Proteomic Interrogation of Complex Biomedical Samples Using the Rapid Denaturing Organic Digestion (DOD) Method

Abstract

Limitations to many current aqueous-based tryptic digestion methods include lengthy digestion times and both relatively high inter- and intra-day variability for both characteristic peptides identified and sequence coverages. This report describes results from digestion of some complex biomedical samples using the rapid Denaturing Organic Digestion method (DOD), an organic solvent-modified digestion method previously optimized for targeted protein digestion. Advantages of the DOD method included a very rapid digestion only requiring inexpensive solvents and reagents generally available in the laboratory, with no requirement for specialized equipment or expensive, specialized consumables. For this study, samples of *E. coli* and murine ileum protein extracts, and K562, a mass spectrometry-compatible human protein extract and reference standard routinely used to evaluate methods, were digested. Sequence coverage and characteristic peptide identification results were compared to those from 18 and 24h conventional aqueous-based digestion methods. Across the samples tested, though the number of characteristic peptides and sequence coverages produced by the 5 min DOD method were very similar to those produced by the aqueous-based digestion methods, the specific characteristic proteins and their corresponding tryptic peptides identified following DOD method digestion included more hydrophilic and less hydrophobic species. In addition, we explored the effect of increasing digestion times with complex samples from 5 to 30 and 90 minutes for the DOD method. Increasing the digestion time to ≥30 minutes resulted in improved intra-day precision and the identification of many more peptide products than the currently used aqueous methods to which it was compared. These results suggest that the DOD organic-modified digestion method could, while markedly reducing protein digestion time, also provide more precise analysis and access to a somewhat different area of the proteome than that provided by current aqueous-based digestion methods.

Introduction

Characterizing the proteome of a biological system is essential to describing and understanding it since proteins, as a class, are the second most abundant mammalian molecule after water [1,2,3], are the most diverse of all the bio-macromolecules [4], and function as enzymes, basic structural building blocks, hormones and cytokines, transport/trafficking vehicles, and primary immune agents, et cetera. Additionally, the identification and characterization of proteins can be very important for clinical diagnostics, medical research, forensics, attribution, and remediation.

Protein identification and quantification can be difficult. Prior to the advent of mass spectrometry-based proteomics, the principal methods used to characterize proteins employed protease digestion followed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) separation or an immunoaffinity separation/detection method. These methods have been time consuming, required relatively expensive reagents that lacked the specificity needed to identify isoforms and post-translationally modified species (PTMs), and were relatively insensitive and irreproducible.

The development of electrospray ionization (ESI) [5] and matrix-assisted laser desorption/ionization (MALDI) [6,7] provided reliable, robust interfaces with mass spectrometers that could be used to measure non-volatile, high molecular weight compounds such as proteins. Traditional mass spectrometric approaches have included primarily bottom-up, middle-up/down and top-down proteomic analyses. The strength of top-down and middle-up/down proteomics as compared to bottom-up approaches has been the high specificity and sequence coverage capabilities of the methods, resulting from the high mass resolution and mass accuracy of the instruments used. In top-down approaches, intact protein ions are isolated and subsequent fragmentation occurs in a high-resolution mass spectrometer. By analyzing intact proteins, complete PTM maps and single amino acid mutations or substitutions can be identified with great accuracy [8,9]. The requirement for very high-resolution instruments like Fourier transform-ion cyclotron resonance (FT-ICR) instruments, due to the initial cost and the resources required to maintain them, is one limitation for top-down and middle-up/down analyses. Recently, higher resolution Orbitraps and time-of-flight (TOF) instruments have been developed capable of performing the analyses. High-resolution instruments are required because of the need to resolve the high charge states of these large biopolymers when ionized by ESI. Another liability associated with these approaches is the inherent relatively long instrumental analysis times, with concomitant relatively poor sensitivities. This makes the identification of intact, low abundance proteins in complex mixtures difficult with these methods.

Bottom-up analysis is the most common approach employed for protein characterization and proteomics [10]. It does not have a requirement for high resolution mass spectrometry, usually requires a much shorter analysis time, and is, in practice, more sensitive. One major liability for bottom-up proteomic analysis is that sequence coverage is often not sufficient to fully characterize identified proteins and elucidate needed information. Another major liability is that lists of peptides from digested proteins cannot be used to define translated polypeptide sequences identified in the sample, *de novo*, since they are usually compared to *in silico* translated sequences. These sequence databases often do not contain sequence modifications or anomalies commonly observed in protein biosynthesis. Additionally, the inherent irreproducibility associated with bottom-up methods, which has been attributed to the tendency for autolysis of the protease during lengthy digestion times, especially at high relative protease concentrations, makes these methods less than optimal for many applications.

Trypsin, the most common protease used in MS-based protein analysis, is a digestive pancreatic serine endopeptidase that hydrolyses peptide bonds on the carboxyl side of lysine or arginine residues, except when either is followed by a proline residue. Previously published aqueous solvent-based methods have required 18-24 h digestion times [11]. Under the conditions used previously, trypsin has been prone to autolysis which is a primary factor contributing to inter- and intra-day variability [12]. Autolysis under these conditions has been directly correlated to trypsin concentration and digestion time.

The effect of organic-modified solvents and temperature on proteolytic digestion has been previously investigated. Fink et al. [13] investigated the denaturation or unfolding of a single protein using methanol-water cryo-solvent systems. They demonstrated that both the addition of methanol (MeOH) to the solvent and elevated temperature increased protein unfolding, ultimately choosing a solvent composed of 50% MeOH as the preferred solvent system for their studies. Welinder et al. [14] studied the activity of eight proteases when digesting reduced and carboxymethylated ribonuclease, using organic solvent-modified systems. The authors investigated four organic modifiers using two digestion temperatures, 22 and 37 ºC, and 2 and 18 h incubations. Amino acid analyses were carried out using N-terminal amino acid sequencing followed by reverse phase HPLC analyses. The analyses were performed using Waters ion-exchange chromatography for separation, followed by post-column derivation with o-phthaldialdehyde in non-halide buffers. The authors concluded that organic modification reduced trypsin activity which was further reduced at the higher temperature.

In 2000 and 2001, Park et al. [15,16] investigated the effect of thermal denaturation of soluble proteins prior to digestion and found that, generally, digestion efficiency increased for single proteins and mixtures of proteins that were otherwise resistant to proteolytic digestion in aqueous solvents. Subsequently, Russell et al. investigated the use of organic-modified solvent systems for proteolytic digestion of 10 target proteins to increase method efficiency in high throughput peptide mapping [17]. Though digestion efficiency improvements were generally very modest, importantly, they discovered that along with digestion times as short as 5 minutes and solvent modification up to 80% organic, there was no need for detergents or other denaturants. Finally, in 2010, Proc et al. [18] evaluated the effects of two different organic solvents on the digestion efficiency of trypsin. Using isotopically labelled characteristic peptides specific to each protein as internal standards, solvent modifications of 40% and 20% AcN and MeOH (v:v), respectively, and a 37°C digestion temperature, they quantitatively measured the "absolute amounts" of peptides produced from the tryptic digestion of 45 clinically important plasma proteins over digestion times ranging from 0.5-23 h. They compared the results to those from an aqueous solvent digestion method employing sodium deoxycholate denaturation. To "prevent protein precipitation", the organic solvent was added to each sample following reduction and alkylation and just preceding the addition of trypsin. The proteins were categorized as "rapidly digested", "moderately digested" or "resistant to digestion" based on the results. Generally, digestion for most proteins was maximized at 4 h followed by a plateau. For a few proteins including fibrinogen γ-chain, however, digestion never plateaued throughout the 23 h and for some proteins the digestion signal decreased after 4 h. Importantly, regarding the use of AcN as a modifier, the authors stated, "a significant reduction in digestion efficiency was observed for all 45 analytes in the presence of 40% v/v acetonitrile, even though no protein precipitation was observed.".

We previously reported on the development of the Denaturing Organic Digestion (DOD) method [19], a 5-min organic solvent-modified tryptic digestion method for targeted protein analysis. The DOD method was compared to two commonly used tryptic digestion methods, the Filter Aided Sample Prep (FASP) [20] and the Flash methods [21]. The FASP method employs a

detergent to disrupt cells exposing cellular proteins to proteolytic digestion. Critical method steps occur in a 30-k filter, using 8 M urea, and a digestion time of 24 h. The Flash method, a rapid digestion method, is a column-based method that employs immuno-affinity immobilized trypsin. The DOD method employs a relatively high temperature, 60 ºC, and an organic solvent-modified system containing a relatively high amount of AcN (60% v:v). Importantly, the use of the Eppendorf Thermomixer C with a heated lid was determined to be essential for consistent method performance. Initially, it was evaluated across protein toxins including ricin, ricin A chain, ricin B chain, *Clostridium botulinum* neurotoxin A, and *Staphylococcus* Enterotoxin B as part of an effort to develop a single highly specific and sensitive rapid screening assay for proteins having a potential for weaponization. Performance measures included protein sequence coverage and the number of characteristic peptides identified. Subsequently, it was employed to digest thyroglobulin and ribonuclease A, two clinically relevant proteins for which routinely used clinical digestion methods have lacked sensitivity, specificity, and reproducibility largely due to digestion method inefficiency and lack of repeatability [22,23]. DOD method performance equaled or bettered that of the aqueous solution-based and surface-immobilized trypsin digestion methods to which it was compared, and, for most of the targeted proteins, performance was substantially improved. The authors postulated that these improvements were in large part due to precise temperature control provided by the Thermomixer C employed.

To further evaluate the utility of the DOD method tryptic digestion performance capabilities for proteomic analysis of complex samples, we digested biological replicates of three complex samples, including *E. coli* and murine ileum protein extracts, and K562, a mass spectrometry-compatible human protein extract and reference standard routinely used to evaluate digestion methods. Results were compared to those from two routinely used proteolytic digestion methods developed to generate peptides from crude cell lysates which are currently used in mass spectrometry-based proteomics. The methods were the FASP method, described previously, and the "NTU Long Digestion Method", an overnight aqueous-based tryptic digestion method employed at the time this work was conducted in the Nottingham Trent University – Clifton Campus John van Geest Cancer Research Centre proteomics laboratory for routine proteomic analysis. Finally, samples of fibrinogen γ-chain, a plasma protein categorized by Proc et. al. as "resistant to digestion" employing two organic-modified tryptic digestion methods [18], was digested using the DOD method and the NTU digestion method and peptide results compared.

Materials and Sample Preparation Equipment

The MS-compatible yeast and human protein extract protein standard reference, K562, was acquired from Promega (catalog #V6951).

FASP Method

The following materials and equipment were used: urea, iodoacetamide (Sigma Aldrich catalog #114959/0681253022), LC/MS grade water, lyophilized Promega Gold mass spectrometry-grade trypsin, NaCl, ammonium bicarbonate, sodium disulfide (SDS), Tris/HCl, dithiothreitol (DTT) (Sigma catalog #D9779), trifluoroacetic acid (TFA), Eppendorf Thermomixer C with heated lid, 1.5 mL Eppendorf snap-cap conical tubes, Speedvac, 30 k filter, and formic acid (FA).

DOD Method

The following materials and equipment were used: AcN, iodoacetamide (Sigma Aldrich catalog #114959/0681253022), LC/MS-grade water, lyophilized reagent-grade trypsin (Sigma Aldrich, catalog # T4799), DTT (Sigma Aldrich, catalog #D0632-25G), Eppendorf Thermomixer C with heated lid, 1.5 mL Eppendorf snap-cap conical tubes, Speedvac, FA, and TFA.

NTU Digestion Method

The following materials and equipment were used: AcN, iodoacetamide (Sigma Aldrich catalog #114959/0681253022), LC/MS-grade water, lyophilized Promega Gold mass spectrometry-grade trypsin, DTT (Sigma Aldrich, catalog #D0632-25G), 1.5 mL Eppendorf snap-cap conical tubes, BIOER MB-102 mixing block, ultracentrifuge, Speedvac, FA, and TFA.

Methods

FASP 24 h Digestion Method

Samples were processed by the filter-aided sample preparation (FASP) proteolytic digestion method according to the protocol described by Wizniewski, *et al*. [20]. Briefly, proteins were reduced with 20 mM DTT (prepared fresh daily) for 1 h at 60 ºC followed by alkylation with 55 mM iodoacetamide for 45 min at room temperature prior to proteolytic digestion. Samples were digested overnight at 37 ºC using 0.02 µg trypsin/µL, prepared in 50 mM aqueous ammonium bicarbonate, at approximately 1:100 trypsin to protein ratio. Following digestion, samples were centrifuged under vacuum to dryness for 40 minutes and stored at -20 ºC until analysis. Prior to analysis, samples were suspended in 100 µL of LC mobile phase.

NTU Long Digestion Method

The NTU method employs an overnight aqueous-based tryptic digestion. Septuplicate biological replicates were digested according to the following regimen. A 30-µg sample of target protein was diluted in 100 µL of 50 mM triethyl ammonium bicarbonate. Sample proteins were reduced by adding 1 µL of 0.5 M DTT (Sigma Aldrich D9779 SLBH59545) with a 20-minute incubation at 56 °C shaking at 600 rpm on a BIOER MB-102 mixing block followed by a pulse centrifugation to collect lid condensate. To alkylate sample proteins, a 2.7 µL aliquot of 0.55M iodoacetamide (Sigma Aldrich 114959/0681253022) was added, and the sample was vortexed and incubated for 15 min at room temperature in the dark. A 1 µL aliquot of 1 µg trypsin /µL was added, and the sample was thoroughly vortexed and again incubated at 37 °C overnight (16.75 hrs). Pulse centrifugation was conducted to collect lid condensate and digestion terminated by adding 0.5 % TFA to a final concentration of 0.5 % (v:v). Using the Speedvac V-AQ setting, samples were vacuum-centrifuged to dryness for 45 minutes and stored at -20°C. Prior to nanoLC/MSMS analysis, samples were thawed at room temperature, suspended in 25 µL of 10% AcN containing 0.1% FA, and shaken for 15 minutes at 5 °C. Sediment particles were removed employing 13 G centrifugation prior to transferring the supernatant to a sample vial for nanoLC/MSMS analysis.

DOD Method

Sample digestion was performed as follows (a fresh solution of 1.0 M dithiothreitol (DTT) in LC/MS-grade water was prepared prior to each digestion). A master mix was formulated consisting of 10-70 µg of targeted protein, 22 µL of 1.0 M DTT (final concentration 20 mM), 660 µL of AcN (60% v:v), and 396 µL LC-MS grade water for a total sample volume of 1100 µL. A 1 µg trypsin/µL LC/MS-grade water solution was prepared fresh daily. Aliquots (50 µL or 450-3180 ng of protein) of the master mix were transferred to separate 1.5 mL Eppendorf snap-cap tubes. Digestion was initiated by addition of trypsin at a 1:3 molar ratio to target protein(s). Vial caps were affixed, and samples were incubated at 60 ºC at 700 rpm in an Eppendorf ThermoMixer C equipped with a heated lid for 5, 30 or 90 minutes. Digestion was terminated by the addition of 1 µL of 50% TFA to each tube. Samples were dried under vacuum for 30 minutes in a Rotovap

under medium heat until dry and stored at -20 ºC until analysis. Samples were then suspended in LC/MSMS mobile phase prior to nanoLC/MSMS analysis.

Note that the DOD digestions involve much higher trypsin to protein ratios: 1:3 as opposed to the 1:100 or 1:50 in conventional digestions. However, the DOD digestions can be undertaken using low cost, reagent grade trypsin instead of the much more expensive grades normally used in protein digestions for LCMSMS analysis. Therefore, the DOD digestions are more cost effective, in terms of the consumables costs, as well as taking less time.

nano-LC/MSMS Analysis

LC/MSMS analyses of the *E. coli* and murine samples were performed on an Orbitrap LC-MSMS. Digest products were suspended in 50 µL of 95/5/0.1 % H_2O/AcN/Formic Acid (v:v:v) as described above, loaded on a 100 µm id x 20 mm x 5 µm 200 Å AQC18 trap column, and separated by reversed-phase nanoLC, on a 75 µm id x 18 cm x 5 µm 100 Å AQC18 analytical column, using 0.1% formic acid in water and 0.1% formic acid in AcN as mobile phases A and B, respectively. The nanoLC was interfaced with a Thermo Fisher Scientific Fusion Orbitrap mass spectrometer or a Q Exactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer. Mass spectrometric data were collected using a 3 sec, top-speed data-dependent acquisition (DDA) method in which full MS scans were collected at a resolving power of 120,000, over a 350-1500 m/z range. Peptide precursors were selected and subjected to collision-induced dissociation (CID), and MSMS spectra were collected from the ion trap, employing a rapid scan rate. The data were processed using MaxQuant software (version 1.5.3.30). Spectra were searched against a trypsin cleavage database for each respective proteome. MaxQuant parameters used to process the data were set to the default with a missed cleavage setting of 2, variable modifications of oxidation on methionine, and fixed modification of carbamidomethylation on cysteine. A 0.5 Da window was used to isolate product ions, and full scan spectra were matched employing a mass tolerance of 10 ppm. Peptide identification was performed with a false discovery rate of 1%.

For the K562 NTU long digestion experiments, 30 µg samples of target protein were analysed on a Sciex QTOF high resolution instrument. A sample was brought to 100 µL in 50 mM tetraethylammonium bromide. Each sample was subsequently reduced and alkylated by adding 1 µL of DTT followed by a 20-minute incubation at 56°C shaking at 600 rpm. Samples were briefly centrifuged to collect lid condensation, and 2.7 µL of 0.55 M iodoacetamide was added. Samples were then vortexed and incubated at room temperature for 15 minutes in darkness. A 1µL aliquot of 1µg/µL trypsin was added to each, and samples were thoroughly vortexed and incubated for 1.5 hours at 47°C (Condition A/medium); or, for 16.7 hours at 37°C overnight (Condition B/long), shaking at 600 rpm. Condensate was collected with a pulse spin, and digestion was terminated by adding TFA to a final concentration of 0.5%. Samples were vacuum-dried for 45 minutes in a Speedvac at 60°C using the "V-AQ" setting, visually checking at 20-minute intervals, then, frozen at -20°C. They were then resuspended in 25 µL of 0.1% FA in 10% AcN and shaken for 15 minutes at 5°C followed by the addition of 25 µL of 0.1% FA and a further 15-minute shaking at 5°C. Sediment particles were removed by a 13 G centrifugation prior to transferring the supernatant from each sample to LC vials for LC/MSMS analysis. Triplicate biological replicates were digested and analysed in triplicate by nanoLC/MSMS on either a SCIEX 5600 or 6600 Triple TOF instrument. The data were processed using Protein Pilot v5.0.3 to identify proteins and evaluate the resulting sequence coverages and characteristic peptide fragments. Data was extracted from the "Protein Summary" and "Peptide Summary" pages in Protein Pilot. For identified proteins with a false discovery rate <1%, the % sequence coverage

and number of characteristic peptides resulting from the DOD and NTU digestion methods were compared. For characteristic peptides, the peptide sequence coverages and peak intensities (areas) were also compared.

Results

E. coli Whole Cell Lysate Digestion

The DOD method was first applied to a protein extract from an *E. coli* whole cell lysate to evaluate optimum digestion time using a complex sample. Without alkylation, a total of 568 characteristic proteins were successfully identified using a 200-ng sample. Replicate 200-ng samples were digested for 5, 15, 30 and 60 minutes, and the data are listed in Table 1. Across digestion times, the number of characteristic peptides and proteins identified, and the proteome sequence coverages, were very consistent at ca. 475 proteins and 12% sequence coverage, respectively. Protein identification was based on the identification of ≥2 characteristic peptides.

Digestion time	5 min	15 min	30 min	60 min
Peptides	3007	3127	3161	2783
Protein identifications	484	496	505	466
Proteome coverage (%)	12.1	12.4	12.6	11.7

Table 1. The effect of digestion time on the number of characteristic tryptic proteins identified with corresponding peptides, and the average proteomic sequence coverage obtained from *E. coli* tryptic digests prepared using the DOD method.

Mouse Ileum Protein Extract Digestion

Subsequently, digestion efficiency of the DOD tryptic digestion method was compared to the commonly used 24 h FASP method using a second highly complex sample – whole mouse ileum protein extract. Initially, quadruplicate biological replicates were digested using the DOD method employing 5-, 15-, 30- and 60-min digestion times; analyses were performed on a Q Exactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer. The total number of protein groups and proteins identified (≥2 characteristic peptides identified) was very consistent across digestion times. In addition, the number of proteins tentatively identified through the identification of only one characteristic peptide was also very consistent across digestion times as was the average number of proteins identified with ≥2 characteristic peptides. Interestingly, the average number of peptides identified per sample seemed to drop with a digestion time of 60 minutes. Quadruplicate biological replicates were then digested using the FASP 24 h method; analyses were also performed on a Q Exactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer. Data from the experiments are listed in Table 2. When compared with the FASP digestions, the DOD digestion produced somewhat fewer protein and peptide identifications (Figure 1). The number of proteins (with ≥2 characteristic peptides identified) common to all DOD method samples across digestion times was 585 with 737 proteins identified across all samples as compared to 607 proteins common to all FASP samples and 801 proteins across quadruplicate biological replicates using the FASP 24 h digestion method.

Identification	5 min DOD	15 min DOD	30 min DOD	60 min DOD	24 h FASP
Total Proteins	660	665	717	663	1008

Total proteins ≥ 2 peptides	590	595	650	569	801
Total single hit proteins	70	70	67	74	207
Average # of proteins ≥ 2 peptides	546	537	552	497	731
Average No. of Peptides	1956	1809	1842	1595	3077
Average sequence coverage (%)	6.3	5.6	5.9	4.8	9.6

Table 2: Comparison of the total number of proteins and peptides identified and their respective sequence coverages from a tryptic digest of 400 ng of mouse ileum protein extract employing 5-, 15-, 30-, and 60-min digestion times using the DOD digestion method and with quadruplicate biological replicates employing the 24 h FASP tryptic digest with alkylation.

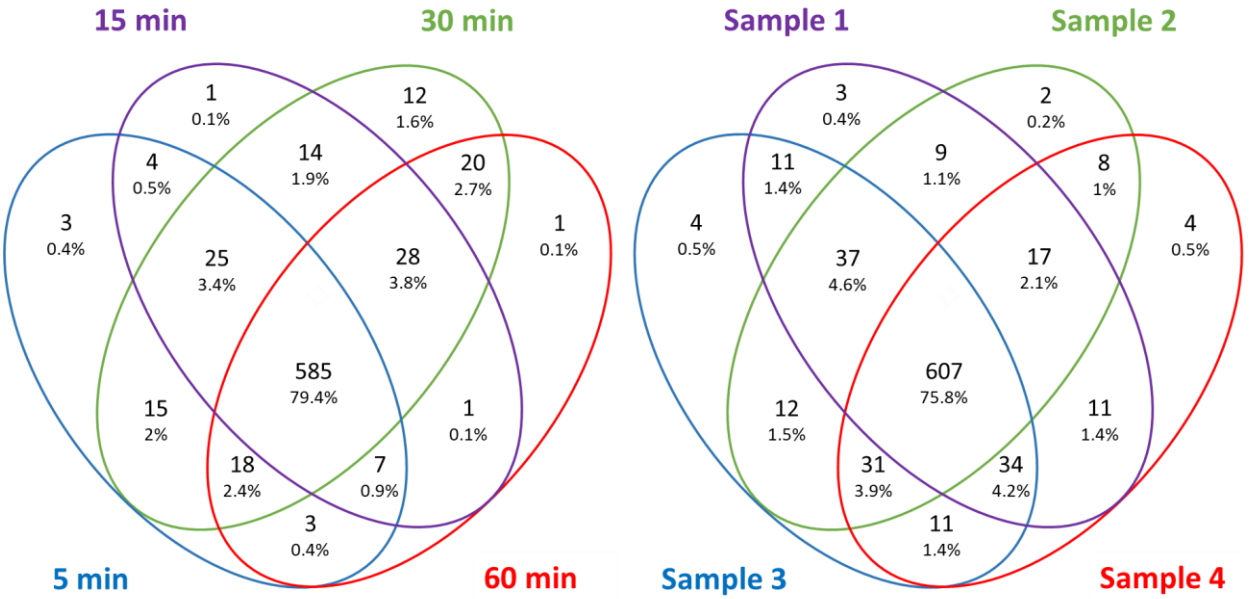

Figure 1: Venn diagrams of the number of mouse ileum proteins identified employing the 5-, 15-, 30-, and 60-min DOD tryptic digestion methods (left panel) and four biological replicates using the 24 h FASP method (right panel).

K562 Human Protein Digestion

The 5-minute DOD digestion method was then used to digest K562, a mass spectrometry-compatible human protein extract and reference standard routinely used to evaluate digestion method performance. Results were compared to those from an overnight aqueous-based tryptic digestion method routinely employed in the Nottingham Trent University (NTU) - John van Geest Cancer Research Centre proteomics laboratory for routine proteomic analysis. Protein sequence coverages, number of characteristic peptides identified, and peptide LC-MSMS peak areas were compared. Proteins and peptides were determined as detected only if they were found in all replicates. The resulting protein sequence coverages and numbers of characteristic peptides were similar (Figure 2) for the two methods with the longer method providing a somewhat greater number of characteristic peptide identifications but with less statistical certainty, as determined by

the precision of peptide LC-MSMS peak area measures. There was substantial overlap in common proteins identified (175), but also significant numbers of unique proteins; with the 5-minute DOD method identifying 115 unique proteins and the NTU method identifying 144 unique proteins. The number of proteins identified when results were combined (434) provided a substantial increase when compared to use of either single method alone (290 and 315) for the DOD and NTU methods, respectively).

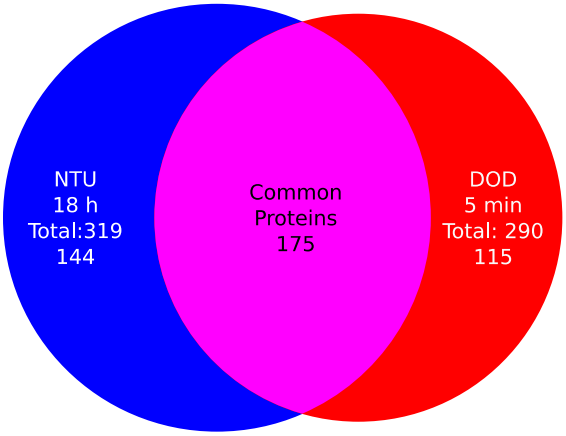

Figure 2. The DOD 5 min and NTU 18 h tryptic digestion method protein profiles for the human protein reference standard K562 (the top number in each area is the total number of proteins identified, and the lower number is the number of proteins identified unique to the digestion method).

For 3 biological replicates, each analysed three times, the overall results are shown in Figure 3. In the data analysis, as described above, proteins and peptides were only classed as detected if they were detected in all replicates for a particular protocol; peptides that were only detected sporadically were ignored.

The results presented in Figure 3 are shown versus the protein or peptide rank because sorting the data in this way makes the trends easier to visualize. Rankings of proteins or peptides are calculated with regards to the other metric being displayed in that graph (i.e. if the metric is peptide intensity then the most intense peptide gets ranked 1, next most intense as 2 etc; if the metric is %RSD, then the best %RSD gets ranked as 1). A greater number of proteins was identified using the 18-h NTU digestion method (315 vs 290), but for those proteins identified, the DOD 5 min digestion generally provided slightly more sequence coverage (Figure 3-A); similarly, for proteins identified in common across the two methods, sequence coverages for the DOD 5 min method are slightly superior for the majority of proteins, only losing out once the sequence coverage drops below 20% (Figure 3-B). The number of peptides per protein identified for the lower ranked proteins, however, was relatively low. Despite the 5-minute DOD digestion method brevity, there were very similar numbers of characteristic peptides (Figure 3-C). However, the replicate peptide LC-MSMS mean peak areas were lower for the 5-minute DOD method (Figure 3-D).

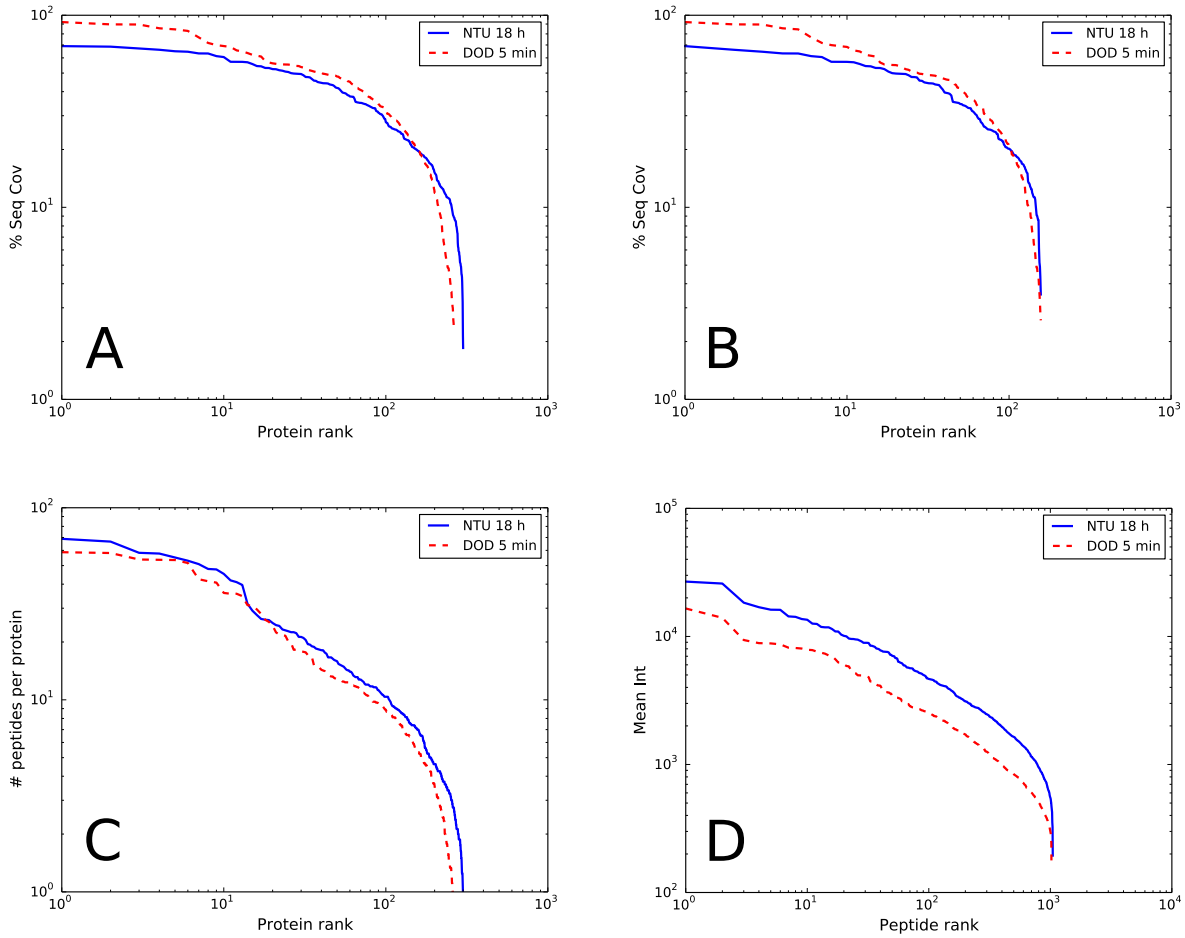

Figure 3. K562 DOD 5-min and NTU 18-h digestion results comparisons for (A) sequence coverages for all proteins identified by each method, (B) sequence coverages for proteins commonly identified across methods, (C) number of characteristic peptides identified per protein, and (D) mean peak intensities of unique characteristic peptides identified per protein.

Though the total numbers of proteins identified, and the sequence coverages were similar for the two methods, LC/MSMS peak area precision (%RSD) was better using the NTU digestion method. (Figure 4). The distribution of peptide lengths was similar but the 5-minute DOD digestion method produced (~20%) longer tryptic peptides, on average.

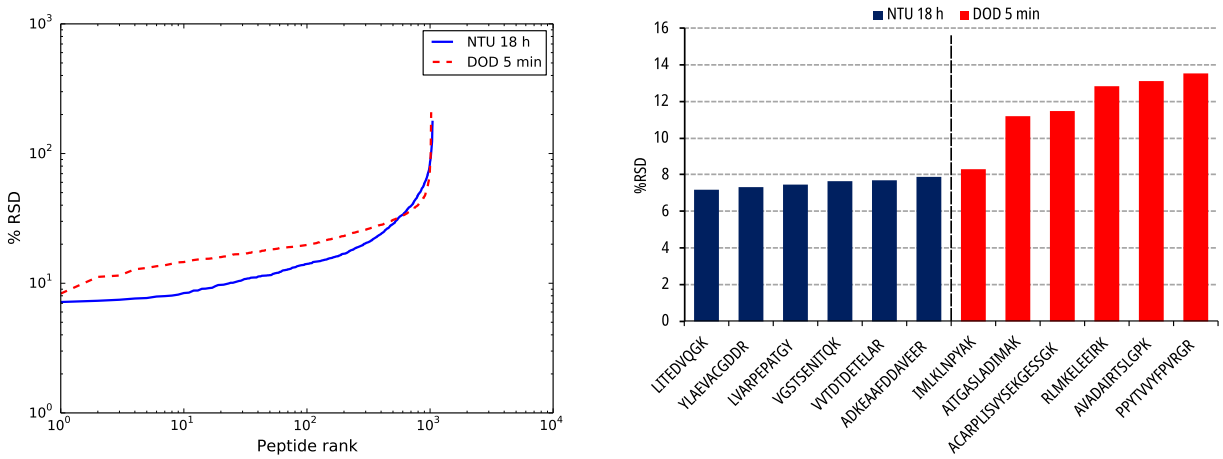

Figure 4. DOD 5-min and NTU 18-h LC/MSMS peak area precision (%RSD for peptide products) for the most precise data sets for each digestion method.

The proteins identified with the two methods were further characterized according to hydrophobicity based on their sequences and the method of Abraham and Leo (Abraham, D.J. and Leo, A.J. (1987), Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients. Proteins: Structure, Function, and Bioinformatics, 2: 130-152. https://doi.org/10.1002/prot.340020207), as is also used by https://web.expasy.org/protscale/. The hydrophobicity profile for proteins identified by each method (Figure 5) illustrated that, though there was considerable overlap in the specific proteins digested, the two methods did not digest the same set of proteins. The proteins unique to the DOD method tended to be slightly more hydrophilic proteins and those for the NTU method more hydrophobic, with the distribution of proteins that were seen by both methods sitting in the middle. In addition, tryptic peptide hydrophobicity profiles from each method were also determined. The distribution of the peptides generated by the DOD method tryptic peptides was also slightly more hydrophilic in character than peptides generated from the NTU method. We therefore conclude that combining the DOD method in parallel with more conventional aqueous digestions might provide even broader coverage of the proteins in complex samples than analysis by either individual method, and while the underlying mechanism is not yet fully understood, the extra information may still be valuable.

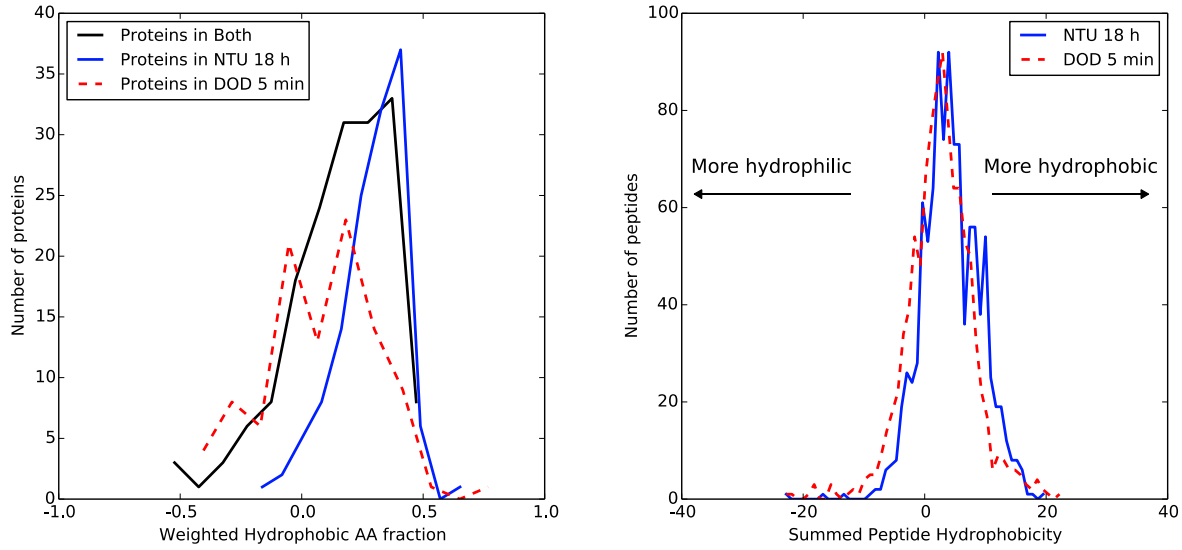

Figure 5. Relative hydrophobicity of proteins digested by the DOD and NTU methods.

The experiments were repeated with K562 increasing the 5-minute DOD method digestion time to 90 minutes - 18x longer than the original digestion time, but still ≥12x shorter than 18h conventional digestion methods (Figure 6). The 90-minute digestion time shows similar performance to the DOD 5 min and conventional digestion in terms of the sequence coverage of the detected proteins (Figure 6-A). But, the DOD 90 min digestion shows the best overall performance when considering the combination of the number of proteins detected (NTU 18h: 319 proteins, DOD 90 min: 330 proteins and DOD 5 min: 290 proteins), number of peptides per protein (Figure 6-B) and quantification precision (%RSD – Figure 6-C).

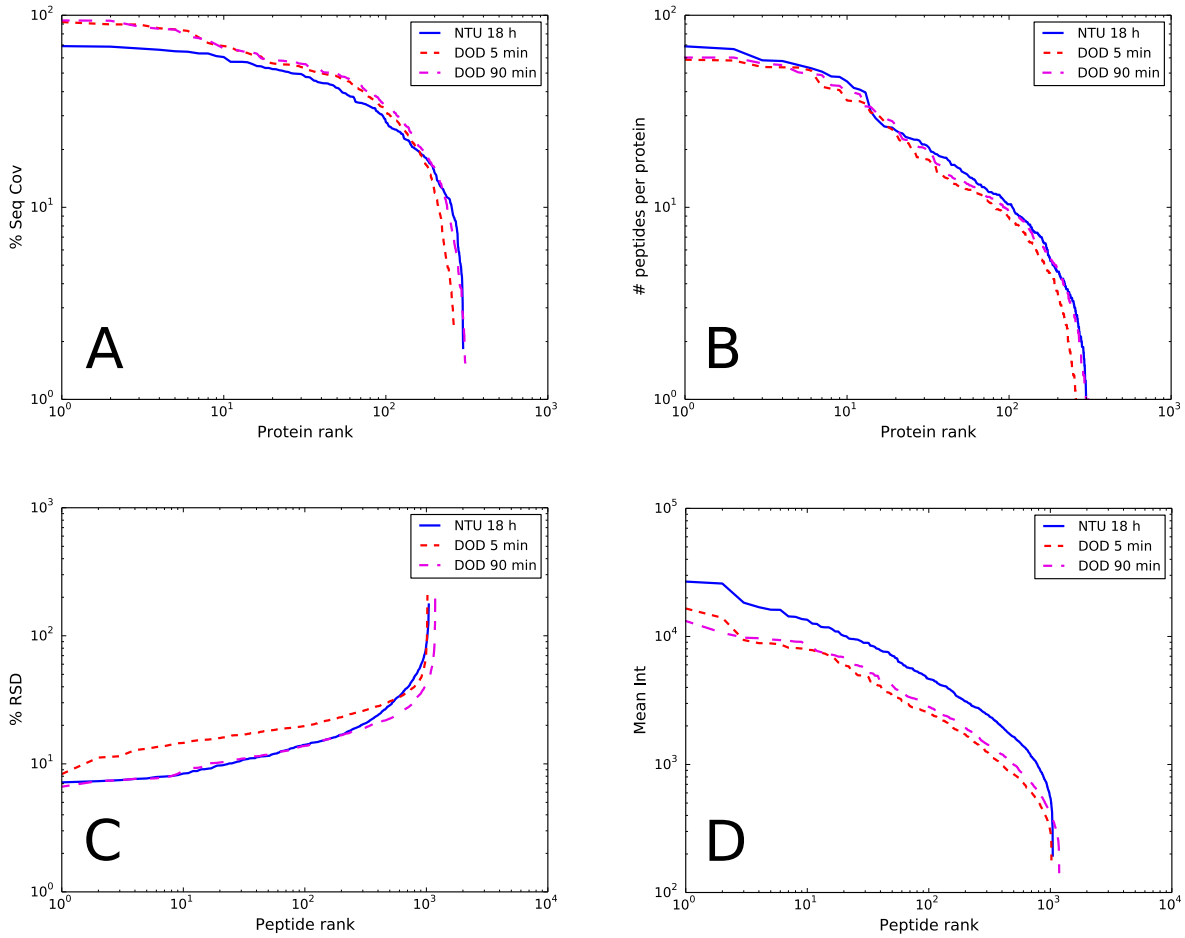

Figure 6. K562 DOD 90 min, 5 min; and 18-h NTU digestion results comparisons for (A) Sequence coverages for all proteins identified by the different methods, (B) number of peptides per protein, (C) the %RSD reproducibility of peptides detected by each method, (D) mean peak areas of unique characteristic peptides identified for each method,.

Figure 7 illustrates select peptide peak area precision differences when comparing the data for the most precisely measured peptides with the 5 and 90 min DOD, and NTU 18-h digestion methods.

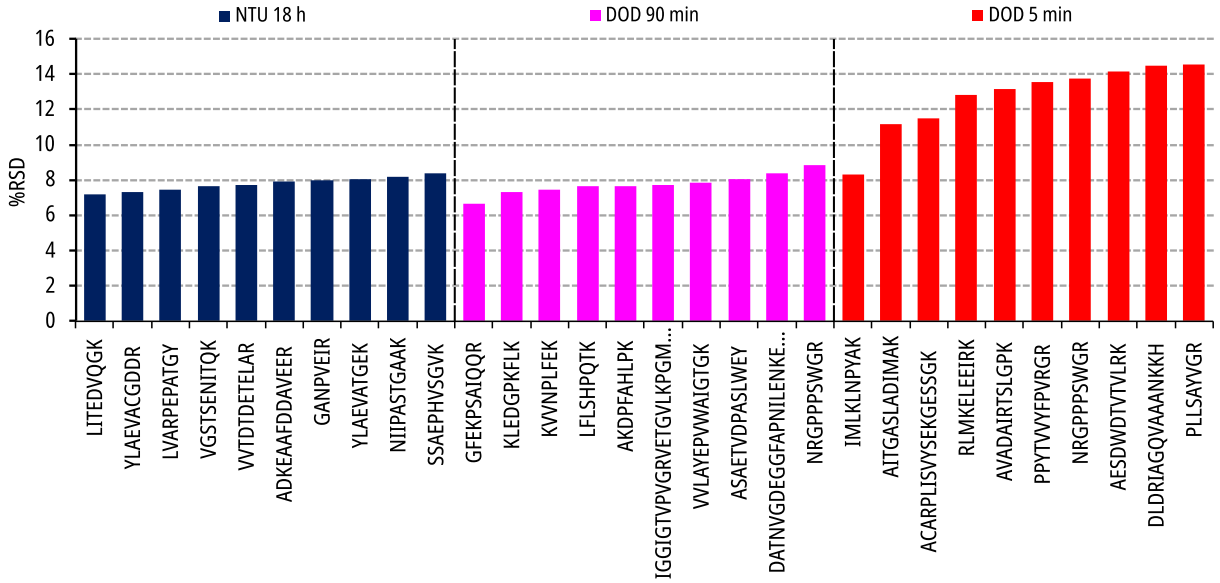

Figure 7. Comparison of the 10 most precise peptide LC-MSMS peak area data sets from digestion of K562 for the 18-h NTU , 90-min DOD and 5-min DOD digestions.

Finally, 3 biological replicates of fibrinogen were analysed in triplicate. Fibrinogen is composed of 3 component proteins: the α, β and γ chains. This protein was selected because the fibrinogen α, β and γ chains were previously studied by Proc et al [18]. In part of that study, the authors were testing the utility of high solvent digestion methods for improving bottom-up digestions; however, the temperatures used were physiological (37 °C) and the trypsin to protein ratios were intermediate (1:20). Under those conditions, Proc et al concluded that high solvent digestion was unlikely to be generally applicable as they identified several proteins that were resistant to solvent assisted digestion . They highlighted fibrinogen γ chain as a particularly difficult case. In the same study, they also tested the α and β chains and identified that these were proteins that did benefit from solvent assisted digestion. Therefore, fibrinogen is a useful test sample for characterizing such digestions as it contains a key target of interest (the γ chain) and two controls (the α & β chains). We aimed to test whether the additional higher temperature and high trypsin loading conditions of the DOD digestion protocol might render improved results for digestion of fibrinogen γ chain.

The results of digesting fibrinogen are shown in Figure 8. As with the other analytes reported previously[19] or here, longer DOD digestions (90 min in this case) exceed the analytical precision of the conventional digestion, even for fibrinogen γ chain. If the aim was number of peptides, then the 90 min DOD digestion provides 45% more fragments for fibrinogen α and similar values for β & γ compared to the conventional, 18 hour digestion. The DOD 90 min digestion does not quite match the sequence coverage achieved by the conventional 18 hour protocol, but it does still provide useful coverage, coming at only 11%, 9% and 6% below the sequence coverage for fibrinogen α, β, & γ respectively. For the shorter digestions (5 min and 30 min) the DOD digestions perform broadly similarly to the conventional 18 hour digestion in terms of the %RSD and hence, for targeted methods, it would be possible to achieve similar levels of precision from the 5 min DOD digestion as for the conventional 18 hour digestion for fibrinogen α & β.

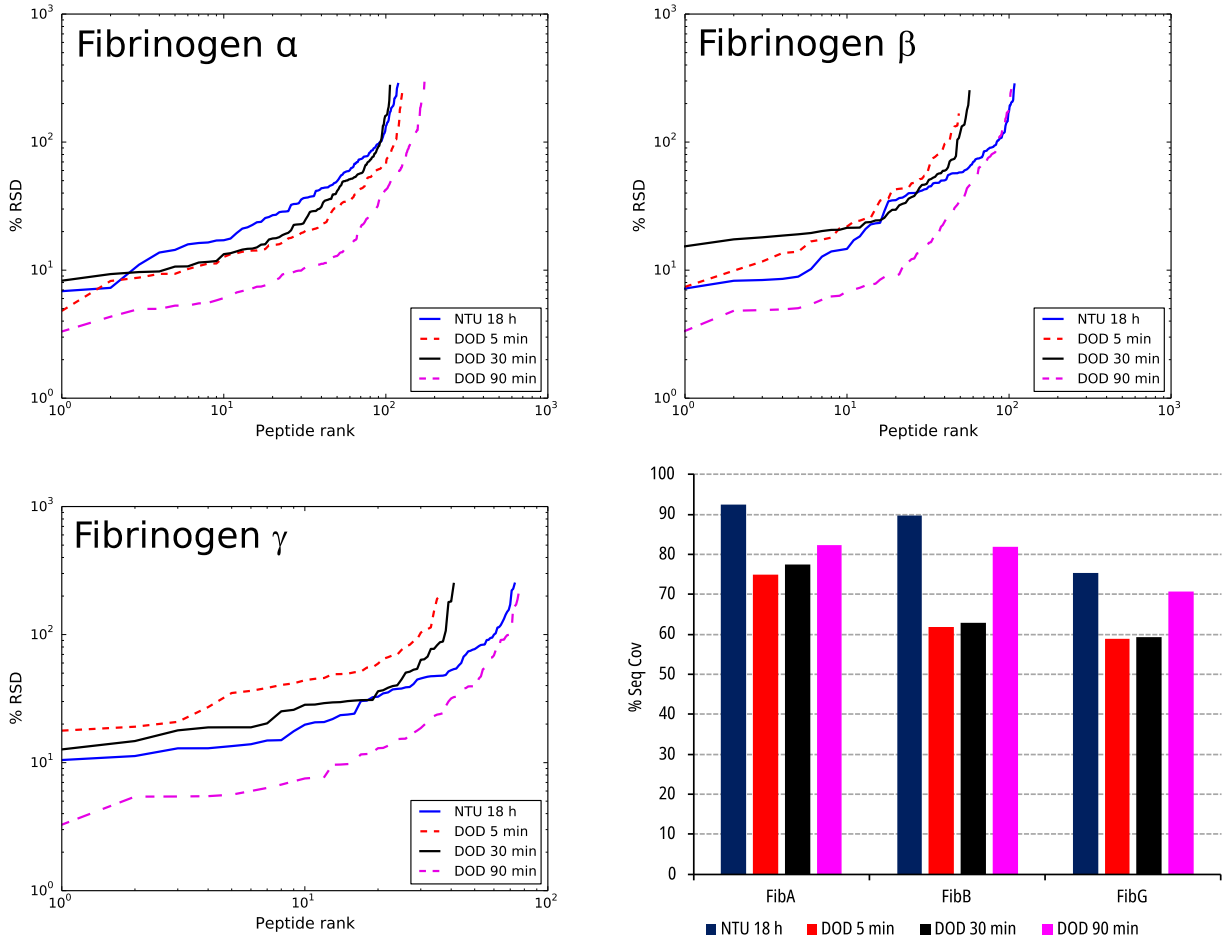

Figure 8. Comparison of the number of peptides seen in every biological replicate produced from digesting fibrinogen α, β and γ-chains, using the 18-h NTU, and the 5-, 30-, and 90-minute DOD tryptic digest methods for human fibrinogen.

Discussion and Conclusion

To date, we have performed bottom-up proteomic analyses using the new, very rapid, simple, and inexpensive organic solvent-modified DOD tryptic digestion method, over a relatively wide range of targeted proteins and with complex samples. Results were compared to several aqueous-based solvent tryptic digestion methods currently in use. Experimental data indicated that the new organic solvent-modified DOD tryptic digestion method offers key performance advantages over conventional aqueous-based protein digestion methods in terms of much shorter digestion times and improved reproducibility, and the new approach generally offers improved sequence coverages as well. As it only uses conventional consumables and widely available hardware, and, in our tests, performs well with cost effective, reagent grade trypsin, it could be easily and inexpensively adopted in laboratories performing targeted protein and proteomic analyses. With its short digest times, capability for identifying relatively high numbers of characteristic tryptic peptides and proteins, consistently relative high sequence coverage capabilities, and relatively good precision, the DOD method could substantially affect not only the analysis of target proteins

of interest but could also be used to great advantage in general proteomics and in the rapid identification of proteins in complex biomedical samples.

As described, we successfully optimized and validated the new rapid digestion method for complex samples (*E.coli* lysate, murine lung and ileum protein extracts, and K562) with proteolytic digestion times as short as 5 minutes. The DOD method was much simpler and substantially reduced digestion time as compared to traditional methods that require hours to more than a day to sufficiently digest targeted proteins. Additionally, in most cases, the DOD method also provided equivalent or better sequence coverages as compared to other methods using a single protease, provided access to a somewhat different area of the proteome, and, importantly, provided substantially better precision across results.

The data indicated that for applications such as clinical and forensic diagnostics, in which the goal is simply the identification or quantification of specific targeted proteins, the 5-minute digestion time would likely be sufficient.

For bottom-up proteomics, in which the goal is the full characterization of component proteins in complex biomedical samples, use of longer digestion times up to 90 minutes or more may be required to optimize method performance. Across the complex biomedical samples subjected to proteomic analysis, especially using digestion times of up to 90-minutes, DOD method performance was generally better in terms of number of proteins identified, the sequence coverages for said proteins, and the number of characteristic peptides produced. In addition, there was improved protein and peptide LC-MSMS peak area precision providing higher confidence in the results as compared to that of the aqueous-based methods we investigated.

DOD digestion, using a relatively high organic solvent content, also tended to digest slightly more hydrophilic proteins and produce slightly more hydrophilic peptides as compared to aqueous-based tryptic digestion methods. This resulted in the identification of a different subset of proteins than the aqueous-based digestion methods, when digesting complex samples, though the effect is relatively minor. The production and identification of more hydrophilic peptides could indicate that in an organic environment and/or a higher temperature environment, different areas of the protein had been accessed.

For complex biomedical samples, if the goal is to identify and fully characterize each individual protein and identify the maximum number of proteins within the proteome, use of the DOD tryptic digestion method in parallel with an aqueous-based method would likely result in the greatest number of proteins identified per sample along with better sequence coverages for those proteins than using any single proteolytic digestion method.

Since we limited our experiments to using the previously published protease-to-sample ratio, a future area to explore would be to increase the amount of trypsin used per sample with complex samples to determine if digestion efficiency was limited due to protease saturation. In addition, digestion pH could also be manipulated to determine whether resulting target protein tertiary structure changes would change the access of the protease to different protein domains. This could result in changes to the protein profile and sequence coverage changes for the proteins identified.

Acknowledgements

This research was made possible by funding provided by the Defense Threat Reduction Agency − Joint Science and Technology Office for Chemical and Biological Defense. Conclusions and opinions presented here are those of the authors and are not the official policy of the U.S. Department of Defense. Information in this report is cleared for public release and distribution is unlimited.

References

[1] Molecular Biology of the Cell. 4th edition. B. Alberts, A. Johnson, J. Lewis, New York: Garland Science; 2002.

[2] J.R. Freitas, A. Robert. (1999). Nanomedicine. Landes Bioscience. Tables 3–1 & 3–2. ISBN 1-57059-680-8.

[3] D. Voet, J.G. Voet. (2004). Biochemistry, Vol 1, 3rd ed. Wiley: Hoboken, NJ.

[4] The Cell: A Molecular Approach. 2nd Ed, G.M. Cooper (editor), Sinauer Associates: 2000.

[5] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse. Electrospray Ionization for Mass Spectrometry of Large Biomolecules, *Science,* 6 Oct 1989, Vol. 246, Issue 4926, pp. 64-71, DOI: 10.1126/science.2675315.

[6] M. Karas, D. Bachmann, F. Hillenkamp. Influence of the Wavelength in High-Irradiance Ultraviolet Laser Desorption Mass Spectrometry of Organic Molecules. *Anal Chem,* 1985, 57 (14), pp. 2935-2939. *doi*:*10.1021/ac00291a042.*

[7] K. Tanaka, H. Waki, U. Ido, S. Akita, Y. Yoshida, T. Yoshida, T. Matsuo. Protein and Polymer Analyses Up to *m/z* 100000 by Laser Ionization Time-of-Flight Mass Spectrometry. *Rapid Commun Mass Spectrom,* 1988, Vol 2, Issue 8, pp. 151-153.

[8] A. Armirotti, G. Damonte. Achievements and Perspective of Top-Down Proteomics, *Proteomics*, 2010 Oct, 10(20), pp. 3566-76.

[9] B. Boqdanov, R.D. Smith. Proteomics by FTICR Mass Spectrometry: Top Down and Bottom Up. *Mass Spectrom Rev*, 2005 Mar-Apr; 24(2), pp.168-200.

[10] E. J. Dupree, M. Jayathirtha, H. Yorkey, M. Mihasan, B.A. Petre, and C.C. Darie. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field. *Proteomes*. 2020 Sep; 8(3): p. 14.

[11] N. Nagaraj, J.R. Wisniewski, T. Geiger, J. Cox, M. Kircher, J. Kelso, S. Pääbo, M. Mann. Deep Proteome and Transcriptome Mapping of a Human Cancer Cell Line. *Mol Syst Biol.*, 8 Nov 2011, 7, p. 548.

[12] T.A. Addona, S.E. Abbatiello, B. Schilling. Multi-Site Assessment of the Precision and Reproducibility of Multiple Reaction Monitoring Based Measurements of Proteins in Plasma. *Nat Biotechnol*, 2009; 27, pp. 633–641.

[13] A.L Fink and B. Painter. Thermodynamics of Protein Denaturation. III. Denaturation of Ribonuclease in Water and in Aqueous Urea and Aqueous Ethanol Mixtures. *Biochem.*, 1987, 26, pp. 1665-1671.

[14] K.G. Welinder. Generation of Peptides Suitable for Sequence Analysis by Proteolytic Cleavage in Reversed-phase High-performance Liquid Chromatography Solvents. *Anal Biochem*, 1988, 174, pp. 54-64.

[15] Z.Y. Park, D.H Russell. Thermal Denaturation: a Useful Technique in Peptide Mass Mapping. *Anal Chem*, 72, pp. 2667-2670.

[16] Z.Y. Park, D.H Russell. Identification of Individual Proteins in Complex Protein Mixtures by High-Resolution, High-Mass-Accuracy MALDI TOF-Mass Spectrometry Analysis of In-

Solution Thermal Denaturation/Enzymatic Digestion. *Anal Chem,* 2001, 73, 11, pp. 2558–2564.

[17] W.K. Russell, Z.Y. Park, D.H Russell. Proteolysis in Mixed Organic-Aqueous Solvent Systems: Applications for Peptide Mass Mapping Using Mass Spectrometry. *Anal Chem,* 2001, 73, pp. 2682-2685.

[18] J.L. Proc, M.A. Kuzyk, D.B. Hardie, J. Yang, D.S. Smith, A.M. Jackson, C.E. Parker, and C.H. Borchers. A Quantitative Study of the Effects of Chaotropic Agents, Surfactants, and Solvents on the Digestion Efficiency of Human Plasma Proteins by Trypsin. *J Proteome Res*, 2010, 9, 10, pp. 5422-5437.

[19] J.M. Oyler, B.Q. Tran and D.P.A. Kilgour. The Rapid Denaturing Organic Digestion Method for Targeted Protein Identification and Characterization. *Anal Chem*, 17 Mar 2021, 93, 12, pp. 5046-5053.

[20] J. Wisniewski and M. Mann. Universal Sample Preparation Method for Proteome Analysis. *Nature Meth*, April 2009, 6, 5, pp. 359-62.

[21] V. Shah, M.E. Lassman, Y. Chen, H. Zhou, O.F. Laterza. Achieving Efficient Digestion Faster with Flash Digest: Potential Alternative to Multi-Step Detergent Assisted In-Solution Digestion in Quantitative Proteomics Experiments. *Rapid Commun Mass Spectrom*, 30 Jan 2017, 31, 2, pp. 193- 199.

[22] A.N. Hoofnagle, and M.Y. Roth. Improving the Measurement of Serum Thyroglobulin with Mass Spectrometry. *J Clin Endocrin Metab*, April 2013, 98, 4, pp. 1343–1352.

[23] C.H. Schein. From Housekeeper to Microsurgeon: the Diagnostic and Therapeutic Potential of Ribonucleases. *Nat Biotech*, June 1997, 15, 6, pp. 529-36.