Urine proteome analysis as a discovery tool in patients with deep vein thrombosis and pulmonary embolism

von zur Mühlen C¹, Koeck T², Schiffer E², Sackmann C¹, Zürlig P², Hilgendorf I¹, Reinöhl J¹,
Rivera J³, Zirlik A¹, Hehrlein C¹, Mischak H^{2,4}, Bode C¹, Peter K³

¹Department of Cardiology and Angiology I, University Heart Center Freiburg

²Mosaiques Diagnostics GmbH, Hannover, Germany

³Baker IDI Heart and Diabetes Institute, Melbourne, Australia

⁴BHF Glasgow Cardiovascular Research, University of Glasgow, UK

Running title: proteomics of thrombosis

Corresponding author:

Constantin von zur Mühlen

Department of Cardiology and Angiology I

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/prca.201500105](#).

This article is protected by copyright. All rights reserved.

University Heart Center Freiburg - Bad Krozingen

Hugstetter Strasse 55

79106 Freiburg, Germany

Tel: +49-761-270-37835; Fax: +49-761-270-33884

email: constantin.vonzurmuehlen@universitaets-herzzentrum.de

Abbreviations: CE-MS, capillary electrophoresis coupled to mass spectrometry DVT, deep vein thrombosis; PE, pulmonary embolism; SVM support vector machine

Keywords: biomarker, discovery tool, proteomics, pulmonary embolism, thrombosis

Number of words: 5130

Statement of clinical relevance

Deep vein thrombosis (DVT) and pulmonary embolism (PE) cause significant morbidity and mortality. Early and accurate detection is a major challenge as the necessary diagnostic procedures are expensive and often not readily available. A reliable noninvasive diagnostic approach could prevent mortality and long-term morbidity.

Aiming towards such an approach, and based on the hypothesis that specific urinary peptide patterns (classifier) may characterize PE and DVT, urine peptide biomarkers were identified, combined into a classifier, and their value in detecting DVT/PE as assessed in a small study.

The results show that this classifier is highly significantly associated with DVT/PE. These data form the basis for a future clinical trial towards demonstration a benefit of this urinary proteome-based classifier in patient management.

Abstract

Purpose: Early and accurate detection of deep vein thrombosis (DVT) is an important clinical need. Based on the hypothesis that urinary peptides may hold information on DVT in conjunction with pulmonary embolism (PE), the study was aimed at identifying such peptide biomarkers using capillary electrophoresis coupled mass spectrometry.

Experimental design: Patients with symptoms of unprovoked/idiopathic DVT and/or PE were examined by doppler-sonography or angio-computed tomography. Urinary proteome analysis allowed for identification of respective peptide biomarkers. To confirm their biological relevance, we induced PE in mice and assessed human *ex vivo* thrombi.

Results: We identified 62 urinary peptides as DVT-specific biomarkers, i.e. fragments of collagen type I and a fragment of fibrinogen β -chain. The presence of fibrinogen α/β in the acute thrombus, and collagen type I and osteopontin in the older, organized thrombus was demonstrated. The classifier DVT62 established through support vector machine (SVM) modeling based on the 62 identified peptides was validated in an independent cohort of 47 subjects (6 cases and 41 controls) with a sensitivity of 100% and specificity of 83%.

Conclusion: Urine proteome analysis enabled the detection of DVT-specific peptides, which were validated in human and mouse tissue. Furthermore, it allowed for the establishment of an urinary-proteome based classifier that is relatively specific for DVT. The data provide the basis for assessment of these biomarkers in a prospective clinical study.

INTRODUCTION

Deep vein thrombosis (DVT) is a major health problem worldwide with an annual incidence of around 80 cases/100000 [1]. It often remains undiagnosed, resulting in significant morbidity and mortality. Potentially fatal pulmonary embolism (PE) occurs in about 25% of patients with DVT [1,2]. Typical signs of venous thrombosis of the lower limb include swelling, cyanosis and pain. However, DVT can also be asymptomatic until manifestation of its complications, such as acute PE after dislodgement of the thrombus into the pulmonary artery, which has a mortality rate of up to 30% [3].

We developed a proteomic approach based on capillary electrophoresis coupled to mass spectrometry (CE-MS) for noninvasive identification of disease-specific urinary markers of pathophysiological relevance in a variety of pathologies, including vascular disease [4-6]. We chose urine as body fluid of interest as it has several important advantages over plasma or serum as a primary source of proteomic biomarkers, such as its accessibility, high stability (over hours) and most importantly low proteolytic activity, which otherwise interferes with proteomic analysis due to protein degradation [7,8]. Fragments of collagen alpha-1 (type I and III) [7], alpha-1-antitrypsin, and endothelial growth factor (EGF) [9] could be established as molecular markers of human and murine atherosclerosis. A urine proteome based classifier could be established that enabled diagnosis of coronary artery disease (CAD) in patients presenting with symptoms suspicious for CAD with an accuracy of 84% [7].

Based on these data, we hypothesized that urine may hold biomarkers for DVT. Clot development and clot organization that occurs in DVT and/or PE is associated with strong protease activity, which would result in the release of small degradation products from clot-derived proteins into the circulation. These peptides would then cross the renal filtration

barrier and become detectable in urine. Therefore, the aim of this study was to apply CE-MS urine analysis to identify specific biomarkers for DVT and PE, which may be used in combination as a novel molecular classifier. In addition, the novel markers may improve our understanding of the molecular mechanisms of this complex disease, and enable identification of novel pathophysiological factors in thrombotic disease.

MATERIALS AND METHODS

Subjects and procedures

Patients with symptoms associated with deep vein thrombosis (DVT), confirmed pulmonary embolism (PE), or both (DVT+PE) were screened for the study. Importantly, only patients with unprovoked events (no previous surgery, trauma, immobilization, or diagnosis of cancer) were included for the establishment of the proteome pattern. For patients with suspected DVT, doppler-sonography was performed to confirm or rule out DVT. A GE LOGIQ with 4 MHz convex and 7.5 MHz linear transducers was used. All pelvic and lower extremity veins were examined in the axial and longitudinal planes by the same sonographer. The diagnosis of DVT was made on the basis of clot visualization with incomplete color filling and compressibility, impaired venous flow, and respiratory phasicity. For evaluation of PE, pulmonary angiogram by computed tomography was performed. Additional control samples from healthy controls were obtained from previous studies [4].

Urine specimens were obtained as spot midstream urine samples and immediately stored at -20°C without any additives or centrifugation steps. D-Dimer testing was performed using a standard laboratory assay (Innovance, Siemens Medical Solutions, Erlangen, Germany).

For further additional assessment of the classifier specificity, data from controls samples collected in previous studies were employed [10,11].

This study complied with the guidelines of the Declaration of Helsinki (<http://www.wma.net/e/policy/b3.htm>). Informed consent was obtained from all patients and ethical approval was obtained from local institutional review boards.

Sample preparation, CE-MS analysis, statistical analysis and peptide sequencing

Sample preparation was performed based on an optimized protocol previously established [12]. A 0.7 mL aliquot of urine was thawed immediately prior to use and diluted with 0.7 mL of an aqueous solution supplemented with 2 M urea, 10 mM NH_4OH , and 0.02 % SDS. To remove proteins of higher molecular mass, the sample was filtered using Centriscart ultracentrifugation filter devices (20 kDa cut-off; Sartorius, Göttingen, Germany) at $3,000 \times g$ for 45 min at 4 °C. 1.1 mL of filtrate was applied onto a PD-10 desalting column (GE Healthcare, Germany) equilibrated in 0.01% NH_4OH in HPLC-grade water to remove urea, electrolytes, and salts. All samples were lyophilized, stored at 4°C, and resuspended in HPLC-grade water shortly before CE-MS analysis. The resuspension volume was adjusted to yield 0.8 $\mu\text{g}/\mu\text{L}$ protein, as measured by BCA-assay (Interchim, Montluçon, France).

CE-MS analysis

CE-MS analysis was performed with a P/ACE MDQ capillary electrophoresis system (Beckman Coulter, USA) coupled online to micro-TOF-MS (Bruker Daltonic, Germany) [13]. The electro-ionization sprayer (Agilent Technologies, USA) was grounded, and the ion-spray interface potential was set between -4.0 and -4.5 kV. Data- and MS- acquisition procedures were automatically controlled by the CE *via* contact closure. Spectra were accumulated every 3 seconds, over a range of m/z 350 to 3,000. The analytical characteristics of the CE-

MS system were extensively described recently [14]. Briefly, the average recovery of the sample preparation was 85%, with a detection limit of ~1 fmol. Data sets were accepted only if the following quality control criteria were met: A minimum of 950 peptides/proteins (mean number of peptides/proteins minus one standard deviation) must be detected with a minimum MS resolution of 8,000 (required resolution to resolve ion signals with $z=6$) in a minimum migration time interval (the time window, in which separated peptides can be detected) of 10 minutes. After calibration, the mean deviation of migration time (compared to reference standards) must be below 0.35 minutes.

Data processing

Mass spectral ion peaks representing identical molecules at different charge states were deconvoluted into single masses using MosaiquesVisu software [15]. Only signals observed in a minimum of 3 consecutive spectra with a signal-to-noise ratio of at least 4 were considered. Signals with a calculated charge of +1 were automatically excluded to minimize interference with matrix compounds or drugs. TOF-MS data were calibrated utilizing FT-ICR-MS data as reference masses applying linear regression. CE-migration time and ion signal intensity were normalized based on reference signals from 29 abundant “housekeeping” peptides generally present in urine, which served as internal standards [16]. Each of these internal standard peptides was present in at least 90% of all urine samples. For calibration, a local regression algorithm was applied with internal standard peptides as reference. A minimum of three reference peptides were required to be present in a dataset in order to accept normalization. The resulting peak list characterized each protein and peptide by its molecular mass and normalized CE migration time. Normalized signal intensity is used as a measure for relative abundance.

Descriptive statistical analysis

Estimates of sensitivity and specificity were calculated based on tabulating the number of correctly classified samples. Confidence intervals (95% CI) were based on binomial calculations performed with MedCalc version 12.16.1.0 (MedCalc Software, Belgium, www.medcalc.be).

Differential statistical analysis

For statistical differential analysis CE-MS data sets were matched to the human low-molecular weight urinary proteome data base, as described in detail [4]. Only peptides with a minimal frequency of 70% in one of the considered groups were included in downstream statistical analysis. Significant differences between groups were investigated using the Wilcoxon rank sum test, as this test has proven to be well suited for CE-MS proteomics data [17]. Adjustment for multiple testing was performed by assessing the false discovery rate [18].

Classification

MosaCluster (version 1.7.0) was employed to develop a classifier for the discrimination between different groups. This software tool allows the classification of samples in the high-dimensional parameter space by using support vector machine (SVM) learning. Each peptide represents one dimension in the n-dimensional parameter space [19]. SVM view a data point as a p-dimensional vector (p numbers of protein used), and they attempt to separate them with a (p-1) dimensional hyperplane. Classification itself was per-

formed by determining the Euclidian distance of the data set to the $n-1$ dimensional maximal margin hyperplane (absolute value of the normal vector) and the direction of the vector.

Sequencing of peptides

Candidate biomarkers were sequenced using CE-MS/MS or LC-MS/MS analysis, as recently described [20]. MS/MS experiments were performed on an Ultimate 3000 nano-flow system (Dionex/LC Packings, USA) connected to an LTQ Orbitrap Velos hybrid mass spectrometer (Thermo Fisher Scientific, Germany) equipped with a nano-electrospray ion source. Peptides were eluted using a binary solvent gradient of 0-50 % B in 180 min and 50-100 % B in 10 min (A: 0.1 % acetic acid, B: 0.1 % acetic acid/70 % acetonitrile) with a flow rate of 100 nL/min. Alternatively, the same MS instrument was connected to a P/ACE MDQ capillary electrophoresis system. The mass spectrometer was operated in data-dependent mode to automatically switch between MS and MS/MS acquisition. Survey full-scan MS spectra (from m/z 300–2,000) were acquired in the Orbitrap. Ions were sequentially isolated for fragmentation in the linear ion trap using collision induced dissociation. General mass spectrometric conditions were: electrospray voltage, 1.6 kV; no sheath and auxiliary gas flow; ion transfer tube temperature, 225 °C; collision gas pressure, 1.3 mTorr; normalized collision energy, 32% for MS. Ion selection threshold was 500 counts for MS/MS. All resultant MS/MS data were searched with the Proteome discovererTM software version 1.2 (Thermo Fisher Scientific, Inc., USA) with the Sequest algorithm for a search in the Swiss-Prot database (Swiss-Prot Number 56.6) against human entries (20413 sequences) without any enzyme specificity. No fixed modification was selected, and oxidation of methionine and proline were set as variable modifications. Accepted parent ion mass deviation was 10 ppm; accepted fragment ion mass deviation was 0.05 Da. For further validation of obtained peptide identifications, the correlation between peptide charge at the working pH of 2 and

CE-migration time was utilized to minimize false-positive identification rates [21]: Calculated CE-migration time of the sequence candidate based on its peptide sequence (number of basic amino acids) was compared to the experimental migration time.

Mouse model of venous thrombosis

Pulmonary embolism was induced in 8-10 week old C57/BL-6 male mice (Charles River, Sulzfeld/Germany) by injection of 5 µg thrombin (Thromborel, Dade-Behring, Erlangen/ Germany) via the tail vein. At 30 min after injection, mice were sacrificed with an overdose of ketamine and xylazine, and the chest opened ("acute PE-model"). The left ventricle was punctured and perfused in a retrograde fashion with 2 mL of normal saline through the aorta while still beating. Thereafter, tracheotomy was performed and the bronchi filled with a 3:1 mixture of Tissue Tec OCT compound (Sakura Finetec, Staufen/Germany) and normal saline. The entire left and right lungs were finally embedded into OCT compound. For immunohistochemistry on collagen, osteopontin, and CD107, animals were sacrificed one week after induction of pulmonary embolism ("subacute PE-model").

All experiments were approved by the local ethics committee (University of Freiburg, Germany, approval-number G-09/47).

Induction of human *ex vivo* thrombi

Blood was obtained by venipuncture from healthy volunteers taking no medication, anticoagulated with citric acid and centrifuged at 1,000 rpm for 10 min. Of the resulting platelet-rich plasma, 1 mL was incubated with 100 µL of ADP (möLab GmbH, Langenfeld, Germany), 88 µL of actin (Dade Behring, Marburg, Germany), and 25 µL of 1 M CaCl₂ to

induce coagulation. Samples were incubated for 12 min at 37 °C in a water bath, and finally embedded into OCT compound.

Immunohistochemistry

Frozen tissue was cut in transversal sections of 10 µm. For the human thrombus, 3 thrombi from different patients were stained, which all showed similar staining in different areas of the tissue. For the mouse lung, we identified representative sections with vascular thrombosis: 5 different animals with 3 samples each were assessed, all of which showed similar staining throughout the lung tissue.

Immunohistochemistry for the identified marker proteins was performed using the following primary antibodies: fibrinogen alpha (clone GTX 74134, GeneTex, Irvine/CA, USA) and fibrinogen beta (clone GTX113694, GeneTex, AAs 250-491), collagen type I (clone GTX112731, GeneTex, AAs 219-478), osteopontin (clone AB8448, Abcam), and CD107b (MCA2293-GA, Serotec). The primary antibody was detected using a corresponding species-specific biotinylated secondary antibody (Vectastain ABC-AP Kit, Vector, Germany) and VectorRed (Vector). Counterstaining was performed using standard hematoxylin/eosin stain (Sigma-Aldrich, Hamburg, Germany). Representative sections were evaluated using AxioVision Software (Carl Zeiss, Germany).

RESULTS

Clinical characteristics for all patients with DVT or DVT+PE as well as controls in the study population are summarized in **Table 1A**. As expected, D-Dimers were higher in patients with DVT/DVT+PE compared to controls. Urine samples from patients and controls were analyzed using CE-MS. To identify novel deep venous thrombosis (VT)-associated urinary peptide biomarkers, CE-MS data from patients suffering from DVT (N = 7) and DVT+PE (N = 10) were compared with profiles of VT negative controls (VTNC; N = 8) and age- and sex-matched healthy volunteers (N = 24) in a 1:2 matched case-control set-up (**Table 1B**). In total, 3368 peptides detected in urine were examined. This resulted in the identification of 62 DVT-associated peptides that were significantly different ($P < 0.05$ in multiple testing corrected statistics, **Table 2**) between the two groups, and that could be identified by MS/MS sequencing. The majority of these peptides consists of fragments of collagen type I and III, whereby excretion of short fragments of collagen type I was mostly decreased, while collagen type III fragments were increased in patients with DVT with and without PE. In addition, two peptides of each from Apolipoprotein A1, Uromodulin, and osteopontin, were found to be significantly upregulated in DVT. Further specific peptides from fibrinogen, gelsolin, Ion protease homolog, peptidoglycan recognition protein 1, plasma protease C1 inhibitor, POTE ankyrin domain family member F, receptor-type tyrosine-protein phosphatase U, retinol-binding protein 4, and WD repeat-containing protein 59 were also detected as being significantly upregulated in DVT, while a peptide from Sodium/potassium-transporting ATPase subunit gamma was significantly reduced in the urine from DVT patients.

Immunohistochemistry for fibrinogen α and fibrinogen β revealed that these proteins as expected are highly expressed in thrombi of mice (**Figure 1A**) and humans (**Figure 1B**).

However, the detection of fragments of collagen type I and osteopontin as DVT-specific peptides was unexpected. Indeed, in animals with fresh pulmonary embolism ("acute PE"-group, see above), we were not able to detect osteopontin or collagen in the thrombus. Since collagens type I to IV support platelet adhesion and the presence of collagen has been described during organisation of a thrombus and its fixation at the vascular wall (23), we also performed immunohistochemistry in subacute pulmonary embolism ("subacute PE-model"). In these organized thrombi, we could detect high abundance of collagen type I via immunohistochemistry (**Figure 2**). The detected antigen corresponded to AAs 219-478 that matched urinary peptide biomarkers with an excretory decrease up to 21.8-fold. Most interestingly, whereas osteopontin could not be detected in the fresh thrombus, we could demonstrate a high level of expression of osteopontin in older, organized thrombi in PE (**Figure 3A**). This expression corresponds to immunohistochemical localization of CD107b, which indicates inflammatory transformation of the older thrombi (**Figure 3B**). Overall, this points to a novel association of osteopontin in thrombus organization.

Besides the identification of novel proteins relevant for the pathophysiology of venous thrombotic disease, a DVT-specific urinary peptide biomarker based classifier could provide a unique opportunity for the non-invasive diagnosis of DVT. The fact that we identified a fragment of fibrinogen β -chain as DVT-specific urinary peptide biomarker indicates that CE-MS of urine is suitable for the detection of thrombotic events in the vascular system. Based on these encouraging findings, the identified peptide biomarkers were combined in a support vector machine (SVM) based classifier. To train the classifier, the same urinary proteome profiles already used for peptide biomarker discovery were utilized (**Table 1B and 2**). The classifier, termed DVT62, performed well in the training set, with a sensitivity of 93% and a specificity of 97% in the complete take-one-out cross validation of the training set.

To demonstrate a true value of the biomarker pattern in clinical applications [22,23], we assessed the performance of the classifier in an open-label case-control approach through the classification of the urine proteome profiles of 47 independent study subjects comprising 6 profiles of patients with DVT and DVT with PE along with 41 control profiles. The classifier reached an AUC of 0.90 (95 % CI 0.81 – 0.99; $P < 0.0001$, **Figure 4A**). Based on the optimal threshold value of -0.85, the classifier thereby discriminated between 6 patients and 41 controls with a specificity of 100% (5/6; 95% CI 35.9 – 99.6%) and a sensitivity of 83% (**Figure 4B**; 30/31, 95% CI 83.3 – 99.9%).

The results on the specificity of the classifier were further confirmed in 146 urine proteomics datasets that were collected in previous studies, from subjects without any evidence of DVT. Of these, 115 samples were negative, indicating a specificity of 79%, highly similar to the specificity observed in the initial validation set.

DISCUSSION

In this study, urine proteome analysis was used as a discovery tool in order to identify proteins associated with DVT with and without PE in conjunction. The identification of 62 excreted DVT-specific peptide biomarkers including a degradation fragment of the fibrinogen β -chain and various degradation fragments of collagen type I indicated that CE-MS-based urine proteomics enables the detection of disease specific proteolytic degradation processes throughout the human body. The identification of these biomarkers along with the histological assessment performed in human and mouse thrombi points to an interesting association of these proteins with organized thrombi.

This further indicates a general potential of CE-MS as a disease-specific discovery tool and also as a diagnostic platform as previously demonstrated for coronary artery disease [7,10,24,25]. Urinary fragments of fibrinogen α were also identified in patients with CAD [10] a disease that is associated with the development of plaque rupture-induced microthrombi. Similarly, an increase of fibrinogen in plasma was observed in patients after acute myocard infarction [26]. Specifically, the fibrinogen α fragment 607-622 was up-regulated in CAD, as it was the case in this study; a finding which could indicate specific peptidase activity, as seen in blood of patients with cancer [27]. The role of individual peptidases in the formation of the specific peptide fragments and their effect on the urinary levels of the resulting fragment is of major interest but cannot be judged at present. For example, it remains unclear why the two identified collagen type I fragments associated with thrombosis demonstrate an opposing regulation (Table 2). A possible explanation might be that they are cleavage products of different proteases, which show opposing expression/activity regulation in the thrombotic microenvironment. Further studies addressing the possible involvement of a complex protease network with modulated

expression and/or activity levels of individual peptidases might provide new insights in the pathophysiology of thrombosis.

One of the most interesting findings in our study is the identification of collagen and osteopontin degradation fragments associated with DVT and/or PE. Their role in acute thrombosis is not intuitive and indeed, we could not detect expression of these proteins in acute thrombi (mouse and human). However, when we assessed older, more organized thrombi we detected the expression of these proteins. Like collagen, osteopontin is an extracellular structural protein and was first discovered in osteoblasts. Osteopontin interacts with multiple cell surface receptors, especially integrins via the tripeptide RGD-dependent and independent binding sites. The term “pontin” describes its potential “bridge building effect”, which is in accordance to effects on thrombus organization. It has been shown to be involved in multiple physiological and pathological processes including wound healing, bone turnover, tumorigenesis, inflammation, ischemia and immune responses [28]. Interestingly, osteopontin was found to be upregulated in swine thrombi, is present in atherosclerotic plaques [29], and in pulmonary tumor-associated thrombotic microangiopathy [30]. As thrombus organization is typically associated with inflammation, we assessed and demonstrated co-expression of CD107b as a general marker of inflammatory activity in organised thrombi. Taken together, these findings suggest a biological relevance of osteopontin in thrombus pathology and its potential role as a marker of an inflammatory response to thrombus formation. The beginning of clinical symptoms does not automatically reflect the time point of thrombus development, indeed this is often delayed, as many patients have gradual emergence of symptoms while others remain asymptomatic. Hence, the exact time line of osteopontin and collagen expression in relation to the initial thrombus development cannot be determined in the current study, and needs to be addressed in further investigations, along with the question whether these fragments become detectable

after the onset of DVT/PE. The immunohistological detection of osteopontin in organized thrombi is one of the central findings of this paper. It shows the suitability of our urine proteome based discovery approach and potentially opens up a new interesting research area in thrombus stabilization/organization.

The major finding of this study is the demonstration that DVT/PE pathology is reflected in changes of the urine proteome. This forms the basis of the discussed discovery approach. This finding also provides the unique opportunity to assess the feasibility of CE-MS of urine as a potential diagnostic biomarker approach for DVT/PE, which is long sought-after and would satisfy a major medical need. At the current stage, overall specificity for thrombosis/embolism detection was 83%, with a sensitivity of 100%. The potential of the urinary-proteome approach to identify DVT/PE in healthy cohorts but also in patient cohorts indicates a high positive predictive value. In comparison, traditional D-Dimer testing has a sensitivity of 85 to 96% and a specificity of 48 to 74% [31]. Therefore in future, urinary proteomics might be developed into a clinically applicable test that provides most benefit being complementary to D-Dimer testing. However, ultimately for such a clinical application the current urinary peptide biomarker-based model for DVT/PE has to be further validated in cohorts that are problematic for D-Dimer testing like liver diseases, inflammatory etiologies, malignancies, and traumata. Based on our promising data, such a development and especially testing in larger cohorts is warranted, but is beyond the aim of the current proof of concept study.

Previous approaches for proteome analysis in patients with venous thromboembolism have used MALDI-TOF MS, reaching a sensitivity of 68% and specificity of 89% [32]. There are numerous advantages of CE-MS over MALDI, among others the much higher resolution, higher dynamic range, consequently ability to assess thousands of peptides with much higher confidence, and also little sensitivity towards signal suppression,

which is observed when certain analytes are preferentially ionized resulting in “biased” detection with very high signal amplitudes at the expense of other analytes that become undetectable [33]. For relative quantification of urinary peptides, we have applied label-free ion counting using highly abundant collagen fragments as internal references [16]. This procedure offers sufficient accuracy compared with methods based on either urinary creatinine or stable isotope labeled synthetic marker analogues (absolute quantification) and has been successfully used in multiple studies.

The urine proteome has only recently been used for proteome analysis, but has attracted major interest as a stable, non-invasive source for proteomic biomarkers [4,5]. Neither prolonged storage nor repeated freeze/thaw cycles interfere with reproducible sample profiling [14,34]. Our study design is in line with current guidelines for clinical proteome analysis [22,23] and includes an independent cohort for clinical validation as well as biomarker sequencing.

The presented work has to be seen as a pilot study aiming towards proof of concept, but with a clear limitation given by the relatively small number of investigated patients. However, the results obtained are of high statistical significance, clearly supporting the validity of the approach. In addition, the data on specificity were confirmed in additional samples from subjects without any apparent evidence of DVT. As a next step, the findings have to be reproduced and verified in a properly powered clinical study in a population at risk, as outlined recently [35]. It will likely be impossible to assess all of the identified biomarkers using immunological assays. Most often antibodies do not enable the specific identification of small peptides. This would require correctly identifying C- and N-terminus, to enable distinguishing the relevant peptide from other peptides derived from the same protein. However, multiple reaction monitoring appears to be well suited to assess specific

urine peptides [36], and may represent a valid alternative for CE-MS in the planned clinical study.

In conclusion, the presented data support the feasibility of CE-MS-based urinary proteomics to reflect localized disease processes. The central finding of our study is that the development and organization of thrombi is reflected by multiple changes in the urine proteome. This allows for the identification of peptide patterns in urine that are specific for DVT/PE. However, further large-scale evaluation is needed to assess the potential role of such a urine peptide based classifier as a clinically usable diagnostic test. An important additional result of our study is the identification of collagen and osteopontin fragments in urine of patients with DVT and/or PE and the conformation of their expression in older organized but not fresh thrombi. These findings describe a novel and highly interesting association of these proteins with organized thrombi.

Acknowledgement

This study was supported in part by the European Union grant HEALTH-2011-278249-EU-MASCARAtO HM and the National Health and Medical Research Council (NHMRC) of Australia to KP.

Conflict of interest

Harald Mischak is the founder and co-owner of Mosaiques Diagnostics, who developed the CE-MS technology for clinical application. Thomas Koeck and Petra Zürgbig are employees of Mosaiques Diagnostics.

References

- [1] Silverstein, M.D., Heit, J.A., Mohr, D.N., Petterson, T.M. *et al.*, Trends in the incidence of deep vein thrombosis and pulmonary embolism: a 25-year population-based study. *Arch. Intern. Med* 1998, 158, 585-593.
- [2] White, R.H., The epidemiology of venous thromboembolism. *Circulation* 2003, 107, 14-18.
- [3] Calder, K.K., Herbert, M., Henderson, S.O., The mortality of untreated pulmonary embolism in emergency department patients. *Ann. Emerg. Med* 2005, 45, 302-310.
- [4] Stalmach, A., Albalat, A., Mullen, W., Mischak, H., Recent advances in capillary electrophoresis coupled to mass spectrometry for clinical proteomic applications. *Electrophoresis* 2013, 34, 1452-1464.
- [5] Pontillo, C., Filip, S., Borrás, D.M., Mullen, W. *et al.*, CE-MS-based proteomics in biomarker discovery and clinical application. *Proteomics Clin Appl.* 2015.
- [6] Carty, D.M., Siwy, J., Brennand, J.E., Züribig, P. *et al.*, Urinary proteomics for prediction of preeclampsia. *Hypertension* 2011, 57, 561-569.
- [7] von Zur Muhlen C., Schiffer, E., Zuerbig, P., Kellmann, M. *et al.*, Evaluation of Urine Proteome Pattern Analysis for Its Potential To Reflect Coronary Artery Atherosclerosis in Symptomatic Patients. *J. Proteome. Res.* 2009, 8, 335-345.
- [8] Kolch, W., Neususs, C., Pelzing, M., Mischak, H., Capillary electrophoresis-mass spectrometry as a powerful tool in clinical diagnosis and biomarker discovery. *Mass Spectrom Rev.* 2005, 24, 959-977.
- [9] von zur Muhlen, C., Schiffer, E., Sackmann, C., Zurbig, P. *et al.*, Urine proteome analysis reflects atherosclerotic disease in an ApoE^{-/-} mouse model and allows the discovery of new candidate biomarkers in mouse and human atherosclerosis. *Mol. Cell Proteomics* 2012, 11, M111.
- [10] Delles, C., Schiffer, E., von Zur, M.C., Peter, K. *et al.*, Urinary proteomic diagnosis of coronary artery disease: identification and clinical validation in 623 individuals. *J. Hypertens.* 2010, 28, 2316-2322.
- [11] Zhang, Z., Staessen, J.A., Thijs, L., Gu, Y. *et al.*, Left ventricular diastolic function in relation to the urinary proteome: A proof-of-concept study in a general population. *Int. J. Cardiol.* 2014.

- [12] Theodorescu, D., Schiffer, E., Bauer, H.W., Douwes, F. *et al.*, Discovery and validation of urinary biomarkers for prostate cancer. *Proteomics Clin. Appl.* 2008, 2, 556-570.
- [13] Mischak, H., Kolch, W., Aivalotis, M., Bouyssie, D. *et al.*, Comprehensive human urine standards for comparability and standardization in clinical proteome analysis. *Proteomics Clin Appl.* 2010, 4, 464-478.
- [14] Mischak, H., Vlahou, A., Ioannidis, J.P., Technical aspects and inter-laboratory variability in native peptide profiling: The CE-MS experience. *Clin. Biochem.* 2013, 46, 432-443.
- [15] Neuhoﬀ, N., Kaiser, T., Wittke, S., Krebs, R. *et al.*, Mass spectrometry for the detection of differentially expressed proteins: a comparison of surface-enhanced laser desorption/ionization and capillary electrophoresis/mass spectrometry. *Rapid Communications in Mass Spectrometry* 2004, 18, 149-156.
- [16] Jantos-Siwy, J., Schiffer, E., Brand, K., Schumann, G. *et al.*, Quantitative Urinary Proteome Analysis for Biomarker Evaluation in Chronic Kidney Disease. *J. Proteome. Res.* 2009, 8, 268-281.
- [17] Dakna, M., Harris, K., Kalousis, A., Carpentier, S. *et al.*, Addressing the challenge of defining valid proteomic biomarkers and classifiers. *BMC. Bioinformatics.* 2010, 11, 594.
- [18] Benjamini, Y., Hochberg, Y., Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc B (Methodological)* 1995, 57, 125-133.
- [19] Girolami, M., Mischak, H., Krebs, R., Analysis of complex, multidimensional datasets. *Drug Discov Today: Technologies* 2006, 3, 13-19.
- [20] Klein, J., Papadopoulos, T., Mischak, H., Mullen, W., Comparison of CE-MS/MS and LC-MS/MS sequencing demonstrates significant complementarity in natural peptide identification in human urine. *Electrophoresis* 2014, 35, 1060-1064.
- [21] Zürgbig, P., Renfrow M.B., Schiffer, E., Novak, J. *et al.*, Biomarker discovery by CE-MS enables sequence analysis via MS/MS with platform-independent separation. *Electrophoresis* 2006, 27, 2111-2125.
- [22] Mischak, H., Apweiler, R., Banks, R.E., Conaway, M. *et al.*, Clinical Proteomics: a need to define the field and to begin to set adequate standards. *Proteomics Clin. Appl.* 2007, 1, 148-156.
- [23] Mischak, H., Allmaier, G., Apweiler, R., Attwood, T. *et al.*, Recommendations for biomarker identification and qualification in clinical proteomics. *Sci. Transl. Med.* 2010, 2, 46ps42.

- [24] Brown, C.E., McCarthy, N.S., Hughes, A.D., Sever, P. *et al.*, Urinary proteomic biomarkers to predict cardiovascular events. *Proteomics. Clin. Appl.* 2015.
- [25] Zhang Z, Thijs, L., Petit, T. *et al.*, Urinary Proteome and Systolic Blood Pressure as Predictors of 5-Year Cardiovascular and Cardiac Outcomes in a General Population. *Hypertension* 2015, 65, in press.
- [26] Cubedo, J., Ramaiola, I., Padro, T., Martin-Yuste, V. *et al.*, High-molecular-weight kininogen and the intrinsic coagulation pathway in patients with de novo acute myocardial infarction. *Thromb. Haemost.* 2013, 110, 1121-1134.
- [27] Villanueva, J., Nazarian, A., Lawlor, K., Yi, S.S. *et al.*, A sequence-specific exopeptidase activity test (SSEAT) for "functional" biomarker discovery. *Mol. Cell Proteomics* 2008, 7, 509-518.
- [28] Wang, K.X., Denhardt, D.T., Osteopontin: role in immune regulation and stress responses. *Cytokine Growth Factor Rev.* 2008, 19, 333-345.
- [29] Shanahan, C.M., Cary, N.R., Metcalfe, J.C., Weissberg, P.L., High expression of genes for calcification-regulating proteins in human atherosclerotic plaques. *J Clin Invest* 1994, 93, 2393-2402.
- [30] Takahashi, F., Kumasaka, T., Nagaoka, T., Wakiya, M. *et al.*, Osteopontin expression in pulmonary tumor thrombotic microangiopathy caused by gastric carcinoma. *Pathol. Int.* 2009, 59, 752-756.
- [31] Geersing, G.J., Janssen, K.J., Oudega, R., Bax, L. *et al.*, Excluding venous thromboembolism using point of care D-dimer tests in outpatients: a diagnostic meta-analysis. *BMJ* 2009, 339, b2990.
- [32] Ganesh, S.K., Sharma, Y., Dayhoff, J., Fales, H.M. *et al.*, Detection of venous thromboembolism by proteomic serum biomarkers. *PLoS. ONE.* 2007, 2, e544.
- [33] Molin, L., Seraglia, R., Lapolla, A., Ragazzi, E. *et al.*, A comparison between MALDI-MS and CE-MS data for biomarker assessment in chronic kidney diseases. *J Proteomics* 2012, 75, 5888-5897.
- [34] Haubitz, M., Good, D.M., Woywodt, A., Haller H *et al.*, Identification and validation of urinary biomarkers for differential diagnosis and dvaluation of therapeutic intervention in ANCA associated vasculitis. *Mol. Cell. Proteomics* 2009, 8, 2296-2307.
- [35] Mischak, H., Ioannidis, J.P., Argiles, A., Attwood, T.K. *et al.*, Implementation of proteomic biomarkers: making it work. *Eur. J Clin Invest* 2012, 42, 1027-1036.

- [36] Mermelekas, G., Vlahou, A., Zoidakis, J., SRM/MRM targeted proteomics as a tool for biomarker validation and absolute quantification in human urine. *Expert. Rev. Mol. Diagn.* 2015, 15, 1441-1454.

Author Manuscript

Figure legends

Figure 1. Immunohistochemistry for fibrinogen α and fibrinogen β on mouse (A) and human tissue (B). In mouse pulmonary embolism, fibrinogen α and β was highly expressed inside the vascular thrombus (A, arrows) in comparison to the control stainings with a non-specific IgG-antibody. In the human thrombus, we also detected fibrinogen α and β inside the thrombus (B).

Figure 2. Immunohistochemistry for collagen in subacute pulmonary embolism (“subacute PE-model”). In organized thrombi, expression of collagen can be detected, as demonstrated inside the thrombus in different magnifications (x4, x40).

Figure 3. Osteopontin expression can be found in the “subacute PE”-model, also demonstrated in different magnifications (x4, x40; A). In parallel, CD107b is expressed in parallel, indicating inflammatory transformation of the older thrombi (B), pointing to a novel association of osteopontin with thrombus organization.

Figure 4. (A) Receiver operating characteristics (ROC) curve analysis of the urinary signature applied to the validation set of 47 samples from patients and controls. 95% confidence intervals are indicated by dashed lines. Diagonal represents AUC value of 50% (random guessing level). (B) VT_{Score} obtained for the validation cohort with 6 patients and 41 controls. Data are depicted as Box and Whisker plots. The boxes depict the quartiles 25% and 75% quartile of each distribution; the statistical medians are shown as horizontal lines in the boxes. The whiskers indicate 3/2 times the interquartile range. P values are based on one-way Anova analysis of variance with Student-Newman-Keuls post-hoc test.

Table 1A. Clinical characteristics of study population

	N	Age (years)	Sex (m/f)	D-dimer (mg/dL) [#]	Platelet count (x10 ³ /μl) [#]
DVT, DVT+PE	23	58±21	15/8	1.8 [0.7;4.9]	110 [0.3;46.6]
VT controls	14	60±19	9/5	1.3 [0.5;3.7]*	100 [0.3;4.9]
HC	59	58±16	32/27	n.a.	n.a.

Data are given as mean ± standard deviation or median [minimum;maximum] as appropriate. P-values are from Student's t-test, Mann-Whitney U-test, chi² test or Fisher's exact test as appropriate. *: P<0.05. [#] Values from patients' citrated plasma (D-Dimers) and EDTA-blood (platelet counts)

Table 1B. Case-control setup in biomarker discovery and validation

Patient cohort	N (total)	N (DVT)	N (Control)	Usage
<i>Biomarker / Classifier</i>	49	17	32	
DVT	7	7		VT markers and SVM modelling
DVT + PE	10	10		VT markers and SVM modelling
VT controls	8		8	VT markers and SVM modelling
HC	24		24	VT markers and SVM modelling
<i>Validation</i>	47	6	41	
DVT	2	2		Validation
DVT + PE	4	4		Validation
HC + VT controls	41		41	Validation
Total	96	23	73	

DVT: deep vein thrombosis, PE: pulmonary embolism, VT: venous thrombosis, HC: healthy controls

Table 2. List of the 62 peptides constituting the venous thrombosis (VT)-specific peptide panel

wilcox-p-value	p-value adj.	Mean DVT	mean control	fold change	Sequence	Prot_Name
0.00000	0.00	141.89	9.13	15.54	SpGPDGKTGPpGP	Collagen alpha-1(I) chain
0.00001	0.01	3725.04	2084.01	1.79	GPpGEAGKpGEQGVpGDLGApGP	Collagen alpha-1(I) chain
0.00001	0.01	183.65	48.94	3.75	QGLpGTGGPpGENGKpGEPpKG	Collagen alpha-1(III) chain
0.00001	0.01	83.61	555.73	0.15	DGKTGPpGPA	Collagen alpha-1(I) chain
0.00004	0.01	137.57	595.69	0.23	DGRpGPpGPpG	Collagen alpha-1(I) chain
0.00004	0.01	2008.98	7132.00	0.28	ApGDRGEpGpP	Collagen alpha-1(I) chain
0.00010	0.02	322.06	53.99	5.97	SpGEAGRpGEAGLpGAKG	Collagen alpha-1(I) chain
0.00010	0.02	134.58	24.98	5.39	GTGGPpGENGKpGEPpGP	Collagen alpha-1(III) chain
0.00013	0.02	605.20	2124.53	0.28	ApGDKGESGPS	Collagen alpha-1(I) chain
0.00016	0.02	2297.16	995.73	2.31	KGNSEpGApGpSKGDTGAKGEpGPVG	Collagen alpha-1(I) chain
0.00016	0.02	145.79	7.15	20.40	SDPRERERDQH	WD repeat-containing protein 59; WDR59
0.00019	0.02	384.30	95.02	4.04	FSVDETQMSATAKGRVR	Retinol-binding protein 4
0.00020	0.03	186.42	90.84	2.05	GpGEAQGPpRGEpGTPGSpGP	Collagen alpha-1(II) chain;
0.00028	0.03	2271.54	3795.57	0.60	ApGDRGEpGPpGPAG	Collagen alpha-1(I) chain
0.00029	0.03	123.41	22.61	5.46	pNGPPpGpAGSRGDGGpPGMTGFPAAGRTGPpGP	Collagen alpha-2(I) chain
0.00033	0.03	4043.94	17647.99	0.23	GNSGEpGpGpSKGDTGAKGEpGPVG	Collagen alpha-1(I) chain
0.00033	0.03	39.90	12.30	3.24	ELTETGVAAAAAISVARTL	Plasma protease C1 inhibitor
0.00035	0.03	579.81	13.77	42.12	GRpGDAGPQKpVPSGAPGEDGRpGPpGPQG	Collagen alpha-1(II) chain;
0.00039	0.03	733.23	32.14	22.81	IPVKQADSGSSEEEKQLYNKYPDAVATW	Osteopontin
0.00040	0.03	113.99	21.11	5.40	GPpGADGQpGAKGEpGDAGAKGDAGpPGPAGP	Collagen alpha-1(I) chain
0.00040	0.03	1291.03	162.93	7.92	pGPQGpLKGpGAPGEPGPQG	Collagen alpha-1(VIII) chain
0.00043	0.03	121.77	13.06	9.33	QGVQGGKGEQpGPpGpGFGQLGPpSGPAGEV	Collagen alpha-2(I) chain
0.00043	0.03	50.00	0.00	∞	VVSLGSPSGEVSHPR	Alpha-2-HS-glycoprotein
0.00043	0.03	76.10	0.00	∞	pPGTGPpGDKGDTGPpGP	Collagen alpha-1(III) chain
0.00043	0.03	98.44	0.00	∞	GPpGpRGNRGERGSESGPHGPpGpGPpPGAPGP	Collagen alpha-1(III) chain;
0.00043	0.03	35.34	0.00	∞	DESGPSIVHRK	POTE ankyrin domain family member F
0.00043	0.03	39.07	0.00	∞	SORFPKAEF	Serum albumin
0.00049	0.03	549.42	263.62	2.08	PpGPpPGpPGpHGPpQGp	Collagen alpha-3(IV) chain;
0.00053	0.03	325.42	106.46	3.06	GpSpVpGpPGLAGERGEQGPpGPTGFQGLPG	Collagen alpha-2(V) chain
0.00058	0.03	62.79	1.13	55.56	GQAGpPGVMGpGPpGP	Collagen alpha-1(XV) chain
0.00059	0.03	86.47	15.86	5.45	GPpGESGREApGAEGSpGRDGSpGAKGDRGET	Collagen alpha-1(I) chain
0.00059	0.03	154.02	22.08	6.98	SpGGPpSDGKpGPpG	Collagen alpha-1(III) chain
0.00063	0.03	1364.22	5489.32	0.25	SpGPDGKTGPpGPA	Collagen alpha-1(I) chain
0.00063	0.03	8660.59	17696.57	0.49	ApGDRGEpGPpGPA	Collagen alpha-1(I) chain
0.00063	0.03	2729.94	4314.11	0.63	VGEpGPAGSKGESGNKGEpGSAGP	Collagen alpha-2(I) chain
0.00065	0.03	119.32	31.88	3.74	GEDGpGKpGRpGERG	Collagen alpha-2(I) chain
0.00065	0.03	63.74	1.17	54.59	VDVLKDSGRDYVSQFEGSALGKQLN	Apolipoprotein A-I
0.00074	0.04	81.80	685.14	0.12	VLNLGPITR	Uromodulin
0.00076	0.04	8.52	268.57	0.03	DDGEAGKpG	Collagen alpha-1(I) chain
0.00082	0.04	8.01	0.67	11.91	ApGHGPpGP	Collagen alpha-1(III) chain
0.00088	0.04	149.59	20.44	7.32	EPTPELPAEVLMEVEN	Lon protease homolog;
0.00097	0.04	110.37	71.94	1.53	GSpSGPpGpGDKTGPpGP	Collagen alpha-1(I) chain
0.00108	0.04	123.96	70.30	1.76	IQNWPHYRSP	Peptidoglycan recognition protein 1
0.00109	0.04	273.05	824.11	0.33	DGGGSPKGDVDP	Sodium/potassium-transporting ATPase subunit gamma
0.00125	0.05	31647.34	44427.77	0.71	ApGEDGRpGPpGP	Collagen alpha-1(II) chain;
0.00125	0.05	511.45	159.75	3.20	GEPGGKGERGApGEKGEpGP	Collagen alpha-1(III) chain
0.00127	0.05	1631.51	1576.19	1.04	IPVKQADSGSSEEEKQLYNKYPDAVAT	Osteopontin
0.00131	0.05	51.79	478.55	0.11	SGSVIDQSRVLNLGP	Uromodulin
0.00137	0.05	114.81	30.80	3.73	pGSDGKpGPpGSGQESGRpGPpGP	Collagen alpha-1(III) chain;
0.00147	0.05	2673.09	1326.80	2.01	PGRpGLDGERGRPGAPpGPpGPSSN	Collagen alpha-6(IV) chain;
0.00148	0.05	9.74	0.00	∞	LSALEEYTKKLT	Apolipoprotein A-I
0.00148	0.05	17.96	0.00	∞	RpGEVGPpGPpGP	Collagen alpha-1(I) chain
0.00148	0.05	72.00	0.00	∞	PpGKAGEDGpGKpGRpGERG	Collagen alpha-2(I) chain
0.00148	0.05	37.15	0.00	∞	GQpGApGVKGEpGApGENTPGGTGARG	Collagen alpha-2(I) chain
0.00149	0.05	179.11	22.41	7.99	LSSHIANVERVPFDAATLHTSTA	Gelsolin;
0.00155	0.05	2115.04	532.19	3.97	EEAPSLRPAPPISGGGY	Fibrinogen
0.00156	0.05	63.70	215.72	0.30	SpGEAGRpG	Collagen alpha-1(I) chain
0.00158	0.05	212.94	129.15	1.65	GLpGTGGPpGENGKpGEPpGP	Collagen alpha-1(III) chain;
0.00159	0.05	2087.54	4066.79	0.51	GPpGKpGDDGEAGKPG	Collagen alpha-1(II) chain;
0.00160	0.05	72.15	1.74	41.57	GEpGSpGENGApGQMGPRG	Collagen alpha-1(I) chain
0.00162	0.05	106.74	23.68	4.51	FGAELAAASLPAmPFTVGDNQTYRGFWNP	Receptor-type tyrosine-protein phosphatase U
0.00168	0.05	279.67	66.83	4.18	PSRSGENGpTGAVGFAGPQGPDGQpGVKGEp	Collagen alpha-2(V) chain

Given are the unadjusted P values, P values after adjustment, the mean relative abundance in cases and controls (for calculating means, values from all samples were used, considering 0 for undetected values*), the sequence, and the parental protein

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

von zur Muehlen, C;Koeck, T;Schiffer, E;Sackmann, C;Zuerbig, P;Hilgendorf, I;Reinoehl, J;Rivera, J;Zirlik, A;Hehrlein, C;Mischak, H;Bode, C;Peter, K

Title:

Urine proteome analysis as a discovery tool in patients with deep vein thrombosis and pulmonary embolism

Date:

2016-05-01

Citation:

von zur Muehlen, C., Koeck, T., Schiffer, E., Sackmann, C., Zuerbig, P., Hilgendorf, I., Reinoehl, J., Rivera, J., Zirlik, A., Hehrlein, C., Mischak, H., Bode, C. & Peter, K. (2016). Urine proteome analysis as a discovery tool in patients with deep vein thrombosis and pulmonary embolism. PROTEOMICS CLINICAL APPLICATIONS, 10 (5), pp.574-584. <https://doi.org/10.1002/prca.201500105>.

Persistent Link:

<http://hdl.handle.net/11343/291142>