Biological Qualification of Biomarkers of Chemical-Induced Renal Toxicity in Two Strains of Male Rat

Ernie Harpur,† Daniela Ennulat,‡ David Hoffman,§ Jean-Charles Gautier,¶ Bjoen Riefke,|| Denise Bounous,||| Kerstin Schuster,|||² Sven Beuhsaen,# Magali Guffroy,** Martin Shaw,†† Edward Lock,§ Syril Pettit,¶¶

on behalf of the HESI Committee on Biomarkers of Nephrotoxicity ⁵

⁎sanofi-aventis, Atnwick, Northumberland, UK; †GlaxoSmithKline, King of Prussia, Pennsylvania 19406-0939; ‡sanofi-aventis, Bridgewater, New Jersey 08920; §Betton ToxPath Consulting LLP, Macclesfield, UK, SK11 0SE; ¶sanofi-aventis, Vitry-sur-Seine 94 400, France; ||Bayer Schering Pharma AG, D-13342 Berlin, Germany; |||Bristol-Myers Squibb Company, Princeton, New Jersey 08540-4000; #Biorn International, Dublin, Ireland; ¶¶Pfizer, St. Louis, Missouri 63141; **Pfizer Inc., Sandwich, Kent CT13 9NJ, UK; ††Argutus Medical, Dublin, Ireland; ‼Liverpool John Moores University, Liverpool L3 3AF, UK; and †††HESI, Washington, DC 20005

¹Present address: Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK.
²Present address: The Ferns, Hilltown BT343WE, Northern Ireland.
³Present address: Aushon Biosystems, Billerica, Massachusetts 01821.
⁴To whom correspondence should be addressed. Fax: (202) 659-3617. E-mail: spettit@ilsi.org.

The ILSI-HESI Technical Committee on Biomarkers of Toxicity, Nephrotoxicity Working Group is a consortium of pharmaceutical companies and government and academic scientists, whose mission is to advance development of biomarkers of target organ toxicity that bridge from the preclinical to the clinical stages of drug development. Participants in the consortium are Argutus Medical (formerly Biotrin International), AstraZeneca, Bayer, Bristol-Myers Squibb, GlaxoSmithKline, Pfizer, sanofi-aventis, and the University of Arizona. ILSI-HESI is an international nonprofit organization that stimulates and supports scientific research that contributes to the collaborative identification and resolution of health and environmental issues of concern to the public, scientific community, government agencies, and industry.

Received February 18, 2011; accepted April 20, 2011

This study reports the evaluation of four urinary biomarkers of renal toxicity, α-glutathione-S-transferase (α-GST), μ-GST, clusterin, and renal papillary antigen-1 (RPA-1), in male Sprague-Dawley and Han-Wistar rats given cisplatin, gentamicin, or N-phenylanthranilic acid (NPAA). Kidney injury was diagnosed histopathologically, according to site/nature of renal injury, and graded for severity. The area under the receiver operating characteristic (ROC) curve was used to compare the diagnostic accuracy of each exploratory renal biomarker with traditional indicators of renal function and injury (blood urea nitrogen [BUN], serum creatinine [sCr] as well as urinary N-acetyl-β-D-glucosaminidase [NAG] and protein). These analyses showed that increased urinary α-GST was superior to BUN, sCr, and NAG for diagnosis of proximal tubular (PT) degeneration/necrosis. Paradoxically, increased α-GST was decreased in the presence of collecting duct (CD) injury without PT injury (NPAA administration). RPA-1 demonstrated high specificity for CD injury, superior to all of the reference biomarkers. The clusterin response correlated well with tubular injury, whatever the location, particularly when regeneration was present (superior to all of the reference markers for cortical tubular regeneration). There was no conclusive evidence for the diagnostic utility of μ-GST. The data were submitted for qualification review by the European Medicines Agency and the US Food and Drug Administration. Both agencies concluded that the data justified the qualification of RPA-1 and increased the level of evidence for, and clarified the context of use of, the previously qualified clusterin for use in male rats. These biomarkers can be used in conjunction with traditional clinical chemistry markers and histopathology in Good Laboratory Practice rodent toxicology studies used to support renal safety studies in clinical trials. Qualification of α-GST must await further explanation of the differences in response to PT and CD injury.

Key Words: urinary biomarkers; α-GST; RPA-1; clusterin; NAG; protein; albumin; proximal tubule; collecting duct; degeneration; necrosis; regeneration; RPN; receiver operating characteristic; biomarker performance; cisplatin; gentamicin; N-phenylanthranilic acid.

Nephrotoxicity is difficult to monitor noninvasively. The most widely used and accepted indicators of kidney injury include plasma/serum markers, such as, blood urea nitrogen (BUN) and serum creatinine (sCr), and urinary markers, such as, urinary volume, specific gravity or osmolality, protein, fractional electrolyte excretion, or sediment examination. Although many of these measurements are valid indicators of renal function, they lack sensitivity and/or specificity in detecting early stages of injury or disease and in some cases may be influenced by prerenal changes, which can make interpretation difficult. Recent efforts have resulted in the biological qualification of seven urinary biomarkers (Dieterle et al., 2010; Vaidya et al., 2010; Yu et al., 2010) for the detection of nephrotoxic injury to either the renal tubules or the glomeruli in preclinical safety studies. These seven biomarkers...
Glutathione-S-transferases (GSTs) are phase II detoxifying enzymes that exist in the kidney as various isoforms (Beckett and Hayes, 1993). Immunohistochemical studies reveal that the distribution and expression of the different isoforms varies along the nephron and between species (Campbell et al., 1991; Harrison et al., 1989; Rozell et al., 1993; Sundberg et al., 1993). Alpha-GST is the predominant isoform in the PT of both rat and humans, whereas μ-GST (GSTYb1) and π-GST are the predominant isoforms in the distal tubule (DT) of rats and humans, respectively. Alpha-GST is constitutively expressed at a high concentration (approximately 2% of soluble protein) in the cells of the PT (Beckett and Hayes, 1993) and it may be induced in response to xenobiotic metabolic demands. The increased presence of GSTs in the urine after nephrotoxic injury to rats has been known for more than 30 years (Bass et al., 1979) and is attributed to leakage from tubular cells into the lumen, secondary to epithelial cell damage. In a study of the effects of volatile anesthetics on the kidney in rats, Kharasch et al. (1998) reported that urinary excretion of α-GST correlated with the extent of PT cell necrosis. Measurement of the GST isoforms in urine also was more sensitive than either BUN or sCr for detection of tubular injury in a study in human volunteers given volatile anesthetics (Eger et al., 1997). Urinary excretion of specific isoforms of GST has been proposed not only as a marker of renal tubular damage in general but also to provide information on the location of the injury along the nephron (Branten et al., 2000; Eger et al., 1997; Sundberg et al., 1994a, b). However, to date there has been no systematic study of the value of measurement of urinary GSTs as biomarkers of renal tubular injury in either preclinical species or humans.

Clusterin, a dimeric glycoprotein, is expressed in the epithelia of many organs. In the kidney, it is highly expressed during early stages of development. In contrast, in the healthy mature kidney, clusterin messenger RNA (mRNA) and protein are not detectable but are upregulated in response to renal tubular injury and in a variety of renal diseases (Rosenberg and Silkensen, 1995). It has been suggested that secreted clusterin suppresses apoptosis and is involved in cell aggregation and attachment (Rosenberg and Silkensen, 1995). Clusterin expression is upregulated in rats following nephrectomy (Correa-Rotter et al., 1992), unilateral ureteral obstruction (Ishii et al., 2007), renal ischemia-reperfusion (Yoshida et al., 2002), or nephrotoxicity (Kharasch et al., 2006; Silkensen et al., 1997) and in dogs with RPN induced by nefiracetam (Tsuchiya et al., 2005). Increased levels of clusterin protein in urine have been detected following ischemic or chemically induced injury in rats (Aulitzky et al., 1992; Eti et al., 1993; Hidaka et al., 2002) or dogs (Tsuchiya et al., 2005). Clusterin was among those urinary proteins recently shown to outperform BUN and sCr for detection of chemical-induced tubular injury (Dieterle et al., 2010). Although increased expression of clusterin (mRNA and/or protein) is seen in humans in a variety of renal disorders (Rosenberg and Silkensen, 1995), to date there has been no clinical study demonstrating the use of clusterin as a diagnostic marker of renal injury.

RPN can be induced experimentally in rats by nonsteroidal anti-inflammatory drugs and chemicals such as 2-bromomethane or N-phenylantranilic acid (NPAA), a biphenyl analog of mafenamic acid (Betton et al., 2005; Price et al., 2010). To develop markers of renal papillary injury, monoclonal antibodies were raised against proteins released in urine from rats treated with nephrotoxicants, and the antibodies were screened for those that correspond to segment-specific antigens (Falkenberg et al., 1996; Hildebrand et al., 1999). This procedure identified an antigen, renal papillary antigen-1 (RPA-1), which is specifically expressed in the collecting duct (CD). Price et al. (2010) found RPA-1 particularly difficult to characterize but showed it to be a very large membrane-bound glycoprotein located specifically within the CDs of rat kidney. They also showed it to be a good early marker of RPN in the rat. RPA-1 was evaluated in the present study as a potential urinary biomarker of CD injury.

The studies reported here utilized cisplatin, gentamicin, and NPAA to induce injury to different parts of the renal tubule. The nephrotoxic properties of cisplatin and gentamicin have been extensively characterized, and they were chosen as good model compounds for inducing PT injury. NPAA was selected to study CD injury as it has been shown to cause RPN in rats after daily oral dosing for 14 days (Price et al., 2010; Williams et al., 2003). Four biomarkers (α-GST, μ-GST, RPA-1, and clusterin) were systematically evaluated and their diagnostic accuracy as markers of renal injury in nephrotoxicant-treated male Sprague-Dawley and Han-Wistar rats assessed. These four biomarkers were selected for study in part because of their potential to provide information on the site of renal injury given their different tubular locations as illustrated by immunohistochemistry (Fig. 1). Also evaluated were albumin (one study only) and the PT brush border membrane enzyme, γ-glutamyl transferase.
GGT has intermittently been advocated as having diagnostic value for renal tubular injury, for example, caused by contrast media (Donadio et al., 1998) and was found to be an excellent PT biomarker in clinical acute tubular necrosis (Westhuyzen et al., 2003). The overall aim of the program was to provide evidence to qualify the exploratory biomarkers for use in nonclinical studies to support safety evaluation in clinical trials of new drugs. This would be based on a comparison of the diagnostic accuracy of each exploratory renal biomarker with traditional indicators of renal function (BUN and sCr) as well as urinary protein and N-acetyl-β-D-glucosaminidase (NAG) using the area under the receiver operating characteristic (ROC) curve (AUCROC). The data were submitted to the European Medicines Agency (EMA) and the US Food and Drug Administration (FDA) for their review and decision concerning the utility of these exploratory biomarkers in nonclinical studies used to support human trials of candidate medicines.

**MATERIALS AND METHODS**

**Experimental approach/study design.** In the studies described here, a total of 340 male Sprague-Dawley or Han-Wistar rats were used. One study was performed with each strain using cisplatin and NPAA; gentamicin was studied only in Sprague-Dawley rats. Sprague-Dawley rats, 10–12 weeks of age, from Charles River Laboratories or Wistar Hannover, substrain Han-Wistar, rats, 8–10 weeks of age, supplied by the AstraZeneca Rodent Breeding Unit were used in the NPAA studies; the animals were fed, respectively, 5002 certified rodent diet or RM1 diet. Sprague-Dawley rats from Charles River Laboratory at approximately 7 weeks of age were used in the gentamicin study and were fed Harlan Diet 2018C. Sprague-Dawley rats at approximately 6 weeks of age or Han-Wistar rats, approximately 7 weeks old, both supplied by Charles River Laboratories were used for the cisplatin studies; the animals were fed, respectively, rodent chow A04-10 pelleted diet or pulverized rodent chow (Sniff R/M-H diet). All animal husbandry and experimental use was conducted in accordance with national legal requirements and all studies were approved in each participating facility by the Institutional Animal Care and Use Committee (or its equivalent).

Dose range finding studies were designed and conducted based on published studies and prior experience. The doses used in the definitive studies were chosen based on the results of the dose range studies in order to generate a range of responses from subthreshold to overtly nephrotoxic. Control data were provided by animals that were administered vehicle only. The study designs are summarized in Table 1.

Animals were dosed once (or once per day) early in the morning (1 dose for cisplatin; 14 daily doses for gentamicin and NPAA). Serum, urine, and tissue samples were obtained at defined time points from nephrotoxicant-treated and control rats. Urine was collected over ice for a timed period (i.e., 16 h overnight) and diluted in buffered salt solution containing bacteriostatic agents. Urine samples were stored at 4–8°C for no more than 48 h, after which they were stored at –20°C.

**Evaluation of liver by clinical and anatomical pathology.** In selected dose-ranging or definitive studies, serum parameters indicative of hepatic injury/function (serum alanine and aspartate aminotransaminases and total bilirubin) were measured and tissue sections from liver were prepared, fixed in neutral buffered formalin, and processed and examined at the discretion of the study pathologist. Hepatobiliary markers and/or liver histopathology were evaluated in definitive studies in both strains for cisplatin. Hepatobiliary markers were assessed in definitive studies of NPAA (Han-Wistar) and gentamicin (Sprague-Dawley), and liver histopathology was evaluated in dose range studies of NPAA and gentamicin studies at doses comparable with, or higher than, those used in the definitive studies.

**Traditional and exploratory biomarker assays for renal function and injury.** Established methods were used to assay traditional urine and serum parameters indicative of renal function (i.e., urine parameters: volume, protein, glucose, NAG, GGT, and creatinine; serum parameters: sCr, BUN, and total protein). Urinary α-GST, μ-GST, clusterin, and RPA-1 were assayed using enzyme immunoassay (EIA) kits that were developed, manufactured, and qualified for use by Argutus Medical Ltd, formerly Biotrin International (Dublin, Ireland). Details of these methods can be found in Gautier et al. (2010). Repeatability (intraassay variability) and intermediate precision...
Cisplatin Han-Wistar and Sprague-Dawley 0 (vehicle
NPAA Han-Wistar 0 (vehicle
eGentamicin Sprague-Dawley 0 (vehicle

conducted according to the published best practices (Burkhardt

histopathology data for ROC analysis.

sections were examined per rat to ensure that papillary injury was not missed in

However, because NPAA-mediated papillotoxicity can be limited to the apex of

additional stains as desired). Histopathological assessment was based on

removed and weighed, fixed by immersion in neutral buffered formalin,
euthanized by anesthetic overdose followed by exsanguination and necropsied

assessment using ROC curve methods (data not shown).

and creatinine-normalized values. Both methods provided similar results as

all values were expressed as a fold change versus the time-matched control group

animal were divided by the mean value for the concurrent control animals. Thus,

there was no change in the amount of creatinine excreted in the urine of

account for variability in urine volume. In these studies, it was first established

transformation and dividing by the observed concentration mean.

Normalization to urinary creatinine is the most commonly used method to

for variability in urine volume. In these studies, it was first established that there was no change in the amount of creatinine excreted in the urine of control or treated animals over the time course of the studies (data not shown).

Therefore, for all urinary markers, analyte concentrations were normalized to the

urine creatinine. These normalized data values for every individual

Thus, all values were expressed as a fold change versus the time-matched control group

mean. A further check on the validity of this approach was done by comparing

variance estimates and confidence intervals were determined and then

converted to coefficient of variation (CV) by applying the square root transformation and dividing by the observed concentration mean.

Preparation of kidney samples for histopathology. Animals were euthanized by anesthetic overdose followed by exsanguination and necropsied using procedures approved at the individual facilities. Both kidneys were removed and weighed, fixed by immersion in neutral buffered formalin, trimmed, embedded, sectioned, and stained with hematoxylin and eosin (or additional stains as desired). Histopathological assessment was based on a single section of each kidney that included cortex, medulla, and papilla. However, because NPAA-mediated papillotoxicity can be limited to the apex of the renal papilla that extends beyond the renal hilus in the rat, up to six-step sections were examined per rat to ensure that papillary injury was not missed in the NPAA studies.

Definition of renal pathology nomenclature and processing of histopathology data for ROC analysis. The histopathology examination was conducted according to the published best practices (Burkhardt et al., 2010; Crissman et al., 2004). For each study, a board certified veterinary pathologist conducted an initial unblinded assessment with full knowledge of the treatment group and, in some cases, with knowledge of group mean (but not individual) data for standard clinical pathology parameters. However, in no case were data relating to the biomarkers under evaluation known to the pathologist. After this initial assessment, the study pathologist performed a targeted masked evaluation of selected slides, without knowledge of treatment group, to reevaluate findings in specific tissues and groups as needed to resolve treatment-related changes from normal variation. A peer review was conducted by a second board certified pathologist initially unblinded to treatment group and then, if considered necessary, a targeted masked evaluation (blinded) was performed to resolve discrepancies.

Histopathology grade was assessed on a scale from 0 (normal tissue) to 4 (severe pathology). For each specific pathological diagnosis, regardless of treatment with vehicle or toxicant, animals with histopathology score ≥ 0 were defined as “negative” and animals with histopathology score > 0 were defined as “positive.” The number of negative and positive animals was determined for each pathology and stratified by rat strain and toxicant.

Slides and histopathology data from all five HESI studies were reviewed by a Pathology Working Group to assess morphological diagnoses and consistency in grading, to identify key treatment-related findings for each toxicant, and to derive a standardized lexicon of morphological diagnoses. Key agreements from this peer review were subsequently integrated with the nascent kidney histopathology lexicon of the C-Path PSTC Nephrotoxicity Working Group and a unified renal histopathology lexicon was prepared (Sistare et al., 2010). Subsequent to the Pathology Working Group and prior to performing the ROC analysis (see “Biomarker performance via ROC curve analysis” section), morphological diagnoses from the five HESI studies were converted to conform to the common lexicon. Histopathology data were further processed to remove redundancies and ensure that each animal had only one diagnosis per pathological process per segment, resulting in identification of 13

TABLE 1
Dosage Regimens, Numbers of Animals, and Timing of Key Procedures by Rat Strains in Cisplatin, Gentamicin, and NPAA Studies

<table>
<thead>
<tr>
<th>Compound</th>
<th>Strain of rat</th>
<th>Dose (mg/kg/day)</th>
<th>Dose volume (ml/kg)</th>
<th>Days of necropsy</th>
<th>Total no. of animals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cisplatin</td>
<td>Han-Wistar and Sprague-Dawley</td>
<td>0 (vehicle)</td>
<td>5</td>
<td>2, 3, 5 (10)</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>5</td>
<td>2, 3, 5 (10)</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>5</td>
<td>2, 3, 5 (10)</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>5</td>
<td>2, 3, 5 (10)</td>
<td>30</td>
</tr>
<tr>
<td>Gentamicin</td>
<td>Sprague-Dawley</td>
<td>0 (vehicle)</td>
<td>1.0</td>
<td>8 or 15 (10)</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>1.0</td>
<td>15 (10)</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>1.0</td>
<td>15 (10)</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>1.0</td>
<td>15 (10)</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>1.0</td>
<td>8 or 15 (10)</td>
<td>20</td>
</tr>
<tr>
<td>NPAA</td>
<td>Han-Wistar</td>
<td>0 (vehicle)</td>
<td>10</td>
<td>8 or 15 (15)</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>10</td>
<td>15 (15)</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>350</td>
<td>10</td>
<td>15 (15)</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>700</td>
<td>10</td>
<td>8 or 15 (15)</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>10</td>
<td>8 or 15 (15)</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>10</td>
<td>15 (15)</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400</td>
<td>10</td>
<td>8 or 15 (15)</td>
<td>30</td>
</tr>
</tbody>
</table>

aDoses (based on the most recent body weight) were administered once by intraperitoneal route for cisplatin, daily for 7 or 14 days by subcutaneous route for gentamicin and by oral route for NPAA and are expressed in terms of pure parent compound for cisplatin and NPAA and as a base for gentamicin.

bVehicle was 1.25% carboxymethylcellulose.

cVehicle was 0.9% saline.

dNo high dose animals survived to study day 15; animals were euthanized as follows (1 animal on day 6, 11 animals with controls on day 7, 4 animals with controls on each of days 8 and 10).

eVehicle was 1.25% carboxymethylcellulose.

fDose was reduced from day 5 to 350 mg/kg/day for animals sacrificed on day 8 and to 500 mg/kg/day for animals sacrificed on day 15.
morphological diagnoses. Morphological diagnoses were further assessed by members of the Pathology Working Group for relationship to treatment and prioritized for ROC analysis. The majority of findings observed were related to degeneration/necrosis and/or regeneration/basophilia involving PT or CD. PT degeneration was never identified in association with degeneration/necrosis of CDs; however, a small subset of rats with CD degeneration had concurrent degeneration of the DT. A diagram summarizing the processing of the histopathology data is shown in Figure 2.

For the ROC analysis, additional subcategories were created to assess concomitant degenerative and regenerative changes in either PT or CD and also to evaluate regeneration/basophilia in the absence of evidence of tubular degeneration. Thus, marker performance was evaluated by ROC analysis for degeneration/necrosis of PT or CD in the presence and absence of regeneration/ basophilia (PT or CD degeneration/necrosis with and without regeneration). In addition, rats with evidence of tubular basophilia/regeneration were split into a cohort in which degeneration/necrosis of either PT or CD was identified (cortical tubular regeneration/basophilia) and one in which there was no evidence of degeneration, interpreted as spontaneous rodent nephropathy (regeneration not otherwise specified [NOS] with no degeneration). Finally, the diagnosis of “tubular cell alteration, vacuolation” was excluded from further analysis because of low incidence and lack of dose response, and findings of “tubular dilation” and “mineralization, papilla” were not used in the ROC analysis because of lack of added value over concurrent diagnoses for CD degeneration/necrosis.

Biomarker performance via ROC curve analysis. The discriminatory accuracy of each marker was assessed using ROC curve methods. AUCROC-a commonly used index of diagnostic accuracy, was used to compare the performance of each marker (Hanley and McNeil, 1982; Pepe, 2003).

Animals for which data were missing for any biomarker (BUN, sCr, NAG, protein, α-GST, μ-GST, RPA-1, and clusterin) were excluded from statistical analyses to enable comparison of all biomarkers using a common subset of samples. Additionally, all animals from the high-dose group (100 mg/kg/day) of the gentamicin study in Sprague-Dawley rats were excluded from statistical analyses. In this group, unscheduled sacrifices due to the poor clinical condition of some animals occurred without collection of clinical chemistry or urinalysis data.

For the ROC analysis, additional subcategories were created to assess concomitant degenerative and regenerative changes in either PT or CD and also to evaluate regeneration/basophilia in the absence of evidence of tubular degeneration. Thus, marker performance was evaluated by ROC analysis for degeneration/necrosis of PT or CD in the presence and absence of regeneration/ basophilia (PT or CD degeneration/necrosis with and without regeneration). In addition, rats with evidence of tubular basophilia/regeneration were split into a cohort in which degeneration/necrosis of either PT or CD was identified (cortical tubular regeneration/basophilia) and one in which there was no evidence of degeneration, interpreted as spontaneous rodent nephropathy (regeneration not otherwise specified [NOS] with no degeneration). Finally, the diagnosis of “tubular cell alteration, vacuolation” was excluded from further analysis because of low incidence and lack of dose response, and findings of “tubular dilation” and “mineralization, papilla” were not used in the ROC analysis because of lack of added value over concurrent diagnoses for CD degeneration/necrosis.

Assessment of Liver Injury No evidence of liver injury either by measurement of hepatobiliary biomarkers or histopathology was detected in control animals or nephrotoxicant-treated animals at the highest doses used during the course of these studies. When liver histopathology was not done in the definitive studies, it was included in the dose-ranging studies at comparable or higher doses. Based on weight of evidence, liver injury was not present as a potential confounding factor in these studies.

Following histopathological assessment of the kidneys, for the purpose of ROC analysis, each animal was assessed as negative or positive for the presence of pathology, whether treated with vehicle or nephrotoxic agent. No account was taken of severity score; however, subsequently AUCROC estimates for the markers and pathologies of greatest interest were calculated using successively reduced ranges of histopathology scores.

For each marker and pathology, nonparametric point estimates and standard error of the AUCROC were calculated (Hanley and McNeil, 1982) by rat strain and across both rat strains. Using pooled data for both rat strains stratified according to pathology, pairwise comparisons of AUCROC were performed for each exploratory marker (α-GST, μ-GST, RPA-1, and clusterin) versus each reference marker (BUN, sCr, NAG, and protein). In addition, pairwise comparisons of AUCROC were performed for NAG and protein versus BUN and sCr. For each pairwise AUCROC comparison, the two-sided p value was calculated (DeLong et al., 1988). For each novel marker separately, raw p values were adjusted for multiple testing (i.e., pathologies and reference markers) via the Hochberg procedure (Hochberg, 1998). No adjustment was made for the number of novel markers considered.

The incremental diagnostic value of each exploratory marker (α-GST, μ-GST, RPA-1, and clusterin) was also assessed, using ROC curves, in conjunction with two reference marker pairs: BUN+sCr or NAG+protein (see Supplementary tables 20–23 for details of the method).

RESULTS

Assessment of Liver Injury No evidence of liver injury either by measurement of hepatobiliary biomarkers or histopathology was detected in control animals or nephrotoxicant-treated animals at the highest doses used during the course of these studies. When liver histopathology was not done in the definitive studies, it was included in the dose-ranging studies at comparable or higher doses. Based on weight of evidence, liver injury was not present as a potential confounding factor in these studies.
Distribution of Renal Pathologies in Nephrotoxicant-Treated Rats. No glomerular changes were observed in any of the studies with cisplatin, NPAA, or gentamicin. Histopathological findings observed in the tubules and interstitium were subclassified into 13 discrete morphological diagnoses based on localization and character of renal injury. All samples were also classified according to histopathological grade. The incidence and distribution of histopathology findings for each rat strain is shown in Table 2, stratified according to positive versus negative response and primary versus secondary pathogenesis. Histopathology findings are shown further stratified by experimental compound in Supplementary tables 1–3. In general, Sprague-Dawley rats had renal histopathological changes of lower incidence and severity than Han-Wistar rats (Figs. 3–5). This is exemplified by CD changes in rats given NPAA where changes in Sprague-Dawley rats were lower in severity than in Han-Wistar rats (Fig. 4) and most notably were unaccompanied by the regenerative changes (medullary tubular regeneration/basophilia) seen in the Han-Wistar rats (Table 2).

Repeatability and Intermediate Precision (intra- and inter-assay variability) of Assays of Exploratory Urinary Biomarkers Data generated to assess assay variability are not shown but summarized as follows. The %CV for intraassay variability was ≤ 10% for all of the exploratory biomarkers across all of the six participating laboratories. Inters assay variability was greater with the observed CV ranging across laboratories from 7–17% (α-GST), 9–13% (μ-GST), 2–14% (RPA-1), and 16–30% (clusterin). This is in line with the suggested acceptance criterion of 20–30%CV for intermediate precision made in the report of the American Association of Pharmaceutical Scientists Biomarkers Workshop (Lee et al., 2005).

Interlaboratory assay variability was also assessed for all assays by exchanging biological samples between all six participating laboratories and the %CV was found to be ≤ 30%.

Biomarker Response in Nephrotoxicant-Treated Rats Summary statistics for each biomarker, stratified by pathology, toxicant, and rat strain are presented (as mean fold-change relative to time-matched control mean) in Supplementary tables 4–11. GGT was systematically evaluated in this program of work but was found to have little diagnostic value, with no or only marginal increases in the presence of most pathologies and estimated AUCROC values generally < 0.60; therefore, no results are presented. Albumin was assessed in only one study and the results have been reported elsewhere (Gautier et al., 2010).

Traditional markers for renal injury including BUN, sCr, and urinary protein were unchanged for most of the pathologies observed in these studies. Where they occurred, increases in BUN, sCr, and urinary protein were generally small, rarely greater than a mean of twofold (threefold in the case of protein). These increases were attributable to marked PT injury with regeneration and cortical tubular hyaline casts seen at later time points in Han-Wistar and Sprague-Dawley rats given cisplatin and Sprague-Dawley rats given higher doses of gentamicin.

Like protein, increases in urinary NAG rarely exceeded an overall mean of threefold but in the case of NAG, the increases were seen not only in response to PT necrosis/degeneration with regeneration but also in the presence of CD injury, with or without regeneration. NAG was moderately increased in toxicant-treated Han-Wistar rats with CD necrosis/degeneration (fourfold), an increase that correlated with the greater severity of CD injury in Han-Wistar compared with Sprague-Dawley rats.

### TABLE 2

<table>
<thead>
<tr>
<th>Pathology</th>
<th>Sprague-Dawley</th>
<th>Han-Wistar</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negative</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>PT degeneration or necrosis</td>
<td>170</td>
<td>34</td>
<td>307</td>
</tr>
<tr>
<td>PT deg/nec with no regen</td>
<td>195</td>
<td>9</td>
<td>153</td>
</tr>
<tr>
<td>PT deg/nec with regen</td>
<td>179</td>
<td>25</td>
<td>176</td>
</tr>
<tr>
<td>Cortical tubular regeneration/basophilia</td>
<td>153</td>
<td>51</td>
<td>146</td>
</tr>
<tr>
<td>DT degeneration or necrosis</td>
<td>184</td>
<td>20</td>
<td>192</td>
</tr>
<tr>
<td>CD degeneration or necrosis</td>
<td>179</td>
<td>25</td>
<td>161</td>
</tr>
<tr>
<td>CD deg/nec with no regen</td>
<td>198</td>
<td>6</td>
<td>179</td>
</tr>
<tr>
<td>CD deg/nec with regen</td>
<td>185</td>
<td>19</td>
<td>174</td>
</tr>
<tr>
<td>Medullary tubular regeneration/basophilia</td>
<td>204</td>
<td>0</td>
<td>178</td>
</tr>
<tr>
<td>Regeneration NOS with no degeneration</td>
<td>197</td>
<td>7</td>
<td>173</td>
</tr>
<tr>
<td>Intratubular casts, granular, cortex</td>
<td>199</td>
<td>5</td>
<td>185</td>
</tr>
<tr>
<td>Intratubular casts, hyaline, cortex</td>
<td>188</td>
<td>16</td>
<td>176</td>
</tr>
<tr>
<td>Inflammation, interstitial, chronic, cortex</td>
<td>142</td>
<td>62</td>
<td>186</td>
</tr>
</tbody>
</table>

**Note.** deg, degeneration; nec, necrosis; regen, regeneration.
Urinary α-GST was increased in the presence of PT degeneration/necrosis with or without regeneration and also with the diagnosis of cortical tubular regeneration. Alpha-GST was also increased in animals with hyaline and granular casts in cortical tubules and cortical interstitial inflammation. These increases were particularly marked in the Han-Wistar rats, where the severity of the histopathological changes was greater. In contrast, α-GST decreased in animals with CD injury.

Overall, increases in μ-GST essentially paralleled the increases seen with α-GST. However, this was heavily influenced by the severe pathology recorded with gentamicin.

**FIG. 3.** Scatterplot of α-GST fold change versus histopathology grade for PT degeneration or necrosis. Animals with no renal pathology (histopathology grade = 0) were stratified according to whether they were given vehicle (Veh) or toxicant (Trt).

**FIG. 4.** Scatterplot of RPA-1 fold change versus histopathology grade for CD degeneration or necrosis. Animals with no renal pathology (histopathology grade = 0) were stratified according to whether they were given vehicle (Veh) or toxicant (Trt).
(study conducted only in the Sprague-Dawley rats) in which the marked PT regeneration/degeneration, intratubular casts and inflammation were accompanied by DT injury.

RPA-1 increased in animals with CD degeneration/necrosis, with a greater magnitude in Han-Wistar (~10-fold) than in Sprague-Dawley rats (~threefold), corresponding to a higher severity of the lesion in the Han-Wistar strain. RPA-1 was also increased in association with cortical or medullary regeneration of CDs.

Clusterin was increased with several diagnoses, including PT degeneration with regeneration, CD degeneration (with or without regeneration), intratubular hyaline and granular casts (the latter only in Sprague-Dawley rats), and interstitial inflammation. The mean increase was greatest for PT degeneration with regeneration (12-fold). Both strains were affected but the magnitude of the clusterin signal was somewhat greater in Han-Wistar than in Sprague-Dawley rats, consistent with the greater severity of injury in the former strain.

The biomarker response in individual animals is illustrated in Figures 3–5, which show the fold change of the biomarkers of principal interest plotted against the severity grade for the primary corresponding histopathological diagnoses. Figure 3 shows fold change in α-GST in animals with PT degeneration/necrosis (with or without regeneration); Figure 4 shows fold change in RPA-1 in animals with CD degeneration/necrosis (with or without regeneration); Figure 5 shows fold change in clusterin in animals with cortical tubular regeneration/basophilia. These scatter plots show that the magnitude of biomarker response correlates positively with severity of renal pathology (i.e., greater fold change in biomarker correlates with increasing histopathology grade). The response is qualitatively similar between strains; the greater fold change in Han-Wistar rats versus Sprague-Dawley rats reflects both the higher incidence and severity of injury in the Han-Wistar rats and the limited ability of a 4-point discontinuous grading system to discriminate fully between individual animals. It is evident from Figures 4 and 5 that, in some nephrotoxicant-treated animals of both strains, urinary RPA-1 and clusterin were raised in the absence of any recorded pathology.

**Performance of α-GST, μ-GST, Clusterin, and RPA-1 Assessed by ROC Curve Analysis** ROC curve analysis was performed and AUCROC values initially calculated and compared pairwise for each novel biomarker and each reference biomarker stratified by rat strain and by pathology. Nonparametric AUCROC estimates for each biomarker and pathology are presented for Sprague-Dawley and Han-Wistar rats as Supplementary tables 12 and 13. From these data, it is clear that the differences between strains in magnitude of response of the exploratory biomarkers to some key pathologies as described in the preceding section are not notable when AUCROC's are compared. Table 3 illustrates the comparable diagnostic performance between strains of α-GST, RPA-1, and clusterin against three key pathologies, based on closely similar AUCROC's.

Estimates of AUCROC were also calculated for strains combined, thereby incorporating variability between strains and providing an average assessment of biomarker performance across the two strains. All the statistical analyses were repeated using the results of strains combined, i.e., pairwise
RPA-1 relative to the four reference biomarkers for CD basophilia:

<table>
<thead>
<tr>
<th>Pathology</th>
<th>Biomarker</th>
<th>Strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT degeneration or necrosis</td>
<td>α-GST</td>
<td>Sprague-Dawley</td>
</tr>
<tr>
<td>CD degeneration or necrosis</td>
<td>RPA-1</td>
<td>Han-Wistar</td>
</tr>
<tr>
<td>Cortical tubular regeneration/basophilia</td>
<td>clusterin</td>
<td></td>
</tr>
</tbody>
</table>

AUCROC Estimates (standard error) for Sprague-Dawley and Han-Wistar Animals for Selected Pathologies and Biomarkers

Comparison for each exploratory biomarker and each reference biomarker for pooled data stratified by pathology (Table 4). All conclusions drawn on the relative performance (diagnostic utility) of the various biomarkers are based on these combined analyses.

ROC curves for the pathologies of principal interest (pooled data) are shown in Figures 6–8. Figures 6 and 8 summarize the performance of α-GST and clusterin relative to the reference biomarkers (BUN, sCr, and urinary NAG, and protein) for PT degeneration/necrosis and cortical tubular regeneration/basophilia, respectively. Figure 7 summarizes the performance of RPA-1 relative to the four reference biomarkers for CD degeneration/basophilia.

Pairwise comparisons of the AUCROC values derived from pooled data (strains combined) for each exploratory biomarker and pathology with each of the reference biomarkers are presented as Supplementary tables 14–17. Also included are pairwise comparisons of the two urinary biomarkers, NAG and protein versus BUN and sCr (Supplementary tables 18 and 19). In addition, as shown in these Supplementary Data, incremental value of the four exploratory biomarkers over the combinations of reference biomarkers in serum (BUN+sCr) or urine (NAG+protein) was assessed by statistical comparison of the AUCROC estimates.

Based on all of these statistical evaluations, several conclusions can be drawn.

Although the results suggest potential utility of increased urinary NAG and protein to report PT degeneration/necrosis with regeneration (AUCROC estimates = 0.87 and 0.89, respectively), neither was statistically superior to BUN or sCr in pairwise comparisons. Furthermore, ROC analysis suggested little diagnostic utility (AUCROC ≤ 0.63) for the modest increases of NAG (overall mean of up to threefold) seen in the presence of CD degeneration/necrosis, with or without regeneration.

Increased urinary α-GST is a sensitive diagnostic biomarker for PT degeneration/necrosis (AUCROC = 0.84). In pairwise comparison (Supplementary table 14), the performance of α-GST was superior to BUN (p < 0.001), sCr (p < 0.001), and NAG (p < 0.001). Although an AUCROC of 0.87 also suggests that α-GST would be a useful biomarker when regeneration is present, at this late stage, it is not superior to the reference biomarkers, which also performed well (AUCROC estimates ranging from 0.79–0.89). The superiority of α-GST is more evident at earlier times when no regeneration is present (p < 0.05 versus both sCr and NAG; Supplementary table 14). The statistical analyses also highlighted urinary α-GST as a sensitive marker of CD injury (AUCROC = 0.92). However, this diagnosis occurred in the absence of PT injury and the direction of change in urinary α-GST was a decrease, not an increase.

RPA-1 demonstrates high specificity for CD injury and, in pairwise comparison, the performance of RPA-1 (AUCROC = 0.93) was superior to all reference biomarkers for CD degeneration/necrosis, particularly with regeneration (p < 0.001).

The highest AUCROC for clusterin (0.94) was in the presence of PT degeneration/necrosis with regeneration but this was statistically significant only in comparison with NAG. Although the clusterin AUCROC for cortical tubular regeneration/basophilia (a diagnosis indicating regenerative response to degenerative change in either PT or CD) was lower (0.81), this was statistically superior (p < 0.001) to all the reference biomarkers. In contrast, in kidneys with a diagnosis of regeneration NOS without regeneration (i.e., normal tissue with a background level of spontaneous nephropathy) AUCROC for clusterin was 0.56. These data indicate that the clusterin response correlates well with tubular injury, particularly when regeneration is present. Clusterin was also superior to BUN and protein (p < 0.001) and NAG (p < 0.05) in diagnosis of CD degeneration/necrosis.

Although the statistical comparisons of the AUCROC estimates indicated better diagnostic accuracy for μ-GST over BUN and/or sCr for PT or CD injury (Supplementary table 15), the magnitude of the AUC values for μ-GST was relatively modest (0.70–0.77) and there was no consistent direction of change in μ-GST. Thus, the data do not indicate utility of μ-GST for detection of PT or CD injury. Furthermore, although the AUC for μ-GST (0.87) was statistically superior to BUN and sCr for detection of DT degeneration/necrosis, this was due to a decrease, rather than an increase, in urinary μ-GST. There were, in fact, very few diagnoses of DT injury in these studies and, with the exception of one rat, all occurred in rats with concurrent CD injury. Therefore, this data set is considered to be insufficient to support any conclusion about the diagnostic value of μ-GST in DT injury.

These conclusions are supplemented by the results of the assessment of incremental diagnostic value (Supplementary tables 20–23). Notably, α-GST and RPA-1 were shown to add diagnostic value to either pair of reference markers (BUN+sCr or NAG+protein) for PT injury or CD injury. The added value for α-GST for CD injury reflects a consistent decrease of urinary α-GST in urine with CD injury in the absence of PT injury, an observation that is without explanation at present. Additionally, these results suggest added value of clusterin over both combinations of reference markers for diagnosis of regenerative changes in the PT and of CD injury.
Although combination of the exploratory biomarkers with the reference (traditional) markers enhanced diagnostic performance of the traditional markers as noted above, the reverse, i.e., combination of traditional markers with the novel urinary markers provided minimal or no improvement in diagnostic accuracy relative to that of the exploratory markers alone.

Table 4 presents an overall summary of the performance of the novel biomarkers versus all of the reference biomarkers for each of the diagnoses, based on statistical comparison of

<table>
<thead>
<tr>
<th>Pathology</th>
<th>Reference Markers</th>
<th>Exploratory Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BUN</td>
<td>SCr</td>
</tr>
<tr>
<td>PT degeneration or necrosis</td>
<td>0.62 (0.04)</td>
<td>0.62 (0.04)</td>
</tr>
<tr>
<td>PT deg/nec with no regen</td>
<td>0.56 (0.05)</td>
<td>0.58 (0.04)</td>
</tr>
<tr>
<td>PT deg/nec with regen</td>
<td>0.79 (0.05)</td>
<td>0.82 (0.05)</td>
</tr>
<tr>
<td>Cortical tubular regeneration/basophilia</td>
<td>0.62 (0.04)</td>
<td>0.64 (0.04)</td>
</tr>
<tr>
<td>DT degeneration or necrosis</td>
<td>0.52 (0.06)</td>
<td>0.67 (0.04)</td>
</tr>
<tr>
<td>CD degeneration or necrosis</td>
<td>0.54 (0.04)</td>
<td>0.57 (0.04)</td>
</tr>
<tr>
<td>CD deg/nec with no regen</td>
<td>0.64 (0.05)</td>
<td>0.60 (0.07)</td>
</tr>
<tr>
<td>CD deg/nec with regen</td>
<td>0.51 (0.05)</td>
<td>0.55 (0.05)</td>
</tr>
<tr>
<td>Medullary tubular regeneration/basophilia</td>
<td>0.59 (0.07)</td>
<td>0.66 (0.07)</td>
</tr>
<tr>
<td>Regeneration NOS with no degeneration</td>
<td>0.52 (0.05)</td>
<td>0.58 (0.05)</td>
</tr>
<tr>
<td>Intratubular casts, granular, cortex</td>
<td>0.62 (0.09)</td>
<td>0.59 (0.08)</td>
</tr>
<tr>
<td>Intratubular casts, hyaline, cortex</td>
<td>0.79 (0.06)</td>
<td>0.82 (0.05)</td>
</tr>
<tr>
<td>Inflammation, interstitial, chronic, cortex</td>
<td>0.63 (0.04)</td>
<td>0.64 (0.04)</td>
</tr>
</tbody>
</table>

*Note.* AUCROC values in bold are significantly different from reference markers (based on statistically significant differences at \( p < 0.05 \) level after prespecified multiplicity adjustment). deg, degeneration; nec, necrosis; regen, regeneration.

**AUCROC** is significantly larger than that for each reference marker except protein.

**AUCROC** is significantly larger than that for all reference markers.

**FIG. 6.** ROC curves for α-GST, BUN, sCr, NAG, and urine protein for classification of PT degeneration or necrosis.
AUC\textsubscript{ROCs} (for details see Supplementary tables 14–17). Values for the exploratory biomarkers AUC\textsubscript{ROCs} that were statistically superior to the reference biomarkers ($p < 0.05$ after multiplicity adjustment) are indicated.

All of the preceding analyses incorporated all animals designated as positive for pathology, irrespective of severity grade. AUC\textsubscript{ROC} estimates for selected markers and pathologies were calculated using successively reduced ranges of

![ROC curves for RPA-1, BUN, sCr, NAG, and urine protein for classification of CD degeneration or necrosis.](image1)

![ROC curves for clusterin, BUN, sCr, NAG, and urine protein for classification of cortical tubular regeneration/basophilia.](image2)
histopathology scores. Results are graphically summarized in Figures 9–11. The AUCROC estimate for RPA-1 (for classification of CD degeneration/necrosis) remains virtually unchanged, regardless of the range of histopathology scores incorporated. AUCROC estimates for α-GST and clusterin (for classification of PT degeneration/necrosis and cortical tubular regeneration/basophilia, respectively) decrease somewhat with successively reduced histopathology ranges. However,

**FIG. 9.** AUCROC estimates for α-GST, BUN, SCr, NAG, and protein for classification of PT degeneration or necrosis versus histopathology score range.

**FIG. 10.** AUCROC estimates for RPA-1, BUN, SCr, NAG, and protein for classification of CD degeneration or necrosis versus histopathology score range.
the α-GST and clusterin AUC\textsubscript{ROC} estimates remain substantially larger than those for BUN, sCr, NAG, and protein in all cases.

**DISCUSSION**

ROC analysis is a well-established analytical tool for assessing and comparing the performance of biomarkers in detecting human pathological states, relative to histopathological reference standards, and the AUC\textsubscript{ROC} is a useful parameter for comparing ROC curves (Zweig and Campbell, 1993). During ROC curve analysis, biomarker performance (diagnostic accuracy) is evaluated for both sensitivity and specificity. Importantly, all sensitivity-specificity pairs are evaluated, and thus performance can be predicted for all possible diagnostic thresholds.

All the conclusions presented here are based on ROC analyses of pooled data from both strains. These analyses incorporate variability between strains and provide an average assessment of biomarker performance across the two strains. Strain differences in response to nephrotoxins are known, particularly when metabolic activation is a prerequisite (Kacew et al., 1995; Mazze et al., 1973; Newton et al., 1983; Tarloff et al., 1989) and were evident in these studies. There was more severe injury and a stronger biomarker response in the male Han-Wistar versus the male Sprague-Dawley rats, particularly for NPAA. As seen in Table 2, this led to differences between strains in the incidences of some pathologies (PT degeneration/necrosis with no regeneration; DT degeneration or necrosis; medullary tubular regeneration/basophilia and inflammation, interstitial, chronic, cortex). The incidence of PT injury resulting from administration of cisplatin was greater in the Han-Wistar than the Sprague-Dawley rats and was characterized by more frequent occurrence of minimal PT injury at early time points and lower dosages in the Han-Wistar compared with the Sprague-Dawley rats. The severity of the CD changes caused by NPAA was much greater in the Han-Wistar than in the Sprague-Dawley rats. This is believed to be attributable both to an intrinsic strain difference in sensitivity to nephrotoxic injury as the Han-Wistar rats sustained greater injury in response to both toxicants, and also to the fact that the top dose of NPAA at the start of treatment in the Han-Wistar rats was too high (700 mg/kg/day vs. 400 mg/kg/day in the Sprague-Dawley rats) and had to be reduced during the course of the study. Thus, the unique diagnosis of medullary tubular regeneration/basophilia in some Han-Wistar rats was secondary to the injury seen in medullary CDs early in treatment of this strain with NPAA. Injury to the CDs in the papilla and medulla, if of sufficient magnitude, is followed by injury to cortical tubules (CDs and/or DTs). In the case of the Sprague-Dawley rats, the overall severity of the injury was much less, which facilitated the discrimination of cortical tubular injury limited to DTs versus cortical CDs. Immunohistochemical staining with nephron segment-specific markers, α-GST (PT), calbindin D28 (DT), Tamm-Horsfall protein (medullary thick ascending limb of the loop of Henle) and aquaporin 2 (CD) was
used to confirm that the cortical tubular injury in the Sprague-Dawley rats was localized to the distal convoluted tubule (data not shown). Thus, once again the unique diagnosis of DT injury in some Sprague-Dawley rats given NPAA is attributable to strain differences in the severity of injury in response to NPAA. The predominance of the diagnosis of interstitial inflammation in the Sprague-Dawley rats has several causes. To a large extent, this is explained by the fact that this was a common finding in the gentamicin study in which only Sprague-Dawley rats were used. Furthermore, there was a higher background incidence of this diagnosis in untreated (control) Sprague-Dawley rats than seen in the Han-Wistar rats and a higher incidence of this diagnosis (mostly minimal in severity) in treated Sprague-Dawley rats in the cisplatin studies.

Nevertheless, these strain differences in nephrotoxic response have no effect on the assessment of diagnostic accuracy of the biomarkers in this program. Relating the biomarker response to histopathological changes (the diagnosis of which was rigorously harmonized during a Pathology Working Group) and expressing all biomarker changes as fold change from time-matched control animals of the same strain provides a high level of assurance in the validity of the interpretation. This is further reinforced by the fact that the AUCROC for each strain alone were remarkably similar (Table 3).

The diagnostic accuracy of these exploratory biomarkers was assessed not only against the current reference standards of BUN and sCr (serum indicators of whole kidney function) but also against the widely used urinary markers of renal injury, protein and NAG. Thus, pairwise comparisons of AUCROC were performed for each exploratory marker (α-GST, μ-GST, RPA-1, and clusterin) versus each reference marker (BUN, sCr, NAG, and protein). In addition, pairwise comparisons of AUCROC were performed for NAG and protein versus BUN and sCr.

Total protein, albumin, and NAG have arguably been the urinary biomarkers most widely used over many years to monitor renal injury in both animal and human studies. The utility of both NAG and Kim-1 to predict outcome in patients with acute kidney injury (AKI) was demonstrated in a study by Liangos et al. (2007). In two recent reviews, Coca et al. (2008) concluded that NAG, together with KIM-1 and IL-18, performed best for mortality risk prediction after AKI, and Parikh et al. (2010) identified NAG among a number of urinary biomarkers with potential utility both for diagnosing AKI and for predicting its severity and outcome.

In this program of work, neither protein nor NAG was shown to have superior diagnostic accuracy to BUN and sCr for any of the tubular pathologies seen. Although proteinuria can signal disease or injury to either the glomerulus or the PT (D’Amico and Bazzi, 2003), to date, urinary protein has been qualified only as a biomarker of glomerular injury (Dieterle et al., 2010). However, the demonstrated promise of NAG as a prognostic marker in AKI in humans suggests that it should be included in any follow-up investigations of the translational value of renal biomarkers. Urinary NAG has the advantage of a widely available assay, which can be automated and applied in different species.

One specific form of proteinuria, albuminuria, is classically considered to arise from damage to the glomerular basement membrane. However, recent studies have challenged this notion in the rat, demonstrating that albumin is freely filtered through the healthy glomerular basement membrane, and efficiently recycled (and a fraction degraded) by the PT (Greive et al., 2001; Russo et al., 2007). These studies imply that albuminuria would more commonly reflect PT damage rather than glomerular damage. In the present program of the work, albuminuria was assessed only in the cisplatin study in the Sprague-Dawley rat (Gautier et al., 2010). The results indicated that it had comparable performance with α-GST and clusterin for detection of PT injury. In the recent C-Path PSTC evaluation of renal biomarkers, albuminuria has been qualified as a biomarker of tubular injury in rodent studies (Yu et al., 2010), and the limited data set generated in context of this program are supportive of that conclusion. However, the fact that albuminuria can also be a response to a variety of physiological and pathophysiological conditions such as exercise, fever, dehydration, diabetes, or hypertension (Bonventre et al., 2010), or could result from inhibition of albumin uptake by PT cells without injury (Chana and Brunskill, 2006), potentially limits its utility as a clinical biomarker of renal tubular injury.

RPA-1 showed clear superiority over all of the reference biomarkers (BUN, sCr, protein, and NAG) for diagnosis of injury to the CD in the rat. The availability of a biomarker reporting injury to the CD is a major step forward because, until now, there has been no specific urinary biomarker of injury to this region of the nephron. The ultimate goal is to characterize these biomarkers for use in human investigations in order to facilitate translation of candidate medicines from preclinical to clinical development. In the case of RPA-1, no human equivalent has yet been identified and thus clinical data are unavailable. Recent work (Sourial et al., 2010) has shown that human CD–specific monoclonal antibodies have been identified and these have potential for the future validation of CD biomarkers in clinical trials and in patients with symptoms of RPN.

The results presented here also provide support for the use of clusterin as a general biomarker of tubular injury, particularly when regeneration is present. This is consistent with current understanding of the biological function of clusterin indicating that, in response to injury, it may be involved in cell aggregation and attachment (Rosenberg and Silkensen, 1995). Clusterin was recently qualified as a biomarker of tubular injury without specification of the specific segment involved (Critical Path Institute, 2008 and 2010). Subsequently, clusterin was characterized as a biomarker of PT injury (Dieterle et al., 2010). This was based on demonstrating upregulation of clusterin in the medulla and in the medullary rays of kidneys from rats treated with vancomycin, corresponding to
compound-induced injury in the S3 segment of the PT and the thick ascending limb of the loop of Henle, respectively. Dieterle et al. (2010) acknowledged that clusterin is not specifically located in the PT and reported that increased levels of urinary clusterin were seen following CD injury induced by lithium. In niferacetam-induced RPN in dogs, Tsujiya et al. (2005) reported increased expression of clusterin and excretion of clusterin precursor in urine. It is our experience that clusterin is not specific to one segment of the renal tubule. There was evidence for its potential utility as a biomarker of CD degeneration/necrosis (AUC ROC = 0.76; statistically superior to BUN, NAG, and urinary protein) and a biomarker of PT degeneration/necrosis with regeneration (AUC ROC = 0.94, although this was statistically superior only in comparison with NAG). The fact that clusterin was not superior to all of the reference biomarkers for PT degeneration/necrosis with regeneration is not surprising. None of the exploratory biomarkers were superior to the reference markers for this diagnosis because this is where the traditional biomarkers performed best (AUC ROC estimates ranging from 0.79–0.89). This is likely due to this diagnosis occurring relatively late and with greater severity. Clusterin performed best relative to all of the reference biomarkers (BUN, sCr, protein, and NAG) in reporting the diagnosis of cortical tubular regeneration/basophilia. Overall, these results suggest that clusterin may have utility wherever the injury is located in the renal tubule and will be particularly useful in tracking regeneration. This is consistent with the finding (Eti et al., 1993) that, following chronic administration of gentamicin over a 2-month period, urinary levels of both clusterin and NAG rose rapidly but clusterin levels remained significantly elevated over the duration of the experiment, whereas NAG levels dropped to within control values within 10 days despite evidence of persistent tubulointerstitial disease. Ozer et al. (2010) found that increased urinary levels of clusterin (and Kim-1) persist during regeneration. Thus, clusterin may add most value in monitoring chronic-active injury and recovery.

The results of these studies do not permit any conclusion to be drawn about the utility of μ-GST. It is known to be primarily located in the DT and there were few diagnoses of injury to this segment of the nephron. Furthermore, where diagnoses of DT injury occurred, 95% were associated with CD injury. In patients in an intensive care unit, π-GST (which has the same localization in the human nephron as μ-GST in the rat) was shown (Westhuyzen et al., 2003) to have comparable sensitivity and specificity with α-GST and comparable positive and negative predictive value for ATN (a condition involving injury to both PTs and DTs). In theory, demonstrating a lack of change in urinary μ-GST (or π-GST) could also be useful to confirm a diagnosis of PT-specific injury based on an increase of α-GST. However, because of both functional interdependencies and proximity of different segments within regions of the kidney, collateral and nonspecific injury to adjacent nephron segments can readily occur with severe nephrotoxicity, as occurred at high doses in the cisplatin and gentamicin studies. As a result, μ-GST was generally, but not consistently, increased in the presence of predominant PT injury in which no concomitant diagnosis of DT injury was made based on light microscopic evaluation.

Alpha-GST was demonstrated to have superior diagnostic accuracy to all tested reference biomarkers except protein (i.e., BUN, sCr, and NAG) for diagnosis of injury to the PT in the rat, most evident with early damage (before regeneration was present). This is consistent with the conclusion of Ozer et al. (2010) who found α-GST to be an excellent biomarker for early tubular necrosis. The potential value of α-GST as a biomarker of PT injury, seen in preclinical studies, is strengthened by the available clinical data. Urinary GSTs are reported to be diagnostic and prognostic markers of AKI in various clinical situations (Branten et al., 2000; Sundberg et al., 1994b; Westhuyzen et al., 2003). Westhuyzen et al. (2003) studied the predictive value of a number of tubular enzymes for the subsequent development of AKI among a group of 26 subjects and found that α-GST and π-GST had the best predictive values on their own. Han and Bonventre (2004) concluded that urinary GSTs might be useful in detecting and discriminating among different types of acute renal injury. Coca et al. (2008) conducted a systematic review of studies that evaluated the diagnostic utility of biomarkers for the diagnosis of AKI and found that the GSTs were among a group of biomarkers that performed best for early diagnosis of AKI.

However, confidence in the utility of α-GST as a reliable biomarker of PT injury is weakened by the unexplained observation that the urinary levels of α-GST were consistently decreased in the presence of CD injury, caused by NPAA, whereas there was no evidence of any injury to the PT, confirmed by immunohistochemistry (data not shown). Although several hypotheses could be postulated to explain the decrease in α-GST, for example, that NPAA is metabolized in the PT leading to consumption of α-GST or to generation of an active intermediate that forms an adduct with α-GST affecting the antigenic site or rendering it liable to degradation, at present the finding remains unexplained.

The data presented and discussed here were also the subject of review by the Biomarker Qualification Review Teams (BQRTs) at EMA and FDA from 2008 to 2010. Although the data were derived from a small set of studies and no specific assessment of specificity was made (e.g., by administration of nonnephrotoxants), they were considered adequate to make qualification decisions. Although the BQRTs were in communication with each other during the review process, according to their respective internal procedures, they arrived at and published their conclusions independently. Their final decisions and recommendations are available for review on the HESI (2010) website. Their conclusions were in broad agreement and are summarized below.

Alpha-GST is not qualified at this time because the opposite behavior of this biomarker in response to PT and CD injury

---

**BIOLOGICAL QUALIFICATION OF RENAL BIOMARKERS**

\[ \text{Dieterle et al. (2010)} \]

\[ \text{Tsuiya et al. (2005)} \]

\[ \text{Eti et al., 1993} \]

\[ \text{Ozer et al., 2010} \]

\[ \text{Branten et al., 2000} \]

\[ \text{Sundberg et al., 1994b} \]

\[ \text{Westhuyzen et al., 2003} \]

\[ \text{Westhuyzen et al., 2003} \]

\[ \text{Han and Bonventre, 2004} \]

\[ \text{Coca et al., 2008} \]

\[ \text{Branten et al., 2000} \]

\[ \text{Sundberg et al., 1994b} \]

\[ \text{Westhuyzen et al., 2003} \]
of chronic injury; to support their use in other nonclinical species during onset and recovery of injury and to study their application in the physiological roles of the biomarkers; to increase confidence in their mechanistic understanding of the physiological and pathophysiological basis of the decreases. Thus, further work is required to refine the various current conclusions, to expand the context of use of the qualified biomarker and, therefore, its diagnostic utility in a wide variety of circumstances. Many more data are required to underpin the usefulness of this biomarker and the mechanistic basis of the decreases.

The data both justify the qualification of RPA-1 and support the prior qualification of clusterin, by increasing the level of evidence and clarifying the context of its use. RPA-1 is qualified for voluntary use in detecting acute drug-induced renal-tubular alterations, particularly in the CD. Clusterin may be used to detect acute drug-induced renal-tubule alterations, particularly when regeneration is present. For both biomarkers, the context of use is specified: where there is prior evidence suggesting nephrotoxicity of a development compound, these biomarkers may be used in male rats in conjunction with traditional clinical chemistry markers and histopathology in Good Laboratory Practice toxicology studies that are used to support renal safety studies in clinical trials. Although neither biomarker is currently qualified for routine monitoring of drug-induced nephrotoxicity in the clinical setting or to define dose-stopping criteria in drug development studies in the clinic, decisions can be made on a case-by-case basis how best to explore and utilize these biomarkers in a clinical development program.

Qualification of biomarkers is complex and has to be considered an incremental process that is contingent upon the development of a full understanding of the behavior of each biomarker and, therefore, its diagnostic utility in a wide variety of circumstances. Many more data are required to underpin the current conclusions, to expand the context of use of the qualified biomarkers and, where necessary, to provide additional biomarkers. Thus, further work is required to refine the various methodologies used in the qualification studies; to provide mechanistic understanding of the physiological and pathophysiological roles of the biomarkers; to increase confidence in their specificity for renal injury; to verify that their use is equally valid in female animals; to evaluate biomarkers for other parts of the nephron; to explore more fully the behavior of the biomarkers during onset and recovery of injury and to study their application in chronic injury; to support their use in other nonclinical species and, most importantly of all, their translation to use in human subjects.

SUPPLEMENTARY DATA

Supplementary data are available online at http://toxsci.oxfordjournals.org/.

FUNDING

In-kind contributions of the contributing authors and their organizations.

REFERENCES


