Chapter 10

Summary

Summary

The field of proteomics has developed rapidly in recent years. The essence of proteomics is to characterize the behavior of a group of proteins, the system rather than the behavior of any single protein component. The successful completion of the human genome project has led to a tremendous increase in our understanding of the molecular basis of diseases. However, a comprehensive understanding of the dynamic protein pathways involved in normal and disease states, and in response to medical treatment, is required if we want to effectively treat diseases. The next major challenge towards this aim is to identify the constituents of the human proteome. Advances in proteomics technology offer great promise in the understanding and treatment of the molecular basis of diseases. The past decade of proteomics research, the study of dynamic protein expression, post-translational modifications, cellular and sub-cellular protein distribution, and protein-protein interactions, has culminated in the identification of many disease-related biomarkers and potential new drug targets.

Proteomic analysis requires the combination of various technologies, including biochemistry, mass spectrometry and bioinformatics. Important techniques for expression analysis of proteins are two-dimensional gel electrophoresis (2-DE) combined with Matrix Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF-MS), Surface-Enhanced Laser Desorption/Ionization Time-of-Flight Mass spectrometry (SELDI-TOF-MS) and/or Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS).

The aim of this thesis was to detect potential biomarkers for sarcoidosis, ankylosing spondylitis, and lacunar stroke using proteomic pattern analysis with SELDI-TOF-MS. In contrast to proteomics, glycomics or glycobiology deals with the structure and function of oligosaccharides (chains of sugars). The identity of the entire repertoire of carbohydrates in an organism is thus collectively referred to as the glycome. The analytical techniques used for biomarker finding can also be used for more extensive analysis of proteins, like protein glycosylation. In this thesis, we look for glycosylation defects in galactosemia. Patients with classical galactosemia, an inborn error of galactose metabolism with secondary glycosylation abnormalities, are at risk for a diminished bone mass since early age. The lack of evidence for the presence of any of the well established risk factors for a diminished bone led us to hypothesize a glycosylation defect of proteins involved in bone metabolism in these patients.

Chapter 1 provides an introduction on proteomics and glycomics. The SELDI-TOF-MS technique is used to detect potential biomarkers for sarcoidosis, ankylosing spondylitis, and lacunar stroke, which are described in this thesis. The background of these diseases is summarized in this chapter. Other technologies, like immunoprecipitation, gel electrophoresis and Western blot are described here

because they were applied in this thesis to look for glycosylation defects in galactosemia.

Chapter 2 gives an overview of the diagnostic value of protein profiles obtained with SELDI-TOF-MS on prostate and ovarian cancer. The pre- and post analytical aspects of proteomics studies are mentioned in this chapter and the pre- and post analytical strategies of the different prostate and ovarian cancer studies are discussed. Some own data are also included in this chapter. After comparing the different studies, there appear to be clear differences in sample pre-treatment and storage conditions. The effects of these pre-treatment steps are highly underestimated. It is essential that sample collection from both the patient and control populations should be completely identical and accurately standardized in future studies. An important conclusion of this chapter is that there is a need for a better and careful description of the methods, including technical details and data analyses, to use protein profiles as a diagnostic tool. A standardized protocol is not only needed for the pre- and post analytical strategies, but also for the quality control procedures.

In **chapter 3** a standard protocol for calibration of the MALDI-TOF-MS part of the SELDI-TOF-MS is described. Acceptance criteria for the independent certified quality control (QC) sample are established. This is also possible for other instrument types. By checking the calibration every week, the QC procedure acts prospectively, while in some studies the quality control acts retrospectively by including the QC samples in the profiling experiments and in some studies there is even no quality control procedure described at all. Actions need to be taken when the quality control samples exceed the acceptance criteria. The acceptance criteria can be exceeded because of different factors, like errors during preparation and handling of the calibration or quality control sample, as well as instrumental problems. Any new technology, particularly one being presented as a potential diagnostic tool, requires stringent quality control to evaluate analytical performance over time.

In **chapter 4** a review on the use of different proteomics techniques to detect potential and/or common biomarkers in chronic inflammatory diseases is given. Different proteomics strategies are explained and also the clinical background of the three chronic inflammatory diseases, multiple sclerosis, rheumatic diseases, and lung inflammatory diseases to which biomarker finding was applied, are described in this chapter. The identified and validated proteins detected in the different studies are compared and discussed to conclude if there are some common markers which can be used in the diagnosis and prognosis of the three chronic inflammatory diseases described in this chapter. A lot of proteins are identified and in some of the studies the identified proteins are also validated with other tests. For multiple sclerosis, the heat shock protein family is entitled to contain biomarkers with potential for further research. In three different rheumatoid arthritis studies using different sample material, myeloid-related

protein 8 is found as a potential marker. Alpha1-antitrypsin is validated in two studies as a marker for sarcoidosis and α1-antitrypsin is found to be a marker for cystic fibrosis (CF), together with myeloperoxidase and Immunoglobulin G.

Chapter 5 is focused on the detection of potential biomarkers in serum for the diagnosis of sarcoidosis using SELDI-TOF-MS. Sarcoidosis is a multi-systemic inflammatory disorder, which affects the lungs in 90 percent of the cases. The main pathologic feature is chronic inflammation resulting in non-caseating granuloma formation. Until now there is no satisfying biomarker for diagnosis or prognosis of sarcoidosis. For detection of potential biomarkers, protein profiles of anion exchange fractionated serum of 35 sarcoidosis patients and 35 healthy controls are compared using SELDI-TOF-MS. Sensitivities and specificities of the potential biomarkers obtained with SELDI-TOF-MS, generated with decision tree algorithm, are compared to the conventional markers angiotensin converting enzyme (ACE) and soluble Interleukin-2 Receptor (sIL-2R). An optimal classification is achieved with metal affinity binding ProteinChip arrays coupled with copper (IMAC-Cu^{2+} ProteinChip array). A single marker with a mass-to-charge (*m/z*) value of 11,955 results in a sensitivity and specificity of 86% and 63%, respectively. A multimarker approach of two peaks, *m/z* values of 11,734 and 17,377, results in a sensitivity and specificity of 74% and 71%, respectively. These sensitivities and specificities are higher compared to measurements of ACE and sIL-2R. Identification of the peak at *m/z* 17,377 results in the alpha-2 chain of haptoglobin. In future studies, we will enlarge the sample group and we will also validate our markers with a blind sample set. After this validation, we will search for disease activity markers. Eventually, implementation of a quantitative immunoassay, is needed, to give a good prediction of the disease state and disease severity.

In **chapter 6** we try to find potential biomarkers for ankylosing spondylitis (AS) using SELDI-TOF-MS. AS is a chronic systemic inflammatory rheumatic disorder that primarily affects the axial skeleton, with sarcoiliitis as its hallmark. Spinal structural damage can be assessed on conventional radiographs as destructive and proliferative lesions ultimately leading to syndesmophyte formation. Sera of 38 AS patients and 38 healthy controls are used to detect potential biomarkers. Serum is separated with an anion exchange fractionation procedure. In the screening experiments, three ProteinChip array surfaces; CM10, IMAC-Cu^{2+}, and hydrophobic, are compared to find out which condition results in the best discrimination of both groups. The optimal discrimination is reached with the following conditions; CM10 array with the organic serum fraction and IMAC-Cu^{2+} array with denatured serum. Analyses of all AS and healthy control samples on CM10 arrays results in a sensitivity of 66% and a specificity of 74% using a multimarker approach of two peaks. *M/z* 4172 is used as first splitter in the decision tree and is up-regulated in the AS group and *m/z* 28,144 is used as second splitter. Analyses of all AS and healthy control samples on IMAC-Cu^{2+} arrays results in a sensitivity and specificity

of 70% using a multimarker approach of two peaks. *M/z* 6644 is used as first splitter in the decision tree and is down-regulated in the AS group and *m/z* 13,875 is used as second splitter. The peaks at *m/z* 28,144 and 13,875 are both successfully identified as apolipoprotein A-I (ApoA1). This is the first study that shows that protein profiling in serum using SELDI-TOF-MS can be used as a diagnostic tool for AS. In future studies, we will enlarge the sample group and we will also validate our markers with a blind sample set. After this validation, we can search for disease activity markers. Eventually, implementation of a quantitative immunoassay, is needed, to give a good prediction of the disease state and disease severity.

Chapter 7 describes a study with the aim to detect differences in protein expression profiles in serum samples of two lacunar stroke subtypes using SELDI-TOF-MS. Lacunar infarcts are small, deeply in the brain located infarcts, mostly caused by occlusion of a small perforating artery. Lacunar stroke patients in whom cerebral imaging shows only a single symptomatic lesion differ clinically from those patients with multiple additional "silent" lacunar lesions. Lacunar stroke patients with multiple lesions have more extensive cerebral white matter lesions on neuro-imaging, have more often hypertension, and have worse prognosis on functional outcome, a higher stroke recurrence rate, higher short- and long-term mortality and higher rate of asymptomatic lesion progression. Two groups are defined according to pre-defined criteria. Group 1 consists of eight patients in whom brain magnetic resonance imaging (MRI) show only one single symptomatic lacunar lesion (type I). Group 2 consists of eight patients in whom brain MRI additionally show multiple (4 or more) asymptomatic lacunar lesions as well as extensive white matter lesions (type II). An anion exchange procedure is used, which allows high-throughput fractionation of all 16 serum samples. All fractions are applied to two different ProteinChip array surfaces (Ciphergen Biosystems Inc); the weak cation exchange (CM10) and the IMAC-Cu^{2+}. The best distinctive pattern is found on the IMAC-Cu^{2+} ProteinChip array with denatured serum. One clearly potential marker at *m/z* 16,122 is up-regulated in type I vs type II with mean intensities of 12.5 and 5.0, respectively. Protein identification is performed by 1-DE and 2-DE followed by MALDI-TOF-MS. The peak at *m/z* 16,122 is identified as the alpha-chain of haptoglobin. The alpha-chain of haptoglobin exists in two variants, alpha-1 (8.9 kDa) and alpha-2 (16 kDa), the latter being compatible with the marker at *m/z* 16,122. The haptoglobin concentration and phenotype distribution are determined. As the total haptoglobin concentration does not differ between the two lacunar groups, the up-regulation of the alpha-2-chain in type I compared to type II represents a higher haptoglobin-2 allele frequency in the former. Yet, in comparison to the reference population, in both lacunar stroke groups haptoglobin-1 outweighs haptoglobin-2 allele frequency. The even higher haptoglobin-1 allele frequency in type II implies a promoting role for haptoglobin-1 in developing multiple silent lacunar lesions and cerebral white matter lesions (WML). The

association between haptoglobin-1 and lacunar stroke brings in a new candidate gene in the study of genetic factors in cerebral small vessel disease etiology. The trend for a difference in haptoglobin-1 association between two lacunar stroke types could be a reflection of a difference in underlying vascular pathology. The results need confirmation in a large group in future studies.

Chapter 8 describes an in-house developed immunoprecipitation method to isolate insulin-like growth factor binding protein-3 (IGFBP-3) and its isoforms from serum to investigate glycan abnormalities in IGFBP-3. Patients with classical galactosemia are at risk for a diminished bone mass without evidence of nutritional factors being the cause. Our hypothesis is that dysglycosylation of glycoproteins of the Growth Hormone/IGF-I (insulin-like growth factor type I) axis play an important role in these disturbances. IGF-I is over 75% bound with IGFBP-3. To detect glycan abnormalities in IGFBP-3, a low abundant protein, isolation followed by glycan analysis is aimed. Our immunoprecipitation method is compared to other existing immuno-precipitation methods. The study of IGFBP-3 isoforms is relevant for further studies on congenital defects in glycosylation, galactosemia, and alcoholic liver cirrhosis. The immunoprecipitation method is validated using Western blotting and enzyme-linked immunosorbent assay (ELISA). The in-house developed immunoprecipitation method using dimethyl pimelimidate as a cross-linker results in improved detection of IGFBP-3 isoforms from serum compared to existing non cross-linking methods. In this study we also perform a clinical validation by using patient serum samples. Because it is known that patients with congenital defects in glycosylation (CDG) type Ia have a deficient synthesis of N-glycans which results in a deficient incorporation of sialic acid, the CDG type Ia samples are used as positive controls. By comparing the serum samples of paediatric CDG type Ia patients and control serum of paediatric healthy persons, we detect a different pattern in the IGFBP-3 isoforms. A shift from more acidic to more basic isoforms for the IGFBP-3 protein in CDG type Ia is detected and will probably be caused by the deficient incorporation of sialic acid, which is also seen in other glycoproteins containing N-glycosylation sites, like transferrin. To confirm that glycan abnormalities in IGFBP-3 can cause diminished bone mass in classic galactosemia, the isoform patterns of IGFBP-3 in galactosemia will also be investigated.

From the investigations in this study it becomes clear that, although several analytical problems have to be overcome, biomarker screening using proteomics followed by identification is a promising tool needing further development.