Comparison of urine and plasma peptidome indicates selectivity in renal peptide handling

Pedro Magalhães[1,2], Claudia Pontillo[3], Martin Pejchinovski[1], Justyna Siwy[1], Magdalena Krochmal[1], Manousos Makridakis[4], Emma Carrick[5], Julie Klein[6,7], William Mullen[5], Joachim Jankowski[8,9], Antonia Vlahou[4], Harald Mischak[1,5], Joost P. Schanstra[6,7], Petra Zürbig[1]*, Lars Pape[2]

1. Mosaiques Diagnostics GmbH, Hannover, Germany

2. Department of Pediatric Nephrology, Hannover Medical School, Hannover, Germany

3. Department of Toxicology and Pharmacology, Hannover Medical School, Hannover, Germany

4. Biotechnology Division, Biomedical Research Foundation, Academy of Athens, Athens, Greece

5. Institute of Cardiovascular and Medical Sciences University of Glasgow, Glasgow, UK

6. Institut National de la Santé et de la Recherche Médicale (INSERM), Institute of Cardiovascular and Metabolic Disease, Toulouse, France

7. Université Toulouse III Paul-Sabatier, Toulouse, France

8. RWTH Aachen University Hospital, Aachen, Germany

9. Department of Pathology, Cardiovascular Research Institute Maastricht (CARIM), University of Maastricht, Maastricht, The Netherlands

***Corresponding author:**

Dr. Petra Zürbig

Mosaiques Diagnostics GmbH

Rotenburger Str. 20

30659 Hannover

Germany

Phone: +49 511 554744 15

FAX: + 49 511 554744 31

E-mail: zuerbig@mosaiques-diagnostics.com

Received: 11 07, 2017; Revised: 02 21, 2018; Accepted: 03 29, 2018

This article has been accepted for publication and 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.201700163 .

Keywords: peptide sequencing, peptidomics, plasma, urine

Abstract: 228 words

Manuscript: 5590 words

Figures: 4

Tables: 2

Statement of clinical relevance

Analysis of body fluids by peptidomics has demonstrated a significant value in the context of biomarker discovery. However, the origin of specific peptides generally remains unclear. With this study, we analyzed and sequenced human plasma and urinary peptides using a combination of techniques (CE-MS, and CE- and LC-MS/MS). As a first step towards predicting the origins of native peptides, we also investigated any overlaps in the data from the two sources. The results indicated surprisingly little overlap between plasma and urine peptides, also supported after investigating additional, previously published, urine and plasma peptidomics datasets. The most plausible hypothesis based on these data is that most plasma peptides may be reabsorbed in the renal tubules. Collagens appear to be an exception displaying a positive correlation and/or a much higher relative abundance in urine versus plasma suggesting the existence of a mechanism for their selective excretion.

Abstract

Purpose: Urine is considered to be produced predominantly as a result of plasma filtration in the kidney. However, the origin of the native peptides present in urine has never been investigated in detail. Therefore, we aimed to obtain a first insight into the origin of urinary peptides based on a side-by-side comprehensive analysis of the plasma and urine peptidome.

Methods: Twenty-two matched urine and plasma samples were analyzed for their peptidome using capillary electrophoresis coupled to mass spectrometry (CE-MS; for relative quantification) and CE- or LC coupled to tandem mass spectrometry (CE- or LC- MS/MS; for peptide identification). The overlap and association of abundance of the different peptides present in these two body fluids were evaluated.

Results: We were able to identify 561 plasma and 1461 urinary endogenous peptides. Only 90 peptides were detectable in both urine and plasma. No significant correlation was found when comparing the abundance of these common peptides, with the exception of collagen fragments. This observation was also supported when comparing published peptidome data from both plasma and urine.

Conclusions and clinical relevance: Most of the plasma peptides are not detectable in urine, possibly due to tubular reabsorption. The majority of urinary peptides may in fact originate in the kidney. The notable exception is collagen fragments, which indicates potential selective exclusion of these peptides from tubular reabsorption. Experimental verification of this hypothesis is warranted.

1. Introduction

Proteomics via mass spectrometry (MS) has become a widely used technique for detecting proteins and characterizing their abundances in a range of biological samples.[1,2] Multiple studies have been published focusing on the urinary[3-6] and plasma proteomes,[7,8] as these two fluids are considered easily accessible, valuable sources of information. After sampling, the urine proteome content is relatively stable (in part due to the fact that urine has normally been "stored" in the bladder for several hours, hence any proteolytic processes are completed), as opposed to blood where proteases are activated after collection.[8-10]

Plasma and urine contain thousands of proteins that are expected to be linked to a large extent. The kidney filters ~ 1500 L of blood daily in the glomeruli, generating ~170 L primary urine (assuming a GFR of 120 ml/min). Most of its components are further reabsorbed in the tubules, while the excreted urine volume represents about 1% of the original ~170 L volume.[11] Consequently, urinary protein composition is the result of a combination of plasma proteins filtered into the urine by glomerular filtration minus tubular reabsorption, and proteins originating from the kidney and urinary tract (*e.g.* secreted from renal tubular epithelial cells, and/or shedding of whole cells along the urinary tract, including shedding of apical membranes of renal tubular epithelial cell, and exosome secretion).[11-13]

Extracting information from online databases, Jia et al.[12] compared plasma and urine proteomes, characterizing their potential overlap. Based on this comparison, they were able to study the potential association of detected proteins with kidney function. The authors proposed that this methodology, based on proteomic techniques, might be relevant for understanding organ (kidney) functions and body homeostasis. In 2014, Farrah et al.[14] compared kidney, urine and plasma proteome based on relative protein abundances, determining potential kidney-specific proteins, *de novo* urine proteins, and proteins filtered from plasma. The resulting data confirmed that the proteome of urine and plasma are associated.[12] It has been suggested that under physiological conditions 70% of the urinary proteins originate from the kidney and the urinary tract, and 30% are plasma proteins.[13] In

kidney disease, the glomerular filtration barrier is altered, affecting directly the above-mentioned distribution.

To date, the majority of studies comparing plasma and urine focused on proteome analysis.[12,14] In the context of peptide profiling, plasma provides a more stable peptidome readout than serum.[15] This is mostly due to the fact that proteolytic activity involved in coagulation is higher in serum than in plasma.[15,16] Recently, Parker et al.[17] described over 5500 endogenous plasma peptides, providing a comprehensive description of the plasma native peptidome. On the other hand, urinary native peptides have been extensively investigated, especially in the context of kidney disease,[18,19] clearly showing that peptides are a rich source of information. In 2008, Coon et al.[20] described the first peptidomic database on naturally occurring urinary peptides, named Human Urinary Proteome database. This database is regularly updated, currently storing datasets from over 50 000 urine samples (healthy controls and patients with different diseases), encompassing endogenous peptide signals and respective sequence information.[21,22] Multiple studies have shown that such endogenous urinary peptides combined in classifiers represent powerful tools in the diagnosis and clinical assessment of several diseases such as kidney diseases,[23-25] different cancers,[26,27] and cardiovascular diseases.[28-30] However, less is known about the origin of these peptides. At first sight, it seems reasonable to assume that the origin of peptides follows the same rule as the proteins: about 30% of urinary peptides may originate from plasma by the process of filtration, while the major fraction originates from the kidney and the urinary tract. However, since peptides are expected to pass the filtration barrier in the glomerulus and in general should all be collected in urine, we hypothesized that a higher fraction (in comparison to proteins) of plasma-derived peptides are present in urine. Unfortunately, no evidence or data (not even indirectly) are readily available to support this (or any other) hypothesis.

As a first step towards investigating the origin of endogenous urinary peptides, we analyzed twenty-two matched urine and plasma samples by CE-MS profiling and tandem-MS. Based on these data, we established a map of the plasma and urinary peptidome dataspaces in these individuals, and determined their overlap, and abundance correlation. These data were further placed in the context of existing literature and available plasma and urine peptidomic datasets. Collectively, the results support that most of the plasma peptides are not detectable in urine, possibly due to tubular reabsorption, and lead to the additional hypothesis of potential selective exclusion of specific peptides (mainly collagens) from this process.

2. Material and methods

2.1. Urine and plasma collection

Plasma and urine samples derived from twenty-two subjects were analyzed to obtain CE-MS peptide profiles and CE-MS/MS and LC-MS/MS peptide sequences. **Table 1** reports the characteristics of the subjects.

Urine samples were collected, stored at -20°C and shipped frozen for subsequent peptidome analysis. All blood samples were taken via vein puncture, collected in anticoagulant-treated tubes (9 heparin and 13 EDTA), immediately cooled to 4°C and centrifuged at 1200×g for 10 min.[31] The plasma was isolated and stored at -80°C. All plasma and urine samples were collected at the same time point.

The study was conducted fulfilling all the laws on the protection of individuals being involved in medical research and in accordance with the principles of the Declaration of Helsinki. All plasma and urine data were obtained anonymized. The local ethics committee from Hannover Medical School (Hannover, Germany) approved the approach, employing anonymized samples and proteomics data (Ethical ID: 3596-2017).

2.2. Plasma and urine sample preparation

For urine samples 0.7 mL and for plasma samples 0.3 mL plus 0.4 mL of H_2O (as described in,[31] without delipidation) was used and diluted with 0.7 mL of 8 M urea, 10 mM NH_4OH, and 0.02% SDS. Ultrafiltration was performed to remove high molecular weight proteins, using Centrisart ultracentrifugation filters (20 kDa MW cut-off; Sartorius) at 2,600×g. To remove urea, electrolytes, salts, and thereby to decrease matrix effects, a volume of 1.1 mL of the filtrate was applied onto a PD-10 desalting column (Amersham Bioscience, Uppsala, Sweden) equilibrated in 0.01% NH4OH in high-performance liquid chromatography (HPLC)-grade H_2O (Roth, Germany). After rinsing the column with 1.9 ml of 0.01% NH4OH in H_2O, 2 mL of HPLC-grade H_2O was applied and the resulting eluate was collected. The eluate was lyophilized, stored at 4°C and resuspended in HPLC-grade H_2O shortly before analysis, as previously described.[32] The generation of this protocol and basic features (reproducibility, peptide recovery, etc.) are reported in detail in Theodorescu et al..[33] In addition, performance characteristics are comprehensively described in Mischak et al..[32]

2.3. CE-MS analysis and data processing

CE-MS analysis was performed as previously described[32] in the same fashion for urine and plasma. A P/ACE MDQ capillary electrophoresis system (Beckman Coulter, Fullerton, CA)

was coupled with a Micro-TOF MS (Bruker Daltronic, Bremen, Germany). A solution of 20% acetonitrile (Sigma-Aldrich, Taufkirchen, Germany) in HPLC-grade water (Roth, Karlsruhe, Germany) supplemented with 0.94% formic acid (Sigma-Aldrich) was used as running buffer. For CE-MS analysis, the electrospray ionization interface from Agilent Technologies (Palo Alto, CA) was set to a potential of -4.0 to -4.5 kV. Spectra were recorded over an m/z range of 350-3000 and accumulated every 3 s. Monoisotopic mass signals were determined for $z \leq 6$ with resolution in the range of 10 000 and a limit of detection of ~1 fmol as previously described.[32]

Mass spectral ion peaks representing identical molecules at different charge states were deconvoluted into single masses using MosaiquesVisu software, which employs a probabilistic clustering algorithm and uses isotopic distribution (for $z \leq 6$) as well as conjugated masses for charge-state determination of peptides/proteins.[34] Only signals with $z>1$ observed in a minimum of 3 consecutive spectra with a signal-to-noise ratio of at least 4 were considered. All detected peptides were clustered, matched and deposited in a Microsoft Structured Query Language (SQL) database.[20] The result is a peak list, characterizing each protein and peptide by its molecular mass [Da] and normalized CE migration time [min]. Migration time and ion signal intensity (amplitude) of the urinary peptides were normalized similarly to previous studies using internal peptides.[35] For normalization of CE-time of the plasma peptides, we used 470 peptides. 20 internal standard peptides present in almost all plasma samples were used for normalization of signal intensity.

Analysis by CE-MS was employed for the relative quantification of the peptides in each sample, whereas MS/MS analysis (coupled to either CE- or LC to maximize coverage) was used for peptide sequencing (described below). Only peptides identified in the cohort described in the present study were reported and further combined into a peptide list. Missing values (in general the result of abundance being below the detection limit) in the individual samples were considered as "zero".

2.4. MS/MS analysis

CE-MS/MS analysis was performed as described.[36] Concisely, the CE was connected to an Orbitrap Velos FTMS (Thermo Finnigan, Bremen, Germany). Sample was ionized using a modified Proxeon nano spray source fitted with a non-grounded Agilent ESI sprayer operating in positive ion mode. Ionization voltage was 3.4 kV and the capillary temperature was 275°C. The mass spectrometer was operated in MS/MS mode scanning from 350 to 1500 amu. The top five multiply charged ions were selected from each scan for MS/MS analysis

using HCD at 40% collision energy. The resolution of ions in MS1 was 60.000 and 7.500 for MS2.

For LC-MS/MS analysis, aliquots of 5 µL were analyzed on a Dionex Ultimate 3000 RSLS nano flow system (Dionex, Camberly, UK) at a flowrate of 5 µL/min. The trap and nano flow column were maintained at 35°C. After loading (5 µL) onto a Dionex 0.1 × 20 mm 5 µm C18 nano trap column, elution was performed on an Acclaim PepMap C18 nano column 75 µm × 15 cm, 2 µm 100 °A at a flowrate of 0.3 µL/min. Typically, samples were eluted with a gradient of solvent A: 97.9% H_2O, 0.1% formic acid, 2% acetonitrile versus solvent B: 80% acetonitrile, 19.9% H_2O, 0.1% formic acid starting at 1% B for 5 min rising to 20% B after 90 min and finally to 40%B after 120 min.

Thereafter, the column was washed and reequilibrated prior to the next injection. The eluant was ionized using a Proxeon nano spray ESI source operating in positive ion mode into an Orbitrap Velos FTMS (Thermo Finnigan, Bremen, Germany). Ionization voltage was 2.6 kV and the capillary temperature was 250°C. The mass spectrometer was operated in HCD MS/MS mode. Using HCD, the top 20 multiply charged ions were selected from each scan for MS/MS analysis and only charge state one was rejected for MS/MS.[36] The detection limit for the LC- or CE-MS/MS analysis using the Orbitrap Velos mass spectrometer, with 60 000 resolution for MS1 and with 7500 resolution for MS2, was in the range of 0.05 – 0.2 fmol.

In order to obtain sequence information, CE- and LC-MS/MS were used as complementary approaches.[36] CE-MS/MS has the advantage of direct matching (mass and CE-time) to the peptides quantified by CE-MS. On the other hand, LC-MS/MS exhibits higher sensitivity due to the increased loading capacity of the LC-column, consequently a better coverage of sequence information. However, the retention time in LC does not directly correspond to the CE migration time, which represents a disadvantage of LC-MS/MS in this approach.

The obtained sequences were matched to the previously quantified peptides assessed by CE-MS (section 2.3).

2.5. Sequence data analysis

Data files were searched against the UniProt human non-redundant database using Proteome Discoverer 1.2 (Thermo) and the SEQUEST search engine. Relevant settings were: no fixed modifications, oxidation of methionine and proline as variable modifications. The minimum precursor mass was set to 790 Da, maximum precursor mass to 6000 Da with a minimum peak count of 10. The high confidence peptides were defined by cross-correlation (Xcorr)

≥1.9 and rank=1. Precursor mass tolerance was 5 ppm and fragment mass tolerance was 0.05 Da.

False discovery rate settings cannot guarantee that the correct sequence is assigned to a fragmentation spectrum generated in an MS/MS experiment.[37] However, a property of CE is that the migration time is dependent on the net positive charge of the peptide. At pH 2 (pH of the running buffer) this (charge) is a function of the number of basic amino acids present.[38] The peptide sequences obtained from the tandem MS analysis were matched to the CE-MS peaks by matching the molecular mass and migration time (in the case of CE-MS/MS) or molecular mass and theoretical migration time based on the number of basic amino acids (in the case of LC-MS/MS). In addition and to further allow matching, the absence of cysteine and absence of hydroxylated proline in non-collagen peptides were required. If a sequence passed all of these criteria it was then assigned to the corresponding CE-MS peak.[38]

2.6. Peptides data processing

The mean peptide abundance was calculated by multiplying the mean of the amplitudes of each individual peptide with its frequency (as defined by the occurrence of each peptide present in a given cohort).

Prior to the statistical analysis, the relative abundance of peptides in the respective body fluid (urine or plasma) was calculated following the equation:

Relative abundance of specific peptide (%) = Mean peptide abundance*100 / Total abundance of all overlapping peptides

2.7. Statistical analysis

Pearson correlation coefficient *r* and regression analysis were performed to evaluate the association of the abundances of respective peptides identified in plasma and urine. P-values less than 0.05 were considered statistically significant. MedCalc software (version 12.7.5.0; MedCalc Software, Mariakerke, Belgium) was used to perform these analyses and to generate the corresponding graphs.

3. Results

3.1. Plasma peptidome profile

The plasma data comprised of 9 heparin and 13 EDTA samples. In previous experiments, we have compared both plasma types and found no significant differences between them.

Analysis of the 22 aforementioned plasma samples by CE-MS resulted in the detection of approximately 700 peptide signals per sample, which could be combined in a total of 1952 unique endogenous peptides. Sequence information was obtained for 561 out of the 1952

(28%) plasma peptides using a combination of CE- and LC-MS/MS analysis (supplementary table 1, 'plasma'). When considering the abundance of the peptides in plasma, these 561 sequenced peptides covered approximately 24% of the total detected peptide-signal.

The intensity across the sequenced plasma peptides corresponded by 55.9% to complement fragments (C3 and C4-A), 25.1% fibrinogen fragments (alpha and beta chain), 5.5% apolipoprotein fragments, 4.1% different collagen fragments, and 9.4% fragments from various proteins (*e.g.* serum albumin, thymosin beta-4, alpha-2-antiplasmin, alpha-1-antitrypsin and inter-alpha-trypsin inhibitor heavy chain H1 and H4) (Table 2A). Moreover, when sorted based on abundance, the 10 most abundant plasma peptide fragments were seven complement C3, two-fibrinogen alpha chain, and one-apolipoprotein A-IV peptide (supplementary table 1, 'plasma').

3.2. Urine peptidome profile

CE-MS analysis of the 22 individual urine samples led to the detection of approximately 1800 peptide signals per sample, which could be combined in a list of 3955 unique urinary endogenous peptides. Sequence information was obtained for 1461 peptides (37%) (supplementary table 1, 'urine'). The abundance of the aforementioned sequenced peptides covered approximately 52% of the total detected urine peptide-signal.

Of the total abundance of identified (sequenced) peptides in urine, collagen fragments comprised 79.5%, uromodulin 8.2%, fibrinogen 6%, and 6.3% were fragments originated from other proteins such as polymeric-immunoglobulin receptor, CD99 antigen, sodium/potassium-transporting ATPase subunit gamma, and membrane associated progesterone receptor component 1 (Table 2B). Along these lines and based on signal intensity (abundance), the 10 most abundant urinary peptides comprised eight collagen fragments (type I, II and III), one uromodulin and one fibrinogen fragment (supplementary table 1, 'urine').

3.3. Comparison of plasma and urine peptidome profiles

3.3.1. Overlap of urinary and plasma peptides

Subsequently, we investigated the overlap of urinary and plasma peptides. This analysis resulted in the identification of 90 common peptide sequences (Figure 1), including collagen and fibrinogen fragments and fragments derived from other proteins such as serum albumin, osteopontin, apolipoprotein (A-1 and L1) and alpha-2-HS-glycoprotein. Corresponding sequence information and relative abundance of these peptides in plasma and urine are reported in supplementary table 2. As shown (supplementary tables 1-2), the majority of

sequenced peptides were exclusively present in plasma or in urine. Among the peptides apparently present only in plasma were peptides derived from plasminogen activator inhibitor 1, H/ACA ribonucleoprotein complex subunit, guanine nucleotide exchange factor VAV3, complement C4-A, apolipoprotein E, alpha-2-antiplasmin, and 5-oxoprolinase. Peptides seemingly restricted to urine, therefore likely originating from kidney or urinary tract were peptides derived from polymeric-immunoglobulin receptor, CD99 antigen, sodium/potassium-transporting ATPase subunit gamma, and membrane associated progesterone receptor component 1.

3.3.2. Correlation of overlapping peptides in urine and plasma

Next, we studied the correlation of the relative abundance (%) of identical peptides in urine and plasma. No significant correlation (r = 0.05; *p-value* = 0.632) could be observed when investigating all 90 overlapping peptides (Figure 2A). However, when investigating only the collagen fragments (n=71, representing 78% of the overlapping peptides), a significant correlation of abundance in urine and plasma could be detected (r = 0.32; *p-value* = 0.007) (Figure 2B). Moreover, investigating only collagen alpha-1 (I) chain fragments (n=50, representing 56% of the overlapping peptides) a correlation of apparently even higher significance was found (r = 0.44; *p-value* = 0.002) (Figure 2C).

3.4. Placing of findings in the existent literature

To investigate if our results are also reflected in published urine and plasma peptidomics datasets, two well-described sequenced peptide lists were further analyzed [17,21,22]. Specifically, for urinary peptides, sequence information was extracted from the Human Urinary Proteome database.[21,22] We were able to retrieve 3366 urinary peptides, where high-confidence sequence information was assigned.[21,22] In the case of plasma, sequence information of 5548 native plasma peptides, previously identified by Parker et al., was extracted.[17] Evaluation of the overlap of the two sets resulted in the identification of 124 common peptides in plasma and urine (Figure 3A). In agreement with the findings above, no significant correlation (p=0.09) of the relative abundance of these peptides could be observed (Figure 3B).

4. Discussion

The urinary and plasma peptide profiles of 22 individuals were evaluated with the aim of generating a map of the urinary and plasma peptidome, and creating a hypothesis on the origin of the former. The data generated expands the current knowledge in the field, which

may also be helpful for further clinical proteomic studies, where peptide-biomarkers may be targeted based on their potential origin.

In the current study, we identified and sequenced 1461 urinary peptides. Surprisingly, most of these peptides were exclusively found in urine such as uromodulin, CD99 antigen, and polymeric-immunoglobulin receptor, suggesting that these are kidney or urinary tract derived. Given that urine represents a concentrate plasma filtrate, we cannot rule out the possibility that some of these peptides may in fact be present in plasma, yet at very low abundance (below our detection limits). Interestingly, many of these aforementioned peptides have been found to be associated with kidney diseases.[23,39,40] In addition, collagen fragments were the most abundant peptides in urine, as also previously reported.[20,23]

In contrast, the plasma peptidome comprised mostly of peptides originating from proteins associated with the circulatory system.[10] In this study, we were able to sequence 561 plasma peptides. Peptide fragments belonging to apolipoprotein E and complement C4-A proteins were uniquely identified in plasma. Furthermore, peptides derived from complement and fibrinogen represented the most abundant plasma peptides. These findings are in line with previous observations.[17,41] In addition, when comparing the peptide relative abundances in our plasma dataset with that of the plasma peptidome described by Parker et al.,[17] a significant correlation was observed (data not shown).

The overlap between plasma and urine peptides (based on peptide sequence) was surprisingly small, considering that plasma peptides are expected to be filtered by the kidney and consequently, are expected to be typically present in urine. Only 90 overlapping (identical) peptides derived from various proteins such as collagen, serum albumin, osteopontin, alpha-1-antitrypsin, and apolipoprotein fragments could be identified, which in the vast majority of cases, did not correlate with abundance. Such low overlap is also observed when investigating previously published urine and plasma peptidomics datasets.[17,21,22] This lack of overlap may in part be associated with the low abundance of some of these peptides in plasma, not allowing their detection. In contrast, the primary urine is concentrated ~100-fold in the tubulus (from about ~150-180 L to 1.5 L daily output);[11] hence, the concentration of peptides in urine also is expected to be increased, facilitating their identification. The most plausible hypothesis for this finding relates to the tubular reabsorption process: under physiological conditions, proteins and peptides in the primary urine are thought to be removed/reabsorbed in the proximal tubule[11,42] by two receptors, megalin and cubilin.[43] This process does not appear to be selective and should apply to all proteins and peptides

equally. As a result, general reduction of the amount of plasma proteins and peptides in urine should be observed. Differential proteolytic activity may also contribute to the observed differences. Plasma and urinary peptides might be a result of different proteolytic processes. Hence, a peptide derived from the circulation may be cleaved by a kidney-specific protease (*e.g.* dipeptidyl peptidase-4 (DPP4)),[44] which could ultimately result in further peptide degradation, revealing a different sequence even if the peptide is from the same protein origin.

Interestingly and as an exception, when investigating collagen-derived peptides, or more specifically, type I collagen-derived peptides, a significant correlation of the relative abundance in urine and plasma could be observed. This finding indicates that collagen derived peptides may be selectively excluded from tubular reabsorption via a yet unknown mechanism (Figure 4). This is a hypothesis meriting further investigation, including studies addressing the existence of biological functions for some of these peptides.

CE and LC coupled to MS present certain advantages and disadvantages, which were previously described in detail.[36,45] Briefly, for CE a constant flow and voltage settings are used. This approach displays high robustness and fast analysis time, as well as high peak resolution. LC has a higher sensitivity, due to the possibility of loading higher amounts of sample. In addition, separation time (and resolution) can be extended by applying a shallow gradient, which allows better coverage (increased number of peptide identifications).

The study has several limitations, but none of these appear to have a severe impact on the main conclusion: that the plasma peptidome bears very little similarity to the urine peptidome, and that our findings indicate selective exclusion of specific peptides from tubular reabsorption. A certain limitation of the study may be the consistently lower number of peptides identified in plasma in comparison to urine. However, this is likely due to the aforementioned differences in abundance. The expected increase in relative abundance of peptides in urine may be attenuated by tubular reabsorption, but apparently not for all peptides equally. An additional limitation of the study may be the relatively low number of subjects included. For further refinement of the peptide content especially of plasma, additional studies in an expanded sample cohort and/or under different sample collection and preparation conditions may be required. However, the observed consistency of our results with earlier reports,[17,41] including the most abundant peptides in plasma, supports the validity of the findings presented here. Further work has to be conducted towards the identification (at the amino acid level) of the yet unidentified peptides, putatively associated

with the presence of post-translational modifications.[46] However, we could identify close to 500 peptides in this study, indicating a quite valid representation of the peptides typically expected in plasma.

To our knowledge, this study represents the first approach to compare plasma and urinary peptidome based on the same methodology and samples from same subjects, with novel and quite unexpected findings. Comparing plasma and urine peptidome may allow a better understanding of the origin of the urinary peptides. Furthermore, this study revealed various interesting findings opening up new research avenues for the future: the presence of specific peptides (collagen fragments), significantly correlated with and/or at much higher relative abundance in urine versus plasma may be suggestive of the existence of an underlying reason and mechanism for their selective excretion - a hypothesis meriting further experimental investigation. Future studies will need to reveal the physiology behind this observation, and utilize this knowledge to increase the specificity of urinary peptide markers for a given disease.

Acknowledgments

The research presented in this manuscript was supported in part by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No. 642937 (RENALTRACT) and by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreements 305507 (HoMAGE), 603288 (SysVasc), and 608332 (iMODE-CKD).

Conflicts of interest

H. Mischak is the founder and co-owner of Mosaiques Diagnostics GmbH, who developed the CE-MS technology for clinical application. P. Magalhães, M. Pejchinovski, J. Siwy, M. Krochmal and P. Zürbig are employees of Mosaiques Diagnostics GmbH.

5. References

[1] R. Aebersold, M. Mann. *Nature* **2003**, 422, 198.

[2] D. Theodorescu, H. Mischak. *World J Urol* **2007**, 25, 435.

[3] J. Adachi, C. Kumar, Y. Zhang, J.V. Olsen, M. Mann. *Genome Biol.* **2006**, 7, R80.

[4] A. Kentsis, F. Monigatti, K. Dorff, F. Campagne, R. Bachur, H. Steen. *Proteomics. Clin. Appl.* **2009**, 3, 1052.

[5] A. Marimuthu, R.N. O'Meally, R. Chaerkady, Y. Subbannayya, V. Nanjappa, P. Kumar, D.S. Kelkar, S.M. Pinto, R. Sharma, S. Renuse, R. Goel, R. Christopher, B. Delanghe, R.N. Cole, H.C. Harsha, A. Pandey. *J. Proteome. Res.* **2011**, 10, 2734.

[6] N. Nagaraj, M. Mann. *J. Proteome. Res.* **2011**, 10, 637.

[7] G.S. Omenn, D.J. States, M. Adamski, T.W. Blackwell, R. Menon, H. Hermjakob, R. Apweiler, B.B. Haab, R.J. Simpson, J.S. Eddes, E.A. Kapp, R.L. Moritz, D.W. Chan, A.J. Rai, A. Admon, R. Aebersold, J. Eng, W.S. Hancock, S.A. Hefta, H. Meyer, Y.K. Paik, J.S. Yoo, P. Ping, J. Pounds, J. Adkins, X. Qian, R. Wang, V. Wasinger, C.Y. Wu, X. Zhao, R. Zeng, A. Archakov, A. Tsugita, I. Beer, A. Pandey, M. Pisano, P. Andrews, H. Tammen, D.W. Speicher, S.M. Hanash. *Proteomics.* **2005**, 5, 3226.

[8] T. Farrah, E.W. Deutsch, G.S. Omenn, D.S. Campbell, Z. Sun, J.A. Bletz, P. Mallick, J.E. Katz, J. Malmstrom, R. Ossola, J.D. Watts, B. Lin, H. Zhang, R.L. Moritz, R. Aebersold. *Mol. Cell Proteomics.* **2011**, 10, M110.

[9] H. Mischak, C. Delles, A. Vlahou, R. Vanholder. *Nat. Rev. Nephrol.* **2015**, 11, 221.

[10] S. Hu, J.A. Loo, D.T. Wong. *Proteomics.* **2006**, 6, 6326.

[11] V. Thongboonkerd, P. Malasit. *Proteomics.* **2005**, 5, 1033.

[12] L. Jia, L. Zhang, C. Shao, E. Song, W. Sun, M. Li, Y. Gao. *PLoS. ONE.* **2009**, 4, e5146.

[13] S. Kalantari, A. Jafari, R. Moradpoor, E. Ghasemi, E. Khalkhal. *Int. J. Proteomics.* **2015**, 2015, 782798.

[14] T. Farrah, E.W. Deutsch, G.S. Omenn, Z. Sun, J.D. Watts, T. Yamamoto, D. Shteynberg, M.M. Harris, R.L. Moritz. *J. Proteome. Res.* **2014**, 13, 60.

[15] W. Kolch, C. Neususs, M. Pelzing, H. Mischak. *Mass Spectrom Rev.* **2005**, 24, 959.

[16] H. Tammen, I. Schulte, R. Hess, C. Menzel, M. Kellmann, T. Mohring, P. Schulz-Knappe. *Proteomics.* **2005**, 5, 3414.

[17] B.L. Parker, J.G. Burchfield, D. Clayton, T.A. Geddes, R.J. Payne, B. Kiens, J.F.P. Wojtaszewski, E.A. Richter, D.E. James. *Mol. Cell Proteomics.* **2017**, 16, 2055.

[18] M. Krochmal, J.P. Schanstra, H. Mischak. *Expert. Opin. Drug Discov.* **2018**, 13, 259.

[19] E. Nkuipou-Kenfack, P. Zurbig, H. Mischak. *Proteomics. Clin. Appl.* **2017**, 11.

[20] J.J. Coon, P. Zürbig, M. Dakna, A.F. Dominiczak, S. Decramer, D. Fliser, M. Frommberger, I. Golovko, D.M. Good, S. Herget-Rosenthal, J. Jankowski, B.A. Julian, M. Kellmann, W. Kolch, Z. Massy, J. Novak, K. Rossing, J.P. Schanstra, E. Schiffer, D. Theodorescu, R. Vanholder, E.M. Weissinger, H. Mischak, P. Schmitt-Kopplin. *Proteomics Clin. Appl.* **2008**, 2, 964.

[21] A. Stalmach, A. Albalat, W. Mullen, H. Mischak. *Electrophoresis* **2013**, 34, 1452.

[22] J. Siwy, W. Mullen, I. Golovko, J. Franke, P. Zurbig. *Proteomics. Clin. Appl.* **2011**, 5, 367.

[23] D.M. Good, P. Zürbig, A. Argiles, H.W. Bauer, G. Behrens, J.J. Coon, M. Dakna, S. Decramer, C. Delles, A.F. Dominiczak, J.H. Ehrich, F. Eitner, D. Fliser, M. Frommberger, A. Ganser, M.A. Girolami, I. Golovko, W. Gwinner, M. Haubitz, S. Herget-Rosenthal, J. Jankowski, H. Jahn, G. Jerums, B.A. Julian, M. Kellmann, V. Kliem, W. Kolch, A.S. Krolewski, M. Luppi, Z. Massy, M. Melter, C. Neususs, J. Novak, K. Peter, K. Rossing, H. Rupprecht, J.P. Schanstra, E. Schiffer, J.U. Stolzenburg, L. Tarnow, D. Theodorescu, V. Thongboonkerd, R. Vanholder, E.M. Weissinger, H. Mischak, P. Schmitt-Kopplin. *Mol. Cell Proteomics* **2010**, 9, 2424.

[24] J. Klein, A. Ramirez-Torres, A. Ericsson, Y. Huang, B. Breuil, J. Siwy, H. Mischak, X.R. Peng, J.L. Bascands, J.P. Schanstra. *Kidney Int.* **2016**, 90, 1045.

[25] J. Metzger, T. Kirsch, E. Schiffer, P. Ulger, E. Mentes, K. Brand, E.M. Weissinger, M. Haubitz, H. Mischak, S. Herget-Rosenthal. *Kidney Int.* **2010**, 78, 1252.

[26] M. Frantzi, J. Metzger, R.E. Banks, H. Husi, J. Klein, M. Dakna, W. Mullen, J.J. Cartledge, J.P. Schanstra, K. Brand, M.A. Kuczyk, H. Mischak, A. Vlahou, D. Theodorescu, A.S. Merseburger. *J. Proteomics.* **2014**, 98, 44.

[27] M. Frantzi, K.E. van Kessel, E.C. Zwarthoff, M. Marquez, M. Rava, N. Malats, A.S. Merseburger, I. Katafigiotis, K. Stravodimos, W. Mullen, J. Zoidakis, M. Makridakis, M. Pejchinovski, E. Critselis, R. Lichtinghagen, K. Brand, M. Dakna, M.G. Roubelakis,

D. Theodorescu, A. Vlahou, H. Mischak, N.P. Anagnou. *Clin Cancer Res.* **2016**, 22, 4077.

[28] C.E. Brown, N.S. McCarthy, A.D. Hughes, P. Sever, A. Stalmach, W. Mullen, A.F. Dominiczak, N. Sattar, H. Mischak, S. Thom, J. Mayet, A.V. Stanton, C. Delles. *Proteomics Clinical Applications* **2015**, 9, 610.

[29] U. Neisius, T. Koeck, H. Mischak, S.H. Rossi, E. Olson, D.M. Carty, J.A. Dymott, A.F. Dominiczak, C. Berry, K.G. Oldroyd, C. Delles. *BMC. Cardiovasc. Disord.* **2016**, 16, 70.

[30] C. von zur Muhlen, E. Schiffer, C. Sackmann, P. Zurbig, I. Neudorfer, A. Zirlik, N. Htun, A. Iphofer, L. Jansch, H. Mischak, C. Bode, Y.C. Chen, K. Peter. *Mol. Cell Proteomics* **2012**, 11, M111.

[31] E. Schiffer, S. Liabeuf, C. Lacroix, M. Temmar, C. Renard, B. Monsarrat, G. Choukroun, H.D. Lemke, R. Vanholder, H. Mischak, Z.A. Massy. *J. Hypertens.* **2011**, 29, 783.

[32] H. Mischak, A. Vlahou, J.P. Ioannidis. *Clin. Biochem.* **2013**, 46, 432.

[33] D. Theodorescu, D. Fliser, S. Wittke, H. Mischak, R. Krebs, M. Walden, M. Ross, E. Eltze, O. Bettendorf, C. Wulfing, A. Semjonow. *Electrophoresis* **2005**, 26, 2797.

[34] N. Neuhoff, T. Kaiser, S. Wittke, R. Krebs, A. Pitt, A. Burchard, A. Sundmacher, B. Schlegelberger, W. Kolch, H. Mischak. *Rapid Commun. Mass Spectrom.* **2004**, 18, 149.

[35] J. Jantos-Siwy, E. Schiffer, K. Brand, G. Schumann, K. Rossing, C. Delles, H. Mischak, J. Metzger. *J. Proteome. Res.* **2009**, 8, 268.

[36] J. Klein, T. Papadopoulos, H. Mischak, W. Mullen. *Electrophoresis* **2014**, 35, 1060.

[37] M. Pejchinovski, J. Klein, A. Ramirez-Torres, V. Bitsika, G. Mermelekas, A. Vlahou, W. Mullen, H. Mischak, V. Jankowski. *Proteomics. Clin. Appl.* **2015**, 9, 531.

[38] P. Zürbig, Renfrow M.B., E. Schiffer, J. Novak, M. Walden, S. Wittke, I. Just, Pelzing M, C. Neususs, D. Theodorescu, C. Root, M. Ross, H. Mischak. *Electrophoresis* **2006**, 27, 2111.

[39] J. Drube, E. Schiffer, E. Lau, C. Petersen, M. Kirschstein, M.J. Kemper, R. Lichtinghagen, B. Ure, H. Mischak, L. Pape, J.H. Ehrich. *Pediatrics* **2012**, 129, e356-e363.

[40] M. Pejchinovski, J. Siwy, J. Metzger, M. Dakna, H. Mischak, J. Klein, V. Jankowski, K.T. Bae, A.B. Chapman, A.D. Kistler. *Nephrol. Dial. Transplant.* **2016**.

[41] Y. Shen, T. Liu, N. Tolic, B.O. Petritis, R. Zhao, R.J. Moore, S.O. Purvine, D.G. Camp, R.D. Smith. *J. Proteome. Res.* **2010**, 9, 2339.

[42] S. Decramer, P.A. Gonzalez de, B. Breuil, H. Mischak, B. Monsarrat, J.L. Bascands, J.P. Schanstra. *Mol Cell Proteomics* **2008**, 7, 1850.

[43] R. Nielsen, E.I. Christensen, H. Birn. *Kidney Int.* **2016**, 89, 58.

[44] U. Panchapakesan, C. Pollock. *Front Immunol.* **2015**, 6, 443.

[45] W. Mullen, A. Albalat, J. Gonzalez, P. Zerefos, J. Siwy, J. Franke, H. Mischak. *Electrophoresis* **2012**, 33, 567.

[46] H. Mischak, J.J. Coon, J. Novak, E.M. Weissinger, J.P. Schanstra, A.F. Dominiczak. *Mass Spectrom. Rev.* **2009**, 28, 703.

Table 1. Characteristics of the study population

Characteristics	
Number of subjects	22
Age (years)	46.59±15.42
Gender (F/M)	10/12

Table 2. Distribution of plasma and urinary peptides, indicating the percentage related to the total number of identified peptides and the coverage (abundance) across the sequenced peptides

A. Plasma peptides

	Number of identified peptides	***% of total number of identified peptides***	***% coverage of total detected peptide-signal (abundance)***
Complement (C3 and C4-A)	49	8.7%	55.9%
Fibrinogens (alpha and beta chains)	182	32.4%	25.1%
Apolipoproteins	82	14.6%	5.5%
Collagens	126	22.5%	4.1%
Others	122	21.8%	9.4%

B. Urinary peptides

	Number of identified peptides	***% of total number of identified peptides***	***% coverage of total detected peptide-signal (abundance)***
Collagens	966	66.1%	79.5%
Uromodulin	29	2%	8.2%
Fibrinogens (alpha and beta chains)	26	1.8%	6%
Others	440	30.1%	6.3%

Figure legends

Figure 1. Comparison between urine and plasma based on the identical sequenced peptides

1371 peptides exclusively identified in urine (e.g. polymeric-immunoglobulin receptor, CD99 antigen, sodium/potassium-transporting ATPase subunit gamma, and membrane associated progesterone receptor component 1); 90 overlapping peptides (collagen fragments, fibrinogen fragments and fragments derived from other proteins (e. g. serum albumin, osteopontin, apolipoprotein and alpha-2-HS-glycoprotein)); 471 peptides only identified in plasma (e.g. plasminogen activator inhibitor 1, H/ACA ribonucleoprotein complex subunit, guanine nucleotide exchange factor VAV3, complement C4-A, apolipoprotein E, alpha-2-antiplasmin, and 5-oxoprolinase)

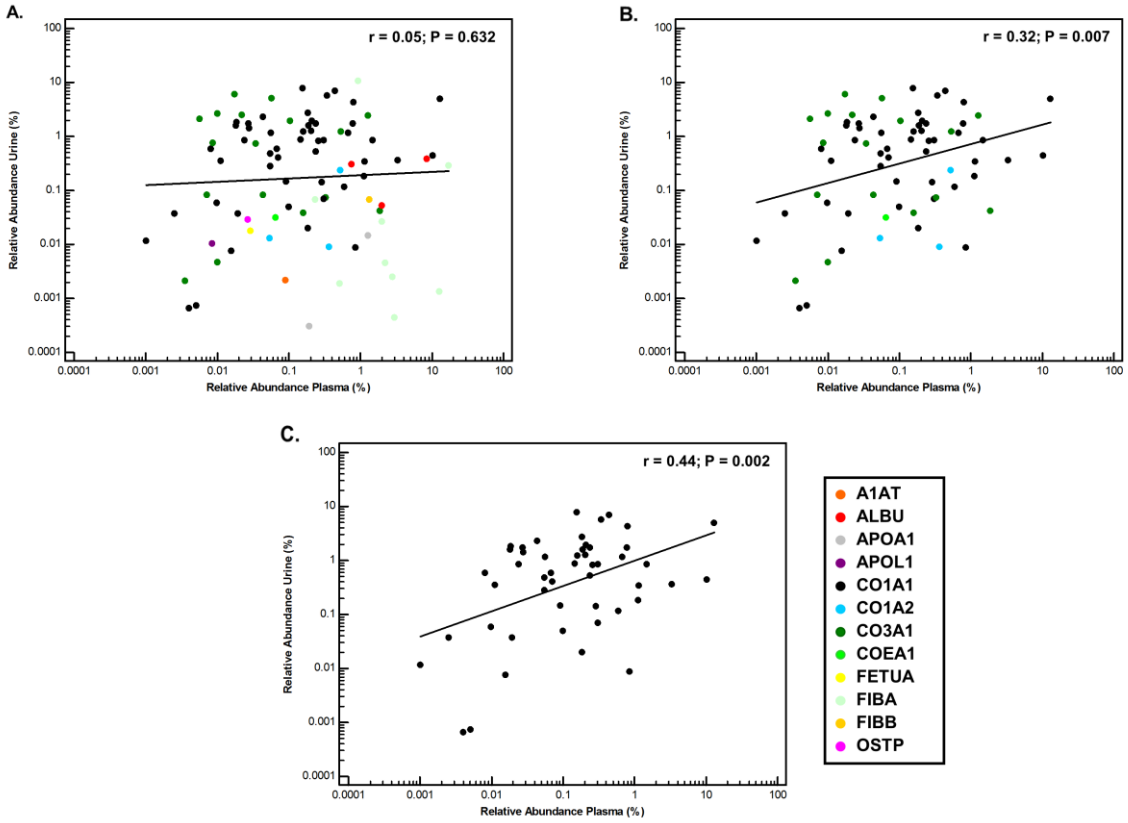
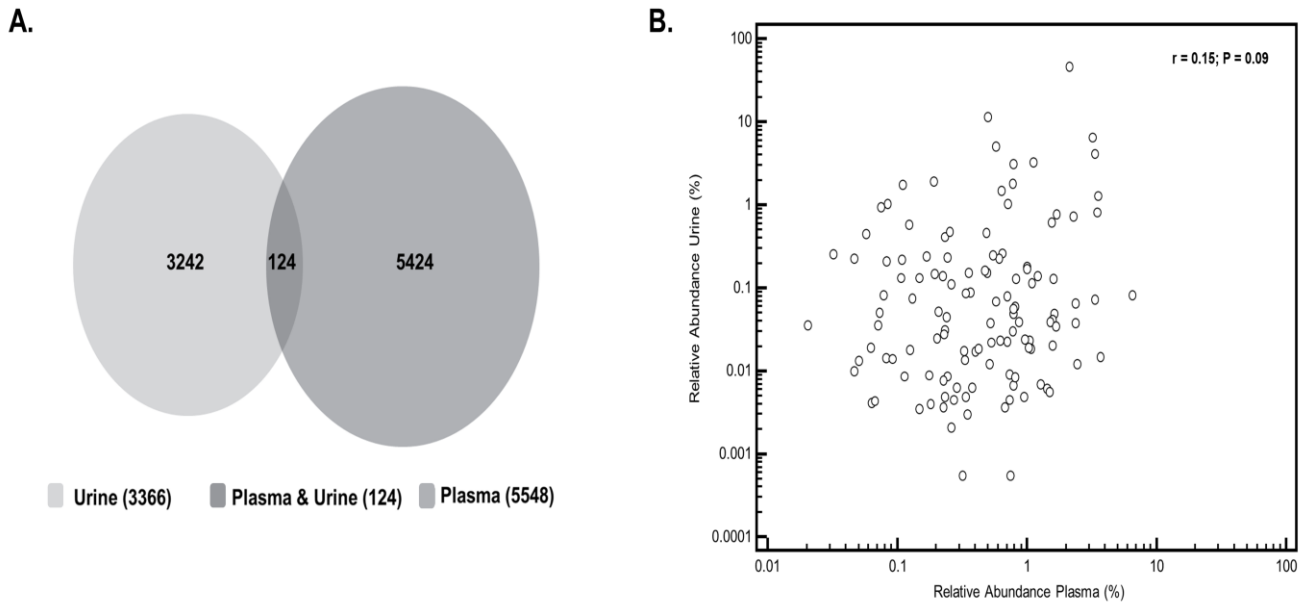


Figure 2. Correlation of the urine and plasma peptides found among the overlap based on their relative abundance in both sources. Correlation A. of all peptides, B. of all collagen peptides, and C. of only peptides from collagen alpha-1(I) chain



*A1AT: Alpha-1-antitrypsin; ALBU: Serum albumin; APOA1: Apolipoprotein A-I; APOL1: Apolipoprotein L-I, CO1A1: Collagen alpha-1(I) chain; CO1A2: Collagen alpha-2(I) chain; CO3A1: Collagen alpha-1(III) chain; COEA1: Collagen alpha-1(XIV) chain; FETUA: Alpha-2-HS-glycoprotein; FIBA: Fibrinogen alpha chain; FIBB: Fibrinogen beta chain; OSTP: Osteopontin

Figure 3. A. Comparison of Human Urinary Proteome database[21,22] and plasma peptidome listed in[17] based on the identical sequenced peptides. B. Correlation of the overlapping peptides based on their relative abundance

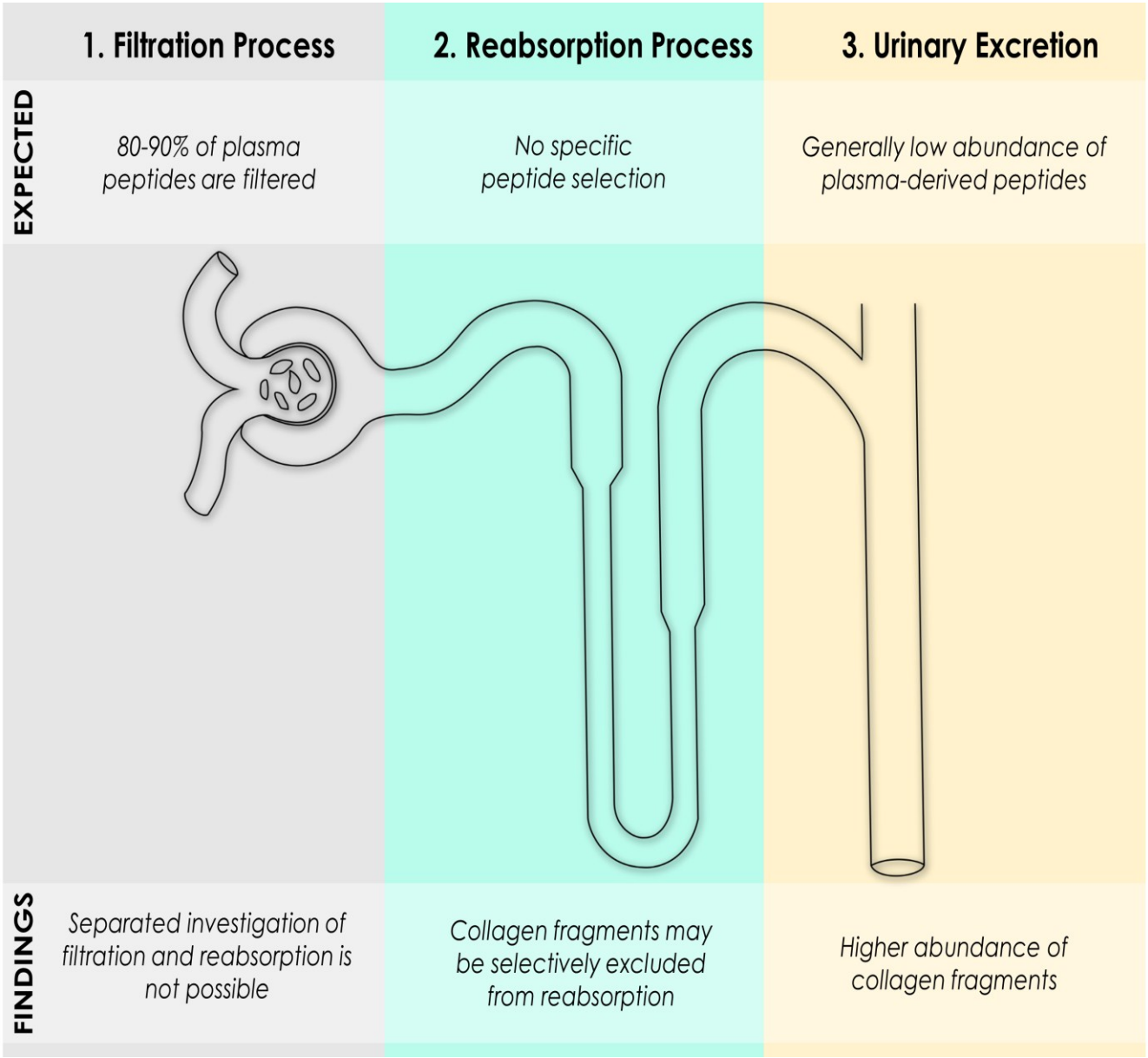
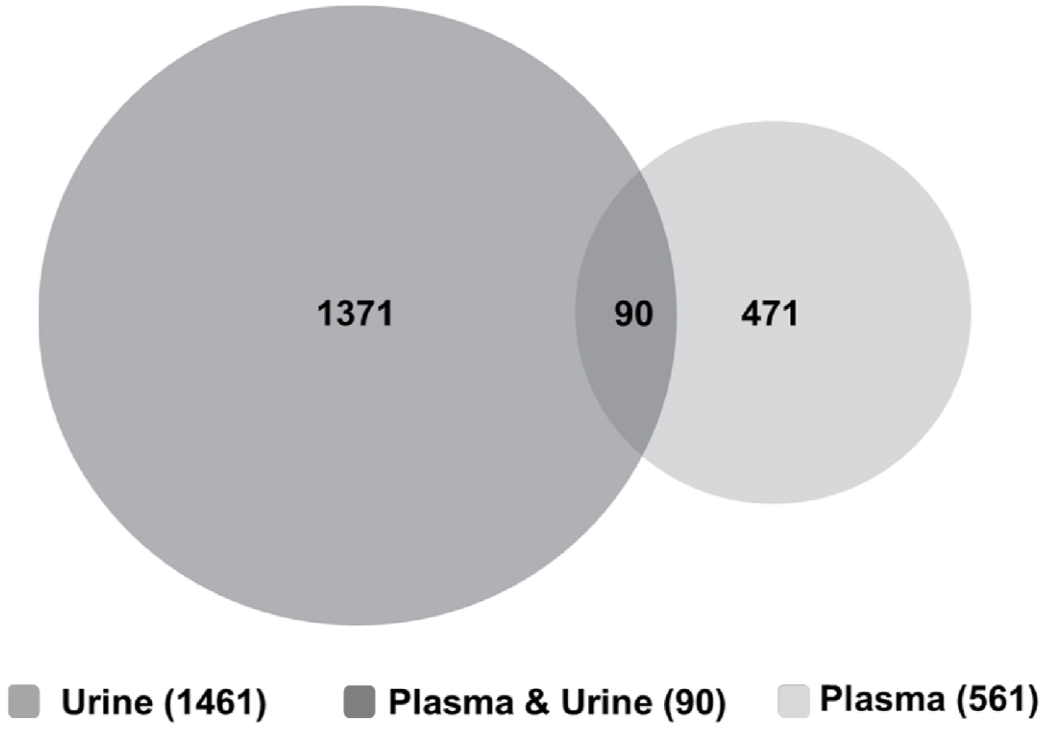


Figure 4. Representation of the expected and our main findings, related to filtration and reabsorption processes, and urinary excretion



Supplementary tables

Supplementary table 1: Spreadsheet "Plasma" - Plasma peptides identified by CE- and LC-MS/MS; Spreadsheet "Urine" - Urine peptides identified by CE- and LC-MS/MS

Supplementary table 2: Peptides commonly identified in plasma and urine