Journal of Proteomics 249 (2021) 104360

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Optimization of proteomics sample preparation for forensic analysis of skin samples

Maryam Baniasad [a], Andrew J. Reed [b], Stella M. Lai [a], Liwen Zhang [b], Kathleen Q. Schulte [c], Alan R. Smith [c], Danielle S. LeSassier [c], Katharina L. Weber [c], F. Curtis Hewitt [c], August E. Woerner [d], Myles W. Gardner [c], Vicki H. Wysocki [a], Michael A. Freitas [b,e,*]

[a] Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA
[b] Mass Spectrometry and Proteomics Facility, Campus Chemistry Instrument Center, The Ohio State University, Columbus, OH, USA
[c] Signature Science, LLC, Austin, TX, USA
[d] Center for Human Identification, University of North Texas Health Science Center, Fort Worth, TX, USA
[e] The Ohio State University Wexner Medical Center, Columbus, OH, USA

ARTICLE INFO

Keywords:
Proteomics
Sample preparation
Protein extraction
Trypsin digestion
Genetically variable peptides
Forensic analysis

ABSTRACT

We present an efficient protein extraction and in-solution enzymatic digestion protocol optimized for mass spectrometry-based proteomics studies of human skin samples. Human skin cells are a proteinaceous matrix that can enable forensic identification of individuals. We performed a systematic optimization of proteomic sample preparation for a protein-based human forensic identification application. Digestion parameters, including incubation duration, temperature, and the type and concentration of surfactant, were systematically varied to maximize digestion completeness. Through replicate digestions, parameter optimization was performed to maximize repeatability and increase the number of identified peptides and proteins. Final digestion conditions were selected based on the parameters that yielded the greatest percent of peptides with zero missed tryptic cleavages, which benefit the analysis of genetically variable peptides (GVPs). We evaluated the final digestion conditions for identification of GVPs by applying MS-based proteomics on a mixed-donor sample. The results were searched against a human proteome database appended with a database of GVPs constructed from known non-synonymous single nucleotide polymorphisms (SNPs) that occur at known population frequencies. The aim of this study was to demonstrate the potential of our proteomics sample preparation for future implementation of GVP analysis by forensic laboratories to facilitate human identification.

Significance: Genetically variable peptides (GVPs) can provide forensic evidence that is complementary to traditional DNA profiling and be potentially used for human identification. An efficient protein extraction and reproducible digestion method of skin proteins is a key contributor for downstream analysis of GVPs and further development of this technology in forensic application. In this study, we optimized the enzymatic digestion conditions, such as incubation time and temperature, for skin samples. Our study is among the first attempts towards optimization of proteomics sample preparation for protein-based skin identification in forensic applications such as touch samples. Our digestion method employs RapiGest (an acid-labile surfactant), trypsin enzymatic digestion, and an incubation time of 16 h at 37 °C.

1. Introduction

Human skin deposits represent a major portion of forensic samples in crime scenes [1]; however, a large proportion of these samples do not contain sufficient DNA for human identification. Alternatively, skin proteins present in these same shed skin/fingerprints/fingertip smears are environmentally more robust and quantitatively more abundant. For such samples, the deposited protein in the form of skin cells is an exploitable biochemical matrix for forensic analysis. The combination of liquid chromatography and tandem mass spectrometry (LC-MS/MS) provides a flexible, dynamic platform for the identification and quantification of proteins in many different matrices, including skin tissue

* Corresponding author at: Mass Spectrometry and Proteomics Facility, Campus Chemistry Instrument Center, The Ohio State University, Columbus, OH, USA.
E-mail address: freitas.5@osu.edu (M.A. Freitas).

https://doi.org/10.1016/j.jprot.2021.104360
Received 30 April 2021; Received in revised form 28 July 2021; Accepted 22 August 2021
Available online 1 September 2021

[2,3]. While MS-based proteomics has typically been applied to solve clinical problems, advances in these methods have found additional applications within forensic science [2]. Currently, analysis of DNA through PCR-based genotyping and genomics sequencing approaches are the gold standard techniques in human forensic science [4]; however, proteomic analysis can overcome some of the challenges of using DNA-based techniques for specific types of forensic samples such as human touch samples with no DNA content or degraded DNA [5–7]. In contrast to nucleic acids, proteins are highly stable molecules that are less vulnerable to oxidation, UV damage, and chemical decomposition [7].

While DNA represents the most common class of forensically informative biological material, protein is considered as an alternative source for human identification. Proteins contain genetic variations in the form of single amino acid polymorphisms (SAPs), which can be detectable within peptides from enzymatic digests of those proteins. Peptides that contain informative SAPs are known as genetically variable peptides (GVPs) [8]. SAPs are the direct products of non-synonymous single nucleotide polymorphisms (nsSNPs) that are present in DNA sequences [9]. These GVPs provide forensic evidence that is complementary to traditional DNA-based profiling and can be used in the context of human identification. Initial efforts in GVP-based human identification have been performed primarily on hair samples [10,11]. However, proteomics-based approaches in forensic science can be extended to any protein-containing sample, particularly those matrices that typically lack DNA, such as tooth, bone, and skin cells [7,12].

While there is value in using skin as a forensically relevant matrix, there are relatively few proteomic studies on human skin compared with other biological tissues [12,13]. One challenge with analyzing skin samples is their high lipid content and insolubility [12,13]. Since protein solubilization is an important step affecting the performance of proteomics analysis, considerable efforts have been dedicated in the past decade to improve protein extraction efficiency by using surfactants to solubilize proteins [14–16]. Different types of surfactants have been used in proteomics studies. Sodium dodecyl sulfate (SDS) is one ionic surfactant that has been used extensively for solubilizing membrane proteins [17]. However, SDS is not compatible with LC-MS/MS downstream analysis as it contaminates LC systems and suppresses the ionization of peptides due to its ready ionizability [18]. Therefore, SDS removal is a critical step required prior to LC-MS/MS, affecting the sample recovery for peptides [18,19]. As an alternative to SDS, acid-labile surfactants, such as sodium-3-[(2-methyl-2- undecyl-1,3-dioxolan-4-yl)-methoxyl]-1-propanesulfonate (RapiGest™) and sodium 3-((1-(furan-2-yl)undecyloxy) carbonylamino)propane-1-sulfonate (ProteaseMAX™) are MS-compatible [20–22]. RapiGest™ (herein RapiGest) is a trypsin-friendly surfactant that undergoes hydrolysis in acidic conditions. This specific feature can be utilized to remove RapiGest from solutions when desired [23]. ProteaseMAX™ (herein ProteaseMAX) is a cleavable surfactant that is sensitive to heat and acid. Both RapiGest and ProteaseMAX improve protein solubility during sample preparation by unfolding the protein structure and exposing its proteolytic sites to enzymatic cleavages. No additional detergent removal steps are required to remove RapiGest and ProteaseMAX. They easily undergo hydrolysis under acidic conditions and can be removed prior to further analysis without sample loss [21]. This unique feature makes them more desirable for MS analyses as compared to SDS.

In the current study, we developed a proteomics sample preparation workflow for human skin samples that can be routinely applied to the analysis of GVPs in forensic labs. To aid the development of an analytical method targeted for measuring a panel of GVPs, e.g., a parallel reaction monitoring or multiple reaction monitoring MS-based method, it is critical to have a robust sample preparation method that reproducibly yields the desired target analytes. For this particular application, critical factors include solubilizing the greatest number of proteins and producing the greatest number of peptides with no missed cleavages. In this study, RapiGest and ProteaseMAX were evaluated for their ability to efficiently solubilize skin proteins (e.g., such as those deposited in touch samples on a surface), and we evaluated enzymatic digestion conditions (e.g., incubation time and temperature) to determine optimal sample preparation conditions for LC-MS/MS analysis.

2. Material and methods

2.1. Human epidermal skin sample collection

Twenty-five adult donors (male and female, over 18 years old) of northern European ancestry used a commercial skin exfoliation product (PedEgg™) on their hands and fingers to collect epidermal skin material. Each PedEgg tool was decontaminated prior to use by rinsing with RNase Away, followed by a rinse with 70% isopropyl alcohol and allowed to dry. Unique, decontaminated PedEggs were used by each donor. Immediately prior to skin collection, donors washed and dried their hands. Donors were then instructed to rub the PedEgg across the palm and fingers of both hands for 100 s; skin particles were collected into the chamber located below the exfoliating grate. After removing the grate from each PedEgg, skin material was collected using a previously decontaminated eyebrow brush (unique for each individual) to brush the skin material into a microcentrifuge tube. All samples were stored at −80 °C prior to further processing. Protocols and informed consents for collecting human subject material were approved by the Institutional Review Boards of the University of North Texas Health Science Center (IRB #00642). Consent was acquired orally and in writing.

2.2. Proteomic sample preparation

Immediately prior to proteomics sample preparation, skin samples were taken out of −80 °C and kept on ice. Three (3) mg of each individual skin sample was transferred to one new protein LoBind microcentrifuge tube (Eppendorf) to prepare a pooled sample of all individuals. To study the effect of different detergents and digestion conditions, 3 mg of the pooled skin sample was transferred to separate microcentrifuge tubes for further preparation.

For each tube containing 3 mg of skin, 300 μL of either RapiGest or ProteaseMax at a concentration of 0.1% (*w/v*) or 0.25% (w/v) in 50 mM ammonium bicarbonate was added to the skin sample. To extract proteins, samples were lysed by probe sonication for 1 min, heated at 95 °C for 5 min (except for samples dissolved in ProteaseMAX), cooled on ice for 1 min, and finally vortexed for 5 min. It should be noted that in the heating step, samples dissolved in ProteaseMAX were heated at 85 °C as ProteaseMAX is vulnerable to degradation at 95 °C. Samples were then centrifuged at 16,000 ×*g* for 30 min to pellet undissolved skin material. Supernatants were then transferred to new LoBind tubes to measure protein concentrations using a Qubit protein assay on a Thermo Fisher Qubit fluorometer per the manufacturer's protocol. For each sample, 30 μg of extracted proteins (an amount comparable to that found in actual fingerprints) [13] were reduced by adding 5.0 μL of 5.0 mg/mL dithiothreitol (DTT) in 50 mM ammonium bicarbonate and incubating for 15 min at 65 °C. Proteins were then alkylated by addition of 5.0 μL of 15.0 mg/mL iodoacetamide (IAA) in 50 mM ammonium bicarbonate and incubation for 30 min in the dark. 1 μg of Trypsin (Promega), reconstituted in 50 mM ammonium bicarbonate, was added at a 1:30 *w/w* ratio (enzyme:protein). To compare the effect of incubation time and temperature on the enzymatic digestion efficiency, samples were then incubated for either 3 or 16 h (overnight) at either 37 °C or 50 °C. Following trypsin enzymatic digestion, surfactant was precipitated by adding 5% trifluoracetic acid (TFA) to a final concentration of 0.5% TFA. The final tryptic peptide samples were vacuum-dried and after evaporation, were reconstituted in 50 mM acetic acid prior to LC-MS/MS analysis.

2.3. Mass spectrometry

Skin samples were analyzed on a Thermo Scientific Q Exactive Plus high resolution, accurate-mass (HRAM) mass spectrometer coupled to a Thermo Scientific Ultimate 3000 nano-LC system configured with a C18 Easy-Spray column (75 μm i.d. × 25 cm, Thermo Scientific) with a column temperature of 55 °C. *Mobile* phase A consisted of water with 0.1% (*v*/v) formic acid, while mobile phase B was comprised of acetonitrile with 0.1% (v/v) formic acid. The mobile phases were maintained at a flow rate of 300 nL/min. The solvent gradient started at 2% B for 5 min, and then separation was achieved using the following linear gradient steps: i) to 20% B over 100 min; ii) to 32% B over 10 min; iii) to 95% B for 10 min; and iv) hold at 95% B for 4 min. The column was then re-equilibrated by a 1-min gradient back to 2% B and held for 15 min before the beginning of the next run. The ion source was operated in positive ion mode. The mass spectrometer was operated in the data-dependent MS mode. Full scan mass spectra were acquired from *m/z* 375 to 1575 at a resolution of 70,000. The fifteen (15) most abundant precursor ions in each full MS1 spectrum were selected for fragmentation. An isolation window of 1.6 *m/z* was used for fragmentation with a normalized collision energy of 30. Tandem mass spectra (MS2) were acquired at a resolution of 17,500.

2.4. Proteomic data analysis

Peptide and protein identifications were acquired using the Thermo Proteome Discoverer software (v1.4), and the Sequest search algorithm against a UniProt human database (release-Nov-2018) appended with a set of GVPs (see Section 2.5). The precursor mass tolerance was set to 20 ppm and the fragment ion mass tolerance to 0.8 Da. Trypsin was set as the enzyme used with a maximum of 2 missed cleavages. Cysteine carbamidomethylation was set as a fixed modification, while oxidation of methionine and deamidation of asparagine and glutamine were set as variable modifications. False discovery rate (FDR) control was performed using Percolator at a threshold of 1% for peptide spectral match (PSM), peptide, and protein identifications. Protein groups were filtered to include a minimum of two peptides per protein group at 1% FDR. For statistical analysis, Python scripts (v3.7) were used to process the outputs of Proteome Discoverer. The mass spectrometry proteomics data have been deposited to the ProteomeXchange [24] Consortium via the PRIDE [25] partner repository with the dataset identifier PXD022720 and https://doi.org/10.6019/PXD022720.

2.5. GVP analysis

To test our skin proteomics sample preparation workflow for GVP analysis, we compared the identified peptides that resulted from each digestion condition with a database of GVPs generated from common human genomic variants predicted to yield changes in protein sequence, i.e., translate genomic variations to amino acids in the proteome. This process was accomplished as described previously [26]. In brief, relatively common non-synonymous SNP and in-frame insertion-deletion (indel) (allele frequency between 1 and 99% in non-Finnish European individuals) were selected. The Variant Effect Predictor (VEP) was used to translate these SNPs and indels into amino acid variations in canonical proteins (Ensembl v85, GRCh37). The predicted protein sequences were enzymatically digested in silico with trypsin allowing for zero missed cleavages. The resulting peptides were filtered to retain those of a peptide length of 7 to 50 amino acids and then converted to FASTA format for proteomic data analysis. The digestion was performed allowing zero missed cleavages, as the goal of our study was to identify common GVPs that could be reproducibly observed [27], and experimental efforts were being performed in parallel to generate digests of skin material with the fewest number of missed cleavages.

3. Results and discussion

Cell lysis and protein extraction are the first critical steps of sample preparation in bottom-up proteomics [28]. Based on the nature of the samples, many different lysis methods have been reported in the literature [15,29]. Extracting proteins from skin cells is more challenging than many other cells and tissues due to the insolubility of skin material [12]. Over the years, detergents or chaotropic agents have been utilized to facilitate protein extraction [14–16,30]. However, due to the incompatibility of most of these reagents with LC-MS, the samples need to undergo clean-up prior to LC-MS analysis, introducing the potential for sample loss and/or sample variability. In the current study, we evaluated the effects of two MS-compatible surfactants, RapiGest and ProteaseMAX, on the proteomic characterization of human skin samples.

The proteolytic digestion step is another critical step to assess during proteomics sample preparation as the duration, temperature, and completeness of digestion can affect the overall performance of bottom-up proteomics analysis [31,32]. While trypsin is the gold standard among proteases used for protein digestion, tryptic digestion is unlikely to be complete due to lysine cleavage inefficiency [33]. Therefore, trypsin digestion remains a variable process that contributes to lower precision and reproducibility. To develop an efficient digestion method for human skin sample analysis, we also assessed the effect of temperature and time on trypsin as a thermostable protease [34].

3.1. Optimizing RapiGest concentrations and digestion conditions for skin protein recovery

We first focused on various digestion conditions using RapiGest. Previous studies have reported different concentrations of RapiGest ranging from 0.05% to 1% (*w/v*) for bottom-up proteomics workflow depending on the sample complexity and protein hydrophobicity [35]. Most of these studies recommended a lower concentration of RapiGest (0.1–0.5%) for protein extraction without negative impact on trypsin activity [35–37]. In this study, two batches of samples were evaluated. Samples from batch one were evaluated at two concentrations, 0.1% and 0.25% (w/v). For batch two, we evaluated digestion efficiency under different lysis conditions, including time and temperature. Although most protein hydrolysis methods that utilize trypsin suggest 37 °C as the optimum temperature for tryptic digestion, trypsin has the advantage of being relatively thermostable and is active over a range of temperatures [34]. Therefore, it is possible that equivalent digestions could be performed in shorter times by increasing the digestion temperature higher than 37 °C. Consequently, we investigated the effect of higher temperature (50 °C) on trypsin digestion efficiency. Moreover, we evaluated the effect of time on enzymatic digestion. Although trypsin digestion protocols have often relied on lengthy digestion times to ensure effective proteolysis, long digestion time is not necessarily desirable and is typically considered the current rate-limiting step for timely analysis. For this reason, we evaluated shorter duration digestion (3 h) and a standard (16 h, overnight) duration typically used in laboratories. In the current study, digestions were performed at all combinations of 3 or 16 h durations and 37 °C or 50 °C. For each sample condition, equal amounts of protein extract following lysis were subjected to shotgun proteomics analyses. Here we chose to use the number of identified proteins and peptides (reference peptides associated with UniProt accession numbers in the FASTA database) as the metric for comparison as the starting materials were identical, and the amount of protein used for digestion was kept the same. We found that 0.1% RapiGest outperformed 0.25% across all conditions, with a 3-h or overnight digestion at 37 °C yielding the greatest number of unique protein and peptide identifications (protein mean = 292 and peptide mean = 2266 for 37 °C, overnight; Fig. 1). In contrast, the 0.25% RapiGest digestion for 3 h at 50 °C resulted in the fewest protein (mean = 183) and peptide identifications (mean = 1468) across all sample groups (Fig. 1). Overall, our results demonstrated that a lower concentration of RapiGest (0.1%)

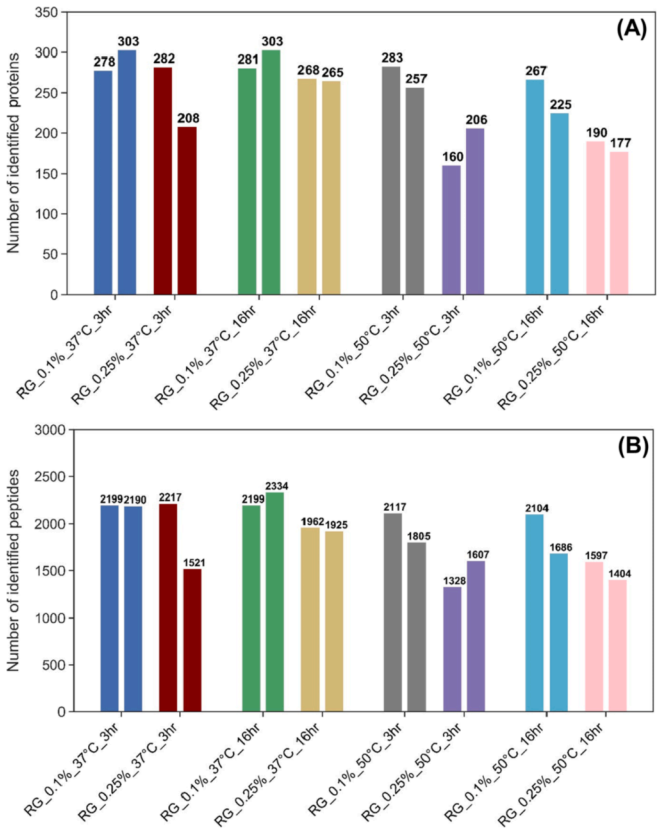

Fig. 1. RapiGest (RG) results for different digestion conditions. (A) Number of proteins identified (based on two or more unique peptides); (B) Number of peptides identified. Bars of the same color indicate each of the two replicates.

outperforms the higher concentration (0.25%), regardless of digestion conditions used for our skin samples. This may be due to denaturation of trypsin at higher RapiGest concentrations, as demonstrated in a previous study [22], wherein moderate reduction in the activity of trypsin was identified at higher concentrations of RapiGest. Notably, this effect is still much smaller compared to MS-incompatible surfactants such as SDS [22]. In addition, our results have shown that increasing the hydrolysis temperature does not improve protein identification and provides further evidence that the optimal trypsin hydrolysis temperature for human skin samples in RapiGest is 37 °C.

Also, we compared the overlap in identified proteins between two replicates of each tested condition (Fig. S1). While there are no statistically significant differences in the overlaps of protein identifications across two replicates, we observed the greatest degree of overlap (81%) across the two experimental replicates for 0.1% RapiGest 16 h (overnight) digestion at 37 °C. In contrast, samples prepared with 0.25% RapiGest demonstrated greater differences in protein identifications between replicates across 3 of the 4 tested digestion conditions. Furthermore, we used Pearson correlation coefficients (r) of the number of peptide spectral matches (PSM) of each protein identified in each replicate to quantitatively assess the correlation between replicates (Fig. S1). It was determined that 0.1% RapiGest in overnight digestion at 37 °C showed the strongest linear correlation (r = 0.973) between replicates.

3.2. Comparing the effect of different surfactants

As mentioned above, RapiGest is one of the more commonly used surfactants that increase trypsin digestion efficiency without the need for physical removal (e.g., filtering) before LC-MS analysis. Our study revealed that increasing the concentration of this surfactant did not improve protein recovery nor yield more reproducible proteomic identifications as compared to a lower concentration. Therefore, 0.1% RapiGest yielded an overall better performance for skin proteomics compared to 0.25% RapiGest. To further assess our protein extraction method, we evaluated another common MS-compatible surfactant, ProteaseMAX. For this comparison, skin samples were solubilized with either Rapigest or ProteaseMAX at 0.1% *w*/w, and then digested using all combinations of 3 or 16 h durations at 37 or 50 °C, with each experiment performed in experimental duplicate. As shown in Fig. 2, the 0.1% RapiGest with a 3 h digestion at 37 °C yielded the greatest number of protein identifications (mean = 332) when comparing all times, temperatures, and surfactants. A similar trend was seen for the number of identified peptides (mean = 2297). Overall, more proteins and peptides were identified using RapiGest compared to ProteaseMAX under the same digestion conditions (Fig. 2), suggesting the RapiGest at 0.1% solubilizes and unfolds a greater number of skin proteins than ProteaseMAX at 0.1%. We further compared the effect of increasing the concentration of ProteaseMAX (0.25%) on trypsin digestion at 37 °C, and our results confirmed that the higher concentration of ProteaseMAX does not improve the protein identifications (Fig. S2), a result that parallels the RapiGest results.

To further compare the effect of RapiGest and ProteaseMAX, the overlap in unique protein identifications from two combined experimental replicates of each of four conditions using either RapiGest or ProteaseMAX were compared. As shown in Fig. 3, there was a greater overlap in the protein identifications between conditions where RapiGest was used as the surfactant compared to ProteaseMAX.

It is worth mentioning that we further investigated the difference between RapiGest and ProteaseMAX on protein extraction from skin samples by merging all of the experimental replicates where RapiGest was used as surfactant and compared the overlap in protein identifications with combined experimental replicates of ProteaseMAX using a

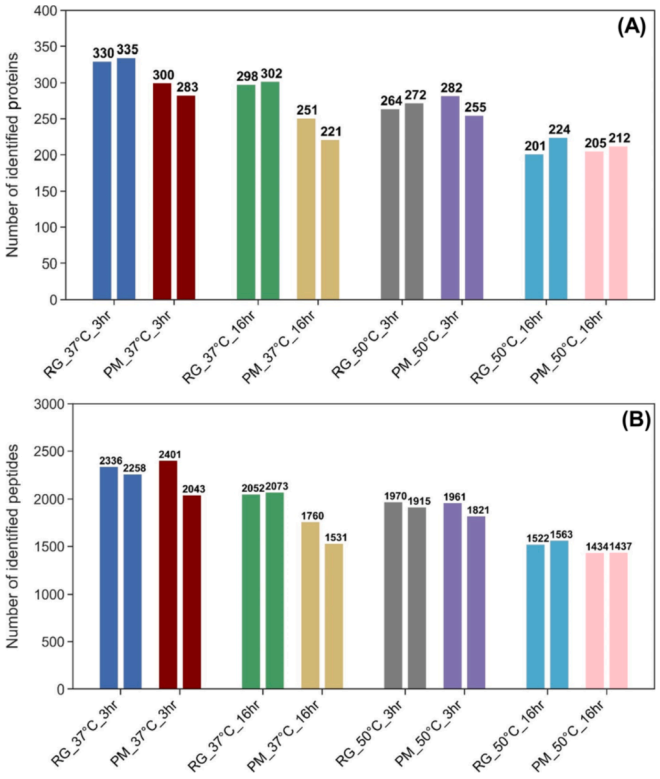

Fig. 2. For each digestion condition using either 0.1% RapiGest (RG) or 0.1% ProteaseMax (PM), (A) Number of proteins identified (based on two or more unique peptides); (B) Number of peptides identified. Bars of the same color indicate each of the two replicates.

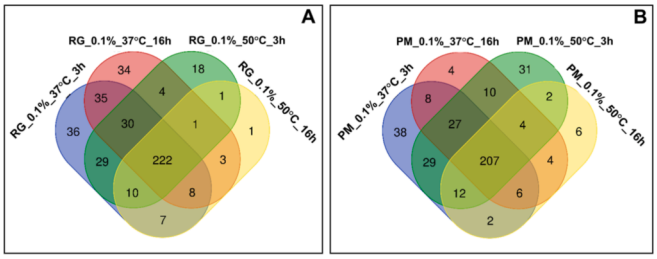

Fig. 3. Overlap in unique protein identifications from two combined experimental replicates for each digestion condition using (A) RapiGest; (B) ProteaseMAX.

Venn diagram (Fig. S3). Moreover, the list of unique proteins extracted from each of these surfactants was provided in Table S2 and their characteristics were compared in Table S3. Our results showed a greater range of molecular weight (MW) and calculated isoelectric point (calc. pI) where RapiGest was used for extraction (Fig. 4 and Table S3).

Another important factor for evaluating different protein extraction and digestion methods is enzymatic digestion efficiency. One of the disadvantages of strong surfactants such as SDS is that they inhibit the activity of endopeptidases such as trypsin. In contrast, acid-labile surfactants such as RapiGest and ProteaseMAX were designed to extract proteins without inhibiting trypsin activity [22,23]. To assess digestion completeness, the percent of identified peptides with missed cleavages was calculated (Fig. 5). We observed that in all tested samples, greater than 75% of the peptides identified were fully cleaved products (i.e., no internal lysine or arginine residues). However, for both surfactants, 16 h performed better than 3 h digestions in yielding a higher percentage of peptides with no missed cleavages, which, along with the convenience of leaving the digestions running overnight, may explain why overnight digestions are still very common in many bottom-up proteomics sample preparation approaches.

The presence of missed cleavages is a potential concern for the development of a targeted analytical method focused on a panel of GVPs that could be applicable for human identification in a similar manner to DNA short tandem repeat (STR) profiling. It is desirable to only target SAPs once in such methods and not in two peptides (one with no missed cleavages and another peptide with a missed cleavage).

Overall, based on our results, 0.1% RapiGest performed better than 0.1% ProteaseMAX yielding the greatest number of protein identifications and overlap of the identified proteins. In addition, 0.1% RapiGest when combined with overnight digestion at 37 °C, resulted in the highest percentage of peptides with no missed cleavages. Therefore, we further combined the overall results of four replicates of 0.1% RapiGest

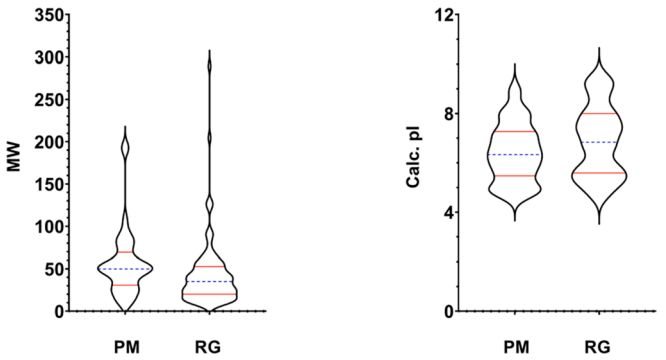

Fig. 4. Violin plots show the differences between molecular weight (MW) and the isoelectric point (Calc. pI) of proteins extracted by ProteaseMAX (PM) and RapiGest (RG). The solid red lines show the upper and lower quartile, and the dashed blue lines show the median. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

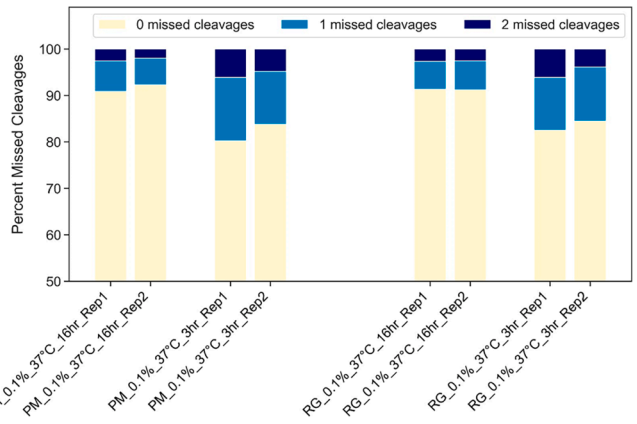

Fig. 5. Comparison of the percentage of missed cleavages in each condition when digestion performed at 37 °C. (RG = RapiGest; PM = ProteaseMax).

in overnight digestion at 37 °C (from batches one and two) to evaluate the reproducibility of our sample preparation method under this condition. For each identified protein, a coefficient of variation was calculated based on the number of peptide spectral matches of that protein across the four replicates of RapiGest 0.1% in overnight digestion at 37 °C. Histograms of coefficient of variations were then plotted, and the gamma function was used as the distribution fit. Fig. 6 shows a density plot (normalized histogram) of the calculated coefficient of variation based on the number of peptide spectral matches for each protein across all replicates ($N = 4$). As shown in Figs. 6, 0.1% RapiGest with overnight digestion at 37 °C showed high reproducibility on the density plot as greater than 71% of the identified proteins had PSMs with a coefficient of variation less than 0.3 across four replicates.

Overall, based on our results, RapiGest dissolution with overnight trypsin digestion at 37 °C, which resulted in the best combination of total IDs, a smaller number of identified peptides with missed cleavages, and high reproducibility, is an optimal method for downstream analysis of GVPs in forensic applications.

3.3. Evaluation of different skin digestion conditions for GVP analysis

To evaluate different digestion conditions for providing informative GVPs, we first filtered the identified peptides against those in the constructed GVP database (as described in Section 2.5). Selected peptides were further filtered based on their allele frequency (<70%) to focus on those peptides that provide human identification value (i.e., those that are not expected to be present in a large majority of humans). As shown in Fig. 7A, 0.1% RapiGest dissolution with overnight trypsin digestion at 37 °C resulted in the highest number of informative GVPs (mean = 88), confirming the potential of these conditions for human identification via

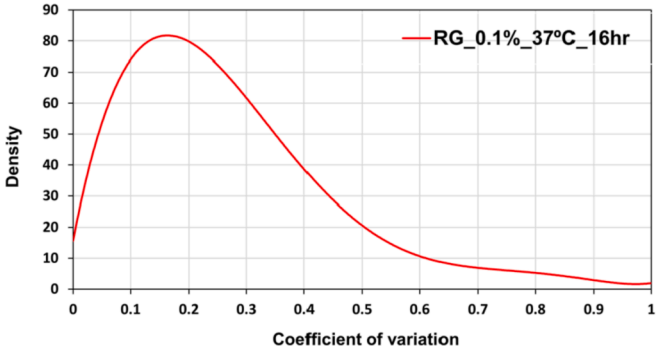

Fig. 6. Density plot of calculated coefficient of variations based on the number of peptide spectral matches for each protein across four replicates of 0.1% RapiGest with overnight digestion at 37 °C.

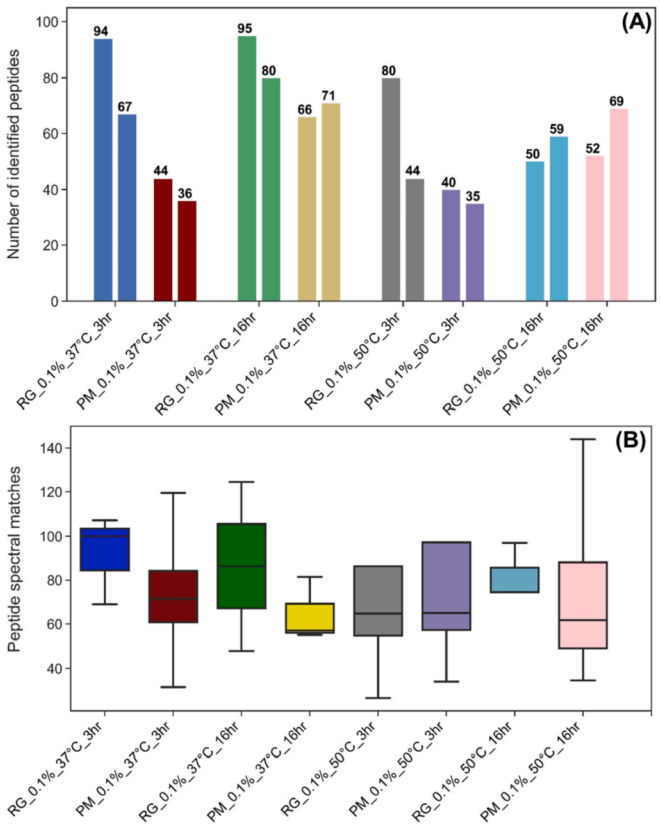

Fig. 7. For each digestion condition using either 0.1% RapiGest (RG) or 0.1% ProteaseMAX (PM), (A) Number of identified peptides (potential GVPs) matched with predicted peptides based on common human genomic variants; (B) Boxplots of peptide spectral matches (PSM) values for potential GVPs.

skin proteomics. The boxplots in Fig. 7B present the range for the number of identified GVP PSMs using each set of digestion conditions.

To further demonstrate the potential utility of this skin preparation workflow (0.1% RapiGest, 37 °C, overnight) for human identification, a histogram of detected GVPs by their allele frequency is presented in Fig. 8. Many low-frequency (< 0.1) peptides are detected. An estimated random match probability (RMP) can be calculated from these

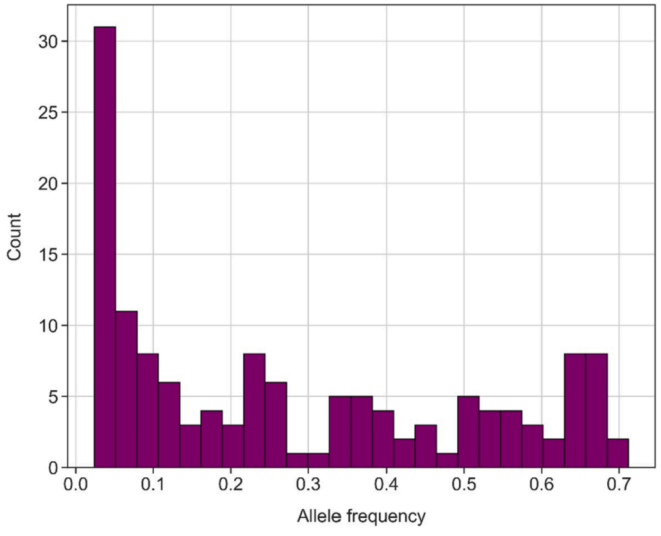

Fig. 8. Histogram of potential genetically variable peptide count as a function of allele frequency using overnight digestion at 37 °C when skin samples were lysed using 0.1% RapiGest.

detections under the assumption that the loci are bi-allelic and are sampled from a population in Hardy Weinberg equilibrium (HWE). Under these assumptions, the sum of the two allele frequencies at a given locus is 1 ($p + q = 1$), where p and q are the allele frequencies for the major and minor alleles, respectively. For diploid organisms such as humans, the sum of the different genotype frequencies must also be 1, and under HWE the genotype frequencies are defined as: $p^2 + 2pq + q^2 = 1$. As each GVP represents a single allele, in the case of major/minor pairs of GVPs detected, the heterozygous genotype frequency (minor/major + major/minor = 2pq) was used. However, in other cases in which only the major or minor form of a GVP was detected, one cannot determine if that GVP was from one or both of the maternal and paternal chromosomes. We thus make a conservative estimate that the genotype frequency can be either heterozygous or homozygous. Therefore, the possible genotypes for a detected major GVP (without detection of its minor form) would be homozygous (major/major = p^2) or heterozygous (minor/major + major/minor = 2pq) for a total genotype frequency of $p^2 + 2pq$. However, the possible genotypes for a detected minor GVP (without detection of its major form) would be homozygous (minor/minor = q^2) or heterozygous (minor/major + major/minor = 2pq) for a total genotype frequency of $q^2 + 2pq$. We note that given the stochastic nature of data-dependent MS/MS analysis, a non-detection of a peptide is not sufficient evidence for the lack of the peptide in a sample. The estimated RMP is then calculated as the product of the individual GVP genotype frequencies. Recognizing that the samples analyzed contained skin from multiple contributors and that this estimate does not take into consideration linkage disequilibrium (i.e., the non-random association of alleles in populations), the estimated RMP for the 126 GVPs detected at 119 loci and presented in Fig. 8 is 3.4×10^{-80}. This RMP value means that one would expect that profile of detected GVPs to be randomly observed in only 1 out of 2.96×10^{80} individuals. While the value of 3.4×10^{-80} overstates the strength of the evidence, it does suggest that many informative alleles can be detected with such an approach. A more conservative estimate would be to take the product of the lowest genotype frequency by chromosome. Such an approach likely serves as an upper bound on the RMP as it uses far fewer alleles (23/126 (18.2%) across 23 chromosomes) derived from a unit of inheritance (the chromosome) that is biologically independent. Using this latter approach yields an RMP of 3.2×10^{-28}. (For a more accurate peptide-based RMP for sets of peptide identifications that can be assumed to be from a single contributor, we refer the reader to Woerner et al., 2020) [26].

The list of detected GVPs shared between all replicates of RG_0.1%_37°C_16hr and their SAP locations in the protein and peptide, as well as allele frequencies are provided in Table S4. In addition, based on our Gene Ontology analysis, the affected proteins mostly fall under the categories of structural, catalytic, and binding proteins (Fig. S4).

Overall, based on our results with shed skin samples, RapiGest dissolution with overnight trypsin digestion at 37 °C, which resulted in more consistent and complete digestion with the fewest missed cleavages, is an optimal method for downstream analysis of GVPs, supporting further development of this technology in forensic applications. We should highlight that our proteomics workflow can be extended to human fingerprints, based on our previous utilization of a similar but unoptimized, proof of concept workflow used to analyze protein markers collected from human fingerprints (Schulte, et al. 2021). [13]

4. Conclusion

We evaluated protein extraction and in-solution digestion conditions (surfactant type, surfactant concentration, digestion duration, and digestion temperature) for bottom-up proteomics analysis of human skin samples. Our results suggest that the optimal lysis buffer was 0.1% RapiGest. In addition, it was determined that overnight digestion at 37 °C was the optimal digestion condition resulting in the greatest digestion efficiency with the fewest missed cleavages. Our optimal workflow for bottom-up proteomic analyses of skin samples is

straightforward and is MS-compatible, with minimal sample loss from additional sample clean-up steps. Finally, the use of our optimized proteomics workflow on a mixed-donor skin sample resulted in the identification of a high number of informative peptides associated with known SNPs in the human genome and therefore serves as a proof-of-concept approach for future GVP identification in forensic skin-proteomics investigations.

Funding

This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA) (https://www.iarpa.gov), via contract number 2018–18041000003. Research was performed in the Comprehensive Cancer Center Proteomic Shared Resource, housed in The Ohio State University Campus Chemical Instrument Center, and funded by the National Institutes of Health under grant number P30 CA016058. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. The funders reviewed and approved the manuscript for publication but had no role in study design, data collection and analysis, or preparation of the manuscript.

Data availability

Data underlying this study have been deposited in the PRoteomics IDEntifications (PRIDE) database via ProteomeXchange with identifier PXD022720 [25]. All other relevant data are within the paper.

Disclosures

MA Freitas a co-owner, co-founder, and Chief Scientific Officer of MassMatrix Inc., a for profit bioinformatics company focusing on developing commercial software for multi-omics analysis.

This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA) (https://www.iarpa.gov), via contract number 2018–18041000003. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. The funders reviewed and approved the manuscript for publication but had no role in study design, data collection and analysis, or preparation of the manuscript.

Data availability

Data underlying this study have been deposited in the PRoteomics IDEntifications (PRIDE) database via ProteomeXchange with identifier PXD022720

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jprot.2021.104360.

References

[1] M. Visser, D. Zubakov, K.N. Ballantyne, M. Kayser, MRNA-based skin identification for forensic applications, Int. J. Legal Med. 125 (2011) 253–263, https://doi.org/10.1007/s00414-010-0545-2.

[2] E.D. Merkley, Introduction to forensic proteomics, ACS Symp. Ser. 1339 (2019) 1–8, https://doi.org/10.1021/bk-2019-1339.ch001.

[3] C.S. Ang, J. Rothacker, H. Patsiouras, A.W. Burgess, E.C. Nice, Murine fecal proteomics: a model system for the detection of potential biomarkers for colorectal cancer, J. Chromatogr. A 1217 (2010) 3330–3340, https://doi.org/10.1016/j.chroma.2009.10.007.

[4] L. Roewer, DNA fingerprinting in forensics: past, present, future, Investig. Genet. 4 (2013) 1–10, https://doi.org/10.1186/2041-2223-4-22.

[5] K.E. Mason, P.H. Paul, F. Chu, D.S. Anex, B.R. Hart, Development of a protein-based human identification capability from a single hair, J. Forensic Sci. 64 (2019) 1152–1159.

[6] R.D. Díaz Martín, Z. Camacho-Martínez, J.R. Ambrosio Hernández, L. Valencia-Caballero, Proteomics as a new tool in forensic sciences, Rev Esp Med Leg 45 (2019) 114–122, https://doi.org/10.1016/j.reml.2018.06.002.

[7] K.E. Mason, D. Anex, T. Grey, B. Hart, G. Parker, Protein-based forensic identification using genetically variant peptides in human bone, Forensic Sci. Int. 288 (2018) 89–96, https://doi.org/10.1016/j.forsciint.2018.04.016.

[8] F. Chu, K.E. Mason, D.S. Anex, P.H. Paul, B.R. Hart, Human identification using genetically variant peptides in biological forensic evidence, ACS Symp. Ser. 1339 (2019) 107–123, https://doi.org/10.1021/bk-2019-1339.ch007.

[9] E.D. Merkley, D.S. Wunschel, K.L. Wahl, K.H. Jarman, Applications and challenges of forensic proteomics, Forensic Sci. Int. 297 (2019) 350–363, https://doi.org/10.1016/j.forsciint.2019.01.022.

[10] Z. Zhang, M.C. Burke, W.E. Wallace, Y. Liang, S.L. Sheetlin, Y.A. Mirokhin, et al., Sensitive method for the confident identification of genetically variant peptides in human hair keratin, J. Forensic Sci. 65 (2020) 406–420, https://doi.org/10.1111/1556-4029.14229.

[11] F. Chu, K.E. Mason, D.S. Anex, A.D. Jones, B.R. Hart, Hair proteome variation at different body locations on genetically variant peptide detection for protein-based human identification, Sci. Rep. 9 (2019) 1–12, https://doi.org/10.1038/s41598-019-44007-7.

[12] D.S. LeSassier, K.Q. Schulte, T.E. Manley, A.R. Smith, M.L. Powals, N.C. Albright, et al., Artificial fingerprints for cross-comparison of forensic DNA and protein recovery methods, PLoS One 14 (2019) 1–14, https://doi.org/10.1371/journal.pone.0223170.

[13] K.Q. Schulte, F.C. Hewitt, T.E. Manley, A.J. Reed, M. Baniasad, N.C. Albright, et al., Fractionation of DNA and protein from individual latent fingerprints for forensic analysis, Forensic Sci Int Genet 50 (2021) 102405, https://doi.org/10.1016/j.fsigen.2020.102405.

[14] M. Waas, S. Bhattacharya, S. Chuppa, X. Wu, D.R. Jensen, U. Omasits, et al., Combine and conquer: surfactants, solvents, and chaotropes for robust mass spectrometry based analyses of membrane proteins, Anal. Chem. 86 (2014) 1551–1559, https://doi.org/10.1021/ac403185a.

[15] X. Zhang, L. Li, J. Mayne, Z. Ning, A. Stintzi, D. Figeys, Assessing the impact of protein extraction methods for human gut metaproteomics, J. Proteome 180 (2018) 120–127, https://doi.org/10.1016/j.jprot.2017.07.001.

[16] E. Patel, M.R. Clench, A. West, P.S. Marshall, N. Marshall, S. Francese, Alternative surfactants for improved efficiency of in situ tryptic proteolysis of fingermarks, J. Am. Soc. Mass Spectrom. 26 (2015) 862–872, https://doi.org/10.1007/s13361-015-1140-z.

[17] Y.-H. Chang, Z.R. Gregorich, A.J. Chen, L. Hwang, H. Guner, D. Yu, et al., New mass-spectrometry-compatible degradable surfactant for tissue proteomics, J. Proteome Res. 14 (2015) 1587–1599.

[18] K.R. Ludwig, M.M. Schroll, A.B. Hummon, Comparison of in-solution, FASP, and S-trap based digestion methods for bottom-up proteomic studies, J. Proteome Res. 17 (2018) 2480–2490, https://doi.org/10.1021/acs.jproteome.8b00235.

[19] M. Hailemariam, R.V. Eguez, H. Singh, S. Bekele, G. Ameni, R. Pieper, et al., S-trap, an ultrafast sample-preparation approach for shotgun proteomics, J. Proteome Res. 17 (2018) 2917–2924, https://doi.org/10.1021/acs.jproteome.8b00505.

[20] F. Wu, D. Sun, N. Wang, Y. Gong, L. Li, Comparison of surfactant-assisted shotgun methods using acid-labile surfactants and sodium dodecyl sulfate for membrane proteome analysis, Anal. Chim. Acta 698 (2011) 36–43, https://doi.org/10.1016/j.aca.2011.04.039.

[21] M. Li, M.J. Powell, T.T. Razunguzwa, G.A. O'doherty, A general approach to anionic acid-labile surfactants with tunable properties, J Org Chem 75 (2010) 6149–6153, https://doi.org/10.1021/jo100954q.

[22] Y.Q. Yu, M. Gilar, P.J. Lee, E.S.P. Bouvier, J.C. Gebler, Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins, Anal. Chem. 75 (2003) 6023–6028, https://doi.org/10.1021/ac0346196.

[23] Y.Q. Yu, M. Gilar, RapiGest SF Surfactant: An enabling tool for in-solution enzymatic protein digestions, Waters Corp, Boston, MA, 2009.

[24] E.W. Deutsch, A. Csordas, Z. Sun, A. Jarnuczak, Y. Perez-Riverol, T. Ternent, et al., The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition, Nucleic Acids Res. 45 (2017) D1100–D1106, https://doi.org/10.1093/nar/gkw936.

[25] Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D. J. Kundu, et al., The PRIDE database and related tools and resources in 2019: improving support for quantification data, Nucleic Acids Res. 47 (2019) D442–D450, https://doi.org/10.1093/nar/gky1106.

[26] A.E. Woerner, F.C. Hewitt, M.W. Gardner, M.A. Freitas, K.Q. Schulte, D. S. LeSassier, et al., An algorithm for random match probability calculation from peptide sequences, Forensic Sci Int Genet 47 (2020) 102295, https://doi.org/10.1016/j.fsigen.2020.102295.

[27] T. Mouchahoir, J.E. Schiel, Development of an LC-MS/MS peptide mapping protocol for the NISTmAb, Anal. Bioanal. Chem. 410 (2018) 2111–2126, https://doi.org/10.1007/s00216-018-0848-6.

[28] M.J. Wither, K.C. Hansen, J.A. Reisz, Mass spectrometry-based bottom-up proteomics: sample preparation, LC-MS/MS analysis, and database query strategies, Curr Protoc Protein Sci 2016 (2016) 16.4.1–16.4.20, https://doi.org/10.1002/cpps.18.
[29] K. Chandramouli, P.-Y. Qian, Proteomics: challenges, techniques and possibilities to overcome biological sample complexity, Hum Genomics Proteomics 1 (2009), https://doi.org/10.4061/2009/239204.
[30] E.I. Chen, D. Cociorva, J.L. Norris, J.R. Yates, Optimization of mass spectrometry-compatible surfactants for shotgun proteomics, J. Proteome Res. 6 (2007) 2529–2538, https://doi.org/10.1021/pr060682a.
[31] J. Norrgran, T.L. Williams, A.R. Woolfitt, M.I. Solano, J.L. Pirkle, J.R. Barr, Optimization of digestion parameters for protein quantification, Anal. Biochem. 393 (2009) 48–55, https://doi.org/10.1016/j.ab.2009.05.050.
[32] Y.Z. Zheng, M.L. DeMarco, Manipulating trypsin digestion conditions to accelerate proteolysis and simplify digestion workflows in development of protein mass spectrometric assays for the clinical laboratory, Clin Mass Spectrom 6 (2017) 1–12, https://doi.org/10.1016/j.clinms.2017.10.001.
[33] L. Tsiatsiani, A.J.R. Heck, Proteomics beyond trypsin, FEBS J. 282 (2015) 2612–2626, https://doi.org/10.1111/febs.13287.
[34] M. Šebela, T. Štosová, J. Havliš, N. Wielsch, H. Thomas, Z. Zdráhal, et al., Thermostable trypsin conjugates for high-throughput proteomics: synthesis and performance evaluation, Proteomics 6 (2006) 2959–2963, https://doi.org/10.1002/pmic.200500576.
[35] Y.Q. Yu, M. Gilar, W. Corporation, RapiGest SF Surfactant: An Enabling Tool for in-Solution Enzymatic Protein Digestions What Is RapiGest SF? Compatibility with tryptic digestion Fast proteolytic digestions. Waters Tech Note, 2002.
[36] M.J. Bailey, A.D. Hooker, C.S. Adams, S. Zhang, D.C. James, A platform for high-throughput molecular characterization of recombinant monoclonal antibodies, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 826 (2005) 177–187, https://doi.org/10.1016/j.jchromb.2005.08.021.
[37] H.Z. Huang, A. Nichols, D. Liu, Direct identification and quantification of aspartyl succinimide in an IgG2 mAb by RapiGest assisted digestion, Anal. Chem. 81 (2009) 1686–1692, https://doi.org/10.1021/ac802708s.