Integrative Urinary Peptidomics in Renal Transplantation Identifies Biomarkers for Acute Rejection

Xuefeng B. Ling,* Tara K. Sigdel,† Kenneth Lau,* Lihua Ying,† Irwin Lau,* James Schilling,* and Minnie M. Sarwal†

Divisions of *Biotechnology Core and †Nephrology and Department of Pediatrics, Stanford University School of Medicine, Stanford University, Stanford, California

ABSTRACT
Noninvasive methods to diagnose rejection of renal allografts are unavailable. Mass spectrometry followed by multiple-reaction monitoring provides a unique approach to identify disease-specific urine peptide biomarkers. Here, we performed urine peptidomic analysis of 70 unique samples from 50 renal transplant patients and 20 controls (n = 20), identifying a specific panel of 40 peptides for acute rejection (AR). Peptide sequencing revealed suggestive mechanisms of graft injury with roles for proteolytic degradation of uromodulin (UMOD) and several collagens, including COL1A2 and COL3A1. The 40-peptide panel discriminated AR in training (n = 46) and test (n = 24) sets (area under ROC curve >0.96). Integrative analysis of transcriptional signals from paired renal transplant biopsies, matched with the urine samples, revealed coordinated transcriptional changes for the corresponding genes in addition to dysregulation of extracellular matrix proteins in AR (MMP-7, SERPING1, and TIMP1). Quantitative PCR on an independent set of 34 transplant biopsies with and without AR validated coordinated changes in expression for the corresponding genes in rejection tissue. A six-gene biomarker panel (COL1A2, COL3A1, UMOD, MMP-7, SERPING1, TIMP1) classified AR with high specificity and sensitivity (area under ROC curve = 0.98). These data suggest that changes in collagen remodeling characterize AR and that detection of the corresponding proteolytic degradation products in urine provides a noninvasive diagnostic approach.


Despite an improvement in renal allograft survival reflecting advances in immunosuppressive medications,1,2 a critical unmet need in patient care is the requirement for sensitive and graft-etioloxy-specific, noninvasive methodologies for monitoring transplant recipients.3 Expression analyses of urine immune mediators,4 peripheral blood samples, and transplant biopsies5,6 support that distinct molecular pathways can define the injury of acute rejection (AR). Some of the concerns relating to biomarker discovery in urine lie with the confounding effect of proteinuria and high-abundance plasma proteins from nonspecific injury (which also occurs in AR). In this study, we have chosen to only analyze naturally occurring peptides in urine samples from transplant patients for three reasons: (1) because the roughly equal mass of protein and peptide in urine translates into at least a ten-fold greater molar abundance of peptides, urinary peptides provide a fertile ground for biomarker discovery; (2) urinary peptide analysis, unlike intact urinary pro-
Peptidomics analysis, is not hampered by the presence of highly abundant urinary proteins that can obscure the discovery of more informative lower abundance biomarker proteins; and (3) analysis of urinary peptides is relatively easier than the analysis of complex tissues such as biopsy and blood because one-dimensional HPLC separation is sufficient for the analysis of >21,000 urine peptides.

An additional important confounder for AR diagnosis and management is BK nephritis. To address these issues, this study performed noninvasive, urine peptidomic analysis of 70 unique urine samples, collected from renal transplant patients and controls, by liquid chromatography and mass spectrometry (LC-MS), followed by multiple reaction monitoring (MRM) verification, on five different cohorts, including samples with nonspecific proteinuria, BK nephritis, and pyuria.

To explore the relevance of altered urinary peptide abundance, we also performed integrated transcriptomic analysis on matching biopsy microarrays, paired with the urine samples, available in the Sarwal Lab (GEO, GSE14328). Quantitative real-time PCR (Q-PCR) verified significant overlapping genes in an independent set of 34 biopsy samples.

Our results indicate that disease-specific alteration of proteolytic and antiproteolytic activities is the underlying mechanism by which these urine peptide biomarkers are generated in graft rejection. To our knowledge, this study represents the first study that analyzed urinary peptidomic and matching renal biopsy transcriptomic analyses, which will help in elucidating the pathophysiological relationships between our nested urine peptide biomarkers and allograft proteolytic networks in vivo in renal allograft diseases.

RESULTS

Sample Characteristics

The overall study design for the peptidomic urine analysis is shown in Figure 1. Seventy unique urine samples were analyzed from the following five cohorts: pediatric kidney transplant patients with biopsy-proven acute allograft rejection (AR, n = 20), stable allograft with normal protocol biopsies (STA, n = 20), BK virus nephropathy with pyuria (BK, n = 10), nonspecific proteinuria with native renal disease (biopsy-proven nephrotic syndrome; NS, n = 10), and healthy age-matched volunteers (HC, n = 10). Samples were split into training sets (n = 46) for urine peptide discovery, and test sets (n = 24) (sample demographics in Supplementary Table 1) for urine peptide prediction and verification.

Discovery of a Urine Peptide Panel for AR by LC-Matrix-Assisted Laser Desorption/Ionization

A total of 20,937 unique peptide peaks with distinct m/z and HPLC fractions were resolved in the 900- to 4000-Da range. Prediction analysis by a nearest centroid method (NSC) algorithm was performed, and 6-fold internal crossvalidation analysis led to the discovery of a set of 630 peptide features with the lowest classification error (Supplementary Figure 1). Discriminant class probabilities and Gaussian linear discriminant analysis (LDA) were performed for each sample (Supplementary Figure 2) in both sample sets and resulted in misclassification of only 2 of the 24 samples in the test set. To find a predictive biomarker panel of optimal feature number, various classifiers were tested for their spread of distribution and goodness of the separation (Figure 1B and Supplementary Figure 3). Linear discriminant probabilities of a biomarker panel of 53 peptide peaks was sufficient for goodness of separation of the clinically relevant transplant categories (AR, STA, and BK) in the training and the test sample sets (Figure 2, A and B). This biomarker panel classified the AR samples with 96% overall agreement with clinical diagnosis of AR in the training set (P = 3.2 × 10⁻⁶ by Fisher exact test) and 83% agreement with clinical diagnosis of AR in the test set (P = 0.0027 by Fisher exact test). When all 70 samples were clustered by unsupervised analysis of their peptide abundance across the 53 peak features,

![Figure 1](image-url)

**Figure 1.** Peptidomics approach for biomarker discovery. (A) Schemas for peptidomic analysis of naturally occurring urinary peptides. (B) Study design for the urine peptide biomarker discovery.
Fisher exact test was to compute the so-called “two-class” prediction) was used to assess the performance of the biomarker panel in the classification of unknown samples. STA and BK were combined into one group as “NON-AR.”

Identification of AR-Specific Urine Peptides

Manual review of the biomarker panel and associated mass spectrometry (MS) spectra interpreted and de-isotoped the 53 MS peak features, which could be mapped to 40 unique urine peptides and were further identified by matrix-assisted laser desorption/ionization (MALDI)-time-of-flight (TOF)/TOF and LTQ Orbitrap MS/MS analysis. In general, the naturally occurring peptides are more difficult to study with current standard mass spectrometric search engines because of the difficulty in complexity reduction in the search space with the knowledge of peptide-ending amino acids. For those peptides unable to be identified by MS/MS analysis, we are currently following up to scale up the purification of these peptides to have sufficient quantity for protein identification using the Edman sequencing approach. We grouped the identified peptides according to their common protein precursor and computed the medians of LC-MS measurements according to sample categories. The peptides were found to map to nine different proteins, eight of which belonged to the collagen family (COL1A1, COL1A2, COL3A1, COL4A3, COL4A4, COL4A5, COL7A1, COL18A1) and uromodulin (UMOD). When MS/MS analysis was extended to the original 630-peptide feature set, 142 urine peptides were identified, again with predominant presence of the non-AR samples cluster together. Modified 2 × 2 contingency tables were used to calculate the percentage of unsupervised clustering that agreed with clinical diagnosis for the biomarker panel. P values were calculated with Fisher’s exact test. Unsupervised clustering based on the peptide biomarker panel was used to construct a heat map in which the colors indicate the intensity of peptide concentration by LC-MALDI: red indicates high peptide abundance and green indicates low peptide abundance in the comparative analysis. It can be seen that by unsupervised analysis, the AR samples, save one, all co-cluster together and all of the non-AR samples cluster together. Modified 2 × 2 contingency tables were used to calculate the percentage of classification that agreed with clinical diagnosis for each sample, a receiver operating characteristic (ROC) curve was constructed to evaluate the testing performance of our peptide biomarker panel⁹,10 and resulted in area under the curve (AUC) values of 0.97 and 0.96 for the training and the test set, respectively (Figure 3A).

Clustering peptide abundance (n=70 samples) with clinical diagnosis

Figure 2. Statistical analyses of the peptide biomarker panel. (A) The discriminant of the peptide biomarker panel for the training (upper) and testing data (lower) probabilities for all transplant samples were calculated from the LDA. The maximum estimated probability for each of the wrongly classified samples is marked with a circle. Two of the 46 samples in the training set and 4 of the 24 samples in the test set were misclassified, giving a correct classification rate of 96% in the training set and 83% in the test set. (B) Left panel: Modified 2 × 2 contingency tables were used to calculate the percentage of classification that agreed with clinical diagnosis for the biomarker panel. P values were calculated with Fisher’s exact test. Right panel: A prediction of AR from the non-AR phenotype (a so-called “two-class” prediction) was used to assess the performance of the biomarker panel in the classification of unknown samples. STA and BK were combined into one group as “NON-AR.” Fisher exact test was to compute the P value for the blind test. (C) Unsupervised clustering based on the peptide biomarker panel was used to construct a heat map in which the colors indicate the intensity of peptide concentration by LC-MALDI: red indicates high peptide abundance and green indicates low peptide abundance in the comparative analysis. It can be seen that by unsupervised analysis, the AR samples, save one, all co-cluster together and all of the non-AR samples cluster together. Modified 2 × 2 contingency tables were used to calculate the percentage of unsupervised clustering that agreed with clinical diagnosis for the biomarker panel. P values were calculated with Fisher’s exact test.
firms the results that UMOD peptides are much lower in individual urine samples taken from patients when the filtering kidney has ongoing AR. Although the significance of these findings is unclear at present, a recent genome-wide association study has identified significant single-nucleotide polymorphism associations with chronic kidney disease at the UMOD locus.12

Interestingly, all of the identified UMOD and collagen urine peptides showed much lower abundance during AR when compared with other samples, with overall lower abundance in transplant patients when compared with nontransplanted patients (NS) and healthy controls (Supplementary Figure 4). Sequence alignment analysis of the collagen and UMOD peptides were found to line up by forming clusters within either the C- or N-terminal end with ladder like truncations at the opposite ends, suggesting that there is likely disease-specific proteolytic degradation of the parent protein. Similar to the proteolytic degradation of urine proteins in AR, serum proteins have also been found to show differences in degradation in cancer.13

**MRM Verification of Selected Urine Peptides**

To verify the presence and quantify differences in peptides between AR and non-AR groups, MRM was performed on two selected peptides14 [UMOD1 (1679.98 Da) and UMOD2 (1911.07 Da)]; Figure 3] on all 70 samples. The box-whisker graphs in Figure 3B illustrate the spread of the distribution of the MRM measurements in AR (n = 20), STA (n = 20), BK (n = 10), NS (n = 10), and HC (n = 10) sample categories for peptides with UMOD 1680.98 and 1912.07 Da, respectively. As seen in Figure 3B (upper panel, left-hand side), similar to the results obtained by LC-MALDI, the abundance of UMOD peptide 1679.98 was significantly lower in AR (P = 0.0003), and as seen in Figure 3B (upper panel, right-hand side), the abundance of UMOD 1911 was also significantly lower in AR (P = 0.0006) when compared with all other non-AR categories. ROC analysis to test the diagnostic ability of the two UMOD peptide biomarkers for AR was seen in terms of AUC. AUCs for UMOD1 and UMOD2 were 0.83 and 0.74, respectively.

**Integrated Analysis of Matched Samples: Transcriptional Analysis of Biopsy AR and Peptidomic Analysis of Urine AR**

Because urine is an ultrafiltrate of the kidney, we hypothesized that the alteration of the urinary proteins and peptides in urine may relate to processes occurring directly in the kidney. To address this we analyzed archived microarray data in the Sarwal Lab (GSE14328) on matched kidney biopsies (20 AR and 20 STA; taken at the time of urine collection, before any treatment intensification for AR) for expression differences between AR and STA samples for the corresponding UMOD and the collagen genes. We also looked for any expression differences in extracellular matrix proteins in AR, because some of these have been previously demonstrated to be differentially expressed in AR.15 We observed that whereas UMOD gene expression in AR biopsy was significantly lower in AR [false discovery rate (FDR) = 0.02%; similar results to the low UMOD peptide abundance in AR urine], the three collagen genes (COL1A2, FDR = 0.18%; COL3A1, FDR = 0.03%; COL4A1, FDR = 0.02%) were upregulated in AR (different from low collagen peptide abundance in AR urine). Gene expression for matrix metalloproteinase-7 (MMP-7; FDR = 0.03%), tissue inhibitor of metalloproteinase 1 (TIMP1; FDR = 24%), and the serpin peptidase inhibitor (SERPING1; FDR = 33%) was higher in AR when compared with STA biopsies, although only MMP7 expression was significant.

We performed Q-PCR in biopsies from a separate set of 34 kidney biopsies (14 AR, 10 STA, and 10 healthy kidney donor biopsies) for UMOD; the most significant collagen genes in rejection, namely COL1A2 and COL3A1; as well as all MMP7, SERPING1, and TIMP1 (Figure 5A). The Q-PCR results vali-
Figure 4. Mapping of collagen and UMOD peptides in the urine. Identified urine peptide biomarkers yielded clusters of overlapping (A) collagen and (B) UMOD peptides (mass/charge ratio, MH+). "P" in red indicates 4-hydroxyproline. Peptides in brackets derive from the same region of the same precursor proteins. Because the genes labeled in red were significantly regulated in microarray data, we tested them by Q-PCR. (C) Human UMOD precursor. Recent MS analyses proved that C-terminal cleavage of the precursor, which has 640 amino acids, occurred after phenylalanine residue 587. Because part of the C-terminal peptide cleaved from the UMOD precursor, the UMOD peptide biomarker cluster (colored in red) discovered in this study spans from serine residue 589, following arginine residue 588, and to lysine residue 607.

Figure 5. A gene panel specific for AR. (A) The distribution of COL1A2, COL3A1, MMP-7, SERPING1, TIMP1, and UMOD genes’ Q-PCR measurements in kidney biopsy were analyzed by box-whisker graphs. (B) ROC analysis was performed to evaluate the performance of the six-member RNA biomarker panel classifying AR from STA. The plotted ROC curve is the vertical average of the 500 bootstrapping runs, and the boxes and whiskers plot the vertical spread around the average.

Proteomic and peptidomic analysis of urine collected from healthy individuals (22 mg peptides in urine per day) and patients with renal disease have identified more than 1500 different proteins and over 100,000 different peptide biomarkers in health and disease. This is the first study of an integrated analysis of the urine peptidome and the biopsy transcriptome in graft rejection that uncovers that overlapping key gene and peptide pathways can be jointly dysregulated in AR. The resultant alterations in the abundance of selected genes and the peptide products of the corresponding proteins can highlight potential mechanisms of graft injury in rejection. Disease-specific alterations of gene transcription in the tissue (by array and Q-PCR) and a change in the balance of proteolytic and anti-proteolytic activities in urine appear to be important mechanisms resulting in an altered pattern of a specific panel of urinary peptides in AR.

There are at least 28 different human collagens that represent approximately 25% of the total protein content of mammals, but in the kidney type I and III collagens are most abundant, whereas type IV collagen is a major component of basement membranes. The increase in the aminoterminal and carboxy terminal propeptides from the procollagen of types I, III, and IV during collagen anabolism...
and later decrease in the collagen-derived urinary naturally occurring peptides during collagen catabolism suggest that increased turnover of renal collagens\textsuperscript{23–26} may be valuable biomarkers for noninvasive diagnosis of the rejection process in the kidney. The upregulation of extracellular matrix regulators (MMP-7, SERPING1, and TIMP1) also supports the hypothesis of tissue remodeling at the time of AR. The observance of high MMP-7 expression in the kidney at the time of AR also has been previously reported in chronic kidney rejection, human kidney aging, and a rat renal AR model. MMP-7 is a collagenase-related connective-tissue-degrading metalloproteinase and plays a role in the breakdown of extracellular matrix in normal physiologic processes, tissue remodeling during injury, and neutrophil influx to sites of injury.\textsuperscript{32} SERPING1 regulates leukocyte trafficking and complement (inactivating C1r, C1s, MASP2, and C3b proteases), which is also locally regulated in the kidney during ischemia reperfusion injury.\textsuperscript{33} Similar to the findings in this study, SERPING1 has previously been shown to be regulated in the graft during AR.\textsuperscript{34} Tissue-specific inhibitors of metalloproteinases are endogenous, specific inhibitors that bind and inhibit matrix metalloproteinases. TIMP1 is a physiological inhibitor of the matrix-degrading enzymes, collagenases, gelatinase, and stromelysin and plays a major role in the inhibition of matrix degradation. Upregulation of TIMP1 mRNA and protein has been previously reported in different models of renal disease\textsuperscript{35–41} and in human sclerotic glomeruli.\textsuperscript{42} The increased expression of TIMP1, a collagenase inhibitor, may be a reason for the reduced activity of collagenases and subsequent reduced breakdown of tissue collagen, leading to the observance of increased graft collagen expression and reduced collagen urine peptides in graft rejection. Thus, altered collagen and extracellular matrix turnover in graft rejection with altered regulation of collagenases in the graft, as seen in independent data sets by microarray and Q-PCR, may be critical pathways that link AR injury with the observed increased downstream clinical risk of chronic injury and graft fibrosis.\textsuperscript{36–44}

**CONCISE METHODS**

**Urine Samples**
Seventy unique urine samples from 50 pediatric renal transplant recipients (20 biopsy-proven AR, 20 STA, 10 BK), 10 age-matched healthy controls (HC), and 10 pediatric patients with nonspecific proteinuria from native renal disease due to nephrotic syndrome (NS; to control for nonspecific renal injury) were collected at Lucile Packard Children’s Hospital at Stanford University from 2004 to 2006. Details on patient age, gender, and other transplantation-related clinical indicators are given in Supplementary Table 1. Informed consent was obtained from all patients and the Stanford University Institutional Review Board approved the study.

**Urine Collection, Storage, and Processing**
Second-morning void midstream urine samples (50 to 100 ml) were collected in sterile containers and were centrifuged at 2000 × g for 20 minutes at room temperature within 1 hour of collection. The details of urine processing and preparation of peptide extraction and fraction are reported elsewhere.\textsuperscript{7}

**Peptidomic Data Analysis**
We used the approach of ion mapping,\textsuperscript{45,46} in which biomarker candidate MS peaks are selected on the basis of discriminant analysis and then targeted for MS/MS sequencing analysis to obtain protein identification. We have developed an informatics platform, “\textit{MASS-Conductor},”\textsuperscript{7} which contains an integrated suite of algorithms, statistical methods, and computer applications to allow for signal processing and statistical analysis in LC-MS-based urine peptide profiling. The peaks are located in the raw spectra of the MALDI data by an algorithm that looks for sites (m/z values) for which the intensity is higher than the estimated average background and the approximately 100 surrounding sites, with peak widths approximately 0.5% of the corresponding m/z value. The binned LC-MALDI MS peak data (20,937 m/z values) obtained for all 70 samples were analyzed separately for the training sample set (n = 46) for discovery of discriminant biomarkers using algorithms\textsuperscript{6} of NSC, 6-fold crossvalidation analyses, and...
Gaussian LDA. The predictive capabilities of the 53 most discriminant peptide peaks were used to blindly test for differentiating AR, STA, and BK samples in the test set (n = 24). To control the number of false significant features found during NSC mining, we permuted the data set 500 times to calculate the global FDR.  

MRM Assay for Peptide Marker Verification

Stable isotope-labeled peptides (with a $^{13}$C-labeled amino acid) were synthesized and used as internal standard peptides. Each urine peptide sample, prepared as described above, was diluted 10-fold with 10% acetonitrile/0.1% formic acid and spiked with the internal standard peptide to a final concentration 0.1 μM. Peptides were resolved in an HPLC equipped with a Polaris C18 column (50 × 20 mm, 3 μM, 6-minute gradient elution; Buffer A: 0.1% formic acid in water; Buffer B: 0.1% formic acid in acetonitrile; flow rate of 200 μl/min). A triple quadrupole mass spectrometer was used. The data were assessed and visualized by an ROC curve ROCR package.  

Integrated Analysis of Peptidomic Data in Urine and Microarray Data from Matched Transplant Biopsies

Affymetrix HU133 plus 2 GeneChips on matched kidney transplant biopsies (20 AR and 20 STA) have been previously performed in the Sarwal Lab (National Center for Biotechnology Information GEO database GSE14328). Raw expression data were preprocessed and normalized using dChip software. Supervised, two-class unpaired Significance Analysis of Microarray analyses were applied to calculate FDR for differences in expression of the corresponding UMOD and the collagen genes in rejection. Additionally, we searched for any differences in the expression of extracellular matrix proteins (TIMP1, SERPING1, and MMP-7) in the rejecting graft.

RNA Preparation and Q-PCR

Total RNA was extracted from kidney biopsy samples using TRIzol reagent (Invitrogen Corporation, Carlsbad, CA) and later was DNase treated and purified using the RNeasy mini kit according to the manufacturer’s protocol (Qiagen, Valencia, CA). cDNA was synthesized from 250 ng of RNA using the RT2 First Strand Kit (SABioscience Corporation, Frederick, MD). Q-PCR reactions were performed on 5 ng of cDNA using RT2 SYBR Green/ROX PCR master mix and commercially available primers: PPH12000A-200 for UMOD, PPH07771A-200 for TIMP1, PPH18747E-200 for SERPING1, PPH00809E-200 for MMP-7, PPH01918B-200 for COL1A2, PPH00439E-200 for COL3A1, PPH20687A-200 for COL4A1, and PPH05666E-200 for 18S rRNA (SuperArray Bioscience Corporation, Frederick, MD). All RNA samples were analyzed in duplicate and normalized relative to 18S rRNA levels.

ACKNOWLEDGMENTS

The work was supported by National Institutes of Health grant RO1-AI-061739 (M.S.), the Deans Fellowship, and the Child Health Research Program (TS, MS). The authors thank Karolina Krasinska at the Stanford University Mass Spectrometry Center for the MRM assay development and applications, Tonya Pekar at Thermo for LTQ Orbitrap MS, and the Stanford University IT group for excellence in Linux cluster support.

DISCLOSURES

None.

REFERENCES
