Urinary Proteomics in Diabetes and CKD

Kasper Rossing,* Harald Mischak,† Mohammed Dakna,† Petra Zürbig,† Jan Novak,‡ Bruce A. Julian,§ David M. Good,¶ Lise Tarnow,* and Peter Rossing,* on behalf of the PREDICTIONS Network

*Steno Diabetes Center, Gentofte, Denmark; †Mosaiques Diagnostics, Hannover, Germany; Departments of ‡Microbiology and §Medicine, University of Alabama at Birmingham, Birmingham, Alabama; and Departments of ¶Chemistry and *Biomolecular Chemistry, University of Wisconsin-Madison, Madison, Wisconsin

ABSTRACT

Urinary biomarkers for diabetes, diabetic nephropathy, and nondiabetic protein-uric renal diseases were sought. For 305 individuals, biomarkers were defined and validated in blinded data sets using high-resolution capillary electrophoresis coupled with electrospray-ionization mass spectrometry. A panel of 40 biomarkers distinguished patients with diabetes from healthy individuals with 89% sensitivity and 91% specificity. Among patients with diabetes, 102 urinary biomarkers differed significantly between patients with normoalbuminuria and nephropathy, and a model that included 65 of these correctly identified diabetic nephropathy with 97% sensitivity and specificity. Furthermore, this panel of biomarkers identified patients who had microalbuminuria and diabetes and progressed toward overt diabetic nephropathy over 3 yr. Differentiation between diabetic nephropathy and other chronic renal diseases reached 81% sensitivity and 91% specificity. Many of the biomarkers were fragments of collagen type I, and quantities were reduced in patients with diabetes or diabetic nephropathy. In conclusion, this study shows that analysis of the urinary proteome may allow early detection of diabetic nephropathy and may provide prognostic information.


Nephropathy is a serious and common complication of diabetes and has become the most prevalent cause of ESRD. Over the years, there has been an ongoing quest to find biomarkers early in the clinical course to identify and treat better individuals at high risk for diabetic nephropathy. Knowledge of the complex molecular and pathophysiologic mechanisms leading to renal disease remains limited, in part because conventional research tools have hampered investigators by restricting their focus to a single or relatively few risk markers at a time. Recent advances in proteomics enable screening of a vast array of proteins simultaneously, aiding assessment of their potential role in the development and progression of disease.1 Urine is well suited for proteomic analysis to identify predictive biomarkers and to unravel the pathogenetic mechanisms of chronic renal disease. The online combination of capillary electrophoresis (CE) and electrospray mass spectrometry (MS) was developed for the rapid (approximately 45 min per sample), sensitive, and automated approach for such an analysis.2

The aim of our study was to examine whether CE-MS can detect differences in the urinary proteome between patients with normo-, micro-, and macroalbuminuric and type 1 diabetes. Furthermore, we sought to evaluate whether CE-MS–defined patterns derived from urinary polypeptides of patients with diabetic nephropathy differ from those of patients with other chronic renal diseases.

Healthy individuals without diabetes and patients with diabetes were matched with respect to age and gender to establish urinary polypeptide patterns characteristic for diabetes and diabetic nephropathy (Table 1). The three groups of patients with diabetes (persistent normoalbuminuria, microalbuminuria, and diabetic nephropathy) had comparable duration of diabetes, severity of diabetic retinopathy, BP, and serum cholesterol levels. Hemoglobin A1c was lower in normoalbuminuric patients, and serum creatinine was significantly higher in patients with diabetic nephropathy. Urine samples from all participants were analyzed using CE-MS, and the data were normalized using
BRIEF COMMUNICATION

internal standards and deposited in a database. Only polypeptides present at a frequency >50% in any of the groups were evaluated. The compiled data for the four groups are shown in Figure 1A. Whereas the urinary peptide profiles in samples from the healthy individuals (STENOcontrol), as well as patients with diabetic normoalbuminuric (STENOnormo) and microalbuminuric (STENOmicro), seemed similar, the urinary polypeptide patterns of patients with diabetic nephropathy (STENOnacro) were markedly differed.

For identification of potential biomarkers for diabetes, urine samples from patients with diabetes and persistent normoalbuminuria. This analysis identified 40 peptides of statistical significance (P < 0.05 in maxT testing; Supplemental Table 1). A support vector machine-based model (SVM-BM) with these 40 polypeptides discriminated healthy individuals and patients with diabetic nephropathy with 89% sensitivity and 93% specificity at cross-validation.

Because SVM-BM do not assign probability, we investigated whether a linear combination of biomarkers may improve disease severity classification. Transformed and calibrated logarithmic amplitudes of all biomarkers were combined for each patient. This approach resulted in 100% sensitivity and 93% specificity.

The diabetic renal disease model was validated in an external data set of 35 patients with diabetes and macroalbuminuria (HANNOVERMacro) and 35 healthy individuals (HANNOVERControl). Using SVM-BM, 34 of 35 controls and all patients were classified correctly. When the linear model was applied, the same control subject and one of the patients were misclassified, resulting in 97% sensitivity and 99% specificity.

When the diabetic renal disease models were applied to the matched 30 patients with diabetes and microalbuminuria in the study (STENOmicro), 21 scored positive for diabetic nephropathy when the linear model was used (six were only marginally positive). On the basis of the same biomarkers in the SVM-BM, 17 of 30 patients scored positive. Eight patients with microalbuminuria later showed an increase in albuminuria of >25% or progressed to macroalbuminuria during a 3-year follow-up interval. Each scored positive in the linear model, and seven scored positive in the SVM-BM. These data indicate the utility of the biomarkers for not only detection of overt nephropathy but also prediction of its development in patients with diabetes and microalbuminuria with statistical significance (P = 0.0359).

Several peptides were sequenced using MS/MS. The sequenced biomarkers that differentiated patients with diabetes from healthy control subjects included fragments of collagen type I (α1 chain) and uromodulin (Supplemental Table 1). For differentiation of patients with and without diabetic nephropathy, the sequences were more diverse, consisting of various collagen types as well as uromodulin fragments (Supplemental Table 2).

We tested the specificity of the diabetic renal disease pattern using urine samples from patients with other chronic renal diseases: Biopsy-proven IgA nephropathy (n = 57), FSGS (n = 30),

Table 1. Characteristics of all participants recruited at the STENO Diabetes Center

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Healthy Controls (n = 30)</th>
<th>Normoalbuminuria (n = 30)</th>
<th>Microalbuminuria (n = 29)</th>
<th>Diabetic Nephropathy (n = 30)</th>
<th>P (Normo- versus Macroalbuminuria)</th>
<th>P (All Groups)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>51 (1.9)</td>
<td>54 (2.0)</td>
<td>56 (1.8)</td>
<td>50 (1.6)</td>
<td>0.120</td>
<td>0.070</td>
</tr>
<tr>
<td>Gender (female/male; mean [SE])</td>
<td>7/23</td>
<td>7/23</td>
<td>7/22</td>
<td>7/23</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td>Diabetes duration (yr; mean [SE])</td>
<td>–</td>
<td>34 (1.9)</td>
<td>35 (2.0)</td>
<td>34 (2.0)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Diabetic retinopathy (nil/simple/ proliferative)</td>
<td>–</td>
<td>2/12/16</td>
<td>1/9/19</td>
<td>2/7/21</td>
<td>0.230</td>
<td>0.990</td>
</tr>
<tr>
<td>Urinary albumin/creatinine (geometric mean [95% CI])</td>
<td>–</td>
<td>5 (4 to 6)</td>
<td>45 (32 to 63)</td>
<td>765 (564 to 1036)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Systolic BP (mmHg; mean [SE])</td>
<td>133 (3)</td>
<td>136 (3)</td>
<td>148 (4)</td>
<td>142 (4)</td>
<td>0.280</td>
<td>0.020</td>
</tr>
<tr>
<td>Diastolic BP (mmHg; mean [SE])</td>
<td>80 (2)</td>
<td>77 (2)</td>
<td>75 (2)</td>
<td>77 (2)</td>
<td>0.980</td>
<td>0.240</td>
</tr>
<tr>
<td>Cholesterol (mmol/L; mean [SE])</td>
<td>5.6 (0.1)</td>
<td>4.6 (0.2)</td>
<td>4.9 (0.2)</td>
<td>4.6 (0.2)</td>
<td>0.370</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Creatinine (μmol/L; mean [SE])</td>
<td>97 (2)</td>
<td>94 (2)</td>
<td>94 (3)</td>
<td>171 (15)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hemoglobin A1c (%) (mean [SE])</td>
<td>5.4 (0.1)</td>
<td>8.0 (0.2)</td>
<td>8.9 (0.3)</td>
<td>8.8 (0.2)</td>
<td>0.0150</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

membranous glomerulonephritis \( (n = 35) \), and minimal-change disease \( (n = 25) \). These patients and data sets have been described previously.\(^{5,7,8}\) The compiled polypeptides of the different chronic renal diseases are shown in Figure 2A. As expected, most \( (n = 104) \) of these 147 samples scored positive for "diabetic renal disease," and only 43 scored as normoalbuminuria, indicating that the "diabetic renal disease" pattern reflects chronic renal damage.

To investigate whether we could define biomarkers to distinguish diabetic nephropathy from the other chronic renal diseases, we randomly selected and compared 104 samples from patients with nondiabetic nephropathy with those from 70% of the patients with diabetic nephropathy (STENOmacro and HANNOVERmacro; \( n = 44 \)). This process identified 37 biomarkers of statistical significance \((P < 0.05 \text{ in maxT testing; Table 2, Figure 2B})\).

Using a take-one-out procedure, the number of biomarkers decreased from 37 to 17. A SVM-BM based on these 17 biomarkers distinguished diabetic renal disease from the four other chronic kidney diseases. This model correctly classified 42 of 44 patients with diabetic nephropathy and 98 of 104 patients with other chronic renal diseases. Cross-validation of this training set showed 91% sensitivity and 89% specificity. To validate these biomarkers further, we analyzed the 30% of initially recruited patients and control subjects who had been omitted. Classification of this masked data set showed 81% sensitivity and 91% specificity (Figure 2C).

In this study we demonstrate urinary proteomic biomarkers distinct for diabetes, diabetic nephropathy, and nondiabetic proteinuric renal diseases. We sequenced only some biomarkers, because the procedure is challenging.\(^{1,9}\) The most striking observation was the decreased excretion of specific collagen fragments in patients with diabetes compared with healthy control subjects; several additional collagen fragments were less common in patients with diabetic nephropathy compared with normoalbuminuric diabetic patients. This observation sup-

**Figure 1.** (A) Protein patterns of the patients with diabetes and control subjects examined in this study. Shown are compiled patterns consisting of all samples from each of the four groups. The molecular mass (0.7 to 25 kD, on a logarithmic scale) is plotted against normalized migration time (17 to 47 min). Signal intensity is encoded by peak height and color. (B) Distribution of potential differential-diagnostic biomarkers for diabetes in the patients with normoalbuminuria and healthy control subjects. All statistically significant biomarkers from Supplemental Table 1 are shown. (C) Distribution of potential differential-diagnostic biomarkers for diabetic nephropathy in the different groups of patients and healthy control subjects. Shown are all statistically significant biomarkers listed in Supplemental Table 2.
ports previous reports and extends earlier findings by using validation in external data sets; CE-MS with ultrafiltration to remove larger proteins; and a new generation of mass spectrometers with better sensitivity, mass accuracy, and resolution.

Collagen increases in renal tissue of patients with diabetic nephropathy, and the decreased proteolysis as a result of increased synthesis of protease inhibitors diminishes excretion of collagen fragments. Furthermore, advanced glycation end products cross-link collagen and thus increase its resistance against proteases. A similar principle was shown for tubulointerstitial elastin deposition in experimental diabetic nephropathy: Decreased activity of elastase IIIB and increased production of elastase inhibitor diminished elastin turnover and increased its quantity.

Urine contains a plethora of collagen fragments. In a study of individuals, 231 of 353 sequences were from collagen. Several fragments were found in >90% of all samples, with similar amplitudes. Other collagen fragments showed significant differences between control subjects and patients. These fragments are likely products of specific proteases and may serve as indicators of the activity of these enzymes.

Patients with diabetic nephropathy showed decreased levels of collagen type I and uromodulin fragments, whereas albumin fragments were highly upregulated. Assessment of these peptides by immunologic technologies, such as Western blot, is difficult because of their

Figure 2. (A) Protein patterns of the patients with chronic renal disease (CRD). Shown are compiled patterns consisting of samples from patients with FSGS (n = 35) IgA nephropathy (IgAN; n = 57), minimal-change disease (MCD; n = 25), and membranous glomerulonephritis (MNGN; n = 29). In comparison with Figure 1A, these compiled data show a much higher degree of similarity to the patients with macroalbuminuria than to any other group with diabetes. The molecular mass (0.7 to 25 kD, on a logarithmic scale) is plotted against normalized migration time (17 to 47 min). Signal intensity is encoded by peak height and color. (B) Distribution of potential differential-diagnostic biomarkers for diabetic nephropathy(DN) in all patients with macroalbuminuria (DN group) and all control subjects with (CRD group) used in the study. All 37 statistically significant biomarkers from Table 2 are shown. (C) Receiver operating characteristic analysis of the performance of the differential diagnostic biomarker pattern for DN. (Left) Data from the training set of 44 case patients and 104 control subjects (sensitivity of 95.5% and specificity of 94.2%; area under the curve was 0.971). (Right) Data from validating the masked test set consisting of 64 samples (81.0% sensitivity and 90.7% specificity; area under the curve was 0.856).
Table 2. The 37 urinary biomarkers that statistically differentiated 104 patients randomly selected from the group of 147 patients with nondiabetic chronic renal diseases from a cohort composed of 70% of patients with clinically confirmed diabetic nephropathy.

<table>
<thead>
<tr>
<th>Polypeptide</th>
<th>Mass (Da)</th>
<th>CE Time (min)</th>
<th>Frequency</th>
<th>Mean Amplitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>Unadjusted</td>
<td>maxT</td>
<td>Bonferroni</td>
<td>Amplitude</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>9</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

- Shown are the protein/peptide identification number in the data set (ID), mass (in Da) and normalized migration time (CE Time), the P values (unadjusted and adjusted using maxT and Bonferroni), and the frequency and mean amplitude in the two groups, diabetic nephropathy (DN) and other chronic renal disease (CRD). Where available, sequence and original protein are shown.

small size. Alternative approaches, such as ELISA, seem inappropriate, because several peptides sharing an antigenic epitope cannot be easily distinguished immunologically.\textsuperscript{19}

In conclusion, we demonstrate that the urinary proteome differentiates healthy individuals from diabetic patients with persistent normoalbuminuria, low-grade albuminuria, or nephropathy. Furthermore, it distinguishes patients with diabetic nephropathy from patients with other chronic renal diseases. This finding could have great clinical value among patients with type 2 diabetes and nephropathy, because they are known to have increased albuminuria on a very heterogeneous background when proteinuria occurs in the absence of diabetic retinopathy.\textsuperscript{20} The data from the blinded prospective evaluation of the patients with microalbuminuria indicated that urinary proteome analysis identified patients who had microalbuminuria and were at greater risk for diabetic nephropathy. On the basis of sequencing of some biomarkers, we speculate that changes in the collagen metabolism are closely linked with renal damage in patients with diabetes.

**CONCISE METHODS**

**Patients, Procedures, and Demographics**

This study included an internal data set used to identify biomarkers of diabetes and diabetic nephropathy by proteomic analysis. The internal data set consisted of patients examined at the Steno Diabetes Center in 2004 and was composed of 30 white, healthy individuals (STENOcontrol) and three groups of white patients who had type 1 diabetes and attended the Steno Diabetes Center: (1) 30 patients with persistent normoalbuminuria (STENOnormo; <30 mg/24 h and long-standing diabetes [>15 yr]), (2) 29 patients with persistent microalbuminuria (STENOmicr; ≥30 but <300 mg/24 h in at least two of three consecutive samples), and (3) 30 patients with diabetic nephropathy (persistent macroalbuminuria [STENOMacro; ≥300 mg/24 h]) and coexistence of diabetic retinopathy or a renal biopsy showing diabetic glomerulosclerosis). The groups were matched for age, gender, and duration of diabetes. The external data sets consisted of 35 healthy individuals without diabetes (HANNOVERcontrol) and 35 patients with diabetic nephropathy, most with type 2 diabetes (HANNOVERmacro). Furthermore, biomarkers of diabetic nephropathy were evaluated in patients with biopsy-verified IgA nephropathy (n = 57), FSGS (n = 30), membranous glomerulonephritis (n = 35), and minimal-change disease (n = 25). Patients in the external data sets had been evaluated in other hospitals.\textsuperscript{5,8,21,22} The local ethics committees approved the study, and all participants gave informed consent. The study was performed in accordance with the Declaration of Helsinki.

**Sample Preparation and CE-MS Analysis**

All analyses and data processing were performed in accordance with the recently published minimum information about a proteomics experiment guidelines\textsuperscript{23} and in agreement with the suggested guidelines for clinical proteome analysis.\textsuperscript{24} All urine samples for CE-MS analyses were from spontaneously voided urine and stored at −80°C until analysis. Sample preparation and CE-MS analysis were performed.\textsuperscript{3} Briefly, a 0.7-ml sample was prepared and resuspended in 10 to 100 μl of H₂O (depending on the peptide concentration) to a concentration of 1 to 5 μg/μl. Between 100 and 500 nl of sample was injected hydrodynamically, aiming for a total amount of 0.5 μg peptide per analysis, corresponding to 0.7 to 14.0 μl of the original sample. The data on all CE-MS analyses of all patients in the study are available in Supplemental Table 3 (downloadable at http://www.mosaiques-diagnostics.de/GenexPivot_12092007.exe password: 99Mischak123). A total of 5616 different polypeptides were tentatively identified (annotated), and, on average, 1632 polypeptides were found in each urine sample; on average, 1185 of these were annotated.

**Data Processing and Analysis**

Mass spectral ion peaks representing identical molecules at different charge states were deconvoluted into a single mass using MosaiquesVisu software.\textsuperscript{3} In addition, the migration time and ion signal intensity (amplitude) were normalized using internal polypeptide standards as described in detail previously.\textsuperscript{25} This approach results in superior comparability of the data sets as compared with normalization of the signal intensity to urinary creatinine or total protein content, because especially the latter will give poor results in patients with proteinuria. Each polypeptide is characterized by its molecular mass (kD), normalized migration time (min), and normalized signal intensity as the measure for relative abundance. All detected polypeptides were deposited, matched, and annotated in a Microsoft SQL database, allowing further analysis and comparison of multiple samples (patient groups). Polypeptides within different samples were considered identical when the mass deviation was <50 ppm and the migration time deviation was <1 min. CE-MS data of all individual samples can be accessed in Supplemental Table 3 (downloadable at http://www.mosaiques-diagnostics.de/GenexPivot_12092007.exe password: 99Mischak123).

**Definition of Biomarkers and Sample Classification**

The unadjusted P values were calculated using the arcsinh-transformed intensities and the Gaussian approximation to the t-distribution. Bonferroni adjustments were made by applying the standard Bonferroni criterion to markers that passed the frequency threshold of 70%. MaxT P values were calculated using the Westfall and Young maxT procedure.\textsuperscript{26} Estimates of sensitivity and specificity were calculated by tabulating the number of correctly classified samples. Confidence intervals were based on exact binomial calculations using MedCalc 8.1.1.0 (MedCalc Software; Mariakerke, Belgium; http://www.medcalc.be). The receiver operating characteristic plot was generated by plotting all sensitivity values (true-positive fraction) on the y axis against their equivalent (1 − specificity) values (false-positive fraction) for all available thresholds on the x axis (MedCalc Software).\textsuperscript{27} Disease-specific polypeptide patterns based on predefined polypeptides were generated using the SVM-BM MosaCluster software.\textsuperscript{28} For linear combination, all normalized signal intensity values of biomarkers were log-transformed. Values <1 were substituted with a value of 1 to avoid negative values. The average signal intensity for a specific biomarker over all cases was compared with the av-
average intensity for the same biomarker over all controls. To avoid artificial weighting of specific biomarkers in the set as a result of the difference in observed signal intensities for case and control, the distance between the two averages (case and control) was set relative to a value of 2. The relative distance of signal intensities between the disease and control samples was provided using the following formula:

\[
\frac{(A_i^c - \text{mean}_{\text{averages}})}{\bar{x}_{\text{control}} - \bar{x}_{\text{case}}}^2
\]

where \(A_i^c\) is the log-transformed signal intensity of the \(i\)th biomarker in the \(k\)th sample in either the test set or the blinded set, mean_{averages} is the average of the mean intensity of all possible markers for test set samples, \(\bar{x}_{\text{case}}\) represents the mean observed signal intensity of the possible biomarker from all case samples, and \(\bar{x}_{\text{control}}\) represents the mean signal intensity of the possible biomarker from the control samples.

Cross-validation was performed and defined by the take-one-out procedure. To reduce the number of biomarkers in a model, a take-one-out procedure was used. Briefly, models were generated that were each based on \(n - 1\) biomarkers. These were evaluated using complete cross-validation and compared with the classification results of the model based on \(n\) biomarkers. A single biomarker that apparently did not improve classification was removed.

Sequencing of Polypeptides
Candidate biomarkers were sequenced using liquid chromatography–MS/MS analysis (on a quadrupole time-of-flight mass spectrometer instrument). Further analysis was performed using instruments with electron transfer dissociation capability.

ACKNOWLEDGMENTS
This work was supported in part by a grant U54DK-076166 from the National Institute of Diabetes and Digestive and Kidney Diseases to H.M. and L.T. and by a EUROTRANS-BIO grant ETB-2006-016 to H.M. via the Urosysteomics consortium. D.M.G. gratefully acknowledges support from a National Institutes of Health predoctoral fellowship: The Biotechnology Training Program (NIH ST32GM08349).

REFERENCES


Supplemental information for this article is available online at http://www.jasn.org/.