

Diss. ETH N° 19088

Innovative methodologies for the proteomic discovery of vascular markers in cancer and kidney disease

A dissertation submitted to

ETH ZURICH

for the degree of

Doctor of Sciences

presented by

TIM FUGMANN

Dipl. Biol. t.o., University of Stuttgart

Born October 06, 1979

citizen of Germany

accepted on the recommendation of

Prof. Dr. Dario Neri, examiner

Prof. Dr. Michael Detmar, co-examiner

Dr. Christoph Rösli, co-examiner

2010

1. SUMMARY

1.1. Summary

State-of-the-art mass spectrometric techniques allow the identification of hundreds of proteins in complex biological specimens. Although mass spectrometry is not inherently quantitative, a multitude of different approaches for the quantification of proteins have been developed in the last years which can be divided into (i) methods applying stable isotope labelling and (ii) label-free methods. While stable isotope labelling is labour intensive and limiting in both sample throughput and economic terms, the main obstacle in applying label-free quantification is the need of sophisticated software for reliable analyses. Unfortunately most software is either limited in sample source and quantity, specific for certain mass spectrometer generations or needs extremely powerful computational resources. On the other hand the quantification of changes in protein abundance in complex biological specimens is essential for proteomic studies in basic and applied research. Here we report on the development and validation of the DeepQuanTR software for the identification and quantification of differentially expressed proteins using LC-MALDI mass spectrometry. Following enzymatic digestion, separation of peptides by reverse phase HPLC and normalization of MALDI-MS signal intensities to the ones of internal standards, the software extracts peptide features, adjusts differences in HPLC retention times and performs the relative quantification of peptides and proteins.

The overall performance of DeepQuanTR was evaluated by analyzing 24 human serum samples spiked with varying amounts of four proteins and eight complex samples of vascular proteins deriving from surgically-resected human tumour bearing kidneys. Besides these proof-of-principle experiments the software tool was applied to three large scale proteomics projects with the aim to identify differentially expressed vascular proteins. For the identification of vascular accessible biomarkers of diabetic nephropathy, streptozotocin treated rats were submitted to *in vivo* perfusion. Following excision of the kidneys biotin-tagged proteins were enriched on streptavidin resin, digested with trypsin, and resulting peptides analysed by LC-MALDI-MS. The subsequent comparative analysis using DeepQuanTR revealed several differentially regulated proteins which were further validated by complementary methods. Additionally the software was used for the identification and quantification of proteins in a mouse model of disseminated lymphoma and in three syngenic mouse models of liver metastasis.

In summary, DeepQuanTR correctly identified the relative regulation of artificially spiked proteins in a complex biological background. Furthermore the experimental validation of proteins identified by DeepQuanTR to be differentially regulated in the above described large scale proteomics projects demonstrated the software's power.

The software developed in the frame of this Ph.D. thesis can easily be integrated in established proteomic workflows and allows the relative quantification of proteins deriving from a wide variety of sample sources. All steps are fully automated and can be performed without limitations on a desktop computer. Finally DeepQuanTR offers a variety of tools for quality control and statistical analysis which facilitate the interpretation of data deriving from large scale proteomics projects.

1.2. Zusammenfassung

Modernste massenspektrometrische Methoden ermöglichen die Identifikation von vielen hundert Proteinen in komplexen biologischen Proben. Obwohl die Massenspektrometrie an sich nicht quantitativ ist, wurde in den letzten Jahren eine Vielzahl von Ansätzen zur Quantifizierung von Proteinen entwickelt. Diese lassen sich in zwei Gruppen unterteilen: (i) Methoden, die Peptide oder Proteine mittels stabiler Isotope markieren und (ii) Methoden, die ohne eine Modifikation auskommen. Während die Markierung mittels Isotopen arbeitsaufwändig ist und die Anzahl der Experimente durch die damit verbundenen Kosten sowie der beschränkten Anzahl erhältlicher Label eingeschränkt ist, stellt die Verfügbarkeit spezifischer Programme zur Datenauswertung die grösste Hürde zur verlässlichen Quantifizierung von Experimenten ohne Markierung mit Isotopen dar. Die bisher vorhandenen Programme sind meistens entweder für spezielle Massenspektrometergenerationen entwickelt, beschränkt in der Anzahl auswertbarer Proben oder in deren Herkunft oder benötigen den Zugang zu Hochleistungsrechnern. Proteomische Studien in Grundlagen- sowie in angewandter Forschung sind jedoch auf die verlässliche Bestimmung von Proteinkonzentrationsänderungen angewiesen. In dieser Arbeit präsentieren wir die Entwicklung und Validierung des Programms DeepQuanTR für die Identifizierung und Quantifizierung von unterschiedlich exprimierten Proteinen mittels LC-MALDI Massenspektrometrie. Nach dem enzymatischen Verdau von Proteinen, der Auftrennung der resultierenden Peptide mittels Reversed Phase Chromatographie und der Normalisierung von MALDI Signalintensitäten hinsichtlich derjenigen von internen Standardpeptiden, extrahiert das Programm Peptidsignale, passt Retentionszeiten zwischen verschiedenen Analysen an und ermittelt das Konzentrationsverhältnis von Peptiden und Proteinen.

Die allgemeine Leistung von DeepQuanTR wurde anhand von 24 Proben humanen Serums evaluiert. Zu diesem Zweck wurden drei verschiedene Mixturen hergestellt bestehend aus gleichen Teilen humanen Serums zu dem vier Proteine in unterschiedlichen Konzentrationen zugegeben wurden. Zusätzlich wurden acht komplexe Proben von vaskulären Proteinen aus chirurgisch entfernten, humanen Tumor-enthaltenden Nieren analysiert. Nach diesen Machbarkeitsstudien wurde das Programm in drei weiteren grossen proteomischen Studien mit dem Ziel vaskuläre Proteine zu identifizieren eingesetzt: (i) Die Perfusion von Streptozotocin behandelte Ratten, ein Modell für Typ 1 Diabetes Mellitus, mittels eines reaktiven Esterderivates von Biotin war die Grundlage für die Identifikation potentieller Biomarker der diabetischen Nephropathie, einer häufigen Komplikation bei

Diabetespatienten. Nach der Entnahme der Nieren wurden Biotin-konjugierte Proteine auf einem Streptavidin-Granulat angereichert, mit Trypsin verdaut und die resultierenden Peptide mittels LC-MALDI-MS analysiert. Die nachfolgende relative Analyse mittels DeepQuanTR führte zu der Identifikation mehrerer unterschiedlich exprimierter Proteine. Zusätzlich wurde das Programm zur Analyse von (ii) einem Mausmodell disseminierter Lymphome und (iii) einem Lebermetastasenmausmodell verwendet.

Zusammenfassend hat DeepQuanTR die Regulation von künstlich zugefügten Proteinen in einem komplexen biologischen Hintergrund korrekt wiedergegeben. Desweiteren konnten die als unterschiedlich reguliert identifizierten Proteine mittels Immunfluoreszenz und Western-blotanalyse validiert werden, was zusätzlich die Qualität der Analysen bestätigt.

Das Programm, das im Rahmen dieser Arbeit entwickelt wurde, kann problemlos in bestehende Arbeitsabläufe integriert werden und ermöglicht die relative Quantifizierung von Proteinen aus einer Vielzahl an Quellen. Alle Arbeitsschritte sind voll automatisiert und können ohne Einschränkungen auf einem Desktop PC durchgeführt werden. Schlussendlich besitzt DeepQuanTR verschiedenste Werkzeuge zur Qualitätskontrolle und zur Durchführung von statistischen Analysen, die die Interpretation von Daten aus grossen proteomischen Projekten ermöglichen.