Novel Automated Biomarker Discovery Workflow for Urinary Peptidomics

Crina I. Balog,1* Paul J. Hensbergen,1 Rico Derks,1 Jaco J. Verweij,1 Govert J. van Dam,1 Birgitte J. Vennervald,2 André M. Deelder,1 and Oleg A. Mayboroda1

BACKGROUND: Urine is potentially a rich source of peptide biomarkers, but reproducible, high-throughput peptidomic analysis is often hampered by the inherent variability in factors such as pH and salt concentration. Our goal was to develop a generally applicable, rapid, and robust method for screening large numbers of urine samples, resulting in a broad spectrum of native peptides, as a tool to be used for biomarker discovery.

METHODS: Peptide samples were trapped, desalted, pH-normalized, and fractionated on a miniaturized automatic reverse-phase strong cation exchange (RP-SCX) cartridge system. We analyzed eluted peptides using MALDI-TOF, Fourier transform ion cyclotron resonance, and liquid chromatography–ion trap mass spectrometry. We determined qualitative and quantitative reproducibility of the system and robustness of the method using BSA digests and urine samples, and we used a selected set of urine samples from Schistosoma hematobium–infected individuals to evaluate clinical applicability.

RESULTS: The automated RP-SCX sample cleanup and fractionation system exhibits a high qualitative and quantitative reproducibility, with both BSA standards and urine samples. Because of the relatively high cartridge binding capacity (1–2 mL urine), eluted peptides can be measured with high sensitivity using multiple mass spectrometric techniques. As proof of principle, hemoglobin-derived peptides were identified in urine samples from S. hematobium–infected individuals, even when the microhematuria test was negative.

CONCLUSIONS: We present a practical, step-by-step method for screening and identification of urinary peptides. Alongside the analytical method evaluation on standard samples, we demonstrate its feasibility with actual clinical material.

© 2008 American Association for Clinical Chemistry

In the last 2 decades, mass spectrometry–based proteomics has emerged as an indispensable tool of modern biomedical science, but it was not before the publication of Petricoin et al. (1) that clinicians fully appreciated the benefits of this new technology. Petricoin and coauthors introduced the concept of protein profiling: the fusion of mass spectrometry with pattern recognition, where specific peak profiles, without knowledge of individual peak identity, were treated as biomarkers. A number of studies have demonstrated the applicability of serum profiling to a range of medical research questions, including diagnostics of a variety of cancers (1–6). Recently, protein profiling was adapted for biomarker discovery studies in cerebrospinal fluid (7, 8) and urine (9).

Apart from providing the possibility of noninvasive diagnostic testing, urine has other advantages over serum or cerebrospinal fluid. Most importantly, many peptides and small proteins occur in urine at nearly the same concentrations as in plasma, whereas the total protein concentration is relatively low (10). Thus, the relative enrichment of low-molecular-weight components makes urine an attractive target for a peptide profiling approach.

Surface-enhanced laser desorption ionization time-of-flight mass spectrometry (SELDI-TOF-MS)3 is one of the preferred platforms for urinary peptide profiling (11–13). It is a robust method easily adjustable to a large number of samples; however, only a limited volume of sample can be used, and direct identification of potential biomarkers is impossible.
Capillary electrophoresis mass spectrometry (CE-MS) is also used for urine peptide profiling. The high separation capability of CE combined with the extended dynamic range and high mass accuracy of electrospray ionization time-of-flight (ESI-TOF) mass spectrometry provides more complex peak patterns than SELDI-based approaches (14). There are reports of CE-MS applications for large clinical studies (15), but owing to its relatively long analysis time, CE-MS remains challenging.

Ideally, a biomarker discovery workflow exhibits the explorative power of a profiling approach and, at the same time, the possibility for identification of the putative biological markers. Here, we present a method that combines rapid automated sample pretreatment with well-established mass spectrometric methods. The method consists of an automatic sample cleanup and fractionation system using a combination of reverse-phase and strong cation exchange cartridges. To test the applicability of our method, we analyzed 65 urine samples collected in the Kaloleni division of the Coast province of Kenya, an area endemic for Schistosoma hematobium.

Materials and Methods

We obtained gradient-grade acetonitrile from Bio-solve, potassium phosphate and potassium chloride from Sigma-Aldrich, and phosphoric acid and trifluor-acetic acid (TFA) from Fluka Chemika. We purchased Peptide Calibration Standard I, BSA standard digest, and 2,5-dihydroxybenzoic acid (DHB) from Bruker Daltonics. We purchased from Sigma-Aldrich, and phosphoric acid and trifluorocate, potassium phosphate and potassium chloride. We obtained gradient-grade acetonitrile from Bio- Materials and Methods

We obtained gradient-grade acetonitrile from Bio-

solve, potassium phosphate and potassium chloride from Sigma-Aldrich, and phosphoric acid and trifluoro-

acetic acid (TFA) from Fluka Chemika. We purchased Peptide Calibration Standard I, BSA standard digest, and 2,5-dihydroxybenzoic acid (DHB) from Bruker Daltonics. We used Millipore C18 ZipTips for manual desalting of the samples.

We selected 65 urine samples from an S. hematobi-
mus–endemic area in Kenya from a previous study performed in our laboratory (16). A discrimination between positive and negative samples was based on urinary egg antigen concentrations, as determined by an antigen ELISA developed in-house (17).

AUTOMATED SAMPLE CLEANUP AND FRACTIONATION

After thawing, urine samples were centrifuged for 10 min at 1500g, 4 °C, for the removal of cellular components. We used 1 mL of the supernatant for subsequent analysis, except sample 651 (Table 1), from which we used 4× 1 mL to test the reproducibility of the cleanup and fractionation method.

To perform automated sample cleanup and fractionation, we used a Prospekt 2 system (Spark Holland) with 2 10- by 2-mm solid-phase extraction (SPE) cartridges, a reverse-phase (RP) cartridge (Hysphere C18 HD), and a strong cation exchange (SCX) cartridge (Isolute). The Endurance autosampler injected 1 mL sample and delivered it to the RP cartridge by a high-pressure dispenser at a flow rate of 2 mL/min. A mixture of 98:2 H2O/acetonitrile, pH 3 (adjusted with phosphoric acid), was used as loading solvent. To ensure sufficient removal of the salts and pH adjustment of the urine samples, the RP cartridge was washed twice with 2 mL loading solvent. At the same time, the SCX cartridge was conditioned with the mobile phase consisting of 30% acetonitrile in water containing 10 mmol/L potassium phosphate, pH 3. Subsequently, the trapped peptides were eluted from the RP cartridge with 2 mL 30% acetonitrile in water containing 10 mmol/L potassium phosphate, pH 3, using the second high-pressure dispenser, and loaded onto the SCX cartridge. Next, the SCX cartridge was washed 2 times with 2 mL SCX washing buffer (20% acetonitrile in water containing 10 mmol/L potassium phosphate buffer, pH 3). The elution of SCX cartridge was performed in 2 consecutive steps using 100 μL SCX washing buffer containing 0.1 mol/L potassium chloride (fraction A) and 100 μL SCX washing buffer containing 0.5 mol/L potassium chloride (fraction B) (Fig. 1).

MASS SPECTROMETRY

Fractions A and B eluted from the RP-SCX fractiona-
tion were dried and redissolved in 100 μL Milli-Q water. We desalted 10 μL of each fraction of the urine samples and 3 μL of the BSA tryptic digest on a C18 ZipTip and eluted them directly onto a 600-μm Anchor-Chip MALDI target plate (Bruker Daltonics) using 2 μL of 2 g/L DHB in 50% acetonitrile and 0.1% TFA.

MALDI-TOF analyses were performed on an Ultraflex II TOF mass spectrometer controlled by the flexControl 3.0 software package (Bruker Daltonics). MALDI Fourier transform ion cyclotron resonance (FT-ICR) spectra were recorded using a 9.4-T FT-ICR mass spectrometer (Apex ultra; Bruker Daltonics) equipped with a combined ESI/MALDI ion source (18). The theoretical monoisotopic masses of the peptides were calculated in www.chemcalc.org using the elemental composition of the peptides.

We performed nanoflow liquid chromatography–
tandem mass spectrometry (LC-MS/MS) using an Ultimate LC system (Dionex). Sample (5 μL) was injected onto a C18 PepMap 0.3- by 5-mm trapping column (Dionex) and washed with 100% solvent A [2% acetonitrile (ACN) in 0.1% formic acid in water, vol/vol] at 20 μL/min for 40 min. Peptides were separated on a C18 PepMap 75-μm by 150-mm column (Dionex) at a constant flow of 200 nL/min. The peptide elution gradient was from 10% to 60% solvent B (95% ACN in 0.1% formic acid in deionized water, vol/vol) over 50 min, then to 90% B in 20 min. The column was washed for another 10 min before returning to original conditions, making a total analysis time of 140 min.
The nanoflow LC system was coupled to an HCT ultra IonTrap (Bruker Daltonics) using a nanoelectrospray ionization source. The spray voltage was set at 1.2 kV, and the temperature of the heated capillary was set to 165 °C. Eluting peptides were analyzed using the data-dependent MS/MS mode over a 300 –1500 m/z range. We selected the 5 most abundant fragments in an MS spectrum for MS/MS analysis by collision-induced dissociation using helium as the collision gas.

DNA ISOLATION AND REAL-TIME PCR

The Schistosoma-specific forward primer Ssp48F (5'-GGTCTAGATGACTTGAGTAGCT-3') and reverse primer Ssp124R (5'-TCCCGAGCGTGTATAATGTCATTA-3') amplified a 77-bp fragment within the ITS2 gene. We used the double-labeled Ssp78T [FAM-5'-TGGGTTGTGCTCGAGTCGTGGC-3'-Black Hole Quencher1 (Biolegio)] to detect Schistosoma-specific amplification. Amplification reaction of each DNA sample was performed in a volume of 25 μL with PCR buffer (HotstarTaq mastermix; Qiagen), 5 mmol/L MgCl₂, 12.5 pmol of each Schistosoma-specific primer, 15 pmol of each Phocin herpesvirus 1 (PhHV-1)–specific primer, 2.5 pmol of each Schistosoma-specific double-labeled probe and PhHV-1–specific Cy5 double-labeled detection probe (Biolegio), and 5 μL DNA sample. Amplification consisted of 15 min at 95 °C followed by 50 cycles of 15 s at 95 °C and 60 s at 60 °C. Amplification, detection, and data analysis were performed using the AB7500 real-time detection system (Applied Biosystems).

For DNA isolation, 200 μL urine was heated for 10 min at 100 °C. After SDS-protease K treatment (2 h at 55 °C), DNA was isolated with the QIAamp DNA mini-kit spin columns (Qiagen). In each sample, 10³ plaque-forming units (PFU)/mL PhHV-1 was added to the isolation lysis buffer to serve as an internal control (19).

### Results

**METHOD DEVELOPMENT: QUALITATIVE AND QUANTITATIVE EVALUATION**

To test the qualitative reproducibility of our workflow, we used a BSA digest. BSA digests were fractionated on the automatic RP-SCX system and the combined sequence coverage found after MALDI-TOF MS analysis of fractions A and B was compared to the sequence coverage observed with an unfractionated BSA digest. We obtained a very reproducible sequence coverage of mean (SD) 63% (0%) (n = 5) after RP-SCX fractionation of BSA digests compared with 73% for the unfractionated BSA digest. For the evaluation of the quantitative reproducibility, we analyzed fractions A and B obtained after RP-SCX fractionation of the 5 replicate BSA tryptic digests on a nano-LC system. In each LC chromatogram, the same 5 randomly selected ultraviolet (UV) peaks were integrated for each fraction (Supplemental Fig. 1, which accompanies the online version of this article at www.clinchem.org/content/vol55/issue1). The calculated relative SDs corresponding to the 5 peak areas for the 2 fractions were 3%–7% (Supplemental Table 1), demonstrating good reproducibility of the automatic RP-SCX fractionation of BSA digests.

Biological matrices such as urine, however, are known to influence chromatographic steps. Therefore,
we also tested our method on 4 replicates of 1 urine sample. After fractionation, fractions A and B were measured on the nano-LC system (Supplemental Fig. 2). Although the integration of the UV peaks within these chromatograms was more challenging, the calculated relative SD for the peak areas was 3%–11%, (Supplemental Table 1).

Even if more extensive fractionation increases the number of observed peptides, we decided to use a 2-step elution, since the method is designed for large studies, requiring a substantial degree of throughput. The advantage of at least a 2-step elution became clear when we compared the MALDI-TOF spectrum of a mixed fraction of A and B with their respective individual spectra (Fig. 2). For example, we observed m/z 2908.5 only after fractionation, and its detection in the mixed fraction was suppressed by the presence of the more abundant peptide at m/z 2912.5.

APPLICATION TO URINE SAMPLES FROM S. HEMATOBium–INFECTED INDIVIDUALS

We tested the applicability of our method on a model system in which we expected to find differences between a control and a patient group. For this purpose, we chose 65 urine samples collected in an area endemic for S. hematobium. A subgroup of 10 samples was used for detailed examination. Five of the 10 selected samples (no. 76, 505, 510, 520, and 565; Table 1) were positive for microhematuria and were positive in the soluble egg antigen (SEA) assay, indicating an active Schistosoma infection.

The comparison of the native peptides patterns of the 5 heavily infected individuals with the other samples revealed a number of native peptides that appeared to be discriminative for these 2 groups. Fig. 3 shows a representative part of the MALDI-TOF MS spectra of fraction B from the RP-SCX–processed urine samples. For example, the native peptide at m/z 1096.6 is predominantly present in the urine samples from heavily infected individuals, whereas the native peptide with m/z 1236.6 is most prominent in the spectra of the other 5 samples.

Using MALDI-TOF MS/MS the peptide at m/z 1096.6 was identified as the hemoglobin peptide VYPWTQRF (Fig. 4A). The identity of this peptide was verified by nano-LC-MS/MS analysis of a urine
sample from an infected individual (data not shown). The identification of the hemoglobin fragment supports the early results of the microhematuria test and demonstrates the feasibility of our approach. In fact, we also observed this hemoglobin fragment in 2 more urine samples (511 and 553), although the intensities were much lower (Fig. 3). Based on the SEA assay, these 2 samples were classified as negative. Furthermore, in the microhematuria test, sample 553 was negative and sample 511 only slightly positive. Thus, our RP-SCX fractionation method, in combination with MALDI-TOF analysis, was able to identify a hemoglobin fragment as a result of *S. hematobium* infection in all samples that were positive in the microhematuria test. Moreover, this peptide was identified even in a sample where the microhematuria test, egg count, and SEA assay were negative.

Using real-time PCR, *Schistosoma*-specific DNA was detected in all 6 urine samples shown to contain *S. hematobium* eggs by microscopy. In addition, *Schistosoma*-specific DNA amplification was detected in 3 of the 4 samples that were negative by microscopy and SEA test. Only 1 sample, 651, was found to be negative using real-time PCR. Samples 663 and 672, in which no hemoglobin was detected by mass spectrometric analysis (Fig. 3), had a very high cycle threshold (Ct) value, indicating a very low level of infection.

We used the same MS/MS strategy for the identification of the peptide at *m/z* 1236.6, predominantly present within in the samples classified as *S. hematobium* negative by the SEA assay (Fig. 4B). Searching the MALDI-TOF TOF spectrum against the mammalian National Center for Biotechnology Information database did not result in a reliable identification. In fact, 3 collagen peptides with similar Mascot scores were identified. The first corresponded to a collagen 1A1 fragment (GQDGPPhGPPhGPh, Mascot ion score 42, Ph being hydroxyproline), the second to a
collagen XXVII α1 fragment (PGLIGDLVGLGPhI, Mascot ion score 44), and the third to a collagen type IX α1 fragment (GPPhGEVGPhRGPQG, Mascot ion score 37). We analyzed 1 of the samples on the nano-LC/MS system, but the iontrap MS/MS fragmentation spectrum did not add sufficient information for a confident identification of this peptide. As the theoretical masses (M+H)+ of the 3 potential peptides are slightly different (1236.5602, 1236.7197, 1236.5967 for collagen 1A1, XXVII α1, and IX α1, respectively), we measured the sample with MALDI FT-ICR MS. A monoisotopic m/z of 1236.5597 was measured, which was within 0.4 ppm of the theoretical monoisotopic mass for the collagen 1A1 fragment, thereby unequivocally identifying this as the peptide present in the urine sample. These results clearly show that our RP-SCX fractionation results in enough material for multiple mass spectrometric measurements, allowing confident identification of peptides.

The subgroup of 10 samples provided the opportunity to perform a detailed examination of samples and pick up a few discriminative features. Using this data as a starting point, we performed a screening of total cohort of 65 subjects. Samples were processed with our workflow, but during analysis of MALDI-TOF MS data we concentrated solely on hemoglobin fragments. The results presented in Table 2 are fully in agreement with our original data. Hemoglobin fragments were detected in 47 of 49 microhematuria-positive samples. At the same time, 12 of 16 microhematuria-negative samples proved to contain hemoglobin fragments.

**Discussion**

We present a novel method for urinary peptide profiling using automated sample cleanup and fractionation combined with well-established mass-spectrometric methods. The method is based on a rapid and robust sample preparation workflow. To avoid the loss of low-molecular-weight components, peptidomic approaches require different sample preparation techniques from those of proteomic analysis (20–23). Furthermore, the relatively high concentration of salts and the variable specimen pH (ranging from approximately pH 5 to 8)
present specific problems in developing reproducible methods for peptidomic analysis of urine. Unless adjusted, the variable pH may lead to variation in recovery during chromatographic fractionation, especially on SCX. Reverse-phase chromatography is less affected by minor pH variations, and therefore we used this in our method as a first fractionation step. In addition to sample concentration, it combines desalting and pH adjustment required for a subsequent, reproducible fractionation on the SCX cartridge. The elution of the reverse-phase cartridges using a 30% organic solvent may result in the loss of very hydrophobic and large peptides. However, we hardly expect the presence of very hydrophobic peptides in the soluble peptide fraction. It has been shown that organic modifier may decrease trapping of peptides on SCX columns (24), but a certain amount of acetonitrile is required to eliminate the secondary interactions of the SCX (25). The smallest peptides and compounds with hydrophilic properties will be easily eluted from the reverse-phase cartridge using 30% acetonitrile, but the bulk of large proteins will remain on the column. Those major abundant urinary proteins, which elute with a high concentration of organic modifier, usually obscure the identification of low-abundance proteins and peptides by mass spectrometry. Moreover, the elimination of large, highly abundant proteins during the RP cleanup step results in an increase in loadability for the native peptides on the SCX cartridge, resulting in improved loadability for the native peptides on the SCX cartridge, resulting in improved

![Graph](image_url)

**Fig. 4.** Identification of potential biomarkers using MALDI-TOF MS/MS and MALDI-FT ICR.

(A) MALDI-TOF MS/MS of \textit{m}/\textit{z} 1096.6 corresponding to the hemoglobin fragment VYPWTQR. (B) MALDI-TOF MS/MS and MALDI-FT ICR MS (inset) analysis of \textit{m}/\textit{z} 1236.6 corresponding to the collagen 1A1 fragment GQDRPhGPPhGPPhG (Ph is hydroxyproline).

**Table 2.** Representation of RP-SCX MALDI-TOF MS method performance.\(^{\text{a}}\)

<table>
<thead>
<tr>
<th></th>
<th>RP-SCX MALDI-TOF MS</th>
<th>PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microhematuria positive</td>
<td>49</td>
<td>47</td>
</tr>
<tr>
<td>Microhematuria negative</td>
<td>16</td>
<td>12</td>
</tr>
</tbody>
</table>

\(^{\text{a}}\) The specificity of our procedure was verified using RT-PCR (a standard method for \textit{S. haematobium} infection monitoring). Microhematuria was evaluated semiquantitatively using Hemastix reagent strips.

* Hemoglobin fragments detected by MALDI-TOF MS; \(^{\text{c}}\) no hemoglobin fragments detected by MALDI-TOF MS.
detection of these peptides during mass spectrometric analysis.

Norden et al. (26) used a similar RP-SCX online purification approach, but our experimental design appears to have some specific advantages. First, the use of cartridges instead of columns results in a short analysis time and permits processing of large number of samples. Second, carryover effects, a major problem that can influence the accuracy and precision of HPLC and LC-MS (27), are substantially reduced in our method by using a SPE system based on exchangeable cartridges for the automatic urine sample fractionation.

Our method combines an analytical setup with a relatively high peptide binding capacity and a state-of-the-art MS platform that allows high-throughput profiling followed by the identification of potential markers. In a pilot study, a hemoglobin fragment was clearly identified in urines from S. hematobium–infected individuals. Moreover, this peptide was identified in 2 additional samples that were microhematuria and SEA negative.

Hematuria has been used in many studies as a means of screening for S. hematobium infection and has been shown to be reliable in diagnosing urinary schistosomiasis (28–32). Mott et al. (33) indicated that active bladder lesions, caused by the passage of S. hematobium eggs through the bladder wall, also result in the leakage of hemoglobin into the urine, this process being probably the main source of blood in urine. However, other sources of blood in urine, such as menstrual blood, may lead to false-positive results. But, as shown by Savioli et al. (31), menstruation in women did not markedly affect the screening results.

Besides hemoglobin, we identified a collagen 1A1 fragment with m/z 1236.6 that was not detectable in the 5 urine samples of heavily infected individuals but was present in the other samples. The proper identification of native peptides often requires fusion of data acquired on different mass analyzers, and the identification of the collagen 1A1 fragment with m/z 1236.6 justifies our aiming for a procedure that provides us with sufficient material to identify and characterize the differentially expressed markers.

In conclusion, we developed a quick and robust method that circumvents time-consuming purification steps and allows automated handling of large amounts of urine samples within clinical studies. Importantly, since there is sufficient material available after the fractionation, the method makes it possible to reanalyze samples with different mass spectrometers, thus leading to a higher chance for a positive identification of potential biomarkers.

**References**


