analytical chemistry

Article

pubs.acs.org/ac

Nontargeted Identification of Reactive Metabolite Protein Adducts

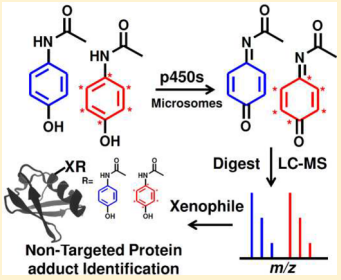
Michael G. Leeming,*[,†] William A. Donald,*[,‡] and Richard A. J. O'Hair*[,†]

[†]School of Chemistry and Bio21 Institute of Molecular Science and Biotechnology, The University of Melbourne, Melbourne, Victoria 3010, Australia

[‡]School of Chemistry, University of New South Wales, Sydney, New South Wales 2052, Australia

Supporting Information

ABSTRACT: Metabolic bioactivation of many different chemicals results in the formation of highly reactive compounds (chemically reactive metabolites, CRMs) that can lead to toxicity via binding to macromolecular targets (e.g., proteins or DNA). There is a need to develop robust, rapid, and nontargeted analytical techniques to determine the identity of the protein targets of CRMs and their sites of modification. Here, we introduce a nontargeted methodology capable of determining both the identity of a CRM formed from an administered compound as well as the protein targets modified by the reactive metabolite in a single experiment without prior information. Acetaminophen (*N*-acetyl-*p*-aminophenol, APAP) and $^{13}C_6$-APAP were incubated with rat liver microsomes, which are known to bioactivate APAP to the reactive metabolite *N*-acetyl-*p*-benzoquinone imine (NAPQI). Global tryptic digestion followed by liquid chromatographic/mass spectrometric (LC/MS) analysis was used to locate "twin" ion peaks of peptides adducted by NAPQI and for shotgun proteomics via tandem mass spectrometry (MS/MS). By the development of blended data analytics software called Xenophile, the identity of the amino acid residue that was adducted can be established, which eliminates the need for specific parametrization of protein database search algorithms. This combination of experimental design and data analysis software allows the identity of a CRM, the protein target, and the amino acid residues that are modified to be rapidly established directly from experimental data. Xenophile is freely available from https://github.com/mgleeming/Xenophile.



Most small organic molecules will be modified in some way through metabolic reactions after they enter the body. While many metabolites are biochemically inert and readily excreted, it is well-established that so-called "bioactivation" metabolic transformations can result in the formation of highly reactive compounds that can lead to toxicity. This concept was first described by Miller and Miller in the 1940s;[1,2] they investigated the binding of the azo dye *p*-dimethylaminoazobenzene to cellular macromolecules in rats and postulated that metabolites of this substance, rather than the substance itself, may be the primary toxicant. Subsequently, many xenobiotics have been found to form reactive metabolites in vivo, including drugs[3,4] as well as insecticides and industrial chemicals.[5,6]

A current theory is that binding of a chemically reactive metabolite (CRM) to cellular macromolecules is a primary event linking bioactivation to the onset of toxicity.[7,8] Indeed, an analysis of the covalent binding characteristics of 42 drugs revealed that the extent of protein binding is broadly correlated with the risk of idiosyncratic adverse drug reactions (IADRs).[9] While the precise mechanistic details remain to be established, clinical observations suggest that most IADRs are immune-related.[8] One proposed mechanism, the hapten hypothesis,[10,11] asserts that modification of a native protein with a CRM, the hapten, triggers the production of antibodies that bind the CRM−protein complex, leading to a potentially severe immune response against the modified protein. Other hypotheses have been put forward,[12] such as the danger hypothesis[13−15] and the pharmacological interaction with immune receptors (p-i) hypothesis.[16,17]

Due to the potential for toxicity and the difficulty in predicting the outcomes of CRM formation, general practice in drug development processes has been to minimize the formation of CRMs through chemical modifications to the drug substrate.[18] To this end, a vast literature has been assembled on the potential bioactivation processes operative for a wide range of organic functional groups.[3,19−21] Assays to determine the global level of protein binding by reactive metabolites are now routinely conducted in drug discovery projects and typically involve incubation of radiolabeled substrate with microsomal fractions or hepatocytes, followed by liquid scintillation counting of washed protein pellets.[18]

However, there are cases of CRM−protein binding that highlight the complexities inherent in this field. For example, when Roberts et al.[22] administered either acetaminophen (*N*-acetyl-*p*-aminophenol, APAP) or the structural isomer *N*-acetyl-*m*-aminophenol (AMAP) to hamsters, similar levels of CRM−protein binding were observed, but APAP treatment induced hepatotoxicity while AMAP did not. Due to the close structural similarity of APAP and AMAP, findings such as this have led to the critical-protein hypothesis,[7] whereby the exact identity of the proteins adducted is a key determinant of the biological

Received: November 21, 2016
Accepted: May 8, 2017
Published: May 8, 2017

DOI: 10.1021/acs.analchem.6b04604
Anal. Chem. 2017, 89, 5748−5756

response and varies with even small changes in the structure of the CRM.[23] It is interesting to note that differences in the protein adduction profiles of APAP and AMAP have been identified by both radiographic and immunochemical analyses of protein fractions derived from various liver preparations.[24,25] Efforts to catalogue the protein targets of reactive metabolites are under way. For example, Hanzlik et al.[26,27] maintain a database of proteins known to be modified by CRMs formed from a range of compounds.

There is a pressing need to develop robust, rapid, and nontargeted analytical techniques to determine the identity of the protein targets of reactive metabolites to provide an approach for elucidating the biochemical origins of IADRs.[28] A range of analytical methods have been employed to identify these adducts.[29,30] For example, antibodies can be raised against protein-bound drug motifs that can be used to detect the modified macromolecules.[31] Alternatively, two-dimensional (2D) gel electrophoresis of protein preparations adducted by radiolabeled drug, followed by phosphorimaging, is used to detect gel spots corresponding to adducted proteins.[24,32,33] Other methods involving affinity capture of protein digestion products and targeted multiple reaction monitoring (MRM)-like experiments have identified modified peptides.[34] However, all of these methods are resource-intensive, laborious, or require some degree of prior knowledge about the metabolites likely to be formed. These factors present significant barriers to the routine identification of protein targets of reactive metabolites and likely contribute to the relative rarity of such studies.

One nontargeted method to detect specific chemicals in liquid chromatographic/mass spectrometric (LC/MS) data is to simultaneously administer equal parts of natural abundance and isotopically enriched xenobiotic, which produces a pair of coeluting peaks that have the same abundance and are separated by an m/z spacing corresponding to the mass of the isotopic label.[35−38] Recently, we have demonstrated that this so-called twin-ion method, in conjunction with automated data analysis software, enables the nontargeted detection of small-molecule metabolites of APAP in rat blood plasma.[39] Here, we extend the twin-ion method to identify the protein targets of CRMs. That is, a nontargeted methodology capable of determining the identity of a CRM formed from an administered compound and the protein targets modified by the reactive metabolite in a single experiment without prior information is introduced. To validate, APAP and $^{13}C_6$-APAP were bioactivated to the reactive metabolite *N*-acetyl-*p*-benzoquinone imine (NAPQI) via incubation with rat liver microsomes.[40,41] Global tryptic digestion followed by LC/MS analysis is used to locate twin-ion peaks associated with peptides adducted by NAPQI, while the identity of the proteins is determined by tandem mass spectrometry (MS/MS) shotgun proteomics. By use of blended data analytics, the identity of the amino acids adducted can be established, which removes the need for specific parametrization of protein database search tools. This novel approach does not require reactive metabolite trapping studies, and the identity of the amino acid residue adducted by a reactive metabolite need not be known in advance.[42] As such, unexpected or previously unidentified metabolites can still be accounted for in the experimental data.

EXPERIMENTAL METHODS

Materials. Acetaminophen (APAP), sequencing-grade modified trypsin, iodoacetamide, triethylammonium bicarbonate (TEAB) buffer, formic acid, urea, and reduced nicotinamide adenine dinucleotide 2′-phosphate tetrasodium salt (NADPH) were from Sigma−Aldrich (St. Louis, MO). Pooled rat liver microsomes (20 mg·mL^{-1} protein, 0.462 nmol of P450·(mg of protein)$^{-1}$, from 83 uninduced, male Sprague-Dawley donor rats, lot number RT053C) were from Life Technologies (Carlsbad, CA), and stored at −80 °C until use. *ring*-$^{13}C_6$-APAP [>99% ^{13}C enrichment] was from IsoSciences (King of Prussia, PA), and tris-(2-carboxyethyl)phosphine hydrochloride (TCEP) was from Thermo Fisher Scientific (Waltham, MA). Oasis HLB solid-phase extraction cartridges (10 mg) were from Waters (Milford, MA). Acetone and acetonitrile were from Merck (Kenilworth, NJ), and 18 MΩ H_2O was obtained from a Milli-Q apparatus

Microsome Incubations. Either aqueous APAP/$^{13}C_6$-APAP solution (1:1; 10 μL, 20 mM) or water (10 μL) was combined with microsomes [10 μL, 20 mg·mL^{-1} protein] in potassium phosphate buffer (170 μL, 100 mM, pH 7.4), and the mixtures were preincubated at 37 °C for 10 min with continuous agitation. Reactions were initiated by addition of NADPH (10 μL, 20 mM) and incubated for 3 h at 37 °C. At both t = 60 and 120 min, an additional aliquot of NADPH (10 μL, 20 mM) was added. Reactions were quenched at t = 180 min via the addition of ice-cold acetone (1 mL). Samples were vortex-mixed for 30 s and then stored at 4 °C overnight. The samples were centrifuged (16000*g*, 10 min, 4 °C), the supernatant was removed, and the remaining protein pellet was gently washed by overlaying acetone (1 mL). After removal of the supernatant, the pellet was resuspended in TEAB buffer (200 μL, 50 mM, pH 8.5) containing urea (8 M) and TCEP (10 mM) and sonicated until dissolution. The samples were incubated at 37 °C for 45 min and allowed to cool to ambient temperature before iodoacetamide (122 μL, 100 mM) was added. Samples were incubated in darkness for 60 min and then diluted to a final urea concentration of 1 M with TEAB (25 mM). Trypsin (4 μg) was added to each sample, and digestions proceeded at 37 °C for 24 h. The samples were acidified with formic acid to a final concentration of 1% (v/v) and purified by solid-phase extraction as follows: (i) the cartridge was preconditioned with 1 mL of 80% acetonitrile containing 0.1% trifluoroacetic acid (TFA), followed by 2 × 1.2 mL of 0.1% aqueous TFA; (ii) sample was loaded onto the cartridge; (iii) the cartridge was washed with 2 × 1.2 mL of 0.1% TFA; (iv) peptides were eluted from the cartridge with 1 mL of 80% acetonitrile containing 0.1% TFA; and (v) eluent was concentrated on a centrifugal evaporator to approximately half the original volume, then lyophilized overnight and resuspended in 0.1% formic acid (100 μL) and stored at −20 °C until the time of analysis via nano-LC/MS/MS experiments. Experiments were performed in triplicate for both APAP and control treatments.

Liquid Chromatographic/Mass Spectrometric and Tandem Mass Spectrometric Experiments. Nano-LC/MS/MS analyses of peptides generated from microsomal proteins were performed on a hybrid linear ion trap mass spectrometer and Orbitrap mass spectrometer (Orbitrap Fusion Lumos; Thermo Fisher Scientific) equipped with a nano-HPLC (UltiMate 3000 RSLC; Dionex). The nano-LC system was equipped with a C_{18} nanotrap column (Acclaim Pepmap, Dionex, 100 Å, 75 μm × 2 cm) and a C_{18} analytical column (Acclaim Pepmap RSLC, Dionex, 100 Å, 75 μm × 50 cm). Peptides were separated by gradient elution with mobile phases of 0.1% (v/v) aqueous formic acid (solvent A) and 0.1% (v/v) formic acid in CH_3CN (solvent B). The gradient

DOI: 10.1021/acs.analchem.6b04604
Anal. Chem. 2017, 89, 5748−5756

timetable was as follows ([time in minutes, %B]): [0, 3], [6, 3], [55, 25], [65, 40], [70, 80], [75, 80], [76, 3], and [86, 3]. Data were collected for each sample in both MS^1 and MS/MS mode. For MS/MS runs, higher energy collision-induced dissociation (HCD) spectra were acquired in a data-dependent fashion. Detailed instrument parameters are given in Supporting Information.

Data Analysis. Raw MS^1 LC/MS data files were converted to the .mzML format with an intensity threshold of 5000 counts by use of MSConvert[43] and scored with HiTIME,[39] using $\Delta m/z$ of 3.01005 and 2.0067 (i.e., the 6.0201 Da difference between APAP and $^{13}C_6$-APAP in the 2+ and 3+ charge states, respectively). For database searching of MS^2 LC/MS runs, Mascot generic format files were produced by use of MSConvert that contained the 100 most intense fragment ions in each HCD spectrum. Peptide and protein identifications were performed with Mascot[44] (v2.4.1), and searches of the Uniprot database[45] (release 2015_07) accounted for the possibility of one missed tryptic cleavage, (carbamidomethyl)-cysteine, and oxomethionine variable modifications. The MS^1 and MS^2 m/z tolerances were set to 20 ppm and 0.2 Da, respectively. Further variable modifications were used in some searches as described in the text. Significance thresholds were set to achieve a global false discovery rate of 1% using a target-decoy approach against sequence-reversed proteins.[46] Proteins with two or more significant matches were retained.

Software Development. Xenophile, custom software to facilitate the analysis of CRM−peptide adducts, can be utilized through a graphical user interface and contains methods that allow the user to perform HiTIME searches, control sample baseline subtraction, postprocessing and review of HiTIME search results, nontargeted reactive metabolite detection, and targeted Mascot-HiTIME correlation (see Supporting Information for details).

RESULTS AND DISCUSSION

Microsomal Bioactivation of Acetaminophen. APAP and $^{13}C_6$-APAP were incubated with liver microsomal preparation as a source of xenobiotic-metabolizing enzymes, including P450, and are known to bioactivate APAP to the reactive metabolite NAPQI.[40,41] Following a 3 h incubation, microsomal proteins were separated by precipitation, digested with trypsin, and analyzed by LC/MS/MS. Mascot analysis of the data against the Uniprot database, allowing variable oxidation of methionine and carbamidomethylation of cysteine, resulted in the identification of 720 and 660 proteins that were common to all three replicates in APAP and control treatment groups, respectively (Figure 1; see Table S1 in Supporting Information for lists of identified proteins). However, adduction of any metabolites formed from APAP would both change the total mass of the peptide and induce an offset in the ladder sequence ions formed in MS^2 experiments, resulting in a mismatch between experimental spectra and theoretically predicted fragments generated by database searching. Adducted peptides will remain unassigned and, accordingly, these data do not account for the possibility of reactive metabolites covalently reacting with microsomal proteins.

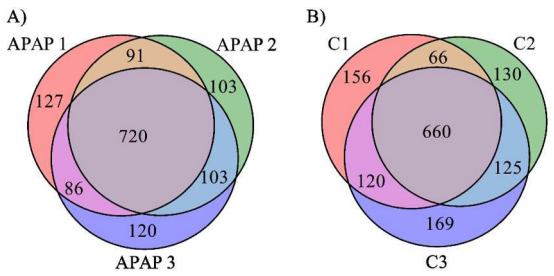


Figure 1. Venn diagrams indicating the number of proteins identified by database searching of LC/MS/MS data sets for (A) acetaminophen and (B) control microsomal protein digests.

Several methods could be used to recover the missed CRM-modified peptides. For example, Mascot searches could be directed to consider variable modifications at specific residues for predicted metabolites. However, this requires knowledge of both the identity of the reactive metabolite and the amino acid residue(s) likely to be modified, which runs counter to the goal of nontargeted identification. Blind post-translational modification (PTM) search algorithms[47−49] that aim to identify modifications by various approaches based on sequence tags may be adequate in some instances; however, protein−CRM adducts are expected to be rare events relative to the thousands of proteins that can be detected in shotgun experiments, the abundances of modified peptides are likely to be low (often leading to noisy MS/MS spectra), and CRM modifications for a compound of interest are unlikely to be recorded in PTM databases such as UniMod.[50] Any search algorithm would therefore be required to handle arbitrary mass offsets and account for modification on any residues, which would result in increased search times, false negatives, and false positives.[51] However, reactive metabolites can be considered a distinct subset of all PTMs present in a given sample, and if these could be selectively detected, then global PTM searching would be unnecessary.

Due to the use of both native and $^{13}C_6$-labeled APAP, this isotopic pattern will be carried through metabolic transformations, providing a signature unique to the products of APAP metabolism. By searching for this signature, peptides modified by APAP CRMs could be detected even though the peptide has not been assigned through database searching and the identity, and site of adduction, of the CRM may be unknown. The MS^1 data were thus processed with HiTIME software capable of identifying twin-ion signals.[39] The heat maps produced after scoring with an m/z change of 3.01005 and 2.0067 (i.e., the 6.0201 Da difference in the 2+ and 3+ charge states) for all replicates are shown in Figures S1 and S2, respectively. HiTIME scoring reveals multiple bright spots in the APAP incubations (Figure 2A) for which comparable points are not observed in the controls (Figure 2B), indicating that the data in those regions closely match the twin-ion signature. Investigation of highly scoring points reveals that these are true twin-ion hits. For example, the HiTIME hit at m/z 752 and 47.8 min suggests twin ions of m/z 752 and 755. The extracted ion chromatogram (EIC) traces of these ions overlap significantly in the region of 47.8 min (Figure 2C), and two peptide signals, both of the 2+ charge state, are evident in the mass spectrum (Figure 2D). These data indicate that this is a true twin ion corresponding to a peptide modified by a reactive metabolite derived from APAP in the 2+ charge state. For control samples, similarly high-scoring data regions as those corresponding to true twin-ion peptides in APAP treatment data were not observed (Figures S1 and S2), and no convincing case of a twin-ion peptide could be found upon inspection of EIC traces and mass spectra following postprocessing (Table S2).

DOI: 10.1021/acs.analchem.6b04604
Anal. Chem. 2017, 89, 5748−5756

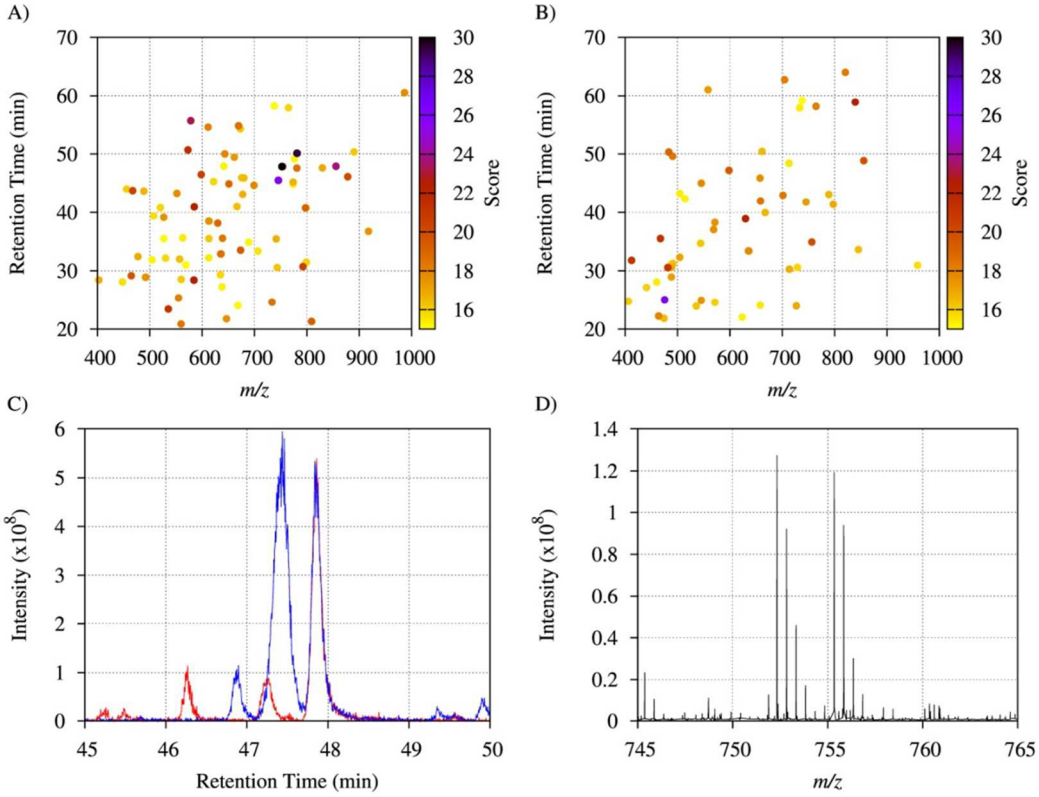


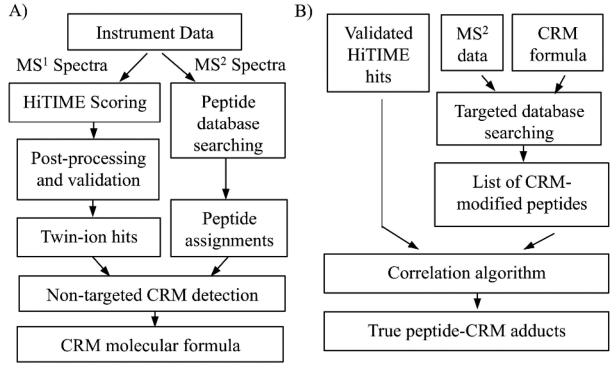
Figure 2. Heat maps produced by HiTIME scoring of microsomal protein digest LC/MS data sets with twin-ion spacing of 3.01005, aimed at mining doubly charged peptides modified by an APAP metabolite. (A) APAP treatment. (B) Control samples. (C) EIC of the light (m/z 752.33, red trace) and heavy (m/z 755.34, blue trace) peaks indicated by the HiTIME hit at retention time (RT) = 47.85 min. (D) Mass spectrum at the region of EIC peak maximum. Heat maps for remaining replicates of APAP and control treatment data are shown in Figure S1.

The observation of numerous twin-ion hits in the drug treatment data indicates the presence of multiple peptides that have been modified by CRMs derived from APAP. Without assuming the identity of the CRM or the reactive site, identification of the modified peptides must be conducted by means other than standard database searching. Since CRM modifications are generally not quantitative and are formed in relatively low abundance, it is reasonable to assume that some quantity of the protein target's unmodified native counterpart will remain. As the native and CRM-modified peptides share the same base structure and differ only in the configuration of various PTMs, MS^2 spectra of these peptides are likely to display sequence ions that are either common to both peptides or offset from one another by a constant amount that is related to the mass of the adducted metabolite. With this in mind, nontargeted peptide−CRM identification software was developed.

Nontargeted Peptide−CRM Identification Algorithm. The structural similarity of CRM-modified peptides and their native counterparts facilitates identification of the site of adduction and the mass of the CRM. However, significant amounts of complex data generated by LC/MS, HiTIME scoring, and Mascot analysis make manual interpretation impractical. Thus, semiautomated computer software has been developed that aids verification of twin ions and comparison of MS^2 spectra for nontargeted analysis.

In the nontargeted CRM identification algorithm (Scheme 1A), a list of Mascot-assigned peptides is produced that fall within user-defined m/z ranges relative to a verified HiTIME hit (Figure 3). The sequences and PTMs of all Mascot-assigned

Scheme 1. Workflow Diagrams for (A) Nontargeted CRM Identification and (B) Twin-ion-directed Targeted Peptide Adduct Identification



peptides in this range are used to calculate the m/z values of a set of sequence ions appropriate for the ion activation method used. The match between the experimental MS^2 spectrum of the HiTIME hit and each theoretical ion sets is then ranked according to Equation S1. This scoring is complicated by the fact that fragment ions in the HiTIME hit spectrum containing the CRM modification will be offset from those of the native spectrum by a mass related to the CRM. To address this problem, a rolling modification model was used that repeatedly scores each pair of MS^2 spectra using m/z offsets for ions in the native peptide corresponding to modification at different sites. A series of theoretical spectra are generated where the m/z offset is iteratively moved along the peptide chain to simulate

DOI: 10.1021/acs.analchem.6b04604
Anal. Chem. 2017, 89, 5748−5756

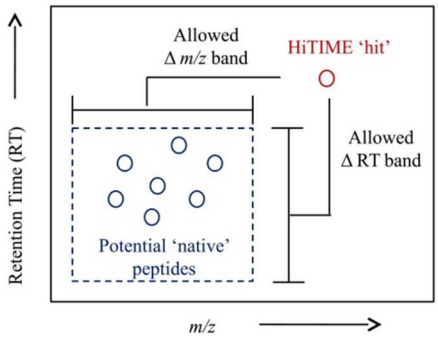


Figure 3. Relationship between HiTIME hits and native peptides.

CRM adduction at different positions (Table S3). The m/z offset is taken to be equal to the mass difference between native and CRM-adducted peptides. Each theoretical spectrum is then independently scored against the experimental HiTIME MS^2 spectrum, with the highest scoring match being the lead candidate.

The Mascot-assigned PTM profile of a peptide is used to account for the possibility that CRM adduction precludes other modifications. For example, carbamidomethylation of cysteine residues upon treatment of proteins with iodoacetamide during sample preparation would be prevented if the Cys residue were already modified with a CRM. This results in the mass of the native peptide being increased by 57 Da, and therefore this has to be accounted for when the mass of the CRM is calculated. The CRM mass is thus equal to the mass difference between the modified and native peptides plus the mass of any PTMs precluded by modification with the reactive metabolite.

The CRM mass determined from the most highly correlated MS^2 spectra is used to determine possible chemical formulas. Although the number of formulas that are determined for a given $\Delta m/z$ may be large, many of these will fall beyond the limits of chemically reasonable compositions. To restrict the search space, only formulas that are within user-specified elemental compositions are calculated. Further refinement can be achieved by noting that the chemical structure and composition of the metabolites should have elements in common with the starting material. The user is given the option of specifying the input structure of the test compound by use of a built-in molecule editor. The molecule is then split into fragments at rotatable bonds in an attempt to mimic metabolic reactions such as heteroatom dealkylation (Equation S2). An example is shown in Figure S4. The list of candidate CRM formulas is then compared to the molecular formulas of input molecule fragments and a residual mass error (RME) is calculated according to Equation S3, which quantifies total mass of atoms in the candidate formula that are not accounted for in the closest-matching input molecule fragment. This metric provides lower scores for formulas that more closely resemble the composition of the input molecule (or fragment thereof) and are thus more likely to be the true stoichiometry of the CRM.

Nontargeted Searching of APAP-Microsome Data. To detect the APAP-derived reactive metabolites, the APAP treatment HiTIME and Mascot data were analyzed with the nontargeted CRM detection algorithm, using the parameters provided in Tables S4 and S5. In this process, MS^2 spectra associated with twin-ion peptides (unassigned by peptide database searching) are correlated with their non-CRM-modified counterparts, and the mass difference between these highest ranked pairs is taken as a potential CRM (Scheme 1A). The three highest-scoring hits for each of the three APAP treatment replicates are provided in Table 1 (replicate 1) and Tables S6 and S7 (replicates 2 and 3). Taken together, eight of these nine results detect a CRM with a mass of 149.05 Da that, in each case, arises from covalent modification of a cysteine residue. Two possible molecular formulas are identified that fit within the allowed atom stoichiometries and parts per million (ppm) error tolerances. These are $C_8H_7NO_2$, identified as a possibility in all eight cases, and $C_6H_5N_4O$, which is identified in five out of the eight cases. The average mass errors for these possible formulas are similar (8.0 and 6.1 ppm, respectively); however, the residual mass errors calculated versus the input APAP structure differ considerably (2 and 86 Da, respectively). This indicates that the former candidate, $C_8H_7NO_2$, is very similar in composition to the APAP input molecule (APAP: $C_8H_9NO_2$). Given that (i) $C_8H_7N_2O$ is consistently identified in multiple HiTIME hits and across all replicates, (ii) this formula is closely related to the APAP input molecule, and (iii) no chemically reasonable structure can be devised for the alternative, $C_8H_7N_2O$ is taken as the stoichiometry of the reactive metabolite formed from APAP in these incubations.

To interrogate the data underpinning this result, MS/MS spectra for the CRM-modified and native peptides were extracted and compared. For example, the native peptide EFTPCAQAAFQK (m/z 699.3) and its twin-ion counterpart (m/z 745.3) that were used to assign a CRM mass of 149.0481 Da are shown in Figure 4. A prominent y-ion series is observed, and the m/z values of these sequence ions are common to both spectra for y_2 through to y_7. However, these diverge beyond the y_7 ion at m/z 763 (assigned as $[AQAAFQK + H]^+$). For the native peptide spectrum (Figure 4A), the y_8 and y_9 ions, which now include a cysteine residue, are located at m/z 923 and 1020, respectively. While no prominent fragments are observed at these positions in the MS/MS spectrum of the twin-ion peptide, two additional fragments appear offset by +92 Da at m/z 1015 and 1112, indicating that the cysteine residue is differentially modified in the two peptides. Because the unmodified peptide is alkylated at the Cys residue (+57 Da

Table 1. Three Highest-Ranked Reactive Metabolite Assignments Produced for One of the APAP Replicates[a] by the Nontargeted Reactive Metabolite Identification Algorithm

hit	CRM mass (Da)	sequence	modification site	formula	mass error (ppm)	residual mass (Da)
1	149.0494	VFANPEDCAGFGK	C (8)	$C_8H_7NO_2$	12	2
2	149.0481	EFTPCAQAAFQK	C (5)	$C_8H_7NO_2$	3	2
				$C_6H_5N_4O$	12	86
3	149.0464	TIQLNVCNSEEVEK	C (7)	$C_6H_5N_4O$	0.4	86
				$C_8H_7NO_2$	−9	2

[a]APAP replicate 1 of 3.

DOI: 10.1021/acs.analchem.6b04604
Anal. Chem. 2017, 89, 5748−5756

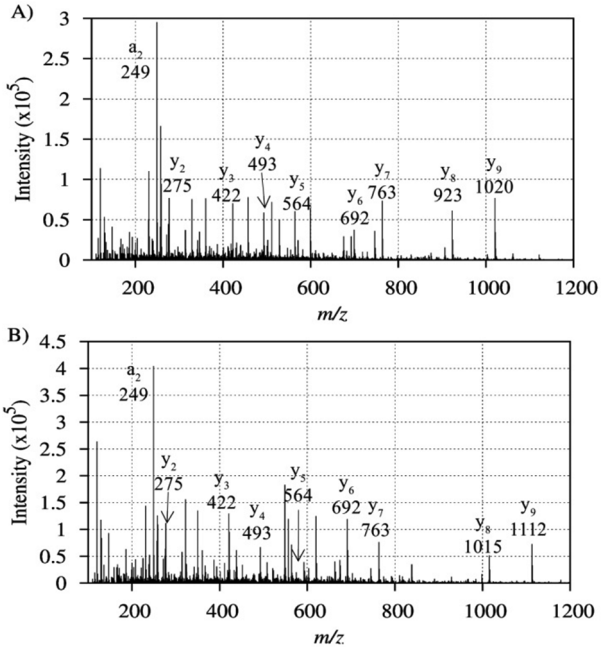


Figure 4. HCD MS^2 spectra of (A) non-twin-ion peptide assigned as (carbamidomethyl-C)-EFTPCAQAAFQK^{2+} at m/z 699.3 and 37.3 min and (B) twin-ion peptide at m/z 745.3 and 43.9 min from LC/MS/MS analysis of APAP-treated microsomal protein digests.

relative to unalkylated peptide) and the twin-ion peptide has a CRM adducted to the Cys residue (+92 Da relative to the alkylated cysteine), the mass of the CRM (149 Da) is calculated from the sum of 57 and 92 Da. The assigned formula, $C_8H_7N_2O$, is the molecular formula of the well-studied electrophilic NAPQI metabolite of APAP, which is known to bind covalently to the side chains of reduced cysteine residues in proteins.[52] It should be emphasized that this CRM has been identified directly from shotgun proteomics data while making essentially no assumptions as to the identity of the metabolite, the protein, or the amino acid target of modification.

Targeted Search of APAP-Microsome MS^2 Data. Having determined the stoichiometry of the reactive metabolite formed from APAP and the amino acid that was modified by this metabolite, these data can be used to direct subsequent protein database searches. Mascot searches of the LC/MS/MS data were replicated with the addition of $C_8H_7NO_2$ and $^{13}C_6$-$C_2H_7NO_2$ (i.e., NAPQI and $^{13}C_6$-NAPQI) as variable modifications of cysteine in order to directly identify peptides carrying the NAPQI modification.

In all cases, a large number of peptides were identified as APAP adducts by Mascot searching. For the APAP treatment samples, an average of 55 peptides were assigned as APAP or $^{13}C_6$-APAP adducts and, interestingly, an average of 41 were assigned for control samples that were not exposed to APAP. It therefore seems highly unlikely that all of these can correspond to legitimate APAP−peptide adducts. For control samples, no APAP assigned peptides displayed the twin-ion signature upon manual inspection of their associated MS^1 spectra, and this was also true of many hits in the drug treatment samples. These data indicate that targeted database searching alone may not always reliably identify the products of CRM−peptide adduction and may in fact give rise to false positives.

However, in these experiments, true modified peptides should be assigned as such by database search algorithms and also display the expected twin-ion shape. To restrict the data to only those peptides that meet these criteria, a targeted data refinement algorithm was developed that correlates CRM modifications assigned by database searching with HiTIME data (Scheme 1B). When this algorithm was run with HiTIME and Mascot data from microsomal protein digests (Table S8), eight peptides were found to be assigned as CRM adducts by Mascot that also appear as twin ions (Table 2) indicating that these are true CRM-modified peptides. For five of these eight cases, both the light and heavy peptides were separately mass-selected during data-dependent MS^2 acquisition and subjected to HCD. For example, the MS^2 spectra of the APAP and $^{13}C_6$-APAP adducts of the peptide EFTPCAQAAFQK^{2+} are shown in Figure S5 panels A and B, respectively. The series of y ions at m/z 1015, 1112, and 1213 in Figure S5A is offset from the analogous series of ions in Figure S5B at m/z 1021, 1118, and 1219 by 6 Da, which is equal to the nominal mass difference between APAP and $^{13}C_6$-APAP. Taken together, these data provide a high level of confidence that the peptides assigned as APAP adducts by database searching and as twin ions in MS^1 data are indeed true products of reactive metabolite adduction of proteins.

These eight modified peptides comprise partial sequences of seven unique proteins (Table 2). Several of these have been identified as the targets of reactive electrophiles by other researchers, providing additional confidence in these assignments. For example, glutathione S-transferase (GST) was first identified as a target for APAP adduction by radiometric counting in 1981,[53] and its constituent peptides VFANPEDCAGFGK and VFANPEDCAGFGKGENAK (and homologous sequences from different species) have since been identified numerous times to be modified at the cysteine residue by NAPQI[34,41,54] and electrophilic metabolites of other drug compounds.[55] Furthermore, the triazine herbicide atrazine has been found to form an adduct with the HBB1 peptide EFTPCAQAAFQK in rats,[56] and the FMO1 peptide

Table 2. Summary of Peptides Assigned by Directed Mascot Search as Carrying NAPQI Modification[a]

peptide	modification	accession no.	protein	unique peptides
SCDLGGLWR	APAP L/H C2	FMO1_RAT	hepatic flavin-containing monooxygenase 1	16
EFTPCAQAAFQK	APAP L/H C5	HBB1_RAT	hemoglobin β-1 chain	13
VFANPEDCAGFGK	APAP L C8	MGST1_RAT	microsomal glutathione S-transferase-I	8
VFANPEDCAGFGKGENAK	APAP L/H C8	MGST1_RAT	microsomal glutathione S-transferase-I	8
HIGDGCHLTR	APAP L/H C6	MET7B_RAT	methyltransferase-like protein 7B	8
TIQLNVCNSEEVEK	APAP L/H C7	BDH_RAT	3-hydroxybutyrate dehydrogenase	12
KHHLCGETEEER	APAP H C5	GSTM1_RAT	glutathione S-transferase Mu 1	18
QVADEGDALVAGGVSQTPSYLSCK	APAP L C23	BHMT1_RAT	betaine-homocysteine S-methyltransferase 1	11

[a]APAPL and APAPH refer to the natural abundance APAP and $^{13}C_6$-APAP, respectively.

DOI: 10.1021/acs.analchem.6b04604
Anal. Chem. 2017, 89, 5748−5756

SCDLGGLWR was identified as a target of reactive metabolites formed from model furan-containing compounds in rat liver microsome incubations.[55]

The number of proteins identified is similar to and slightly greater than the number of proteins identified in closely related studies that used more targeted methodologies to identify APAP−protein adducts from rat liver microsome incubations.[34] In addition, many compounds such as diclofenac and halothane are known to form specific adducts with selected targets despite the presence of large numbers of other proteins.[27] These results are consistent with our results, which indicate that NAPQI forms relatively specific adducts with a limited number of microsomal proteins under these conditions.

The global biological and toxicological effects of reactive metabolite adduction to these specific proteins are largely unclear. Previous studies have demonstrated dose-dependent inhibition of the glutathione conjugation activity of GST by APAP which has been linked to NAPQI formation and adduction to Cys47[57] and may have an impact on GSH/GSSG balance. Other proteins identified have roles in drug metabolism (FMO1), methionine synthesis (BHMT), and fatty acid catabolism and energy balance (BDH).

False Assignments and Limitations. To investigate the ability of Xenophile to identify twin-ion signals and correctly assign CRM masses and elemental compositions in a controlled manner, semisynthetic data sets were produced by superimposing artificially generated twin-ion signals onto experimental shotgun LC/MS data from the three control treatments for each cysteine-containing peptide, resulting in the incorporation of 560, 610, and 620 signals, respectively (see Supporting Information for details).

These data were scored with HiTIME (Figure S6), and comparison of the locations of synthetic twin ions with local maxima in the HiTIME data revealed that ∼95% of the synthetic signals were identified (Table S9). Many of the remaining 5% were found to heavily overlap with experimental signals, resulting in large distortions to the target peak shape. False positive rates were found to be less than 10% for approximately the top 90% of targets (Figure S7). Running the nontargeted CRM detection algorithm resulted in ∼88% of the synthetic targets being assigned as NAPQI modifications to within 20 ppm, which increases to 98% upon relaxing this threshold to 100 ppm (Figure S8 and Table S9). The confidence in these assignments was investigated by analyzing the difference in correlation score between the first and second ranked peptide, which provides a metric of the margin by which the top-ranked hit was chosen over the next candidate (Figure S9). Confidence in CRM assignments was ca. 90% or better for the highest-scoring hits and decreased to ca. 60% for lower-scoring hits, indicating that the correct peptide correlations are well separated from their closest counterparts. For large CRMs, many chemical formulas may have the same nominal mass. To profile the number of formulas as a function of molecular weight, various modifications were computationally applied to 1337 marketed drugs, and formulas were generated for the products in a manner analogous to that of the Xenophile software. The number of hits within 100 ppm of the target mass rises rapidly with molecular weight, reaching ∼100 candidates for a mass of 500 Da. However, this decreases to ∼20 hits when tolerances are tightened to 20 ppm (Figure S10). This number can be further refined by use of residual mass values, which quantify the proportion of a CRM formula that is not explained by the administered xenobiotic. These should lie within certain ranges corresponding to common metabolic reactions. By conservative application of this rationale, the number of hits is reduced to ca. 5 for a 500 Da compound, which is feasible for manual user review.

It should be noted that reproducing the complexity and nuances of experimental data with synthetic spectra is difficult, and therefore these results likely represent a best-case scenario. However, fundamentally, the detection limit for CRM-modified peptides using the software approach presented here should be similar to that for any standard peptides using typical workflows. That is, detection is limited by the ability of the mass spectrometer to ionize, fragment, and detect a peptide. As an approximation of the sensitivity of this experiment, ca. 4 μg of protein was injected onto the column, resulting in the detection of 1021 proteins in each sample (on average) with a mean molecular mass of 52 794 Da. This equates to an average protein loading of ca. 75 fmol. It should be noted that CRM-modified peptides identified here were likely of substantially lower abundances, as CRM adduction generally results in low yields. For example, the APAP-modified peptide VFANPED-CAGFGK (Figure 2C,D) is detected at ca. 20-fold lower abundance than its native counterpart.

One limitation of the software presented here is that reactive metabolites with a heavy/light offset of ca. < 4 Da will produce twin-ion signals with overlapping isotope distributions, which may hinder twin-ion assignments. Computational strategies are currently being planned that may account for these cases and will be included in future Xenophile updates.

CONCLUSION

A new nontargeted method to identify protein adducts of chemically reactive metabolites from LC/MS data without prior knowledge of the CRM identity has been developed. This method involves two distinct steps. The first step identifies the molecular formula of a CRM as well as the amino acid residue site of adduction on the peptides by searching for pairs of native and CRM-modified peptides that have similar MS^2 fragmentation behavior. The second step uses this information to direct further peptide database searches that are specifically parametrized to include the possibility of residue modifications identified in the first step. This methodology was demonstrated for the bioactivation of acetaminophen to the reactive metabolite NAPQI in hepatic microsomes. Nontargeted searching identified a reactive metabolite adducted to cysteine residues with a formula of $C_8H_7NO_2$, which is consistent with the molecular formula and known residue reactivity of NAPQI. Upon subsequent Mascot searches, eight unique NAPQI-modified peptides were identified that arise from seven distinct proteins, including three that are known targets of NAPQI adduction. This approach should prove useful for identifying the protein targets of chemically reactivate metabolites to within a single amino acid residue, without prior knowledge of the metabolite(s) or protein target(s). It is anticipated that this powerful method should be generally applicable to study the mechanisms of toxicity induced by many different compounds.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.6b04604.

DOI: 10.1021/acs.analchem.6b04604
Anal. Chem. 2017, 89, 5748−5756

Additional text with experimental details; 10 figures including HiTIME heatmaps for additional replicates of APAP and control microsomal protein digests, histograms of HiTIME scores, and sample MS^2 spectra of peptides identified by targeted correlation; eight tables listing possible CRM formulas identified by nontargeted searching and parameters used for nontargeted CRM searching and targeted results correlation; three equations used to score MS2 spectra correlation, identify rotatable bonds, and calculate residual mass error values; six screenshot images of the analysis software user interface (PDF)

One table with lists of proteins identified (XLSX)

AUTHOR INFORMATION

Corresponding Authors

*E-mail m.leeming@student.unimelb.edu.au (M.G.L.).
*E-mail w.donald@unsw.edu.au (W.A.D.).
*E-mail rohair@unimelb.edu.au (R.A.J.O.).

ORCID

Michael G. Leeming: 0000-0001-8981-0701
Richard A. J. O'Hair: 0000-0002-8044-0502

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Financial support from the University of Melbourne Interdisciplinary Seed Grant program is gratefully acknowledged. M.G.L. thanks the Elizabeth and Vernon Puzey Foundation for the award of a Ph.D. scholarship and The University of Melbourne for a Norma Hilda Schuster scholarship. W.A.D. thanks the Australian Research Council for a Discovery Early Career Researcher Award (DE130100424) and a Discovery Project (DP160102681). We gratefully acknowledge helpful discussions with Ching-Seng Ang. We thank James Ziogas, Andrew Isaac, Bernard Pope, and David Perkins for useful discussions.

REFERENCES

(1) Miller, E. C.; Miller, J. A. *Cancer Res.* **1947**, *7*, 468−480.
(2) Miller, E. C.; Miller, J. A. *Cancer Res.* **1952**, *12*, 547−556.
(3) Kalgutkar, A. S.; Gardner, I.; Obach, R. S.; Shaffer, C. L.; Callegari, E.; Henne, K. R.; Mutlib, A. E.; Dalvie, D. K.; Lee, J. S.; Nakai, Y.; O'Donnell, J. P.; Boer, J.; Harriman, S. P. *Curr. Drug Metab.* **2005**, *6*, 161−225.
(4) Nelson, S. D. *J. Med. Chem.* **1982**, *25*, 753−765.
(5) Guengerich, F. P.; Liebler, D. C.; Reed, D. L. *CRC Crit. Rev. Toxicol.* **1985**, *14*, 259−307.
(6) Dipple, A.; Michejda, C. J.; Weisburger, E. K. *Pharmacol. Ther.* **1985**, *27*, 265−296.
(7) Park, B. K.; Kitteringham, N. R.; Maggs, J. L.; Pirmohamed, M.; Williams, D. P. *Annu. Rev. Pharmacol. Toxicol.* **2005**, *45*, 177−202.
(8) Uetrecht, J.; Naisbitt, D. J. *Pharmacol. Rev.* **2013**, *65*, 779−808.
(9) Nakayama, S.; Atsumi, R.; Takakusa, H.; Kobayashi, Y.; Kurihara, A.; Nagai, Y.; Nakai, D.; Okazaki, O. *Drug Metab. Dispos.* **2009**, *37*, 1970−1977.
(10) Langman, R. E.; Cohn, M. *Semin. Immunol.* **2000**, *12*, 159−162.
(11) Landsteiner, K.; Jacobs, J. *J. Exp. Med.* **1935**, *61*, 643−656.
(12) Uetrecht, J. *Annu. Rev. Pharmacol. Toxicol.* **2007**, *47*, 513−539.
(13) Matzinger, P. *Annu. Rev. Immunol.* **1994**, *12*, 991−1045.
(14) Séguin, B.; Uetrecht, J. *Curr. Opin. Allergy Clin. Immunol.* **2003**, *3*, 235−242.
(15) Pirmohamed, M.; Naisbitt, D. J.; Gordon, F.; Park, B. K. *Toxicology* **2002**, *181−182*, 55−63.
(16) Zanni, M. P.; von Greyerz, S.; Schnyder, B.; Brander, K. A.; Frutig, K.; Hari, Y.; Valitutti, S.; Pichler, W. J. *J. Clin. Invest.* **1998**, *102*, 1591−1598.
(17) Pichler, W. J. *Curr. Opin. Allergy Clin. Immunol.* **2002**, *2*, 301−305.
(18) Evans, D. C.; Watt, A. P.; Nicoll-Griffith, D. A.; Baillie, T. A. *Chem. Res. Toxicol.* **2004**, *17*, 3−16.
(19) Stepan, A. F.; Mascitti, V.; Beaumont, K.; Kalgutkar, A. S. *MedChemComm* **2013**, *4*, 631−652.
(20) Stepan, A. F.; Walker, D. P.; Bauman, J.; Price, D. A.; Baillie, T. A.; Kalgutkar, A. S.; Aleo, M. D. *Chem. Res. Toxicol.* **2011**, *24*, 1345−1410.
(21) Leung, L.; Kalgutkar, A. S.; Obach, R. S. *Drug Metab. Rev.* **2012**, *44*, 18−33.
(22) Roberts, S. A.; Price, V. F.; Jollow, D. J. *Toxicol. Appl. Pharmacol.* **1990**, *105*, 195−208.
(23) Hanzlik, R. P. *Drug Metab. Rev.* **1982**, *13*, 207−234.
(24) Myers, T. G.; Dietz, E. C.; Anderson, N. L.; Khairallah, E. A.; Cohen, S. D.; Nelson, S. D. *Chem. Res. Toxicol.* **1995**, *8*, 403−413.
(25) Matthews, A. M.; Hinson, J. A.; Roberts, D. W.; Pumford, N. R. *Toxicol. Lett.* **1997**, *90*, 77−82.
(26) Hanzlik, R. P.; Koen, Y. M.; Theertham, B.; Dong, Y.; Fang, J. *BMC Bioinf.* **2007**, *8*, 95.
(27) Hanzlik, R. P.; Fang, J.; Koen, Y. M. *Chem.-Biol. Interact.* **2009**, *179*, 38−44.
(28) Liebler, D. C. *Chem. Res. Toxicol.* **2008**, *21*, 117−128.
(29) Gan, J.; Zhang, H.; Humphreys, W. G. *Chem. Res. Toxicol.* **2016**, *29*, 2040−2057.
(30) Tailor, A.; Waddington, J. C.; Meng, X.; Park, B. K. *Chem. Res. Toxicol.* **2016**, *29*, 1912−1935.
(31) Bartolone, J. B.; Sparks, K.; Cohen, S. D.; Khairallah, E. A. *Biochem. Pharmacol.* **1987**, *36*, 1193−1196.
(32) Ikehata, K.; Duzhak, T. G.; Galeva, N. A.; Ji, T.; Koen, Y. M.; Hanzlik, R. P. *Chem. Res. Toxicol.* **2008**, *21*, 1432−1442.
(33) Koen, Y. M.; Sarma, D.; Hajovsky, H.; Galeva, N. A.; Williams, T. D.; Staudinger, J. L.; Hanzlik, R. P. *Chem. Res. Toxicol.* **2013**, *26*, 564−574.
(34) Golizeh, M.; LeBlanc, A.; Sleno, L. *Chem. Res. Toxicol.* **2015**, *28*, 2142−2150.
(35) Kluger, B.; Bueschl, C.; Neumann, N.; Stückler, R.; Doppler, M.; Chassy, A. W.; Waterhouse, A. L.; Rechthaler, J.; Kampleitner, N.; Thallinger, G. G.; Adam, G.; Krska, R.; Schuhmacher, R. *Anal. Chem.* **2014**, *86*, 11533−11537.
(36) Bueschl, C.; Kluger, B.; Berthiller, F.; Lirk, G.; Winkler, S.; Krska, R.; Schuhmacher, R. *Bioinformatics* **2012**, *28*, 736−738.
(37) Knapp, D. R.; Gaffney, T. E.; McMahon, R. E.; Kiplinger, G. *J. Pharmacol. Exp. Ther.* **1972**, *180*, 784−790.
(38) Hoi, K. K.; Daborn, P. J.; Battlay, P.; Robin, C.; Batterham, P.; O'Hair, R. A. J.; Donald, W. A. *Anal. Chem.* **2014**, *86*, 3525−3532.
(39) Leeming, M. G.; Isaac, A. P.; Pope, B. J.; Cranswick, N.; Wright, C. E.; Ziogas, J.; O'Hair, R. A. J.; Donald, W. A. *Anal. Chem.* **2015**, *87*, 4104−4109.
(40) Yukinaga, H.; Iwabuchi, H.; Okazaki, O.; Izumi, T. *J. Pharm. Biomed. Anal.* **2012**, *67−68*, 186−192.
(41) Shin, N.-Y.; Liu, Q.; Stamer, S. L.; Liebler, D. C. *Chem. Res. Toxicol.* **2007**, *20*, 859−867.
(42) Leeming, M. G.; Gamon, L. F.; Wille, U.; Donald, W. A.; O'Hair, R. A. J. *Chem. Res. Toxicol.* **2015**, *28*, 2224−2233.
(43) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; et al. *Nat. Biotechnol.* **2012**, *30*, 918−920.
(44) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551−3567.
(45) The UniProt Consortium.. *Nucleic Acids Res.* **2008**, *36*, D190−D195.
(46) Elias, J. E.; Haas, W.; Faherty, B. K.; Gygi, S. P. *Nat. Methods* **2005**, *2*, 667−675.
(47) Dasari, S.; Chambers, M. C.; Slebos, R. J.; Zimmerman, L. J.; Ham, A.-J. L.; Tabb, D. L. *J. Proteome Res.* **2010**, *9*, 1716−1726.

DOI: 10.1021/acs.analchem.6b04604
Anal. Chem. 2017, 89, 5748−5756

(48) Tabb, D. L.; Saraf, A.; Yates, J. R. *Anal. Chem.* **2003**, *75*, 6415−6421.
(49) Na, S.; Jeong, J.; Park, H.; Lee, K.-J.; Paek, E. *Mol. Cell. Proteomics* **2008**, *7*, 2452−2463.
(50) Creasy, D. M.; Cottrell, J. S. *Proteomics* **2004**, *4*, 1534−1536.
(51) Na, S.; Bandeira, N.; Paek, E. *Mol. Cell. Proteomics* **2012**, *11*, No. M111.010199.
(52) Josephy, P. D. *Drug Metab. Rev.* **2005**, *37*, 581−594.
(53) Wendel, A.; Cikryt, P. *Res. Commun. Chem. Pathol. Pharmacol.* **1981**, *33*, 463−473.
(54) Zhang, H.; Gan, J.; Shu, Y.-Z.; Humphreys, W. G. *Chem. Res. Toxicol.* **2015**, *28*, 775−781.
(55) Tzouros, M.; Pähler, A. *Chem. Res. Toxicol.* **2009**, *22*, 853−862.
(56) Dooley, G. P.; Prenni, J. E.; Prentiss, P. L.; Cranmer, B. K.; Andersen, M. E.; Tessari, J. D. *Chem. Res. Toxicol.* **2006**, *19*, 692−700.
(57) Jenkins, R. E.; Kitteringham, N. R.; Goldring, C. E. P.; Dowdall, S. M. J.; Hamlett, J.; Lane, C. S.; Boerma, J.-S.; Vermeulen, N. P. E.; Park, B. K. *Proteomics* **2008**, *8*, 301−315.

DOI: 10.1021/acs.analchem.6b04604
Anal. Chem. 2017, 89, 5748−5756