

Improvement in Speed and Reproducibility of Protein Digestion, and Peptide Quantitation Utilising Novel Sample Preparation Technology in a Full Solution Workflow

by Jon Bardsley¹, Joanne Jones¹, Matthew Styles², Phillip Humphries¹, John Griffiths³, Yvonne Connolly³, Ken Cook¹, Kevin Meyer⁴, Duncan Smith³, Tony Edge¹

¹ - Thermo Fisher Scientific, Runcorn, UK

² - Department of Biology and BioChemistry, University of Bath, UK

³ - Cancer Research UK Manchester Institute, University of Manchester, UK

⁴ - Perfinity Biosciences Inc., Indiana, USA

The analysis of proteins, whether it is in the development of the latest biopharmaceuticals or the identification of a protein biomarker within the field of proteomics, has resulted in the introduction of new workflows. Part of this workflow is to get a thorough understanding of the building blocks of the protein, specifically the series of constituent peptides. By performing a bottom up analysis, it is possible to determine the active components and thereby determine the nature or purity of the compound. This has traditionally been performed by cleaving the protein using specific enzymes at certain predetermined groups within the peptide chain that are unique to a specific protein molecule [1-5]. One such enzyme is trypsin which is commonly used to digest proteins into smaller peptides which are easier to handle both in terms of the chromatography and also the mass spectrometry [6].

However, the use of trypsin and indeed other enzymes, presents analytical scientists with many challenges such as:

- Autocatalysis of the enzyme
- Variable digestion rates for different proteins
- Time consuming processes
- Limited information on optimised methods for specific proteins

This article will investigate this problem statement by looking at data derived from traditional solution based digestion procedures and investigate a novel approach that can be used to address it. This approach will be combined into a larger workflow process that can be readily used both in the field of proteomics and also in the quality control procedures used for the development of new protein based drugs.

Introduction

The drug industry has seen a substantial change in emphasis in the development of new therapeutic drugs over the past two decades [7]. In particular the

development of protein based therapies, whether from monoclonal antibodies derived from cell lines or proteins derived from recombinant human DNA using biomolecular engineering. Unlike small molecules that formed the dominant form of new drugs developed within the pharmaceutical industry, which typically comply with the Lipinski rule of 5 [8], proteins have substantially higher molecular masses and also have a substantial number of active moieties within the amino acid chain. Characterisation of these molecules does, therefore, require strategies that are not employed in the small molecule environment. In a small molecule environment the use of NMR and mass spectrometry is commonplace to elucidate the structure of the compound; however applying these approaches to protein elucidation is technically very challenging.

It is not only the pharmaceutical industry that is interested in determining the structure of proteins, many biologists are now entering into the realm of proteomics, to obtain a better understanding of how the protein structure affects biological systems.

The approach that the biologists take is the same as that employed in the quality control laboratories of the biopharmaceutical industry. It is, therefore, important to ensure that the workflow associated with the determination of the protein structure is optimised both in terms of time and also in terms of the quality of the data that is produced.

Approaches to analysing proteins

There are two classical approaches to determining the structure of a large molecule that are routinely employed within the pharmaceutical QC laboratories, and are referred to as 'top down' and 'bottom up' analysis. The bottom up analysis looks at breaking the protein into smaller more manageable blocks, peptides, which can be analysed using mass spectrometry. One rather elegant solution often applied to generate the peptide building blocks, is to digest the protein using specific enzymes. The preferred enzyme for this is trypsin, which selectively cleaves after the amino acid residues arginine and

lysine, reducing proteins to predictable component peptides, providing the protein sequence is already known. Several of these peptides can be selected to ensure greater confidence in the identification of a specific protein. The peptides are typically analysed using high resolution mass spectrometry.

Assuming the digestion is 100% efficient, there will be a one-to-one ratio between a specific isolated peptide and the parent protein (Figure 1). Selected peptides are chosen that can only carry a small number of charges (typically two or three) meaning there is only a small distribution of charge states and thus higher sensitivity for the m/z used in the analysis. These 'signature' peptides are effectively representing the protein, and so must be chosen carefully. This presents a substantial challenge to the analytical scientists as the resultant peptide mixture becomes very complex, with 50-100 times more peaks being generated compared to the original protein mixture. It can be readily assumed that in a purity assay 10 or more proteins can exist within a sample, however for proteomic samples derived from a biological system the number of protein molecules will be substantially more, making the use of high resolution separation techniques coupled to high resolution mass spectrometry essential.

Chromatography

The development of UHPLC has meant that such complex samples can be readily analysed and an example of the power of UHPLC is given in Figure 2, where a protein has been digested and the resultant mixture has been analysed on low and high resolution columns obtained using a combination of the Acclaim™ C18 column portfolio (Thermo Scientific, Runcorn, UK) and a UHPLC system (Thermo Scientific, Germering, Germany). It is very evident that there is a greater degree of separation with the higher performance columns and monitoring the performance of the peak capacity shows it increases from 290 to 380 as the particle size decreases from 5 to 2.2 μm , while the analysis time is simultaneously reduced. In all cases the gradient was from 5 to 55% ACN with 0.5% TFA in the aqueous mobile phase. The gradient time was altered in accordance with the method optimisation in Chromeleon 6.8.

Although it is evident that the introduction of UHPLC does allow for the greater identification of signature peptides, it is routine when looking for specific proteins, using a bottom up approach, to look for a

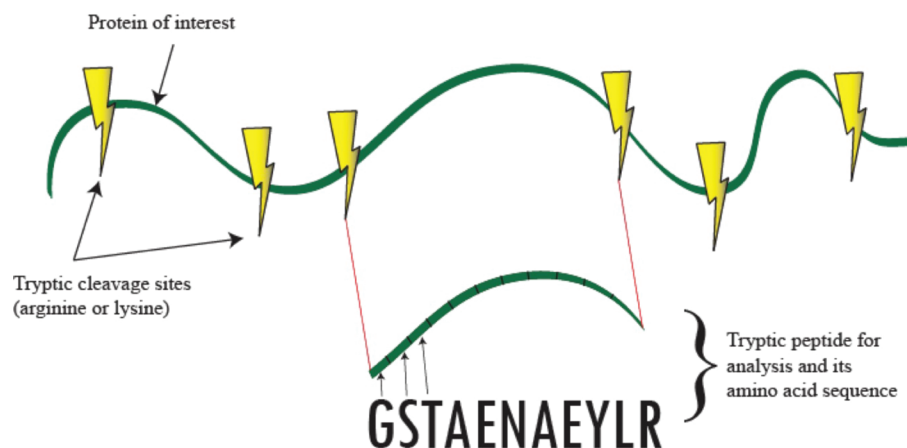


Figure 1: Schematic diagram of protein digestion

signature peptide, or a group of signature peptides, and so this does not necessarily require such a high chromatographic resolving power. It is still important to ensure that the greatest chromatographic resolution is obtained as this will ensure that the quantitative analysis does not suffer from ion suppression effects due to co-eluting components. It is also important that the nature of the preparation of the peptides needs to be robust and that this has to be able to be performed in a quantitative manner if the amount of protein is to be determined.

Digestion challenges

The analysis of proteins is clearly a complex issue and this is further complicated by the digestion step itself, where the developed protocols are time consuming and can be prone to a degree of unacceptable error. To understand why this is the case, it is necessary to understand the digestion process and the various steps that are

employed and why they are employed. Although this process is referred to as protein digestion, and to the uninitiated this would suggest that there is a single process occurring, this process of digesting a protein into constituent peptides involves several stages.

As was previously mentioned proteins are complex structures and can exist in a variety of complex shapes due to the different modes of interaction that are at play within a molecule of this size. This can result in parts of the molecule effectively becoming protected from the enzymatic digestion procedure, since there is considerable steric hindrance to overcome before the enzymatic protein can attack all of the specific parts of the protein molecule. To overcome this it is necessary to completely unfold the protein, which is achieved by denaturing, after which any other bonds which may restrict access to potential cleavage sites also need to be broken. The resulting protein is energetically unstable and from a kinetic perspective it

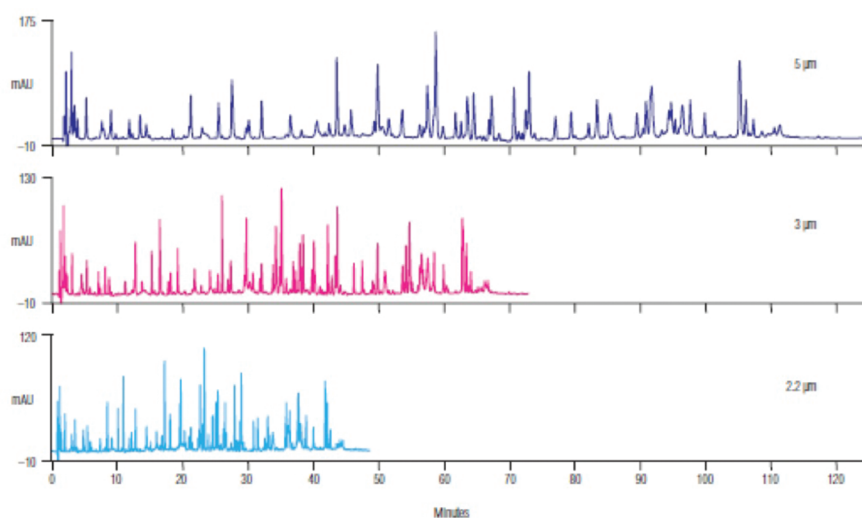


Figure 2: Importance of using UHPLC for the separation of complex mixtures such as obtained in a protein digestion.

is very easy to reform certain bonds that would result in some tertiary structure being re-introduced into the protein structure, and so certain groups are capped to stop this happening. These steps involve the addition of a wide range of reagents which need to be removed from the sample to ensure that the trypsin is not affected in an adverse manner.

There is a basic set of procedures that must be undertaken to ensure that digestion is both successful and efficient. These procedures are discussed below.

- Denaturing the protein

Proteins can be denatured by a myriad of factors, including high temperatures and the addition of chaotropic agents such as urea, guanidine hydrochloride and acetonitrile. Denaturing the protein allows the trypsin or other protease enzyme to access the whole protein backbone resulting in better peptide recovery, and higher sequence coverage upon analysis.

- Reduction of disulphide bonds

Dithiothreitol (DTT) reduces the disulphide bonds between cysteine residues without affecting other amino acids in the protein. This allows the protein to become more fully unfolded.

- Alkylation

An alkylating agent such as iodoacetic acid is added to alkylate all of the cysteine residues preventing the formation of disulphide bonds, so that the protein remains unfolded. For both the reducing and alkylating reaction steps the solvent environment should be modified to a reducing environment. This is typically achieved through the use of a buffer such as TRIS, or ammonium bicarbonate.

- Desalting

Salts and other reagents, added in the reduction and alkylating steps that may denature the enzyme need to be removed or diluted to ensure successful digestion. Some common contaminants and their threshold concentrations for trypsin functionality are listed in Table 1.

- Digestion

The final stage of the workflow is the addition of a digestion reagent, which is an enzymatic reaction, typically trypsin, although other reagents can be used but their specificity is less. Trypsin activity is highest between pH 8-9 and hence the solution is generally buffered to this pH range with TRIS or ammonium bicarbonate.

Reagents	Maximum allowable concentration
Acetonitrile	40%
Detergents	1%
DTT	20mM
Salts/buffers	250mM
Urea	1M
Guanidine hydrochloride	1M

Table 1: Maximum allowable concentrations for trypsin activity.

Solution based Approach to Digestion

Some generic protocols [9-11] can take up to one and a half days to complete, and involving the multiple steps listed previously, and then followed by evaporation and reconstitution to allow analysis by LC-MS. This adds potential for a high degree of variability, which is further extenuated since the enzyme can digest itself, which results in the production of a different enzymatic protein, which will not cleave the analyte protein at the same specific points. It is therefore possible to have a very complex peptide mixture which becomes difficult to chromatograph and almost impossible to deconvolute to the original protein. If the analytical procedure is being used for quantitative analysis, then the quantification of the original protein will invariably be incorrect.

The next part of this article will look at some of the challenges associated with the use of a solution based digestion procedure which is common in many laboratories. The robustness of this approach will be investigated as will the experimental sensitivity to variations in the experimental conditions. This will be reviewed both in terms of the absolute response for a particular peptide, but also in terms of how the selectivity of the assay varies for different signature peptides. Subsequent to this a novel approach which uses immobilised enzymes as opposed to solution enzymes will be reviewed.

Experimental investigation into stability of digestion process in a solution based system

In order to determine the sensitivity of a solution based protein digestion procedure, a series of experiments were performed to assess how each part of the work flow process affects the overall quantitative determination of a known protein. The protein that was being investigated was bovine serum albumin, BSA, which has a

very well characterised digestion process due to the availability of the specific protein. This protein produces several signature peptides, and the one that was chosen to determine the sensitivity of the assay was YLYEIAR [12].

Experimental

Albumin acetylated from bovine serum, bovine serum albumin (BSA), proteomics grade trypsin, DTT, guanidine hydrochloride, urea, iodoacetic acid, TRIS (7-9), 0.1 N Hydrochloric acid and ammonium bicarbonate were all obtained from (Fisher Scientific, Loughborough, UK).

Several stock solutions were made up prior to each digestion experiment. These were:

- 100 mM ammonium bicarbonate buffer (100 ml, pH 8).
- 6 M guanidine HCl (5 ml, made up in ammonium bicarbonate buffer).
- 200 mM DTT (1 ml, made up in ammonium bicarbonate buffer).
- 1 M iodoacetic acid (200 µl, made up in ammonium bicarbonate buffer).

The standard protocol used was as follows:

Denaturing

A 100 µl solution of 500 µg/ml of BSA was prepared in 6 M guanidine HCl and 100 mM ammonium bicarbonate. The solutions were left to denature at 22°C room temperature over 30 minutes.

Reduction

The reduction was carried out by the addition of 5 µl of DTT stock solution, vortexed, and left for 30 minutes at 37°C.

Alkylation

Alkylation was carried out by addition of 4 µl of iodoacetic acid stock solution for alkylation, vortexed and left for 30 minutes at 37°C. The alkylation reaction was then stopped by addition of 20 µl of DTT stock solution to quench the reaction and prevent alkylation of other amino acid residues.

Desalting

The solution was then diluted with 800 µl of the ammonium bicarbonate buffer to reduce the concentration of the chaotropic agent (guanidine HCl) to below 1 M.

Digestion

20 µg of lyophilised trypsin was reconstituted and activated in 20 µl of 0.01% Trifluoroacetic acid (TFA). This solution was diluted down so that the enzyme concentration was reduced to 20 µg/ml by addition of 980 µl of ammonium bicarbonate buffer. The ideal ratio of enzyme to protein

is between 1:20 – 1:50, so in this case 1000 μ l of the trypsin solution was added to the sample, at a ratio of 1:50 (m/m) and given a gentle shake to aid reconstitution. The sample was left to digest over night at 37°C.

Desalting

This second desalting process removes any salts and buffers that could potentially cause ion suppression effects in the mass spectrometer. The resulting digests were then acidified by addition of 20 μ l of 10% formic acid, so that the total concentration of formic acid in the samples was 0.2%. The samples were applied to the 3mL HyperSep™ 100 mg C18 solid phase extraction cartridges (Thermo Scientific, Runcorn, UK) for desalting under positive pressure. The loaded cartridge was washed with 3 ml of 0.1% formic acid to remove the salts and the sample was then eluted with 400 μ l 50/50/0.1 acetonitrile/water/formic acid (v/v/v). The peptides of interest all elute in lower organic concentrations than this. The extracts were dried down and reconstituted ready for analysis using LC-MS.

Results

Four steps were investigated, namely:

- Nature of the denaturing reagent, with urea and guanidine both being trialed, this included looking at the repeatability with N=3 for one set of conditions.
- Temperature and length of the denaturing step
- Reduction time
- Alkylation temperature

Denaturing reagent

Figure 3a, b shows the data obtained by using different denaturing reagents and also the degree of variability obtained from three samples processed in nominally the same manner. It can be clearly seen that in these experiments there is a great

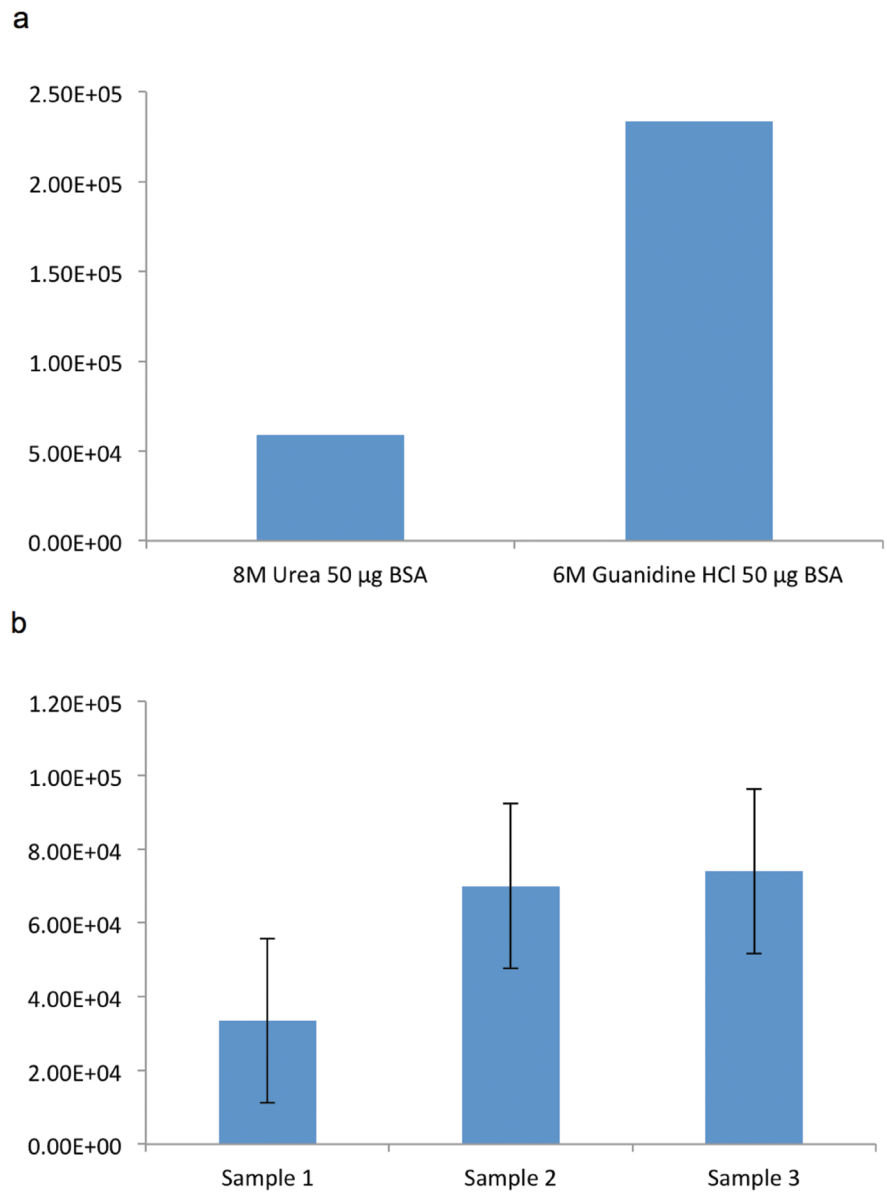


Figure 3a,b: A - Effect of varying the denaturing reagent B - Variability observed with fixed experimental condition. Based on three samples the %RSD is nearly 40%.

deal of variability both in the use of different denaturing reagent and also in terms of reproducibility under one set of experimental conditions, with the % RSD calculated to be above 35%, although it should be noted that the number of samples tested is low, N=3.

Figure 4 shows another interesting observation associated with the denaturing step. It can be seen that altering the denaturing reagent not only affects the quantitative response for a particular marker peptide, but it also changes the relative response that is observed between two

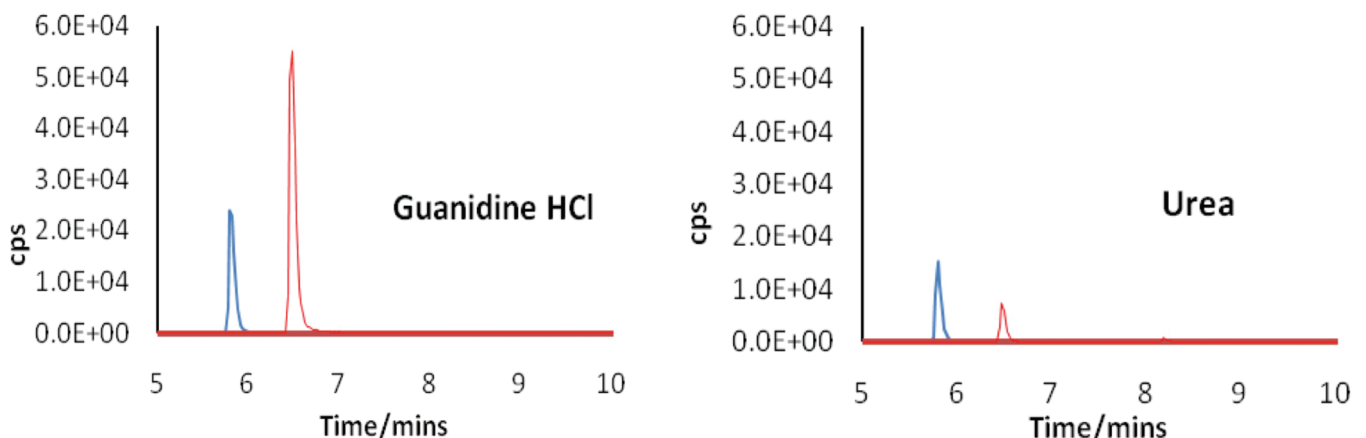


Figure 4: The chromatograms obtained using the different denaturing agents

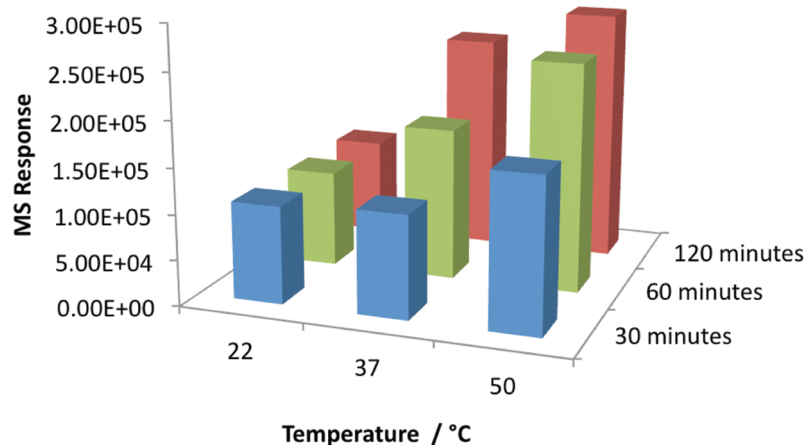


Figure 5: The effect of changing the time and temperature at which the protein is denatured on peptide recovery, using YLYEIAR as the representative peptide.

different marker peptides for the same protein.

Temperature and time

Figure 5 shows the effect of varying either the temperature of the denaturing step or the duration of that step. It can be seen that increasing the duration or increasing the temperature has a beneficial effect on the recovery of the signature peptide. Increasing the conditions from thirty minute incubation at 22°C to a two hour incubation at 50°C increases the recovery of the signature peptide by a factor of three.

Immobilised Enzyme Approach to Digestion

It is obvious then that the current approaches to protein digestion are fraught with issues, so what is the alternative? Previous approaches used to reduce the digestion time include microwave-assisted

digestion [13], the use of enzyme friendly surfactants [14,15] and immobilised trypsin [16,17]. However, a recent introduction has seen the use of temperature stable immobilised trypsin [18], referred to as SMART Digest™ (Thermo Scientific, Runcorn, UK). This has several advantages:

- Temperature is used to denature the protein, so reducing the number of solvents needed.
- The trypsin is immobilised eliminating autocatalysis.
- The use of temperature to denature the proteins means that there is less requirement to use the reduction and alkylation steps.

Experimental

In order to evaluate the robustness of the SMART Digest a comparison was made on the digestion of a cell pellet. Three approaches were employed for the

comparison, SMART Digest, overnight solvent digestion and overnight solvent digestion using solid phase extraction (SPE) to clean up the sample. Cell pellets were lysed using ice-cold SMART Digest buffer containing 0.1% w/w RapiGest™ surfactant and 2 $\mu\text{L mL}^{-1}$ benzonase. Buffer (550 μL) was added to the cell pellet on ice, which was allowed to lyse for 30 minutes with vortexing every 10 minutes. Lysate was passed several times through a 23 gauge (0.6 mm) needle to shear any remaining DNA rendering the sample suitable for accurate pipetting. 100 μL aliquots of lysate were removed for either SMART or overnight digestion. SMART Digests were carried out on a polymerase chain reaction (PCR) heater shaker set at 70°C with constant agitation at 1400 revolutions per minute (rpm). After a suitable optimised digestion time, RapiGest was degraded by the addition of 400 μL of 0.1% TFA at 37°C for 40 min. Particulates were removed by centrifugation at 16,100 rcf for 10 minutes. The supernatant was transferred to an autosampler vial for direct injection.

To allow comparison between the SMART Digest approach and a solution based approach, overnight solution digests were carried out on the PCR heater shaker set at 37°C with agitation at 500 rpm. After digestion, RapiGest was removed as above, and 200 μL supernatant was transferred to an autosampler vial for direct injection. A further 200 μL was removed for clean-up. SPE clean-up procedures as described in Villen and Gygi [19]. The resulting solutions were then analysed using UHPLC-MS.

Results

Five samples were prepared using the three different approaches. From these five different samples five replicate injections for each of the three methods were performed, resulting in 25 sets of data for each technique. The five injections per sample were combined into three concatenated files and analysed using an on-line search engine for protein identification using mass spectrometry data (e.g. Mascot search engine) with the data presented in Figure 6. The aim of the experiment was two-fold, one to determine the identification of as many of the signature peptides of the cell lysate as possible, and also to determine the technical variability within

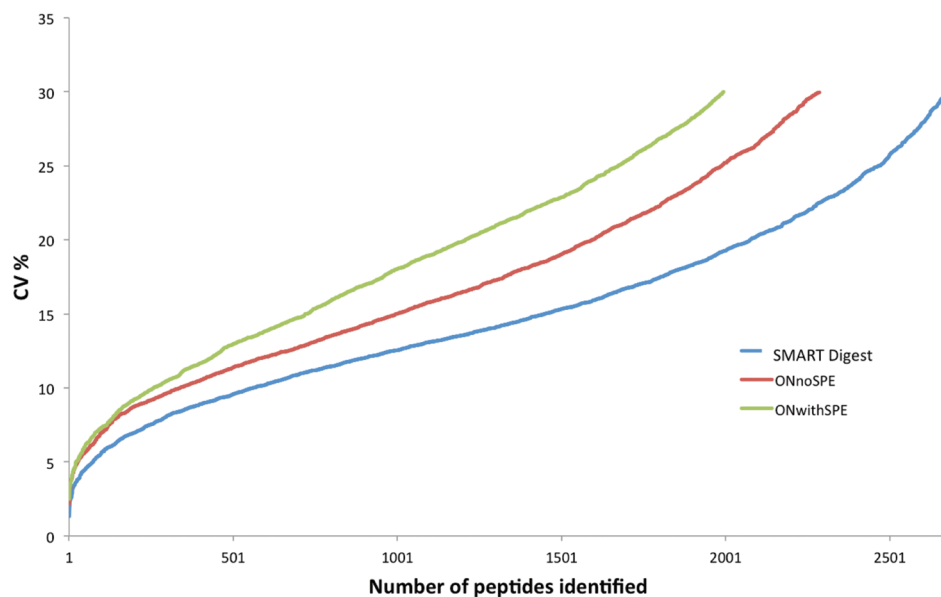


Figure 6: Comparison of three approaches to protein digestion (two based on solution reactions and one using the SMART Digest). Reprinted with kind permission from [18].

an experiment. This is an important consideration for quantitative analysis of proteins or where the identification of biomarkers is sought, where small differences in the up and down regulation of a particular protein can be of great significance.

Figure 6 contains both aspects of data required for this experiment. It is evident that the SMART Digest approach produces more peptides, and also with a reduced variability between sample sets than the solution based approaches. At 15% CV (coefficient of variation), there are approximately 1500 peptides identified using the SMART Digest method, whereas only 1000 and 700 are identified in the overnight digest and overnight digest plus SPE methods respectively. It can therefore be concluded that the SMART Digest method is more precise than the other two, and that the technical variability associated with this approach is significantly lower.

Conclusion

It has been demonstrated within this article that the current approaches to bottom up protein analysis are inherently unstable, time consuming and potentially provide inaccurate information regarding the protein structure and also the amount of protein present. A smarter way has been introduced which uses a thermal denaturing step rather than using a chemical chaotropic reagent which results in much tighter precision and much shorter analysis times. Coupling this approach with ultra high resolution chromatography and high resolution mass spectrometry provides the analyst with a much greater degree of confidence both in qualitative and quantitative assays.

It is very evident that the nature of the pharmaceutical world is changing, which coupled to the growth of proteomics from within the academic communities means that the world of chromatography is undergoing some fundamental changes in the types of molecules that are being separated, moving from the analysis of small to large molecules. The analysis of small molecules will still continue to challenge analytical scientists, but the bigger challenges will invariably be derived from the analysis of protein molecules. Currently quantitative analysis of intact proteins using chromatographic approaches has many challenges and the use of signature peptides as a marker for protein analysis has been shown to have some considerable benefits, particularly when associated with the quantitative analysis of proteins using mass spectrometry. As the analysis of proteins grows so the introduction of new workflows will aid the separation scientist perform more sensitive and more robust assays.

References

1. X. Zhang, D. Wei, Y. Yap, L. Li, S. Guo, & F. Chen, *Mass Spec. Rev.* 2007, 26, 403
2. N. Lundell & T. Schreitmuller, *Anal. Biochem.* 1999, 266, 31
3. D. F. Hunt, J. R. Yates III, J. Shabanowitz, S. Winston & C. R. Hauer, *Proc. Nati. Acad. Sci. U.S.A.*, 1986, 83, 6233.
4. A. J. Link, J. Eng, D. M. Schieltz, E. Carmack, G. J. Mize, D. R. Morris, B. M. Garvik & J. R. Yates, *Nat. Biotechnol.* 1999, 17, 676
5. G. J. Opitck & J. E. Scheffler, *Expert. Rev. Proteomics*, 2004, 1, 57.
6. S. P. Gygi, B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb & R. Aebersold, *Nat. Biotechnol.* 1999, 17, 994.
7. IMS Health, BCC Research, Pharmaceutical Technology, Biopharm International, Reuters, PhRMA
8. C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, *Adv. Drug Deliv. Rev.* 46 (1-3) (2001) 3–26
9. E.L. Kilpatrick, D.M Bunk, *Anal. Chem.* 81 (2009) 8610-8616.
10. *Proteomics: A Cold Spring Harbor Laboratory Course Manual*, A.J. Link and Joshua LaBaer. CSHL Press, Cold Spring Harbor, NY, USA, 2009.
11. <http://www.biochem.uwo.ca/wits/bmsl/protocols.html>
12. R. Cunningham, J. Wang, D. Wellner, L. Li, *J Mass Spectrom.*, 47(10) (2012) 1327–1332.
13. J.R. Lill, (2009) *Microwave Assisted Proteomics*. Royal Society of Chemistry.
14. A.R.S. Ross, P.J. Lee, D.L. Smith, J.I. Langridge, A.D. Whetton, S.J. Gaskell, *Proteomics* 2 (2002) 928-936.
15. Y.Q. Yu, M.Gilar M, P.J. Lee, E.S. Bouvier ES, J.C. Gebler, *Anal Chem* 75 (2003) 6023-6028.
16. J. Ma, J. Lui, L. Sun, L. Gao, Z. Wang, L. Zhang, Y. Zhang, *Anal. Chem.* 81 (2009) 6534-6540.
17. L. Sun, G. Zhu, X. Yan, S. Mou, N.J. Dovichi, *J Chromatogr A* 1337 (2014) 40-47.
18. R. Griffiths, Y. Connolly, K. Cook, K. Meyer, D.L. Smith, *J Anal Bioanal Tech* 5 (2014) 1-6.
19. Villén J, Gygi SP, *Nat. Protoc.* 3 (2008) 1630-1638.