

Analysis of formalin-fixed, paraffin-embedded (FFPE) tissue via proteomic techniques and misconceptions of antigen retrieval

Matthew B. O'Rourke and Matthew P. Padula

Proteomics Core Facility, University of Technology Sydney, Cnr Harris and Thomas St, Ultimo, NSW, Australia

BioTechniques 60:229-238 (May 2016) doi 10.2144/000114414

Keywords: proteomics, mass spectrometry, formaldehyde fixation, antigen retrieval

Since emerging in the late 19th century, formaldehyde fixation has become a standard method for preservation of tissues from clinical samples. The advantage of formaldehyde fixation is that fixed tissues can be stored at room temperature for decades without concern for degradation. This has led to the generation of huge tissue banks containing thousands of clinically significant samples. Here we review techniques for proteomic analysis of formalin-fixed, paraffin-embedded (FFPE) tissue samples with a specific focus on the methods used to extract and break formaldehyde crosslinks. We also discuss an error-of-interpretation associated with the technique known as “antigen retrieval.” We have discovered that this term has been mistakenly applied to two disparate molecular techniques; therefore, we argue that a terminology change is needed to ensure accurate reporting of experimental results. Finally, we suggest that more investigation is required to fully understand the process of formaldehyde fixation and its subsequent reversal.

Formaldehyde fixation is a technique that dates back to the late 19th century (1). Originally seen as a possible bactericidal agent, formaldehyde was not used for preserving tissues until some years later. A full history of formaldehyde fixation and its discoverer Ferdinand Blum can be found in Fox et al. (1). Clinical samples can be preserved in formaldehyde, embedded in paraffin wax, and stored at room temperature for decades without fear of degradation or decay. This has resulted in huge tissue libraries that contain thousands of pathologically interesting human and animal tissue samples (2). Formalin-fixed, paraffin-embedded (FFPE) tissue preservation has traditionally been the domain of histologists and immunohistochemists, however, in recent years proteomics experts have begun to see the potential wealth of information that could be discerned from preserved samples using proteomics (3).

The biggest issue in trying to extract protein from FFPE samples lies in the nature of formaldehyde crosslinking. The first attempt to determine the effect of formaldehyde crosslinking on proteins was reported by Conrat et al. in 1948 (4). From this seminal work, a number of chemical

interactions were identified, as well as the likely order in which they occur.

Formaldehyde reacts preferentially with amino groups via condensation to form aminomethylol groups. That methylol group will then condense with a guanidyl, primary amide, or secondary amine to form stable methylene bridges (Figure 1). Conrat et al. noted that at 70°C, formaldehyde will also react with guanidyl groups to form methylene groups; this was in concurrence with the finding of Middlebrook and Phillips in a 1946 study (5), who also determined that at 70°C, formaldehyde can condense with a thiol reduced from a disulfide group to form a methylene crosslink (5). Characterization of this set of reactions was not again revisited until 2004, when Metz et al. (6) decided to repeat this work using mass spectrometry (MS) to analyze the reaction products.

In their study, Metz and colleagues concluded that the following rules apply to formaldehyde crosslinking of proteins: (i) formaldehyde will form crosslinks through the creation of Schiff bases and methylol groups with a primary amino or thiol group, and (ii) these will react with other free residues such as arginine, asparagine, glutamine, histidine, tryptophan, and

tyrosine. The authors determined that two primary amine groups would not react with one another.

These conclusions follow very closely to and expand on the work of Conrat et al. (4), and together form a solid basis for understanding the diversity of possible sequence-dependent chemical modifications of proteins in a complex proteome (6). The work of Conrat et al. and Metz et al. is a prelude to the real problem facing proteomic analysis of FFPE samples, which is either the difficulty in completely reversing these formaldehyde-induced modifications or the characterization of the chemical modifications that remain after sample treatment to break crosslinks. Here we focus on the advancement of FFPE analysis techniques, including MS and associated technologies (i.e., gel based and non-gel based workflows).

Extraction of proteins from FFPE samples

1-D SDS-PAGE

One of the earliest attempts to extract protein from FFPE samples was published by Ikeda et al. in 1998 (3). Their focus was

on the ability to bind antibodies to “unlocked” proteins in order to perform Western blot analyses when searching for disease biomarkers. This approach used radio-immunoprecipitation assay (RIPA) buffer (150 mM NaCl, 1.0% IGEPAL CA-630, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris, pH 8.0) with 2% SDS and incubation at 100°C for 20 min followed by 60°C for 20 h to extract protein from whole-cell lysates of the FFPE tissue. Using this method, Ikeda et al. were able to extract 121 µg of protein from a single 5 × 5 mm tissue section. However, despite successfully binding antibodies to six target molecules at the correct molecular weight (MW) indicators, Coomassie blue-stained SDS-PAGE gels of FFPE tissue extracts did not show the same protein profiles as extracts from un-fixed, fresh-frozen tissue.

Recent attempts to extract proteins from FFPE samples have also focused on increasing binding of antibodies in either fluorescence systems or Western blotting. In 2005, Chu et al. (7) extracted protein for use in Western blot analysis via a technique dubbed non-destructive molecular extraction (NDME). NDME uses a proprietary buffer that, when coupled with heating at high temperature, increases the effectiveness of immunoblotting assays.

A similar proprietary buffer was used by Becker et al. (8) in 2007 to extract protein from FFPE tissue sections for Western blotting. The authors were able to quantitatively determine, through densitometry of bound fluorescent antibodies, that this buffer (recipe unknown) could extract proteins from FFPE tissue as efficiently as traditional methods using un-fixed samples (i.e., the level of fluorescence for both NDME fixed and non-fixed samples was the same).

Although Becker et al. (2007) and Chu et al. (2005) both used 1-D SDS-PAGE, their focus was on the specific detection of extracted proteins with Western blot analysis: 8 proteins/antibodies for Becker et al. and 12 for Chu et al. It wasn't until 2009 that Addis et al. (9), following a short report from Shi et al. (2006) (10), attempted to design a complete workflow for the proteomic analysis of FFPE samples. Addis et al. developed a protocol whereby FFPE samples of muscle tissue were incubated in 20 mM Tris-HCl, pH 8.8, 20 mM dithiothreitol (DTT) at 100°C for 20 min followed by 2 h at 80°C. Their rationale was that heating in a strong reducing environment would break the crosslinks formed by exposure to formal-

dehyde, thereby significantly improving protein recovery rates. 1-D SDS-PAGE data revealed similar protein profiles between fixed and un-fixed samples, although the intensity of bands of the same MW varied greatly. Western blot analyses also showed that antibody binding to actin, GAPDH, and E-Cadherin was improved in the FFPE tissue, such that fixed and un-fixed samples were indistinguishable. Liquid chromatography–mass spectrometry (LC-MS) analysis, however, revealed that there were still major differences between the fixed and un-fixed proteins, with 18% of detected proteins in the FFPE samples not appearing in the un-fixed replicate. The authors provided no explanation for this phenomenon; however, it was noted that due to the nature of the likely incomplete breakage of crosslinks, fragments from the most common muscle tissue peptides (i.e., actin, myosin, and tropomyosin) were found bound to other peptide fragments. This created a pseudo-background noise that had to be accounted for when performing database searches of the fragment data. Closer inspection of the Addis et al. article shows that only 10 bands were selected for analysis and comparison; thus, it is highly likely that the extra proteins detected in the FFPE samples, the identity of which were not reported, are crosslinked to the dominant cytoskeletal proteins in the excised bands. This conclusion is further supported by the observation of only one or two peptide hits per protein, indicating that they are low abundance, and only a small number of molecules are randomly crosslinked to the dominant protein in the band. The overall conclusion from this work was that, despite some continuing issues, proteins extracted from FFPE samples could be of high enough quality for proteomic analyses.

2-D PAGE

While 1-D SDS-PAGE is useful for proteomic analysis, 2-D PAGE provides a more complete proteomic profile (11). In 2008, using the work by Shi et al. (2006) (10) as a starting point, Ono et al. (12) published the first 2-D PAGE gels from an FFPE sample. Their method made special note of the necessity of a stepwise heating protocol, as mentioned previously (9,10), of 100°C for 20 min followed by 60°C for 2 h. This procedure showed a marked improvement in the apparent breakage of crosslinks.

The next attempt to analyze FFPE tissue using 2-D PAGE was made by Addis et al.

in 2009 (13), following up on their work from earlier that year (9). Here, the authors described a protocol that used their earlier method of protein extraction from FFPE tissue, involving a strong reducing environment mediated by DTT and SDS, combined with a stepwise heating method similar to that described by Shi et al. (2006) (10). The Addis et al. protocol produced successful 2-D gels with a high level of similarity to the comparative un-fixed samples but a reduction in the number of spots detected in the FFPE gel profile. This approach facilitated downstream in-gel trypsin digestion followed by either MALDI-TOF or liquid chromatography tandem mass spectrometry analysis (LC-MS/MS). MS data revealed that corresponding spots between the fixed and un-fixed gels contained the same proteins, although variations in the number of matched queries were observed but without a trend toward one sample type. Addis et al. concluded that reproducible analysis of FFPE samples using 2-D PAGE was possible with this protocol. The authors further noted that the phenomenon seen in their earlier work (9), namely peptide fragments of low-MW proteins staying attached to medium- and high-MW peptides, was still present and provided an additional challenge when searching the accompanying fragment spectra against known databases.

The next major step forward in the development of 2-D PAGE analysis of FFPE samples was the application of 2-D difference gel electrophoresis (DIGE) to FFPE sheep muscle tissue, as described by Tanca et al. in 2011 (14). DIGE is a differential display technique using fluorescent tags to show differences in separately labeled and then pooled samples separated on a single 2-D gels. Tags can then be illuminated in a fluorescence scanner to reveal a differential display, by color, of various samples.

Tanca et al. (14) used protein extracted as per their earlier work (9) to test the effectiveness of a DIGE approach using FFPE tissue. Interestingly, the method was modified to include two proprietary buffers for the actual isoelectric focusing (IEF) and 2-D PAGE parts of the protocol. First, the 2-D Clean-Up Kit (GE Healthcare) was used for protein extraction and precipitation instead of the trichloroacetic acid (TCA)/acetone method used previously. Secondly, DeStreak Rehydration Solution (GE Healthcare) was added to the IPG buffer (GE Healthcare) prior to IEF. There was no explanation as to why these changes were made to the protocol,

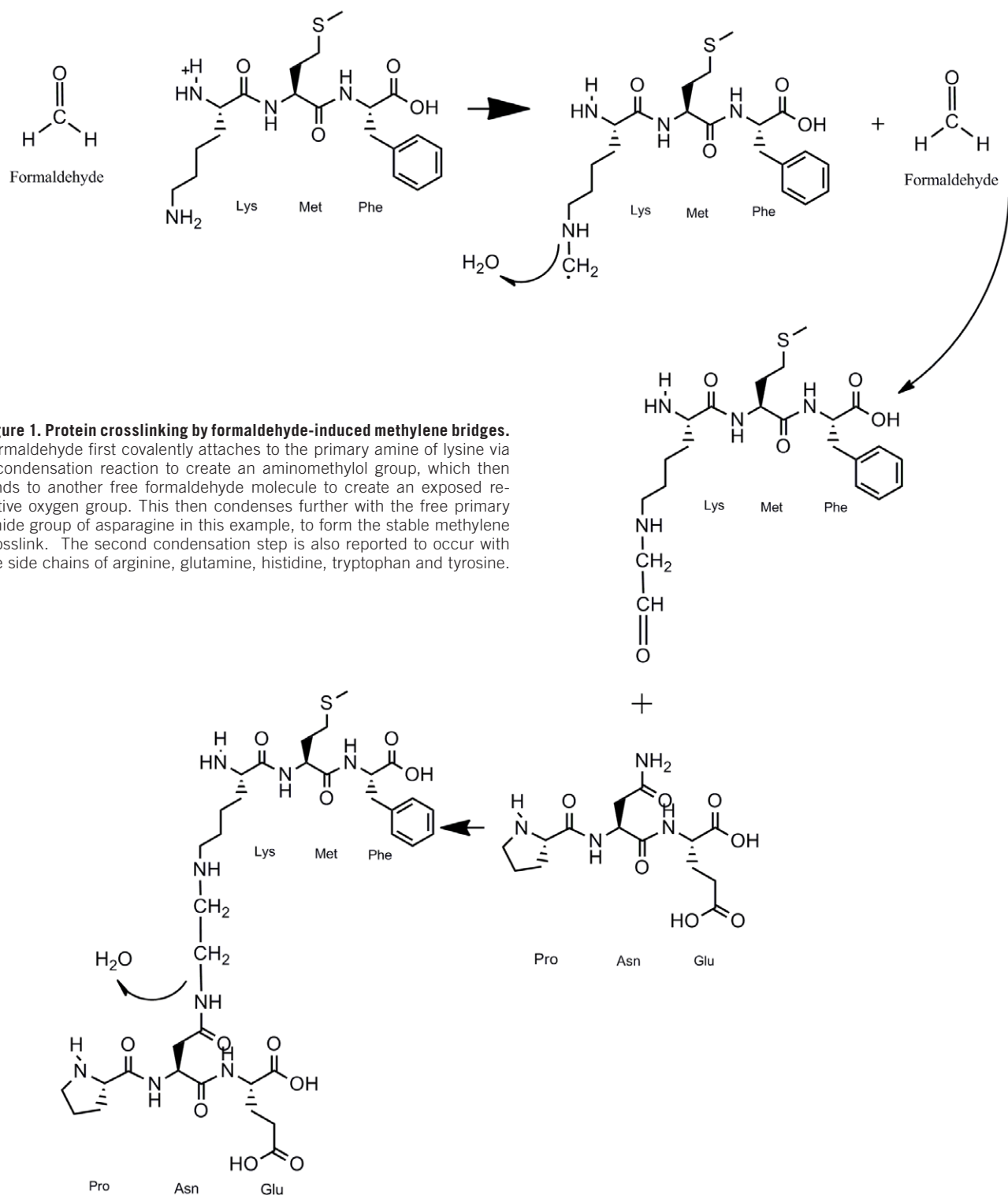


Figure 1. Protein crosslinking by formaldehyde-induced methylene bridges.

Formaldehyde first covalently attaches to the primary amine of lysine via a condensation reaction to create an aminomethylol group, which then binds to another free formaldehyde molecule to create an exposed reactive oxygen group. This then condenses further with the free primary amide group of asparagine in this example, to form the stable methylene crosslink. The second condensation step is also reported to occur with the side chains of arginine, glutamine, histidine, tryptophan and tyrosine.

nor any indication as to the chemical composition of the solutions. The results of this experiment did reveal some interesting phenomena, however. Upon first inspection of the completed DIGE gels it appeared that the FFPE and fresh-frozen samples had completely different 2-D spot profiles. When this was investigated further by excision of corresponding spots from replicate 2-D

PAGE gels, it was determined that the spot profiles were in fact highly similar, but the FFPE samples showed an isoelectric point (pI) shift toward the acidic end. A coefficient for this pI shift was determined, and when used to compare the remaining 2-D profiles of FFPE and fresh-frozen tissue, it was determined that in the low-, medium-, and medium-high- MW ranges, the profiles

were very similar. There was still a high level of difference in the high-MW range, and it was mentioned that this is likely due to high levels of modification, by formaldehyde, of high-MW proteins; greater amounts of amino groups hence have a greater potential for crosslinking and side chain modification (6). It was also mentioned that resolving this issue was a challenge that had yet to be

overcome, as it was most likely the result of the incomplete breakage of crosslinks. This was repeated in 2013 using clinical samples, and the results were comparable (15).

Shotgun LC-MS/MS

While 1-D and 2-D PAGE are common and useful proteomic analysis techniques, with next-generation instrumentation and the advent of data independent acquisition (DIA) technologies such as SWATH (16), it has become beneficial to develop whole-proteome approaches to complex sample analysis. In this section, we focus on the development of so-called “shotgun” proteomics methods for the analysis of FFPE tissue samples.

Shotgun proteomics, in this context, refers to the proteolytic cleavage of a complex protein sample by addition of an enzyme, generally trypsin, to an unfractionated whole-cell lysate in the presence of appropriate buffers. This technique is long-standing in general proteomic workflows. However, it was not until 2005 that a shotgun-based workflow for the analysis of FFPE tissue was proposed by Hood et al. (17), using human prostate cancer as a model. This protocol involved protein extraction by heating at 95°C for 90 min then cooling the sample and adding trypsin, which was then allowed to digest at 37°C overnight. Following this protocol, Hood et al. were able to identify 702 proteins from benign prostatic hyperplasia, with 159 identified with 2 or more unique peptides. The authors were also able to identify 1153 proteins from prostate carcinoma, with 282 identified with 2 or more unique peptides. It is important to note that the researchers only included truly tryptic peptides in their results, as peptides containing formaldehyde adducts or modifications could not be accurately identified. The authors concluded that they had provided the first example of a shotgun proteomics workflow for the analysis of FFPE tissue and that the issue of identifying peptides containing formaldehyde modifications needs to be addressed for this type of technique to truly be useful.

Matrix-assisted laser desorption/ionization imaging mass spectrometry

Imaging mass spectrometry (IMS) is a recently established technique that can visually display MALDI MS data. In the case of FFPE tissue, a tissue section can be analyzed via IMS to determine where individual proteins or peptides are

on a pixel-by-pixel basis. Standard IMS workflows involve coating a slide-mounted tissue section that has been either deparaffinized and digested with trypsin (18) or simply washed (19) before addition of an organic acid matrix and recrystallizing the matrix and analyte together. The matrix allows for absorption of UV energy from the laser, thus generating ions that are measured by their time of flight. The laser moves in an incremental pattern to generate ions at each X/Y coordinate on a predefined grid. These accumulated spectra can then be processed to display a spatial map of any given molecule (20). This technique has been applied for cancer biomarker discovery (21), drug research, microbiology (22), and the high-throughput analysis of tryptic digests (23). Its foremost clinical application lies in the ability to screen multiple FFPE tissue samples for specific cancer biomarkers in a fast and highly accurate way using tissue microarrays (24). Processing of FFPE tissue for use with MALDI-IMS is identical to that of shotgun approaches, except that delocalization of the proteins and peptides must be avoided. In short, the sample is deparaffinized, heated in an appropriate buffer, and then digested with trypsin (20).

Since the first reports of IMS of FFPE sections in 2005, there has been a steady and continuous history of minor improvements and modifications to what is essentially the same protocol. Table 1 summarizes the publications that have made innovative steps toward the successful development of a universal FFPE processing protocol. This table clearly demonstrates that the same basic steps are necessary to extract proteomically useful quantities of peptides from FFPE tissue, namely the deparaffinization of the tissue section with xylene or a xylene substitute followed by heating in an appropriate buffer for a specific amount of time.

Variations in sample preparation and protein extraction

While the majority of reports agree on the necessity of certain basic steps for protein extraction from FFPE samples, there exists a high degree of variation in what is considered to be the ideal set of conditions.

Buffers and additives: Extraction buffers can be easily divided into two categories: (i) proprietary/commercially available and (ii) homemade. Commercially available buffers such as liquid tissue and acid-labile surfac-

tants (e.g., RapiGest SF Surfactant from Waters Corporation) have shown reliable and reproducible results with regard to their protein extraction efficiency (3,25–27). However, due to their proprietary nature, the chemical compositions of the pre-made buffers are unknown; thus, further development or improvement is difficult.

Homemade buffers have a wide range of pH values and chemical compositions; however, the buffers that result in the highest levels of protein extraction are generally those containing SDS (28–32). SDS is a detergent commonly used in traditional proteomic workflows that assists with unfolding and solubilizing proteins. Boiling an FFPE sample in SDS is purported to help with protein unfolding as the methylene crosslinks are hydrolyzed by the high temperature; however, no direct evidence for this has been reported to date. The biggest issue with the use of SDS is its incompatibility with both electrospray and MALDI ionization (due to ion suppression), as well as chromatography. In spite of this, SDS is, ironically, the best candidate for a high-quality extraction buffer. This was first acknowledged by Wisniewski et al. in 2011 (33) and 2013 (34), and then again by Tanca et al. 2014 (32). These two teams employed filter-assisted sample preparation (FASP), which uses membrane filters to facilitate SDS/urea exchange of FFPE samples to allow further downstream processing via tryptic digestion and LC-MS/MS analysis. Here, soluble peptides produced by digestion are allowed to pass through the membrane, which retains insoluble material, while SDS is exchanged for urea. Another method for removing SDS before LC-MS/MS analysis is to dialyze the sample for a period of time (35).

Helander et al. (1994) (36) made an interesting observation: when washed for 26 days in water, almost 90% of formaldehyde crosslinks formed during 4 days of fixation could be removed. However, it should be noted that this result was never confirmed by MS analysis. The proposed mechanism involves the formation of methylene glycol with formaldehyde (36), a reaction reported to be heavily favored (1,36). Since then, there have been no further reports of water-only based protocols; we believe this approach should be revisited.

Heating: There is conjecture regarding the most appropriate temperature and duration for heating samples. The majority of studies

Table 1. The progression of shotgun proteomic analysis of formalin-fixed, paraffin-embedded (FFPE) tissue.

Reference	Year	Tissue type	Identifications from FFPE tissue	Identifications from fresh-frozen tissue	Additional notes
Crockett et al. (47)	2005	SudHl-4* human cells	514 total from trypsin and glutamic C endopeptidase (GluC)	324 (total from trypsin and GluC)	Method did not include any form of sample heating for protein extraction.
Prieto et al. (25)	2005	Colon cancer with apparent adenocarcinoma	350 unique proteins (trypsin)	No comparison provided	Method utilized a proprietary extraction buffer (Liquid Tissue) and heating at 95°C for 1.5 h prior to tryptic digestion.
Palmer-Toy et al. (28)	2005	Soft tissue and bone from inner ear canal	123 (42 with >1 peptide)(trypsin)	94 (31 with >1 peptide) (trypsin)	Protein extraction was done in a buffer consisting of 2% SDS/100 mM ammonium bicarbonate / 20 mM DTT; sample was ultrasonicated and heated at 70°C for 1 h.
Jiang et al. (29)	2007	Mouse liver	130–827 (57–470 with >1 unique peptide) (trypsin)	976 (480 with >1 unique peptide) (trypsin)	Six different methods were tried, some with heating and some without. (This paper is discussed further in the "Antigen Retrieval" section.)
Hwang et al. (26)	2007	Assorted prostate cancers in tissue microarray	402 proteins detected from 4 banks of pooled samples (trypsin)	Indicated via silver stain densitometry as equivalent to FFPE	A common protocol was used that included heating at 90°C for 30 min then incubation at 60°C for 3 h. The inclusion of RapiGest buffer resulted in a 77% increase in protein extraction efficiency.
Guo et al. (48)	2007	Glioblastoma	2845 proteins detected from sample from a single patient (trypsin)	3902 proteins detected from sample from a single patient (trypsin)	Method followed protocol from Shi et al. (10), with the inclusion of overnight dialysis of the sample in 100 mM Tris-HCl, pH 8.2 at 4°C.
Patel et al. (27)	2008	Head and neck squamous cell carcinoma (HNSCC)	391–866 for pooled patient samples (<i>n</i> = 4) (trypsin)	No comparison provided	Method utilized a proprietary extraction buffer (Liquid Tissue) and heating at 95°C for 1.5 h prior to tryptic digestion.
Nazarian et al. (49)	2008	Paediatric brainstem glioma	188 total proteins (trypsin)	No comparison provided	Method involved heating at 95°C for 30 min followed by heating at 65°C for 3 h*.
Sprung et al. (30)	2009	Colon adenoma	2838 total proteins (trypsin)	2900 total proteins (trypsin)	Method used utilized heating at 80°C for 2 h in a custom buffer** followed by ultra-sonication and heating again at 60°C for a further 2 h prior to tryptic digestion.
Balgley et al. (2)	2010	Sarcoma and leiomyoma	1593–2895 total proteins	No comparison provided	Method followed the work by Guo et al. (48) very closely. Protein extraction was performed in the same way.
Alkhas et al. (50)	2011	Endometrial cancer	1544–1720 (735–863 with >1 unique peptide)(trypsin)	No comparison provided	Samples were heated at 90°C for 1 h followed by 65°C for 2 h in 100 mM ammonium bicarbonate/20% acetonitrile.
Donadio et al. (31)	2011	Parathyroid glands	Eleven from gel spots, 28–95 for gel free methods. (70 proteins with >1 unique peptide)	No comparison provided	Samples were heated at 100°C for 20 min followed by 60°C for 2 h in 20 mM Tris, 2% SDS, 0.2 M glycine, pH 4.0.
Gamez-Pozo et al. (51)	2013	Triple negative breast cancer	1681 (1061 with >1 unique peptide) (trypsin)	No comparison provided	Samples were heated for 20 min at 100°C followed by 80°C for 2 h in 40 mM Tris-HCl, 2% SDS, pH 8.2.
Lai et al. (52)	2014	Rat liver, lung, kidney, bladder	1676 total proteins from all 4 tissue types (trypsin)	No comparison provided	Method used a novel approach of repeated suspension of the sample in 100 mM ammonium bicarbonate and heating at 90°C.
Tanca et al. (32)	2014	Liver	Filter-assisted sample preparation (FASP) = 1693, direct tissue trypsinization (DT) = 1510, and in solution digestion (ISD) = 1358 proteins	No comparison provided	Method followed the protocol previously described in Addis et al. 2009 (9) with the addition of filtration for SDS/urea exchange.

* Human transformed follicular lymphoma-derived cell line.
 ** See main text for more detailed information.
 # Buffer used is not explicitly described but likely contains 100 mM ammonium bicarbonate and 30% ACN.

suggest that a stepwise method is most effective: the sample is heated following deparaffinization at a high temperature (90°–100°C) for a short time (20–30 min) followed by 2–3 h at a lower temperature (60°–80°C) (Table 1). Heating appears to be the most important aspect of any protein extraction workflow, as the heat is necessary to allow for hydrolysis of methylene crosslinks.

It has also been suggested that the presence of a strong reducing environment, mediated by DTT, helps to reverse crosslinks (9). This assumption is not supported by the known chemistry nor is any explanation based on chemical characterization or reaction given for its use. As previously stated, methylene bridges are formed from

the linkage of amino groups via the condensation of methylol groups. To hydrolyze these bridges, any chemical additive would need to interact with the methylene group. DTT acts to reduce disulfide bonds in traditional proteomic workflows. The breakage of disulfide bonds does not have anything to do with methylene crosslinking nor does it facilitate other interactions. Further study is required to determine if and how DTT assists in the recovery of protein from FFPE samples.

Finally, in addition to buffers and heating, some studies have concluded that increasing the pressure under which sample preparation occurs increases the effectiveness of heating the sample. This idea follows a very basic principle of chemistry:

A reaction will proceed more energetically in the presence of heat and pressure. Fu et al. in 2013 (37) showed that increased pressure did increase the amount of protein extracted from FFPE tissue. However, due to the expensive and possibly dangerous nature of the pressure chambers used, it is difficult for this technique to be routinely employed in large-scale proteomic studies.

Despite variations between protocols, the single step that is universally used is the heating of samples for protein extraction. This technique, first proposed in 1986, has been dubbed heat-induced antigen retrieval (HIAR). This method of extracting proteins and uncovering antigenic sites is ubiquitous in the wider FFPE literature. But HIAR is also a term that incorrectly encom-

passes more than one discreet technique (see “Misconceptions surrounding antigen retrieval” below).

Post-acquisition processing of mass spectrometry data

As mentioned previously, formaldehyde crosslinking can produce both intra- or inter-protein linkages. Depending on the distance between reactive groups, numerous different proteoforms could be linked together multiple times, displaying a variety of different crosslinks with different molecules of the same proteoform linked in different ways. Such linkages greatly complicate the interpretation of MS data, as the heterogeneity reduces the number of molecules of a specific proteoform or peptide, reducing sensitivity. The apparent inability to reverse all of these crosslinks means that enzymatic digestion and peptide-centric shotgun LC-MS/MS-based proteomics would give the most comprehensive catalog of the proteome of the sample, as peptides without amino acids residues participating in crosslinks should be solubilized and analyzed, while many crosslinked peptides could remain insoluble. However, it is likely that individual proteins will be identified by fewer peptides than if the protein was not treated with formaldehyde for a number of reasons. First, current popular search algorithms are not able to assign an MS/MS spectrum containing two linked peptides to two sequences, assuming the crosslinked peptides are soluble and able to be analyzed in the first place. In the MS scan, the detected mass of the crosslinked peptides is the sum of the peptide masses and mass of the methylene bridge between them, while the charge state of this ion would be 4+ or higher. No algorithm has the capacity to create the search space necessary to include every tryptic peptide crosslinked to every other tryptic peptide that possesses amino acids susceptible to reacting with formaldehyde, which is necessary to create the initial list of crosslinked peptides that have the same mass as the ion detected in the MS scan. The situation is further complicated in MS/MS scans where fragments from two peptides with different sequences are present, and the algorithm is expected to assign fragment ions to the correct sequence.

A possible solution to this problem is the use of MS/MS spectral libraries and sequence tags (38). The spectral library,

routinely created for data independent analysis techniques such as SWATH (39), would need to contain MS/MS spectra for all ionizable peptides. The MS/MS spectra of the crosslinked peptides could then be compared with the library, and the library spectrums for a number of peptides with the most b- and y-ions matching would be collated into a list. These peptides would then be mathematically tested to see if the introduction of a methylene bridge causes the mass of the peptide combinations from the library to be the same as the mass selected in the MS scan. We are unaware of any software that could perform this analysis, although spectral library matching has been shown to be a reliable and extremely fast method for peptide identification in proteomics workflows (40).

Metz et al. (6) demonstrated which amino acid side chains are able to react with formaldehyde but did not extensively investigate the reaction of the attached methylene group to the side chain of another amino acid shown previously to react with formaldehyde with the exception of the bridging of lysine and arginine by formaldehyde in the same peptide. No MS/MS spectra demonstrating the fragmentation of this peptide were reported. A lack of published reports regarding which chemical residues remain on formaldehyde reactive amino acids means that it is impossible to specify the post translational modifications to be considered in a database search. Extending on the work of Metz et al. by comprehensive MS/MS analysis of model peptides subjected to formaldehyde crosslinking and attempted reversal of crosslinking would allow these modifications to be defined and therefore result in better assignment of MS/MS spectra to the correct peptide sequence.

Misconceptions surrounding antigen retrieval

Antigen retrieval is a technique, first described by Battifora et al. in 1986 (41), for increasing epitope antigenicity for studies involving antibody binding to fixed tissue. Initially, the term was used to describe the unmasking of epitopes and antigenic sites through enzymatic digestion (41), based on the theory that antigens were being retrieved from crosslinked matrixes previously thought to be irretrievable.

It was found that while tryptic digestion for an optimal amount of time did increase antibody binding efficiency to keratins,

there was little if any increase in binding to other antigens. The biggest drawback with enzymatic antigen retrieval was that its efficiency varied depending on the length of time that the tissue had been exposed to formalin fixative, where a longer fixation time meant that a longer digestion time was required to reach maximum antigenicity (41).

It was not until 1991 that the term “antigen retrieval” was coined (42). Shi et al. aimed to improve on the work of Battifora et al. (1986) (41) by creating an antigen retrieval methodology using heavy metal solutions heated in a microwave. The result was a significant increase in antigenicity and binding affinity for 39 proteins, while 13 proteins showed either no change or a decrease in binding affinity. The key difference between the work of Shi et al. and Battifora et al. was the lack of proteolytic processing—the sample was simply boiled in an inorganic solution.

Since this early work, there have been a number of papers focused on improving either form of the antigen retrieval technique. Table 1 provides a comprehensive list of buffers used for solution/heat-based antigen retrieval in proteomics. It should be noted that buffers containing citric acid or urea (43,44) and Laemmli buffer (45) have also been used. Progress has also been made in development of better proteolytic methods of antigen retrieval with additions such as EDTA to trypsin solutions (46). Taken together, this has resulted in a strong body of literature aimed at improving immunoreactivity and protein recovery from FFPE tissue. However, there is no reported difference between these two discretely different molecular techniques; they both fall under the umbrella of antigen retrieval, yet the molecular mechanisms responsible for these processes are completely different.

Enzymatic digestion works through either the specific (trypsin) or non-specific (pepsin, chymotrypsin) cleavage of proteins. Simply, enzyme molecules attach to the amino acid chain, moving along that chain until a relevant amino acid is placed in its active site, thereby eliciting cleavage. In the case of trypsin, this is at the locations of lysine and/or arginine residues.

HIAR using inorganic buffers causes the methylene bridges formed between one amino group and a subsequent amino, amido, guanidyl, or sulfhydryl group (6) to hydrolyze, thereby severing the bond. Currently, there is no protocol that is able to complete this reaction to 100% and break every crosslink.

Augmented Microscopy™



- ▶ capture
- ▶ analyze
- ▶ annotate
- ▶ video

Lionheart™ FX Automated Live Cell Imager enables superior digital microscopy. From simple fixed cell assays and slide scanning to advanced, environmentally controlled time-lapse movies, Lionheart FX provides qualitative and quantitative data in a compact automated microscopy system.

Think Possible



www.biotek.com

The use of a single term when referring to these two techniques is likely attributable to initial experiments that only focused on the end result: increased binding efficiencies of antibodies. Looking retrospectively, it is easy to see how methylene bridge hydrolysis (i.e., HIAR) and partial digestion of a sample could be mistaken to have the same mode of action. Early results published by Battifora et al. (1986) (41) demonstrated that trypsin digestion increases antigenicity. The mode of action was never explored and to date has not been truly investigated. Despite this, the assumption was made that trypsin uncovers antigenic sites. The work by Shi et al. (1990) (42) demonstrated the first HIAR method and concluded that this form of antigen retrieval was more effective, but once again there was no mention of a possible mode of action. It is from this publication that subsequent literature seems to have adopted the term antigen retrieval to indicate the undefined unmasking of antigenic sites in FFPE tissue.

With the knowledge contained in a substantial body of literature that aims to utilize FFPE tissue in proteomic workflows (Table 1), it is clear that HIAR hydrolyzes methylene bridges in FFPE tissue, and instead of uncovering specific antigenic sites, the aim is to reverse formaldehyde crosslinks and restore protein structures with minimal chemical modifications.

We propose distinguishing between the techniques of methylene hydrolysis (MH) and enzymatic antigen retrieval (EAR) using terms specific to either proteomics (MH) or histology (EAR). The adoption of these terms in their correct context will allow new research to be categorized more accurately, thus eliminating confusion between the proteomic and histological fields. As such, methodologies that aim to increase antibody binding efficiency in FFPE tissue may not be intrinsically useful to proteomics researchers attempting to retrieve proteins for downstream LC-MS analysis.

Conclusions

The ability to analyze FFPE tissue via proteomic techniques has tremendous potential to further our understanding of many diseases. The existence of tissue banks containing thousands of human specimens makes the continued development of FFPE analysis techniques acutely important. The current state of proteomic FFPE tissue analysis is still emerging. Issues of removing crosslinks and accounting for formaldehyde

modifications bioinformatically have not yet been fully addressed. We also believe that a clear separation of histological and proteomic contexts for antigen retrieval needs to be established with all workflows referring to the breaking of crosslinks as “methylene hydrolysis” and histological workflows referring to the uncovering of antigenic epitopes through proteolysis as “enzymatic antigen retrieval”

Author contributions

M.O'R. investigated the topic and wrote the text. M.P. contributed the “Post Processing” section, and proofread and edited the text.

Competing interests

The authors claim no competing interests.

References

1. **Fox, C.H., F.B. Johnson, J. Whiting, and P.P. Roller.** 1985. Formaldehyde fixation. *J. Histochem. Cytochem.* 33:845-853.
2. **Balgley, B.M., T. Guo, K. Zhao, X. Fang, F.A. Tavassoli, and C.S. Lee.** 2009. Evaluation of archival time on shotgun proteomics of formalin-fixed and paraffin-embedded tissues. *J. Proteome Res.* 8:917-925.
3. **Ikedo, K., T. Monden, T. Kanoh, M. Tsujie, H. Izawa, A. Haba, T. Ohnishi, M. Sekimoto, et al.** 1998. Extraction and analysis of diagnostically useful proteins from formalin-fixed, paraffin-embedded tissue sections. *J. Histochem. Cytochem.* 46:397-403.
4. **Fraenkel-Conrat, H. and H.S. Olcott.** 1948. The reaction of formaldehyde with proteins; crosslinking between amino and primary amide or guanidyl groups. *J. Am. Chem. Soc.* 70:2673-2684.
5. **Middlebrook, W.R. and H. Phillips.** 1947. The action of formaldehyde on the cystine disulphide linkage of wool; the conversion of subfraction A of the combined cystine into combined lanthionine and djenkolic acid and subfraction B into combined thiazolidine-4-carboxylic acid. *Biochem. J.* 41:218-223.
6. **Metz, B., G.F. Kersten, P. Hoogerhout, H.F. Brugghe, H.A. Timmermans, A. de Jong, H. Meiring, J. ten Hove, et al.** 2004. Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. *J. Biol. Chem.* 279:6235-6243.
7. **Chu, W.S., Q. Liang, J. Liu, M.Q. Wei, M. Winters, L. Liotta, G. Sandberg, and M. Gong.** 2005. A nondestructive molecule extraction method allowing morphological and molecular analyses using a single tissue section. *Lab. Invest.* 85:1416-1428.
8. **Becker, K.F., C. Schott, S. Hipp, V. Metzger, P. Porschewski, R. Beck, J. Nahrig, I. Becker, and H. Hoffer.** 2007. Quantitative protein analysis from formalin-fixed tissues: implications for translational clinical research and nanoscale molecular diagnosis. *J. Pathol.* 211:370-378.
9. **Addis, M.F., A. Tanca, D. Pagnozzi, S. Crobu, G. Fanciulli, P. Cossu-Rocca, and S. Uzzau.** 2009. Generation of high-quality protein extracts from formalin-fixed, paraffin-embedded tissues. *Proteomics* 9:3815-3823.

10. Shi, S.R., C. Liu, B.M. Balgley, C. Lee, and C.R. Taylor. 2006. Protein extraction from formalin-fixed, paraffin-embedded tissue sections: quality evaluation by mass spectrometry. *J. Histochem. Cytochem.* 54:739-743.
11. Oliveira, B.M., J.R. Coorssen, and D. Martins-de-Souza. 2014. 2-DE: the phoenix of proteomics. *J. Proteomics* 104:140-150.
12. Ono, A., T. Kumai, H. Koizumi, H. Nishikawa, S. Kobayashi, and M. Tadokoro. 2009. Overexpression of heat shock protein 27 in squamous cell carcinoma of the uterine cervix: a proteomic analysis using archival formalin-fixed, paraffin-embedded tissues. *Hum. Pathol.* 40:41-49.
13. Addis, M.F., A. Tanca, D. Pagnozzi, S. Rocca, and S. Uzzau. 2009. 2-D PAGE and MS analysis of proteins from formalin-fixed, paraffin-embedded tissues. *Proteomics* 9:4329-4339.
14. Tanca, A., D. Pagnozzi, G. Falchi, R. Tonelli, S. Rocca, T. Roggio, S. Uzzau, and M.F. Addis. 2011. Application of 2-D DIGE to formalin-fixed, paraffin-embedded tissues. *Proteomics* 11:1005-1011.
15. Tanca, A., S. Pisanu, G. Biosa, D. Pagnozzi, E. Antuofermo, G.P. Burrai, V. Canzonieri, P. Cossu-Rocca, et al. 2013. Application of 2-D-DIGE to formalin-fixed diseased tissue samples from hospital repositories: Results from four case studies. *Proteomics Clin. Appl.* 7:252-263.
16. Schubert, O.T., L.C. Gillet, B.C. Collins, P. Navarro, G. Rosenberger, W.E. Wolski, H. Lam, D. Amodè, et al. 2015. Building high-quality assay libraries for targeted analysis of SWATH MS data. *Nat. Protoc.* 10:426-441.
17. Hood, B.L., M.M. Darfler, T.G. Guile, B. Furusato, D.A. Lucas, B.R. Ringeisen, I.A. Sesterhenn, T.P. Conrads, et al. 2005. Proteomic analysis of formalin-fixed prostate cancer tissue. *Mol. Cell. Proteomics* 4:1741-1753.
18. Gustafsson, O.J., J.S. Eddes, S. Meding, S.R. McColl, M.K. Oehler, and P. Hoffmann. 2013. Matrix-assisted laser desorption/ionization imaging protocol for in situ characterization of tryptic peptide identity and distribution in formalin-fixed tissue. *Rapid Commun. Mass Spectrom.* 27:655-670.
19. O'Rourke, M.B., B.B.A. Raymond, S.P. Djordjevic, and M.P. Padula. 2015. A versatile cost-effective method for the analysis of fresh frozen tissue sections via matrix-assisted laser desorption/ionisation imaging mass spectrometry. *Rapid Commun. Mass Spectrom.* 29:637-644.
20. Weaver, E.M. and A.B. Hummon. 2013. Imaging mass spectrometry: From tissue sections to cell cultures. *Adv. Drug Deliv. Rev.* 65:1039-1055.
21. Kang, H.S., S.C. Lee, Y.S. Park, Y.E. Jeon, J.H. Lee, S.Y. Jung, I.H. Park, S.H. Jang, et al. 2011. Protein and lipid MALDI profiles classify breast cancers according to the intrinsic subtype. *BMC Cancer* 11:465.
22. Watrous, J.D. and P.C. Dorrestein. 2011. Imaging mass spectrometry in microbiology. *Nat. Rev. Microbiol.* 9:683-694.
23. Kussmann, M., E. Nordhoff, H. Rahbek-Nielsen, S. Haebal, M. Rosell-Larson, L. Jakobsen, J. Gobam, E. Mirgorodskaya, et al. 1997. Matrix-assisted laser desorption/ionization mass spectrometry sample preparation techniques designed for various peptide and protein analytes. *J. Mass Spectrom.* 32:593-601.
24. Chughtai, K. and R.M. Heeren. 2010. Mass spectrometric imaging for biomedical tissue analysis. *Chem. Rev.* 110:3237-3277.
25. Prieto, D.A., B.L. Hood, M.M. Darfler, T.G. Guile, D.A. Lucas, T.P. Conrads, T.D. Veenstra, and D.B. Krizman. 2005. Liquid Tissue: proteomic profiling of formalin-fixed tissues. *Biotechniques (Suppl)*:32-35.
26. Hwang, S.I., J. Thumar, D.H. Lundgren, K. Rezaul, V. Mayya, L. Wu, J. Eng, M.E. Wright, and D.K. Han. 2007. Direct cancer tissue proteomics: a method to identify candidate cancer biomarkers from formalin-fixed paraffin-embedded archival tissues. *Oncogene* 26:65-76.
27. Patel, V., B.L. Hood, A.A. Molinolo, N.H. Lee, T.P. Conrads, J.C. Braisted, D.B. Krizman, T.D. Veenstra, and J.S. Gutkind. 2008. Proteomic analysis of laser-captured paraffin-embedded tissues: a molecular portrait of head and neck cancer progression. *Clin. Cancer Res.* 14:1002-1014.
28. Palmer-Toy, D.E., B. Krastins, D.A. Sarracino, J.B. Nadol, Jr., and S.N. Merchant. 2005. Efficient method for the proteomic analysis of fixed and embedded tissues. *J. Proteome Res.* 4:2404-2411.
29. Jiang, X., X. Jiang, S. Feng, R. Tian, M. Ye, and H. Zou. 2007. Development of efficient protein extraction methods for shotgun proteome analysis of formalin-fixed tissues. *J. Proteome Res.* 6:1038-1047.
30. Sprung, R.W., Jr., J.W. Brock, J.P. Tanksley, M. Li, M.K. Washington, R.J. Slebos, and D.C. Liebler. 2009. Equivalence of protein inventories obtained from formalin-fixed paraffin-embedded and frozen tissue in multidimensional liquid chromatography-tandem mass spectrometry shotgun proteomic analysis. *Mol. Cell. Proteomics* 8:1988-1998.
31. Donadio, E., L. Giusti, F. Cetani, Y. Da Valle, F. Ciregia, G. Giannaccini, E. Pardi, F. Saponaro, et al. 2011. Evaluation of formalin-fixed paraffin-embedded tissues in the proteomic analysis of parathyroid glands. *Proteome Sci.* 9:29.
32. Tanca, A., M. Abbondio, S. Pisanu, D. Pagnozzi, S. Uzzau, and M.F. Addis. 2014. Critical comparison of sample preparation strategies for shotgun proteomic analysis of formalin-fixed, paraffin-embedded samples: insights from liver tissue. *Clin. Proteomics* 11:28.
33. Wisniewski, J.R., P. Ostasiewicz, and M. Mann. 2011. High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. *J. Proteome Res.* 10:3040-3049.
34. Wisniewski, J.R., K. Dus, and M. Mann. 2013. Proteomic workflow for analysis of archival formalin-fixed and paraffin-embedded clinical samples to a depth of 10 000 proteins. *Proteomics Clin. Appl.* 7:225-233.
35. Xu, H., L. Yang, W. Wang, S.R. Shi, C. Liu, Y. Liu, X. Fang, C.R. Taylor, et al. 2008. Antigen retrieval for proteomic characterization of formalin-fixed and paraffin-embedded tissues. *J. Proteome Res.* 7:1098-1108.
36. Helander, K.G. 1994. Kinetic studies of formaldehyde binding in tissue. *Biotech. Histochem.* 69:177-179.
37. Fu, Z., K. Yan, A. Rosenberg, Z. Jin, B. Crain, G. Athas, R.S. Heide, T. Howard, et al. 2013. Improved protein extraction and protein identification from archival formalin-fixed paraffin-embedded human aortas. *Proteomics Clin. Appl.* 7:217-224.
38. Küster, B., P. Mortensen, J.S. Andersen, and M. Mann. 2001. Mass spectrometry allows direct identification of proteins in large genomes. *Proteomics* 1:641-650.
39. Huang, Q., L. Yang, J. Luo, L. Guo, Z. Wang, X. Yang, W. Jin, Y. Fang, et al., 2015. SWATH enables precise label-free quantification on proteome-scale. *Proteomics*. 15:1215-1223.
40. Lam, H., E.W. Deutsch, J.S. Eddes, J.K. Eng, N. King, S.E. Stein, and R. Aebersold. 2007. Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* 7:655-667.
41. Battifora, H. and M. Kopinski. 1986. The influence of protease digestion and duration of fixation on the immunostaining of keratins. A comparison of formalin and ethanol fixation. *J. Histochem. Cytochem.* 34:1095-1100.
42. Shi, S.R., M.E. Key, and K.L. Kalra. 1991. Antigen retrieval in formalin-fixed, paraffin-embedded tissues: an enhancement method for immunohistochemical staining based on microwave oven heating of tissue sections. *J. Histochem. Cytochem.* 39:741-748.
43. Shi, S.R., B. Chaiwun, L. Young, R.J. Cote, and C.R. Taylor. 1993. Antigen retrieval technique utilizing citrate buffer or urea solution for immunohistochemical demonstration of androgen receptor in formalin-fixed paraffin sections. *J. Histochem. Cytochem.* 41:1599-1604.
44. Gustafsson, J.O., M.K. Oehler, S.R. McColl, and P. Hoffmann. 2010. Citric acid antigen retrieval (CAAR) for tryptic peptide imaging directly on archived formalin-fixed paraffin-embedded tissue. *J. Proteome Res.* 9:4315-4328.
45. Azimzadeh, O., Z. Barjaktarovic, M. Aubele, J. Calzada-Wack, H. Sarioglu, M.J. Atkinson, and S. Tapio. 2010. Formalin-fixed paraffin-embedded (FFPE) proteome analysis using gel-free and gel-based proteomics. *J. Proteome Res.* 9:4710-4720.
46. Ghatak, A. and C.K. Combs. 2014. Iba1 immunoreactivity is enhanced following an antigen retrieval treatment with EDTA, pH 6.0. *MethodsX.* 1:269-274.
47. Crockett, D.K., Z. Lin, C.P. Vaughn, M.S. Lim, and K.S. Elenitoba-Johnson. 2005. Identification of proteins from formalin-fixed paraffin-embedded cells by LC-MS/MS. *Lab. Invest.* 85:1405-1415.
48. Guo, T., W. Wang, P.A. Rudnick, T. Song, J. Li, Z. Zhuang, R.J. Weil, D.L. DeVoe, et al. 2007. Proteome analysis of microdissected formalin-fixed and paraffin-embedded tissue specimens. *J. Histochem. Cytochem.* 55:763-772.
49. Nazarian, J., M. Santi, Y. Hathout, and T.J. Macdonald. 2008. Protein profiling of formalin fixed paraffin embedded tissue: Identification of potential biomarkers for pediatric brainstem glioma. *Proteomics Clin. Appl.* 2:915-924.
50. Alkhas, A., B.L. Hood, K. Oliver, P.N. Teng, J. Oliver, D. Mitchell, C.A. Hamilton, G.L. Maxwell, and T.P. Conrads. 2011. Standardization of a sample preparation and analytical workflow for proteomics of archival endometrial cancer tissue. *J. Proteome Res.* 10:5264-5271.
51. Gámez-Pozo, A., N.I. Ferrer, E. Ciruelos, R. Lopez-Vacas, F.G. Martinez, E. Espinosa, and J.A. Vara. 2013. Shotgun proteomics of archival triple-negative breast cancer samples. *Proteomics Clin. Appl.* 7:283-291.
52. Lai, X. and B.P. Schneider. 2014. Integrated and convenient procedure for protein extraction from formalin-fixed, paraffin-embedded tissues for LC-MS/MS analysis. *Proteomics* 14:2623-2627.

Received 02 December 2015; accepted 20 February 2016.

Address correspondence to: Matthew P. Padula, Proteomics Core Facility, University of Technology Sydney, Cnr Harris and Thomas St, Ultimo, NSW, 2007 Australia. E-mail: matthew.padula@uts.edu.au

To purchase reprints of this article, contact: biotechniques@fosterprinting.com