Microwave-assisted acid hydrolysis for whole bone proteomics and paleoproteomics

Caitlin Colleary[1], Nicole C. Little[1], Timothy P. Cleland[1*]

[1]Museum Conservation Institute, Smithsonian Institution, Suitland, MD 20746

*Corresponding author: clelandtp@si.edu

Keywords: bone proteomics, mass spectrometry, microwave, hydrolysis

Abstract

Rationale: Whole bone proteomic analyses rely on lengthy sample preparation including demineralization and digestion to break bone down into peptides to recover using mass spectrometry. However, microwave-assisted acid hydrolysis, a technique used in proteomic analyses on other soft tissues and cells, will combine both demineralization and digestion and only take minutes.

Methods: To test microwave-assisted hydrolysis on whole moose bone, we microwaved 5 concentrations of acetic and formic acid (15%, 12.5%, 10%, 7.5% and 5%) for three times (10, 20 and 30 minutes) at 140^{o} C using an ETHOS UP High Performance Microwave Digestion System. Peptides were injected and separated using Thermo BioBasic C18 columns and detected on an LTQ Orbitrap Velos mass spectrometer. We searched the raw data on PEAKS 8.5 against the white-tailed deer database.

Results: Formic acid hydrolysis led to the most complete digestion, and therefore the highest number of PSMs, more protein groups and better sequence coverage for collagenous proteins. However, in the formic acid samples there is a tradeoff with digestion completeness and a higher incidence of *in vitro* modifications (i.e., formylation) that are not induced using acetic acid. Acetic acid has greater cleavage specificity and higher sequence coverage for non-collagenous proteins.

Conclusions: Depending on the goals of analysis, there are benefits and drawbacks to using both acetic and formic acid. Overall, microwave-assisted acid hydrolysis was successful in demineralizing and digesting bone fragments to considerably speed up the preparation for bottom-up proteomic analysis.

Introduction

Proteomic analyses on whole bone often demineralize bone samples before enzymatic digestion, a process that first requires powdering samples and then an incubation period of up to several days. [1] Then, additional cleaning, digestion and concentration steps can lead to a total sample preparation time of one to two weeks before mass spectrometry. Recent studies have shown advantages to digesting demineralized bone pellet, resulting in detection of less soluble parts of the bone proteome (e.g., glycosylation).[2, 3] However, despite the advantage of digesting the solid, demineralized matrix, the demineralization process remains slow. A possible avenue to overcome the long incubation time is microwave-assisted acid hydrolysis that can both demineralize bone and hydrolyze the proteins to peptides in a short time period. [4] This chemical cleavage method has been shown to target acidic residues and consistently cleave peptides on either side of aspartic acid (Asp, D). [5, 6] This method also minimizes handling and therefore may reduce the introduction of common contaminants such as keratin. [7] Microwave-assisted hydrolysis has been used in proteomic studies on ribosomal proteins, [5] bacteria, [7] human adenovirus, [8] ovalbumin and several model peptides, [9] but has not been evaluated in bone. Results from these studies have shown that this method successfully identifies a portion of the peptidome and protein modifications with good sequence coverage. [5]

To determine the utility of microwave-assisted acid hydrolysis for whole bone proteomics and paleoproteomics, we tested 5 different acid concentrations of acetic and formic acid for three different microwave times to compare the tradeoff in the quality of proteomic data collected and reducing the necessary time for sample preparation.

Materials and Methods

Bone fragments from an American moose femur (*Alces alces*) were cut into small pieces (~8 mm^3) using end-cutting pliers. ~50mg of bone was added to PTFE reaction vessels and 1 mL of formic or acetic acid was added for hydrolysis in an ETHOS UP High Performance Microwave Digestion System. Using Canon *et al.* 2010 as a baseline for laboratory microwave parameters, we set the microwave to 140°C and then we varied the acid concentration and time as follows: five acid concentrations (15%, 12.5%, 10%, 7.5%, 5%) for both formic and acetic acid; three lengths of time (10, 20, 30 minutes).

LC-MS

Peptides were injected (single) onto a self-packed 3 cm Thermo BioBasic C18 column (75 µm i.d., 4.6 um) then separated on a 22 cm Thermo BioBasic C18 analytical column (75 µm i.d., 4.6 um) using an Ultimate 3000 UHPLC (ThermoScientific). Peptides were separated with the following gradient: 2% B (0 – 8 minutes), 2 – 55% B (8 – 98 minutes), 55 – 90% B (98 – 100 minutes), 90% B (100 – 103 minutes), 2% B (104 – 120 minutes). Buffer A was 0.1% formic acid in water, buffer B was 99.9% acetonitrile, 0.1% formic acid. Separated peptides were detected on an LTQ Orbitrap Velos mass spectrometer with the following parameters: MS 1: 60k resolution, max inject 100 ms, AGC 1E6. MS 2: top 8, 15k resolution, max inject 250 ms, 5E5, HCD at 30% NCE with 5 m/z isolation window, dynamic exclusion repeat count 3 within 30 seconds, exclusion for 180 seconds, monoisotopic precursor selection enabled, rejecting first charge state.

Data Analysis

Raw data were searched in PEAKS 8.5 against the white-tailed deer (*Odocoileus virginianus)* National Center for Biotechnology Information (NCBI) database (downloaded on 02 May 2019). Because we used acid instead of enzymatic digestion, we created "complex acid digestion" with cleavage sites after X and before DE and after DE and before X. The following parameters were used for the de novo search: 10.0 ppm parent mass error tolerance, 0.02 Da fragment mass error tolerance; variable modifications: deamidation (0.98), oxidation M (15.99), hydroxylation P (15.99), max variable 3 PTMs per peptide, 12 data refine dependencies, report 5 peptides; for Peaks DB: monoisotopic precursor mass search type, 4 maximum missed cleavages, 1 non-specific cleavage, White-tailed deer NCBI database, contaminant database, FDR estimation enabled; for Peaks PTM: 313 variable modifications, de novo score (ALC%) threshold: 15, peptide hit threshold (-10logP): 30.0, Peaks run ID: 15; For SPIDER: 0.02 fragment ion tolerance, and the Peaks default L equals I, Q equals K . All peptides were filtered at 1% FDR at the PSM level with at least 1 unique peptide.

Results and Discussion

By combining demineralization and digestion, microwave-assisted acid hydrolysis considerably shortens sample preparation time and the steps required before analysis using mass spectrometry in whole bone proteomic studies. Therefore, to assess the potential to replace time-intensive bone demineralization/extraction techniques, we evaluated a series of criteria to examine bone digestion using acid hydrolysis [1, 3].

PSM and Peptide Mass

Whereas traditional sample preparation that include powdering bone samples and separate demineralization and digestions can take days or weeks, microwave-assisted acid hydrolysis makes it possible to do preparation and mass spectrometry in the same day. After microwave hydrolysis, bone fragments in all formic acid conditions were completely solubilized, whereas bone remained in all acetic acid samples indicating incomplete hydrolysis. In traditional bone protein extraction techniques, a pellet typically remains after demineralization and solubilization [1, 3]. This incomplete digestion is reflected in the number of PSMs (acetic acid: 632.47 ± 391.11; formic acid: 1305.13 ± 138.07) and average peptide mass (acetic acid: 2075.13 Da; formic acid: 1842.32 Da). More specifically, peptide spectrum matches (PSMs) (Figure 1) were considerably higher in the samples of both acids that were microwaved for 30 minutes at higher concentrations (15%, 12.5%). In the acetic acid samples, there is a substantial drop off in the number of PSMs based on concentration and based on microwave time (Figure 1). The average number of PSMs for the acetic acid samples are 934.8 at 30 minutes, 682.4 at 20 minutes and 280.2 at 10 minutes. The formic acid samples, however, only show a marginal decrease in the number of PSMs by concentration and microwave time. After 30 minutes, the average number of PSMs was 1425.8, followed by 1368.6 for 20 minutes and 1121 for 10 minutes. The number of PSMs decreases slightly with microwave time, but concentration does not have as much of an effect as it did for the acetic acid.

As a proxy for completeness of hydrolysis, peptide mass was evaluated (i.e., larger average/median peptide masses indicate more incomplete hydrolysis) for both acids. Formic acid samples had smaller peptide masses than the acetic samples at all

microwave times reflecting the completeness of hydrolysis (Figure 2). At 30 minutes, the average peptide mass in the acetic acid samples is 2061.44 Da ± 770.72 with an average median of 1937.37 Da. The average mass in the formic acid samples after 30 minutes is 1774.05 Da ± 648.37 with an average median of 1689.02 Da. The peptide mass in the acetic acid samples remains consistent at all concentrations (15%: average 2038.72 Da, median 1928.97 Da; 12.5%: average 2032.38 Da, median 1915.99 Da; 10%: average 2030.83 Da, median 1981.95 Da; 7.5%: average 2160.17 Da, median 2017.95 Da; 5%: average 2013.95 Da, median 1931.98 Da), whereas the peptide mass increases with decreasing concentration in the formic acid samples (15%: average 1607.01 Da, median 1536.2 Da; 12.5%: average 1645.36 Da, median 1564.73 Da; 10%: average 1812.63 Da, median 1721.86 Da; 7.5%: average 1874.68 Da, mean 1793.43 Da; 5%: average 1905.59 Da, mean 1828.88 Da) (Figure 2A). In the 20 minute samples, the average peptide mass remains similar in the acetic acid samples at all concentrations (2054.24 Da ± 757.28), although larger peptides are detected. The peptide mass also remains similar in the formic acid samples at 20 minutes – the average mass across all concentrations is 1849.62 Da ± 678.96 – and shows a similar trend to the 30 minute hydrolysis time of increasing size with decreasing acid concentration (15%: average 1765.33 Da, median 1649.75 Da; 12.5%: average 1799.67 Da, mean 1662.86 Da; 10%: average 1873.22 Da, median 1805.96 Da; 7.5%: average 1917.91 Da, mean 1874.43 Da; 5%: average 1893.14 Da, mean 1811.94 Da) (Figure 2B). After 10 minutes, the average mass across all acetic acid concentrations is 2130.11 Da ± 755.411 and peptides show a decrease in size with decreasing acid concentration (Figure 2C), but considerably fewer peptides are detected (average 280.2 PSMs in 10 minute acetic acid samples, compared to average 934.8 PSMs

in 30 minute acetic acid samples). At 10 minutes in the formic acid samples (average 1902.19 Da ± 700.22), peptide mass continues to increase with decreasing acid concentration (Figure 2C), but different acid concentrations are more similar than in the longer microwave times (15%: average 1851.74 Da, mean 1773.90 Da; 12.5%: average 1884.74 Da, mean 1805.96 Da; 10%: 1904.21 Da, mean 1847.86 Da; 7.5%: average 1936.11 Da, mean 1915.99 Da; 5%: average 1963.44 Da, mean 1867.87 Da). Increasing mass with decreasing acid concentration and microwave time reflects more incomplete hydrolysis, therefore the formic acid samples, with a total average mass of 1838.65 Da ± 676.80 Da, shows more complete hydrolysis than the acetic acid samples, with a total average mass of 2070.46 Da ± 763.80 Da.

Proteins and Sequence Coverage

We also evaluated the numbers and types of proteins detected for the different treatments (Table S1-S30). Both acetic and formic acid samples have a similar average number of protein groups detected (9.4 ± 3.5 in the acetic acid samples and 11.1 ± 3.1 in the formic acid samples) (Figure 3). The acetic acid samples show a decrease in protein groups based on both concentration and microwave time (30 minutes: average 11.8 ± 4.38; 20 minutes: average 9.6 ± 2.61; 10 minutes: average 6.8 ± 1.30), whereas the formic acid samples show less of a clear pattern and on average are more similar across concentration and microwave time (30 minutes: average 13.6 ± 4.04; 20 minutes: average 10.2 ± 1.79; 10 minutes: average 9.6 ± 1.34). The overall average of the protein groups is skewed by the 30 minute formic samples, which are the highest, and the 10 minute acetic samples, which are the lowest, otherwise the numbers are similar.

Collagen I alpha 1 and alpha 2, collagen XI alpha 1 were detected in all samples. Collagen V alpha 1 was detected in all samples except for the 12.5% concentration formic acid samples at 30 and 20 minutes. Collagen II alpha 1 was only detected in the 7.5% concentration formic acid sample that was microwaved for 30 minutes. Collagen XI alpha 2 was only detected in three acetic acid samples and the 5% concentration of formic acid that was microwaved for 10 minutes. Collagen III alpha 1 was only detected in three samples (15% and 12.5% acetic acid at 10 minutes and 7.5% formic acid at 20 minutes). Several non-collagenous proteins were detected, primarily osteocalcin which was detected in all samples. Alpha 2-HS glycoprotein (AHSG) was detected in half of the acetic acid samples (all concentrations at 30 minutes, 15% and 12.5% at 20 minutes and 12.5% at 10 minutes) and was detected in all of the formic acid samples except three (15% and 12.5% at 30 minutes and 7.5% at 10 minutes). Additional non-collagenous proteins that were detected in over half of the samples include thrombospondin and biglycan.

Sequence coverage is high for collagen I alpha 1 in all of the samples (Figure 4). 30 minute acetic acid samples have an average of 68% sequence coverage (corrected to reflect mature collagen instead the lower procollagen value), 20 minute samples have an average of 58% coverage and 10 minute samples have an average of 45% coverage. These averages decrease with microwave time, but there is also a clear decrease within each microwave time based on acid concentration (Figure 4). For example, at 30 minutes, 15% acetic acid has 80% sequence coverage for COL1A1, whereas 5% acetic acid at the same time drops to 53% sequence coverage. Collagen I alpha 1 sequence coverage for the formic acid samples remains consistent across different microwave times, with an

average 78% coverage at 30 minutes, 78% at 20 minutes and 73% at 10 minutes. Additionally, sequence coverage actually increases with a decrease in concentration in the 30 minute samples, with 72% sequence coverage for COL1A1 in the 15% formic acid samples and 85% coverage at 5% concentration (Figure 4). The formic acid samples show a general trend of higher sequence coverage among the collagenous proteins detected, however acetic acid has higher sequence coverage in non-collagenous proteins such as osteocalcin (Figure 5). The average osteocalcin sequence coverage for the acetic acid samples is 68% and 62% for formic acid. 12.5% concentration of acetic acid shows particularly high sequence coverage for osteocalcin at 94%.

Cleavage Specificity and Missed Cleavages

Acid hydrolysis cleaves at acidic residues with the most likely cleavage on the c-terminal of aspartic acid (D), followed by the n-terminal of D. Cleavage on glutamic acid also occurs, but not nearly as frequently. The change in cleavage specificity in the samples is dictated by the kinetics of acid hydrolysis and the change in non-specific cleavages reflects the progression of the digestion. Therefore the shift seen in cleavage specificity (Figure 6) shows that digestion in the acetic acid samples is happening more slowly than the digestion in the formic acid samples. This suggests that formic acid digestion time could be further reduced to less than 10 minutes leading to even greater specificity. The majority of the cleavages are on the c-terminal of aspartic acid (D), followed by on the n-terminal of aspartic acid. A small proportion of the cleavages are non-specific (Table S31). In the 30 minute samples across concentrations, just over half (52.67%) of peptide cleavages are on the c-terminal of D, with 17.33% on the n-terminal

of D. In the 20 minute samples across concentrations, there are more cleavages on the c-terminal of D (59.35%) and the number of cleavages decrease on the n-terminal of D (14.95%). In the acetic acid samples, specificity increases in the 30 and 20 minute samples with decreasing acid concentration. For example, in the 30 minute samples, at 15% concentration, 39.73% of the peptides are cleaved on the c-terminal of D, whereas at 5% concentration, 63.23% of the peptides are cleaved on the c-terminal of D. This trend is also seen in the 20 minute samples (15% concentration: 50.52%; 5% concentration: 70.16%). In the 10 minute samples, the specificity remains constant (average across concentrations at the c-terminal of D is 69.8%). The 10 minute acetic acid sample at 7.5% concentration has the highest specificity amongst all of the samples, with 75.65% of the peptides having cleaved at the c-terminal of D. Non-specific cleavages decrease with decreasing microwave time (30 minutes: average 17.14% across concentrations; 20 minutes: 14.76%; 10 minutes: 8.11%).

The formic acid samples have much more non-specific cleavage. The average number of peptide cleavages at the c-terminal of D across concentrations at 30 minutes was 32.22%, 39.74% at 20 minutes and 47.32% at 10 minutes. Cleavages on the n-terminal of D across concentrations were 22.49% at 30 minutes, 19.15% at 20 minutes and 14.79% at 10 minutes. Similar to the acetic acid samples, the specificity of the cleavage in the formic acid samples increases with decreasing acid concentration. For example, in the 30 minute formic acid samples at 15% concentration, 25.68% of peptides are cleaved at the c-terminal of D, whereas at 5% concentration, 40.76% of peptides are cleaved at the c-terminal of D. Non-specific cleavage is higher in the formic acid samples, with the averages across concentrations at 35.01% after 30 minutes, 32.38%

after 20 minutes and 30.00% after 10 minutes. The majority of the specific cleavage in the formic acid samples is on aspartic acid, with slightly more cleavages on the c-terminal. Cleavage is detected less often on glutamic acid (E) in formic acid than acetic acid. The decrease in specificity with microwave time is very clear in the acetic acid samples (Figure 6), but is not the case in the formic acid samples. The cleavage specificity reflects the rate of the reaction, demonstrating that the samples in acetic acid is reacting more slowly and can therefore be used as proxy for the progression of the bone digestion.

Additionally, after normalizing to the number of PSMs, the acetic acid samples had fewer missed cleavages (Figure S1) than the formic acid samples. On average, the acetic acid samples had missed cleavages on 12% of peptides across acid concentrations. More specifically, after 30 minutes, missed cleavages were on 13% of peptides, and 12% after 20 and 10 minutes. In the formic acid samples, however, the average across concentrations was 19% missed cleavages, with 22% after 30 minutes, 19% after 20 minutes and 16% after 10 minutes. We therefore hypothesize that because acetic acid has more specific cleavage and fewer missed cleavages, there is a correlation between the two.

Induced modifications from acid hydrolysis

Formylation can occur from incubation in formic acid and has been previously observed for microwave assisted hydrolysis. [7] Consistent with previous studies, we detect formylation in all of the formic acid treated samples (Figure 7). The amount of formylation increases with increasing acid concentration. No formylation was detected

for acetic acid (Figure 7) and acetylation is present in both the formic and acetic acid samples (S2). This modification may complicate future analyses in paleoproteomics by splitting peptide signals, but accounting for it may allow for successful application of formic acid digestion in fossil remains. Dehydration was also observed for both acids (Figure S3) reflecting the hypothesized intermediates in the hydrolysis process. [6]

Conclusions

Microwave-assisted acid hydrolysis simplifies sampling and preparation to reduce processing time for bone proteomic studies from days to a few minutes. Depending on the goal of the analysis, formic acid and acetic acid have different benefits and drawbacks. Whereas acetic acid has greater cleavage specificity and fewer modifications, overall digestion is much better using formic acid at all microwave times and acid concentrations. However, acetic acid has better sequence coverage for non-collagenous proteins (e.g., osteocalcin). Formic acid digests bone better than acetic acid, leading to smaller peptides, more PSMs and better sequence coverage for collagenous proteins. Both acids have similar numbers of identifiable protein groups. Overall, bone microwaved in 10% formic acid for 20 minutes has the most PSMs and protein groups, with the best tradeoff for modifications and cleavage specificity.

Acknowledgements

This project was generously supported by Smithsonian's Museum Conservation Institute Federal Funds (CC, NCL, TPC) and Smithsonian Office of Fellowships and Internships (CC). We also thank the Smithsonian Office of Research Information Services for High Performance Computing

(Hydra) for access, Greg Streveler and the National Park Service (NPS) for providing the moose bone, and two anonymous reviewers for helpful comments on this manuscript.

Data Availability

Raw data and PEAKS.8 searches available on ftp://massive.ucsd.edu/MSV000084109.

Reviewer access: ftp://MSV000084109@massive.ucsd.edu; Username: MSV000084109

Password: @cidBone!2019

References

[1] E. R. Schroeter, C. J. DeHart, M. H. Schweitzer, P. M. Thomas, N. L. Kelleher, *PeerJ* **2016**, *4*, e2603.

[2] M. Terajima, I. Perdivara, M. Sricholpech, Y. Deguchi, N. Pleshko, K. B. Tomer, M. Yamauchi, *Journal of Biological Chemistry* **2014**, *289*, 22636.

[3] T. P. Cleland, *Journal of proteome research* **2017**, *17*, 536.

[4] J. R. Lill, E. S. Ingle, P. S. Liu, V. Pham, W. N. Sandoval, *Mass spectrometry reviews* **2007**, *26*, 657.

[5] J. Cannon, K. Lohnes, C. Wynne, Y. Wang, N. Edwards, C. Fenselau, *Journal of proteome research* **2010**, *9*, 3886.

[6] A. Li, R. C. Sowder, L. E. Henderson, S. P. Moore, D. J. Garfinkel, R. J. Fisher, *Analytical Chemistry* **2001**, *73*, 5395.

[7] L. Hua, T. Y. Low, S. K. Sze, *Proteomics* **2006**, *6*, 586.

[8] C. Fenselau, O. Laine, S. Swatkoski, *International journal of mass spectrometry* **2011**, *301*, 7.

[9] S. Swatkoski, P. Gutierrez, C. Wynne, A. Petrov, J. D. Dinman, N. Edwards, C. Fenselau, *The Journal of Proteome Research* **2008**, *7*, 579.

Figure Captions

Figure 1: Peptide spectrum matches (PSMs) for both acetic and formic acid. Acetic acid samples have fewer PSMs overall compared to formic acid samples. In the acetic acid samples, there is a trend of decreasing PSMs with decreasing acid concentration that is not present in the formic acid samples.

Figure 2: Peptide mass as a proxy for digestion. A. At 30 minutes, formic acid samples have smaller peptides than acetic acid samples and the peptides in the formic acid samples increase in size with decreasing acid concentration. Acetic acid samples are larger on average and stay similar sizes regardless of acid concentration. B. The peptides are similar sizes after 20 minutes, with the same trend of little change in acetic acid based on concentration and increasing size with decreasing concentration in the formic acid samples. C. After 10 minutes, the acetic acid samples show a decrease in peptide mass with decreasing acid concentration, whereas the formic acid samples show less variation based on concentration.

Figure 3: Protein groups detected by acid type. Formic acid samples have a higher average number of protein groups, with the highest number detected with 10% formic acid after 30 minutes. Acetic acid samples have fewer protein groups overall and show a trend of decreasing proteins with decreasing acid concentration, which is not present in the formic acid samples.

Figure 4: Collagen I alpha 1 sequence coverage. Formic acid sequence coverage is higher than the sequence coverage for acetic acid. Coverage for formic acid is more stable across concentrations, whereas sequence coverage decreases with decreasing acetic acid concentration.

Figure 5: Osteocalcin sequence coverage. On average, acetic acid has higher sequence coverage for osteocalcin, particularly 12.5% after microwaving for 30 minutes.

Figure 6: Cleavage specificity in acetic and formic acid samples. Acid cleaves on acidic residues aspartic acid (D) and glutamic acid (E), with the majority of cleavages seen on D. Cleavage specificity is higher in the acetic acid samples, with the most cleavages on the C-terminus of D. The majority of cleavages on formic acid are also on the C-terminus of D, but the formic acid samples have considerably more non-specific cleavages across concentrations and times.

Figure 7: Formylation is a modification induced by formic acid hydrolysis. It is absent in the acetic acid samples and decreases in the formic acid samples with time and acid concentration.