28. Vincent, J.P.; Lazdunski, M. Trypsin-pancreatic trypsin Inhibitor association. Dynamics of the interaction and role of disulfide bridges. *Biochemistry* **1972**, *11*, 2967–2977. [CrossRef] [PubMed]
29. Chowdhury, S.K.; Chait, B.T. Analysis of mixtures of closely related forms of bovine trypsin by electrospray ionization mass spectrometry: Use of charge state distributions to resolve ions of the different forms. *Biochem. Biophys. Res. Commun.* **1990**, *173*, 927–931. [CrossRef]
30. Ashton, D.S.; Ashcroft, A.E.; Beddell, C.R.; Cooper, D.J.; Green, B.N.; Oliver, R.W.A. On the analysis of bovine trypsin by electrospray-mass spectrometry. *Biochem. Biophys. Res. Commun.* **1994**, *199*, 694–698. [CrossRef] [PubMed]
31. Cunningham, L.W. Molecular-kinetic properties of crystalline diisopropyl phosphoryl trypsin. *J. Biol. Chem.* **1954**, *211*, 13–19. [PubMed]
32. Günther, A.R.; Santoro, M.M.; Rogana, E. pH titration of native and unfolded β-trypsin: Evaluation of the $\Delta\Delta G^0$ titration and the carboxyl pK values. *Braz. J. Med. Biol. Res.* **1997**, *30*, 1281–1286. [CrossRef] [PubMed]
33. Buck, F.F.; Vithayathil, A.J.; Bier, M.; Nord, F.F. On the mechanism of enzyme action. LXXIII. Studies on trypsins from beef, sheep and pig pancreas. *Arch. Biochem. Biophys.* **1962**, *97*, 417–424. [CrossRef]
34. Travis, J.; Liener, I.E. The crystallization and partial characterization of porcine trypsin. *J. Biol. Chem.* **1965**, *240*, 1962–1966. [PubMed]
35. Ianucci, N.B.; Albanesi, G.J.; Marani, M.M.; Fernández Lahore, H.M.; Cascone, O.; Camperi, S.A. Isolation of trypsin from bovine pancreas using immobilized benzamidine and peptide CTPR ligands in expanded beds. *Sep. Sci. Technol.* **2007**, *40*, 3277–3287. [CrossRef]
36. Matthews, B.W.; Sigler, P.B.; Henderson, R.; Blow, D.M. The three-dimensional structure of tosyl-α-chymotrypsin. *Nature* **1967**, *214*, 652–656. [CrossRef] [PubMed]

37. Birktoft, J.J.; Blow, D.M. Structure of crystalline α-chymotrypsin: V. The atomic structure of tosyl-α-chymotrypsin at 2 Å resolution. *J. Mol. Biol.* **1972**, *68*, 187–240. [CrossRef]
38. Bode, W.; Schwager, P. The refined crystal structure of bovine β-trypsin at 1.8 Å resolution. II. Crystallographic refinement, calcium binding site, benzamidine binding site and active site at pH 7.0. *J. Mol. Biol.* **1975**, *98*, 693–717. [CrossRef]
39. Bode, W.; Fehlhammer, H.; Huber, R. Crystal structure of bovine trypsinogen at 1.8 Å resolution. I. Data collection, application of Patterson search techniques and preliminary structural interpretation. *J. Mol. Biol.* **1976**, *106*, 325–335. [CrossRef]
40. Fehlhammer, H.; Bode, W.; Huber, R. Crystal structure of bovine trypsinogen at 1.8 Å resolution. II. Crystallographic refinement, refined crystal structure and comparison with bovine trypsin. *J. Mol. Biol.* **1977**, *111*, 415–438. [CrossRef]
41. Sandler, B.; Murakami, M.; Clardy, J. Atomic structure of the trypsin-aeruginosin 98-B complex. *J. Am. Chem. Soc.* **1998**, *120*, 595–596. [CrossRef]
42. Page, M.J.; Di Cera, E. Combinatorial enzyme design probes allostery and cooperativity in the trypsin fold. *J. Mol. Biol.* **2010**, *399*, 306–319. [CrossRef] [PubMed]
43. Inagami, T.; Sturtevant, J.M. Nonspecific catalyses by α-chymotrypsin and trypsin. *J. Biol. Chem.* **1960**, *235*, 1019–1023. [PubMed]
44. Nakano, M.; Tanizawa, K.; Nozawa, M.; Kanaoka, Y. Efficient tryptic hydrolysis of aryl esters with a cationic center in the leaving group. Further characterization of "inverse substrates". *Chem. Pharm. Bull.* **1980**, *28*, 2212–2216. [CrossRef]
45. Seydoux, F.; Yon, J. On the specificity of tryptic catalysis. *Biochem. Biophys. Res. Commun.* **1971**, *44*, 745–751. [CrossRef]
46. Keil-Dlouhá, V.; Zylber, N.; Tong, N.T.; Keil, B. Cleavage of glucagon by α- and β-trypsin. *FEBS Lett.* **1971**, *16*, 287–290. [CrossRef]
47. Ghalayini, M.K.; Dong, Q.; Richardson, D.R.; Assinder, S.J. Proteolytic cleavage and truncation of NDRG1 in human prostate cancer cells, but not normal prostate epithelial cells. *Biosci. Rep.* **2013**, *33*, e00042. [CrossRef] [PubMed]
48. Kirkpatrick, D.S.; Gerber, S.A.; Gygi, S.P. The absolute quantification strategy: A general procedure for the quantification of proteins and post-translational modifications. *Methods* **2005**, *35*, 265–273. [CrossRef] [PubMed]
49. Nigam, A.; Subramanian, M.; Rajanna, P.K. Non-specific digestion artifacts of bovine trypsin exemplified with surrogate peptides for endogenous protein quantitation. *Chromatographia* **2018**, *81*, 57–64. [CrossRef]
50. Krause, E.; Wenschuh, H.; Jungblut, P.R. The dominance of arginine-containing peptides in MALDI-derived tryptic mass fingerprints of proteins. *Anal. Chem.* **1999**, *71*, 4160–4165. [CrossRef] [PubMed]
51. Blavet, N.; Uřinovská, J.; Jeřábková, H.; Chamrád, I.; Vrána, H.; Lenobel, R.; Beinhauer, J.; Šebela, M.; Doležel, J.; Petrovská, B. UNcleProt (Universal Nuclear Protein database of barley): The first nuclear protein database that distinguishes proteins from different phases of the cell cycle. *Nucleus* **2016**, *8*, 70–80. [CrossRef] [PubMed]
52. Havliš, J.; Thomas, H.; Šebela, M.; Shevchenko, A. Fast-response proteomics by accelerated in-gel digestion of proteins. *Anal. Chem.* **2003**, *75*, 1300–1306. [CrossRef] [PubMed]

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Molecules Editorial Office
E-mail: molecules@mdpi.com
www.mdpi.com/journal/molecules

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34
Fax: +41 61 302 89 18

www.mdpi.com

ISBN 978-3-03897-827-5