Accepted Manuscript

Title: Biomedical analysis of formalin-fixed, paraffin-embedded tissue samples: The Holy Grail for molecular diagnostics

Authors: Boglarka Donczo, Andras Guttman

PII: S0731-7085(18)30405-9
DOI: https://doi.org/10.1016/j.jpba.2018.03.065
Reference: PBA 11896

To appear in: Journal of Pharmaceutical and Biomedical Analysis

Received date: 15-2-2018
Revised date: 30-3-2018
Accepted date: 31-3-2018

Please cite this article as: Boglarka Donczo, Andras Guttman, Biomedical analysis of formalin-fixed, paraffin-embedded tissue samples: The Holy Grail for molecular diagnostics, Journal of Pharmaceutical and Biomedical Analysis https://doi.org/10.1016/j.jpba.2018.03.065

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Biomedical analysis of formalin-fixed, paraffin-embedded tissue samples: The Holy Grail for molecular diagnostics

Boglarka Donczo¹ and Andras Guttman¹ ²

¹Horváth Csaba Laboratory of Bioseparation Sciences, Research Centre for Molecular Medicine, Faculty of Medicine, University of Debrecen, Debrecen, Hungary; ²MTA-PE Translational Glycomics Research Group, Research Institute for Biomolecular and Chemical Engineering, University of Pannonia, Veszprem, Hungary

Boglarka Donczo – Horváth Csaba Laboratory of Bioseparation Sciences, Research Centre for Molecular Medicine, Faculty of Medicine, University of Debrecen, 4032 Debrecen, Nagyerdei krt 98. Hungary donczo.boglarka@med.unideb.hu

Corresponding author: Andras Guttman – Horváth Csaba Laboratory of Bioseparation Sciences, Research Centre for Molecular Medicine, Faculty of Medicine, University of Debrecen, 4032 Debrecen, Nagyerdei krt 98. Hungary andrasguttman@med.unideb.hu

Highlights

- Formalin fixation with paraffin embedment is most widely used for tissue fixation
- Formalin fixation differently affects various biomolecules
- Biomarker discovery form FFPE samples is the Holy Grail of molecular diagnostics
Abstract

More than a century ago in 1893, a revolutionary idea about fixing biological tissue specimens was introduced by Ferdinand Blum, a German physician. Since then, a plethora of fixation methods have been investigated and used. Formalin fixation with paraffin embedment became the most widely used types of fixation and preservation method, due to its proper architectural conservation of tissue structures and cellular shape. The huge collection of formalin-fixed, paraffin-embedded (FFPE) sample archives worldwide holds a large amount of unearthed information about diseases that could be the Holy Grail in contemporary biomarker research utilizing analytical omics based molecular diagnostics. The aim of this review is to critically evaluate the omics options for FFPE tissue sample analysis in the molecular diagnostics field.

Abbreviations

AMeX: acetone-methylbenzoate-xylene; CE: capillary electrophoresis; FFPE: formalin-fixed, paraffin-embedded; HOPE: Hepes-glutamic acid buffer-mediated organic solvent protection effect; LC: liquid chromatography; MALDI: matrix-assisted laser desorption/ionization; MS: mass spectrometry; MSI: mass spectrometry imaging; MW: microwave; NBF: neutrally buffered formalin; PCR: polymerase chain reaction; SDS: sodium dodecyl sulfate;

Keywords

FFPE, genomics, proteomics, glycomics, metabolomics, molecular diagnostics
Introduction

Formalin-fixed, paraffin-embedded (FFPE) tissue specimens represent an extremely valuable sample source for prospective and retrospective biomedical and biopharmaceutical studies, therefore, could be the Holy Grail for analytical omics based molecular diagnostics. FFPE samples are banked in hospital archives worldwide for decades, thus appropriate number of samples can be quickly cherry-picked for specific retrospective studies with the great advantage of having the associated clinical history of the patients. This largely unexploited resource represents an extensive repository of tissue material with clinical follow-up information, providing a valuable resource for translational clinical research, especially with the use of modern analytical omics tools at the molecular level. Therefore, the use of FFPE samples in biomarker discovery could be of high interest.

There is an increasing need for validated biomarkers in the diagnostics field to stratify patients for targeted therapies and to monitor therapeutic responses for a wide spectrum of diseases. Blood or tissue specimens are frequently used in the omics field to discover new biomarkers. FFPE tissue specimens can represent a valuable alternative in the field of molecular diagnostics [1] as they are routinely prepared for pathological studies [2]. Moreover, the archived material is kept for decades in most countries, and usually accompanied with the clinical outcome information [3].

In the past decade, many molecular level FFPE studies have been reported, mainly covering the following three categories: 1) studies investigating FFPE DNA and protein extraction methods (this latter is also referred to as antigen retrieval), very often accompanied with fresh frozen (fr/fr) versus FFPE tissue comparison [4, 5]; 2) articles determining the suitability of FFPE specimens in biomarker research [6]; and 3) studies looking for the influence of preanalytical factors such as fixation time and protocol or archival time [7]. This last one is important because tissue preparation parameters may result in some degradation of the sample, e.g., differences were found in rRNA patterns between 4 °C and 37 °C incubation, and the pattern almost completely disappeared after 12 hours incubation at 37 °C [8].

This review compares the most frequently used methods of FFPE tissue preparation, analysis and their potential utilization in molecular diagnostics. Many studies have been published on the recovery of biomolecules from FFPE tissue samples with the goal of biomarker discovery.
via analytical omics. One of the major challenges was to develop special protocols to recover the biomolecules of interest in their intact or close-to intact forms. The next challenge will be to reveal the biological information they possess for utilization in the everyday clinical practice, e.g., to detect diseases in early stages.

2. Tissue fixation and embedding techniques

2.1 Formaldehyde fixation

Histology is an essential discipline since Rudolph Virchow has established the principal values in his pioneering work [9]. Formaldehyde as a chemical was first reported in 1859 by Butlerov, however, the practical method of its synthesis was invented only 10 years later. Due to its disinfectant properties, formalin was proposed for medical applications by the German physician Ferdinand Blum in 1893 [10]. During his work towards the end of the 19th century, Blum realized that tissues treated with formalin hardened, even more than with other commonly used agents at that time and summarized the results of his findings in 1910 [11]. Based on his reports, contemporary histologists confirmed the unequally advantageous properties of formalin as a fixative, preserving tissues without significant shrinkage or distortion, in contrast to alcohol-based reagents. Moreover, classic staining procedures with hematoxylin and eosin dyes were compatible with this fixation method. Since then, formalin fixation procedure has become the gold standard for pathological tissue preservation. Other advantages of this conservation method were good architectural preservation of tissue structures and cellular shape stabilization. [12].

From chemical point of view, formaldehyde reacts with the free amino groups of proteins and peptides during the fixation process, via the formation of methylol groups as shown in Figure 1. The reaction with formaldehyde consists of two steps. First, when the primary amine (nucleophilic group; R) and the aldehyde produce a hydroxymethyl form. After losing a water molecule, a Schiff base is formed. In the second step, during the stabilization, other proteins are involved in the reaction and linked with methylene bridges. It is a stable crosslink between amino acids such as arginine, glutamine, tryptophan, histidine, asparagine, and tyrosine in the form of methylene bridges within or in between proteins and their subunits [13]. Formalin fixative solutions are usually prepared by dissolving 37 – 40% w/w formaldehyde in water, containing 10-15% methanol as stabilizer. Tissues are routinely fixed for 24–48 hours in 10% v/v solution of formalin (3.7 – 4.0% formaldehyde, buffered with phosphate to pH 7.0). Fixation in this way yields samples, which can be stored at ambient condition for decades.
Combined with paraffin embedding (described below), the resulting formalin-fixed, paraffin-embedded samples are routinely used in histological and histopathological laboratories. FFPE samples can be stored long-term at room temperature that is easier and simpler that of dealing with the storage of fresh/frozen tissues, which require -80 °C or even in liquid Nitrogen in some instances. Figure 2 shows the main steps of the formalin fixation and paraffin embedding process. The first step is sampling, which is usually done by a surgeon. The sample then either gets frozen or fixed immediately. Fixation takes place in a 10% buffered formalin solution for 24 to 48 hours at room temperature. If samples are spaded, penetration of formalin into the tissue is more sufficient. The next step is paraffin embedding that requires dehydration of the sample by washing twice with ethanol and xylene, respectively. The dehydrated sample is embedded in melted paraffin and allowed to solidify at room temperature. The formalin-fixed, paraffin-embedded block is then cut into 3-5 µm sections for pathologic and molecular diagnostic investigation. Research groups from all around the world have recently started evaluating the huge archival FFPE sample collection from molecular diagnostics point of view because of their clinical-pathological significance [13].

2.2 Tissue fixation by microwave irradiation

Tissue fixation can also be accomplished by alternative techniques to alleviate some of the non-desirable effects of the formaldehyde, e.g., the direct reaction with nucleotides, which could result in degradation of the nuclear material. One of the substitute methods for tissue fixation is microwave irradiation (MW) with a frequency range between 300 MHz and 300 GHz, first reported by Mayers [14]. He described that direct exposure to MW irradiation resulted in satisfactory fixation of mouse and human postmortem tissues without the addition of formalin. MW fixation depends on the chemical environment surrounding the specimen during the irradiation process, the duration of the microwave exposure, and the sequence of microwave irradiation (before or after the chemical fixation) as well as chemicals present such as formaldehyde, methanol or glutaraldehyde. As of today, five major MW assisted fixation methods are used: 1) For stabilization, the samples are irradiated in situ or in a physiological salt solution to avoid structural degradation otherwise might happen with the use of a chemical fixatives. 2) Fast or ultrafast primary MW combined with chemical fixation in which specimens are exposed to MW irradiation in a formaldehyde containing chemical environment for very short time periods, ranging from milliseconds to seconds. 3) MW irradiation followed by chemical fixation to improve fixation uniformity. 4) Primary chemical fixation followed by MW irradiation to ease the penetration of the fixatives inside the sample.
5) MW irradiation applied together with freezing to further promote the preservation of the samples [15]. While, the great advantage of microwave stabilization is to avoid or minimize the involvement of chemicals, it has got some disadvantages including constriction, sponginess of tissues, and collapse of red blood cells [16]. On the other hand, microwave irradiation can increase diffusion of the fixation reagents deep into the tissue and accelerate the chemical crosslinking process. Primary microwave irradiation is an attractive alternative fixation method for ultrastructural and genetic studies because it provides easy access for the cross-linking reagent to penetrate the tissue. However, using MW irradiation on tissues previously fixed in formalin often results in poor outcome since the surface of the tissue fixes so quickly that it apparently restrains further diffusion of the fixative into the inner parts of the sample. Therefore, a combined procedure is recommended, which starts by formalin fixation of the tissue blocks for a few hours at room temperature, followed by microwave irradiation for 1 to 2 minutes at 55°C [17]. The speed by which microwaves can help to complete fixation of both large and small biopsy specimens is a major advantage, thus MW fixation considerably reduces processing time before paraffin embedding [18, 19].

2.3 Other fixatives
In some instances, the use of alternative chemical fixatives overcomes the possible problems of formaldehyde treatment mentioned earlier and could be better for the special task in hand, e.g., preservation of nucleic acids and proteins. Similar to formaldehyde, glutaraldehyde also acts as a cross-linking agent for tissue sample fixation as was first introduced by Sabatini et al. [20]. Glutaric dialdehyde, the aqueous solution of glutaraldehyde contains 70% hemiacetal, 16% monohydrate, 9% dihydrate and ~4% free aldehyde. Glutaraldehyde preferably reacts with nucleic acids and the ε-amino groups of amino acids through its aldehyde functional groups. Douglas et al. [21] reported that high-molecular weight DNA can be better preserved by 1% glutaraldehyde at pH 7.0 than in 10% formalin. It is important to note that the diffusion rate of 4% glutaraldehyde (at 4°C) is approximately half that of for the same concentration formaldehyde solution. Albeit, glutaraldehyde is widely used as a fixative for standard electron microscopy, the need for periodic purification to maintain the functional aldehyde levels and the slower penetration rate have greatly limited its use as pathological/ biological fixative.

Genipin (C_{11}H_{14}O_{5}), extracted from Gardenia jasminoides is an aglycon and acts as a natural cross-linker, which can spontaneously react with amino acids. This fixative is excellent for
cross-linking collagen, gelatin, and chitosan. It has low acute toxicity (LD50 i.v. 382 mg/kg in mice), therefore, much less toxic than many other commonly used synthetic cross-linking reagents. It was reported, however, that after genipin fixation, the resistance of the tissue against collagenase degradation significantly increased. The effect of genipin fixation on tissue nucleic acids is not known [22].

Methanol and ethanol was first used as a fixative by Camillo Golgi [23]. For nucleic acid fixation it is recommended to use alcoholic reagents, which are non-crosslinking [24, 25]. Prior to use of nucleases, nucleic acids are precipitated from the alcohol in the presence of salts, however, the procedure tends to cause shrinkage, thereby making the sample less suitable for morphological studies [21]. 100% methanol and ethanol are perceived as adequate fixatives to prevent high-molecular weight DNA and RNA from degradation. On one hand methanol, ethanol and acetone preserved relatively high molecular weight DNA (≈ 440 000 Da) equally well, compared to formalin. On the other hand, histological analyses revealed that regional tissue shrinkage differences occurred in ethanol and acetone fixed specimens, compared to methanol-fixed ones [24, 25].

Since no particular fixative is adequate to preserve all types of tissue samples for various applications, mixtures of several fixatives have also been attempted to alleviate this handicap. One common reagent is the mixture of 60% ethanol, 30% chloroform and 10% glacial acetic acid, known as Carnoy’s fixative, that proved to offer optimal preservation of tissue nucleic acids [26]. In comparison to neutrally buffered formalin fixation (NBF), RNA was easily extractable from Carnoy’s-fixed tissues and only some slight degradation was observed in the high-molecular weight RNA range [25]. Application of post-fixation formaldehyde vapor reduced cellular RNA loss from Carnoy’s-fixed tissues, which was important before in situ hybridization (ISH) [19]. By all means, alcohol-based fixatives showed great efficiency to preserve the molecular integrity of nucleic acids in tissue samples [19, 27].

Another frequently used fixative mixture is 60% methanol, 30% chloroform and 10% glacial acetic acid, called as methacarn. Replacing ethanol with methanol resulted in an outstanding alternative to preserve RNA in tissue samples. Methacarn and Carnoy’s fixatives were introduced by Puchtler et al. [28, 29]. Using cultured cell lines and rodent tissues, Shibutani et al. [30] have demonstrated that the integrity of the extracted total RNA content, and also the efficiency of the extraction from methacarn-fixed tissues, produced identical results to unfixed frozen cells and tissue samples. Genomic DNA contaminant concentration was very
low, and 300-700 base fragment long species of both low copy number RNA and abundant mRNA were amplified successfully from methacarn-fixed tissues [30]. It has also been suggested that methacarn fixation induced ribosomal protein precipitation, thereby reduced endogenous RNase activity. Methacarn fixation was also better than NBF in preserving antigen immunoreactivity requiring no antigen retrieval, therefore, recommended for prospective immunohistochemical tissue studies [31].

Acetone is widely used as a fixative in the so called AMeX (acetone-methylbenzoate-xylene) technique [32]. The procedure requires overnight tissue fixation in acetone at −20 °C, then a xylene and methylbenzoate wash prior to paraffin embedding to maintain the morphology of the sample and also to preserve the immunoreactivity of unstable proteins. AMeX-treated tissues provided intact high-molecular weight DNA in good quality, ideal for molecular diagnostic analysis. Moreover, signals of c-myc mRNA and albumin mRNA were comparable to the ones from fresh/frozen tissues. Dot-blot hybridization-based detection was also successfully carried out from isolated RNA of previously AMeX-fixed tissues [33].

The HOPE method (Hepes-glutamic acid buffer-mediated organic solvent protection effect) involves incubation of fresh tissues in a protecting solution that contains mixtures of amino acids with pH ranging from 5.8 to 6.4 [34]. Dehydration was accomplished by incubation in acetone between 0 and 4°C. The protective solution accelerated dehydration by enhancing the infusion of acetone. As the last step, the specimens were placed into paraffin. This new fixation method was suitable for full-scale pathological analysis, consequently adequate for both immunohistochemistry and molecular pathology. HOPE-fixed sections tend to show formalin fixation type morphology and preserve proteins and antigenic structures for differential analysis by immunohistochemical and/or enzyme histochemical techniques. HOPE-fixed specimens provided eligible quality RNA and DNA, even after 5 years of storage. The DNA and RNA in those samples were subject to polymerase chain reaction (PCR), reverse transcription polymerase chain reaction (RT-PCR) analysis and fluorescence in situ hybridization (FISH). The noncross-linking feature and the high amount of recovered nucleic acids clearly showed the valuable alternative HOPE fixation offered [34].

Other fixation additives like mercury (II) chloride (Helly’s, Zenker’s, Ridley’s solutions or B5), picric acid (Bouin’s solution) and tannic acid all improved tissue penetration. Metal salt solutions, like zinc sulfate or phenol containing reagents (they have less toxicity) increased protein precipitation efficiency. Some of the aforementioned preservatives have been proved
to be valid alternatives in various immunohistochemical studies, but only limited data is available about their use as nucleic acid preserving agents [27]. Table 1 summarizes the most frequently used fixatives for tissue preservation.

2.4 Paraffin embedding
Formalin or other fixation techniques in combination with paraffin embedding retains the cellular details and morphology of tissue samples. Therefore, formalin-fixation, paraffin embedding quickly became the standard preservation method for diagnostic pathology. In clinical practice, biopsy samples usually treated with formalin as described above then further prepared for paraffin embedding by washing with ethanol and xylene to ensure the removal of all water content. Upon full dehydration, melted paraffin is added to the dry formalin fixed tissue samples and let solidify at room temperature [35]. Pathology departments routinely archive large quantities of FFPE specimens worldwide, since long-term storage of frozen tissues demands extra low temperatures, thus keeping them for long term is more complicated than the storage of FFPE blocks.

3 Analytical Omics

3.1 Genomics
Formalin fixation parameters of time and temperature are critical in maintaining nucleic acid (DNA and RNA) integrity [36], but fortunately only small fragments of mRNAs are required to reconstruct an expression profile of a given gene. Degradation might not always be distinctly observable when quantifying mRNA, but low expression profile genes with short half-life might be significantly affected [37], thus should be investigated with greater scrutiny.

DNA isolated from FFPE blocks, even if highly fragmented are still useable for PCR amplification of shorter products up to 250 bp in length [38]. In addition, partially fragmented DNA was applicable for downstream analysis by sequence-specific oligonucleotide (SSO) probe procedures in combination with PCR. Ribeiro-Silva et al. found that RNA in FFPE samples may degrade over time into smaller fragments, indicated by the much lower RNA quality indicators than in fresh / frozen tissues, however, could still be successfully used in gene expression analysis [39]. DNA and miRNA extracted from FFPE blocks have currently reached levels that allow their application as essential and primary approaches in the search for new genomic disease markers [40].
DNA is a very stable type of cellular macromolecule however, it often loses the quality needed for high throughput analysis methods like microarrays or sequencing. Thus, special extraction methods have been developed to alleviate this issue [41]. RNA, on the other hand, is prone to cross-linking to proteins during the formalin fixation process. To offset the problems in isolation, amplification, and quantification Specht et al. [42] introduced a method, which combined laser-capture microdissection (LCM) and quantitative reverse transcription polymerase chain reaction (QRT-PCR), using proteinase-K for digestion and extraction of RNA. This has made quantitative gene expression analysis possible from archived tissue banks. Several publications reported about the use of FFPE tissue samples in RNA analysis [43-56]


3.2 Proteomics

Proteomics profiling of FFPE archival tissues holds the promise to find potential disease biomarkers at the protein level for diagnostic and therapeutic follow-up. However, this option has not been realized yet, due to such challenges as limitations of high sensitivity analytical instrumentations, lack of efficient antigen retrieval methods, or even restricted sample accessibility [13].

Formaldehyde fixation affects proteins at different levels. The first is in their primary structure through amino acid modification including lysine, arginine, serine, tyrosine, asparagine, histidine and glutamine [57]. The principal cross-linking between the side chain amino groups of the above mentioned amino acids are manifested as methylene bridges. Cross-linking can also occur between the aminomethylol groups and phenol, indole and imidazole side chains via Mannich reaction. Cross-linking modifies the $\alpha$-helixes and $\beta$-sheets of the secondary protein structures, also causing alterations in their tertiary structures, even link multiple proteins. These changes may hinder protein identification when trying to match with *in silico* databases after mass spectrometry (MS) analysis [13].
Recent attempts aimed investigating protein identification options from FFPE blocks after long-term storage. The results showed that low abundance proteins might suffer an archival effect suggested by the less than ten spectral counts in mass spectrometry. These proteins were difficult to retrieve from tissue blocks older than 10 years [58]. Magdeldin et al. [13] investigated the effect of archival time by K-means cluster analysis of FFPE tissue proteomes from samples stored for 9 to 21 years. Their findings revealed that, at the level of protein and distinct peptide identification, more peptides were identifiable in the blocks that were stored for 9 years compared to 21 years.

Proteins isolated from FFPE blocks have been successfully used in Western blot analysis, reverse-phase protein arrays, and surface-enhanced laser desorption/ionization (SELDI) time-of-flight MS [59]. The reverse phase protein array (RPPA) method is similar in fashion and structure to Western-blot analysis because of the use of target specific primary antibodies and fluorescently labeled secondary antibodies. Different proteins can be analyzed by simultaneously comparing their relative fluorescence levels [60].

Ambiguous results were frequently obtained in global proteome analysis of FFPE specimens due to the failure of matching the experimentally found peptides to those in the existing databases, such as EBI (European Bioinformatics Institute), Proteome Analysis, InterPro and CluSTr [61], just to list the most important ones. The possibility of correct peptide identification depends on the degree of modification and cross-linking caused by the fixation process, thus, can be categorized accordingly, i.e., high, moderate or low probability for successful identification. Peptides in the first group are mainly non-modified ones and usually match protein database entries. Peptides with a moderate likelihood to be properly identified are non-crosslinked peptides with known modifications. The identification of this second peptide group mainly depends on the database search engine used and the precise knowledge about the modifications occurred during formalin fixation. The last group represents formalin cross-linked peptides, which may also be modified by endopeptidase digestion. These peptides have low chance to be correctly identified and the modifications caused by the fixation may also lead to shifts in their physicochemical properties from native forms, thus identification of this group of peptides is very challenging [13].

Effective proteomic analysis of FFPE samples relies on proper extraction, denaturation, and digestion with extensive knowledge about all possible peptide modification(s). By applying proper antigen retrieval techniques, Shi and associates successfully applied immunohistochemistry (IHC) on FFPE samples [62]. As a matter of fact, IHC on FFPE
blocks became a widely used procedure in routine diagnostic pathology. Recent technical advances in the field of molecular biotechnology made possible the extraction of glycoproteins from FFPE blocks for downstream analysis. Near the end of the last century, one of the pioneers in this extraction field, Ikeda et al. [63] studied the effect of several special buffers and temperatures and found that combining lysis buffers, e.g., RIPA (radio immunoprecipitation assay buffer) with high concentration sodium dodecyl sulfate (SDS) at appropriate temperatures resulted in efficient protein recovery. After its introduction, several groups attempted to improve the method by applying other buffer systems to increase the extraction yield. Tris-HCl was the general starting buffer at different pHs, mostly containing a number of chaotropes and detergents. The generally utilized extraction solutions of 50 mM Tris-HCl at pH 7 [4] and 20 mM Tris-HCl at pH 9 have been used with 2% SDS in combination with heat-induced retrieval conditions (20 min at 100ºC, then 120 min at 60ºC) [64]. Reducing agents of beta-mercaptoethanol [65] or dithiothreitol (DTT) were reportedly added to these buffers for antigen retrieval from skeletal muscle or liver samples [66, 67], as well as heart tissues [68] along with the use of octylglucoside detergent. Chaotropic agents like guanidin-HCl [69] were also used for protein retrieval from FFPE tissues [70]. RIPA buffer in combination with high concentration sodium chloride, Triton X-100, sodium deoxycholate, phenylmethylsulfonyl fluoride, aprotinin, leupeptin (addition of protease inhibitors is optional; aprotinin inhibits serine proteases, such as trypsin, chymotrypsin, plasmin, trypsinogen, urokinase, and kallikrein; leupeptin inhibits both serine and cysteine proteases, such as calpain, trypsin, papain, and cathepsin B) and SDS was used on colorectal carcinoma [63], lymphoma [71], and prostate cancer FFPE specimens [72]. Addition of organic solvents were also attempted at different concentrations, e.g., 30% acetonitrile was applied to pancreatic [73] and colon cancer tissues [74], while 50% trifluoroethanol was tested in lung tissue processing [75].

Commercial reagents are also available for FFPE tissue processing, like the acid-deteriorating detergent RapiGest [76], Liquid Tissue MS protein Prep kit [77, 78], Qproteome FFPE tissue kit [79], NDME-U [80], and the FFPE Protein Extraction Solution [81]. All of these reagent kits can be applied with a digestion step accomplished on various molecular filtration units [81]. Such Filter Aided Sample Preparation (FASP) procedure was elaborated by Mann and coworkers [82] to recover proteins from FFPE tissues using complex lysis solutions composed of detergents (e.g., SDS), chaotropes (e.g., urea), denaturing solutions (e.g., DTT) and alkylating agents (e.g., iodoacetamide). The filter-based approach was particularly
efficient in removing or changing buffers and rinsing the samples; however, it was important to ensure that the treatments did not damage the filter. Subsequently, protein digestion could be performed directly on the filter to recover the tryptic peptides, which were then subject to analysis after desalting [83]. This procedure was apparently very efficient in proteomics studies of microdissected tissue samples as well for subsequent LC-MS/MS analysis that identified up to 10,000 proteins [84]. Among these studies, researchers tried to evaluate the efficiencies of the extraction procedures on FFPE tissues in comparison to fr/fr tissue samples. Generally, analysis of fr/fr tissues may identify a higher number of proteins compared to FFPE tissues [85], albeit, some studies reported better protein identification from FFPE tissues [76, 86, 87]. These comparisons were of vital importance in finding the adequate sample preparation methods for efficient of FFPE specimen analysis [81].

Currently available tools for tissue sampling and analysis are extremely diverse and allow researchers to address many questions using the most relevant approaches. Matrix-assisted laser desorption / ionization (MALDI) MS imaging, e.g., allows mapping of hundreds of compounds on a tissue section in a differential manner, thus represents strong illustration of the synergy between “classical” and “molecular” histology [81]. MALDI imaging mass spectrometry with ultra-highspeed matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) and high mass resolution matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance imaging mass spectrometry (MALDI FTICR IMS) platforms are powerful tools to visualize proteins in fixed tissues. Spraggins et al. [88] have used high spatial resolution MALDI-TOF technique to obtain protein images of rat brain tissues and cystic fibrosis lung tissues with image acquisition at rates of >25 pixels/s. Structures as small as 50 μm were spatially resolved and the proteins associated with the host immune response were identified in cystic fibrotic lung tissue.

Yin et al. [89] investigated different stage colorectal cancer FFPE tissues by mass spectrometry. As a result of their study, they provided large scale systematic quantification for the differentially expressed proteins during cancer development from early stage to stage IIIC. Approximately two-thirds of the detected proteins were quantified. Label free quantification and iTRAQ (using isobaric tag for relative and absolute quantitation) was also reported [90].

3.3 Glycomics
Carbohydrates play vital roles in biological processes such as cell–cell and cell–extra cellular matrix interactions, protein degradation, inflammation, and activation of the immune
system. They participate in motility and adhesion, intracellular signaling, and also control protein folding, stabilize the conformation of proteins, modify immunological properties and act as cell surface receptors for lectins, antibodies and toxins [91]. Any alterations in these functions (e.g., normal cells transforming into malignant ones) can appear as changes in the cell surface glycan profile in many diseases [92]. Glycosylation changes have been shown in cancer, cardiovascular, autoimmune and neurodegenerative diseases [93], just to list a few important ones. Therefore, specific structural changes in glycans may serve as biomarkers [94], and the type of the change can be correlated to the disease [95].

The major glycoconjugate classes in vertebrates are glycoproteins, proteoglycans, glycolipids, glycosaminoglycans and O-GlcNAcylation. Most glycoconjugates are located in the cell surface membranes. Glycoproteins have the two major types based on the way their sugar moieties bound to the polypeptide chain: asparagine (N-) and serine/threonin (O-) linked glycans.. Around 1% of the human genome participates in the biosynthesis of glycans [96], resulting in one of the most complex co- and post-translational modifications of proteins. The great variability in glycan structures is contributing the ability to fine-tune the chemical and biological properties of glycoproteins. For N-linked glycosylation, after the initial co-translational attachment of the GlcNAc2Man9Glc3 structure during the synthesis of the polypeptide backbone, the final processing of carbohydrates occurs in the Golgi apparatus and the endoplasmic reticulum, but in some instances may also in the cytoplasm and nucleus [97]. All asparagine linked glycans share a common paucimannose core structure of Manα1–6(Manα1–3)Manβ1–4GlcNAcβ1–4GlcNAcβ1 at an Asn-X-Ser/Thr consensus sequence (X cannot be Ser, Thr or Pro). N-linked glycans can be classified into three categories: (1) high mannose type, in which only mannose residues are attached to the core; (2) complex type, in which all sugar types are attached but mannose; and (3) hybrid type, in which only mannose residues are attached to the Manα1–6 arm of the core and one or two complex type antennas are to the Manα1–3 arm. O-linked glycans are attached to serine or threonine residues with currently known 8 core structures [97].

Histochemical staining using lectins is one of the common methods to analyze the glycan moieties of FFPE tissue sections. Concanavalin A (ConA) and Aleuria Aurantia Lectin (AAL) are the two most frequently used lectins in histostaining to detect oligomannose and fucosylated glycans, respectively [98]. Unfortunately, lectin affinity-based methods cannot reveal enough structural information about the epitopes. For instance, ConA can bind to both terminal and internal α-mannoses [99]. Moreover, lectins cannot differentiate between glycan...
subgroups, e.g., both N-linked and O-linked glycans possessing fucose residues bound by AAL. In addition, just a few lectins can be stained at a time on a tissue section due to possible steric hindrance of different glycan epitopes, which makes overall glycan imaging difficult. Another drawback of lectin based histostaining is quantification inaccuracy. Alternative imaging methods are being developed to obtain the required information from lectin stained assays [98]. Carbohydrate specific antibodies are more specific with respect to glycan recognition however, the number of currently available carbohydrate specific monoclonal antibodies is far from covering the entire palette of mammalian glycans.

Utilizing the recent developments in MS technology, in one approach, mass spectrometry imaging (MSI) was utilized to acquire spatial information and molecular profiling of glycans from tissue sections [100]. Given the high specificity and sensitivity of mass spectrometry, MSI has addressed some of the handicaps of the common histostaining techniques. Prior knowledge of the glycans here is not essential, otherwise required for affinity-based immunohistochemistry staining. Moreover, hundreds of compounds can be detected in a single run, producing a vast amount of molecular information in a spatial manner. MSI can also support semi-quantitative and quantitative modes by sampling directly from the tissue sections. Albeit, advances of analytical techniques and data interpretation methodologies have extended the limits of glycomics studies [101], the low ionization efficiency of native glycans makes it difficult to study them from complex biological mixtures. Therefore, glycan analysis by mass spectrometry usually requires isolation, separation, extraction and chemical modification (e.g., permethylation) prior to analysis [102]. Using this approach, however, the spatial glycomic information is lost due to sample homogenization. Yang et al. [103] demonstrated that if glycans were released from solid phase immobilized glycoproteins, direct mass spectrometry analysis was applicable. This method did not require glycan purification to alleviate interferences from peptides and proteins representing a critical step during glycan imaging platform development to provide information about complex carbohydrates from samples with high biological complexity, like FFPE tissue specimens [98].

Everest-Dass and Briggs [104] developed a novel method using MALDI MSI on formalin-fixed paraffin-embedded tissue sections to search for tissue specific glycan markers. They investigated ovarian cancer tissue samples with porous graphitized carbon liquid chromatography connected to electrospray ionization mass spectrometry (PGC-LC-ESI-MS/MS) and MALDI-MSI techniques to reveal structural glycan profile information. After the enzymatic release of N-glycans with peptide-N-glycanase, the sugars were separated by
porous graphitized carbon liquid chromatography (PGC-LC) and analyzed by collision induced electrospray MS fragmentation. The authors detected 40 individual N-glycan structures from the FFPE tissue section. The N-glycans found were high mannose, hybrid and complex (neutral and sialylated) type core fucosylated carbohydrates. The presence of these three generally known glycan structures was also verified by high-resolution MALDI-MS. Tissue-specific distribution of N-glycan structures detected in distinct regions of the tissue samples, i.e., high mannose glycans were mainly expressed in the tumor region, while complex or hybrid type N-glycans were abundant in the intervening stroma. Therefore, tumor and non-tumor tissue regions were clearly distinguished by their N-glycan structure allocations [104]. Powers et al. used MALDI imaging mass spectrometry and tissue microarrays to profile N-linked oligosaccharides from human and murine FFPE tissue samples and found 28 N-glycans in mouse kidney and another 13 in human pancreatic cancer FFPE tissues [78, 105]

As there is a general interest to utilize the huge FFPE sample collection archives, novel, efficient and readily applicable glycan analysis methods should be developed. So far, LC/MS and MALDI techniques were mostly used for such investigations as described above. To the best of our knowledge, our group was the first to analyze glycan profiles from FFPE samples by capillary electrophoresis with high sensitivity laser induced fluorescence detection (CE-LIF). We emphasize the global analysis aspect of the N-glycans from the FFPE tissue samples in contrast to the histological approach. To find possible formalin fixation mediated structural alterations at the glycome level, N-linked sugars of FFPE treated standard glycoproteins, as well as PNGase F released human serum and mouse tumor tissue sample glycans were analyzed. The liberated N-glycans were labeled by 8-aminopyrene-1,3,6-trisulfonic acid (APTS) fluorescent dye and separated by CE-LIF. The results revealed no apparent changes between the N-glycosylation patterns before and after the formalin fixation and paraffin embedding processes in all sample types examined. Thus, it was suggested that FFPE tissue samples could be readily utilized to discover disease related N-glycosylation changes at the molecular, cellular and tissue levels in prospective and retrospective studies [106].

3.4 Metabolomics and lipidomics

Metabolomics is the global study of low molecular mass molecules (metabolites) found within biological systems [107]. Metabolites are precursors, intermediates or end products of biochemical processes, and information about the whole metabolic pathway represents additional information of the cellular state at a specific time point and represent the final level
of the omics cascade of genomics, transcriptomics, proteomics and glycomics [108]. Molecular pathways and interactions are generally studied through the analysis of metabolites [109]. However, additional factors, like the tumor microenvironment in case of cancers, may influence the metabolic profiles. Frequently observed polyclonality generates further hindrance to fully comprehend the metabolomic data. Metabolic profiling of intact tissue samples is becoming more popular in clinical research, as it can potentially identify novel prognostic or predictive metabolic biomarkers or explore the abnormal biochemical activity aiming to identify novel therapeutic targets [110].

Nuclear magnetic resonance (NMR) spectroscopy and MS or LC/MS are the most rapidly developing analytical techniques in the metabolomics field. NMR spectroscopy is usually used in the analysis of metabolic compounds that cannot be examined by GC/MS or LC/MS techniques. In high-throughput analysis of human liquid samples (e.g., blood), NMR is a very applicable method for quantitative and qualitative analysis of low-molecular-weight metabolites and lipids [111].

Lipids represent a significant part of the metabolome, so lipidomics research shares its main concepts and methods with the metabolomics field. For the time being, mass spectrometry is the dominant analytical technique in this field [112], and the two major MS-based approaches presently used are: 1) combined with separation techniques such as GC/MS or LC/MS, which are used for analyzing complex mixtures; and 2) shotgun approaches by which samples are directly transferred into the mass spectrometer. Hyphenation with LC is important in lipidomics because isomers could not be differentiated by MS alone. Recent biochemical studies have used deparaffinizing and lysing procedures for lipid analysis of FFPE samples. Gaudin et al. [113] assessed the reliability of FFPE human brain samples for lipidomic analysis. Their study demonstrated that isolating brain lipids from formalin fixed tissue samples was a feasible approach.

As in other omics fields, metabolomics and lipidomics would greatly benefit from the ability to interrogate FFPE tissue specimens, acquired during routine medical examinations. Because of their widespread availability and long-term stability, accurate profiling of the metabolite content of these samples could accelerate the rate to discover clinically useful metabolomic biomarkers. Kelly et al. [107] were the first who described the analysis of polar metabolites from FFPE samples by LC/MS and found differences between malignant and non-malignant tissue samples based on more than a hundred molecules. Richter et al. used the same
analytical approach to assess the succinate / fumarate ratio in fresh frozen and FFPE samples [114] to reveal the effect of mutation in the succinate dehydrogenase gene (SDHx) in the sample. In other words, the succinate level increased, while fumarate decreased in the malignant state in individuals with SDHx mutation. Wojakowska et al. developed and validated a method to analyze primary metabolites from FFPE tissue samples using GC/MS technique and detected almost 70 compounds from the samples including even such sugars as glucose, galactose, erythrose and xylose [115].

4 Conclusion and future prospective
Surgical tissue fixation by formaldehyde has unique advantages, e.g., simple fixation process and easy storage; however, alternative fixatives can be also useful in special cases such as preserving high-molecular weight DNA. FFPE samples represent a rich source for prospective and retrospective biomedical and biopharmaceutical studies. In comparison to fresh frozen samples, they have longer storage stability and simple storage conditions. Many reports have been recently published about extraction of nucleic acids, proteins, complex carbohydrates and metabolites/lipids from FFPE samples for biomarker discovery. DNA isolated from tissue FFPE blocks proved to be prudent material for the genomics field with PCR amplification and in microarray studies. Proteins isolated from FFPE specimens have been used in Western blot analysis and reverse-phase protein arrays, in spite of the fact that they are cross-linked during the formalin fixation process. Options to analyze the carbohydrate moieties of glycoproteins include liquid chromatography, capillary electrophoresis and mass spectrometry after sugar release and derivatization or by lectin arrays or imaging mass spectrometry in their intact form. Powerful metabolomics methods are based on magnetic resonance spectroscopy and mass spectrometry, the latter technique dominants in lipidomics also. MALDI imaging is also a recently emerging technique in metabolomics/lipidomics. Figure 3 summarizes the analytical techniques mentioned in this review. Since molecular level analysis of FFPE samples have long been a challenge in the omics field, specific protocols have been developed for the investigation of such tissue samples as critically discussed in this review. In all omics subfields, the first major challenge with FFPE tissue samples was to create appropriate archiving and testing methods for tissue samples. The next challenge will be to reveal the biological information from these samples and to use them in daily routine in clinical practice. To accomplish this mission, it is necessary to properly elaborate, apply and control/inspect the examination methods/analysis [116].
With all the technological advances in the 21st century, we have a promising new opportunity to fully explore the biological information of FFPE specimens both in prospective and retrospective studies. Comprehensive analysis of this huge number of samples could reveal early diagnostic markers for a plethora of diseases including cancer and heart diseases, representing the Holy Grail for contemporary molecular diagnostics.

Acknowledgement
This work was supported by the MTA-PE Translational Glycomics program (#97101), the NKFIH K116263 grant and the BIONANO_GINOP-2.3.2-15-2016-00017 project. This is contribution #140 from the Horvath Csaba Laboratory of Bioseparation Sciences.

References


29


Figure legends

Figure 1: Crosslinking of the polypeptide chains during formalin fixation. Adapted from [117]
First step: the primary amine (nucleophilic group; R) forms a hydroxymethyl structure with the aldehyde. After losing a water molecule, a Schiff base is formed. Second step: a stable crosslink is produced between the proteins via methylene bridges.

1) \[ R - NH_2 + O = C \overset{\text{H}}{\text{H}} \rightarrow R - NH - C \overset{\text{H}}{\text{H}} \rightarrow R - N = C \overset{\text{H}}{\text{H}} + H_2O \]

2) \[ R - N = C \overset{\text{H}}{\text{H}} + NH_2 - R' \leftrightarrow R - NH - CH_2 - NH - R' \]

Figure 2: The formalin fixation and paraffin embedding process. i: sampling – it is usually done by a surgeon. ii: fixation – it takes place in a 10% buffered formalin solution for 24 to 48 hours at room temperature. If samples are spaded, penetration of formalin into the tissue is more sufficient. iii: paraffin embedding – after dehydration of the sample by consecutive double washing with ethanol and xylene, it is embedded in melted paraffin and allowed to solidify at room temperature. iv: The block is then cut into 3-5 µm slices for pathologic and molecular diagnostic investigation.
Figure 3: Summary of the analytical techniques cited in this paper. CE: capillary electrophoresis; GC/MS: gas chromatography/ mass spectrometry; HPLC: high-performance liquid chromatography; LC/MS: liquid chromatography / mass spectrometry; MALDI: matrix-assisted laser desorption/ionization; MALDI FTICR IMS: matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance imaging mass spectrometry; MS: mass spectrometry; MSI: mass spectrometry imaging; NMR: nuclear magnetic resonance spectroscopy; PGC-LC-ESI-MS/MS: porous graphitized carbon - liquid chromatography - electrospray ionization - tandem mass spectrometry; PCR: polymerase chain reaction; QRT-PCR: quantitative reverse transcription - polymerase chain reaction; RPPA: reverse phase protein array; SELDI-TOF-MS: surface-enhanced laser desorption/ionization time-of-flight mass spectrometry.
Table 1: The most frequently used fixatives for tissue preservation.

<table>
<thead>
<tr>
<th>Year of introduction</th>
<th>Fixative</th>
<th>Type of reaction</th>
<th>Target</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1893</td>
<td>Formalin</td>
<td>Crosslink</td>
<td>Protein, DNA</td>
<td>[11]</td>
</tr>
<tr>
<td>1908</td>
<td>Ethanol/Methanol</td>
<td>Coagulation</td>
<td>DNA</td>
<td>[23, 24]</td>
</tr>
<tr>
<td>1963</td>
<td>Glutaraldehyde</td>
<td>Crosslink</td>
<td>Nucleic acid/amino group</td>
<td>[20, 21]</td>
</tr>
<tr>
<td>1968</td>
<td>Carnoy’s fixative</td>
<td>Coagulation</td>
<td>RNA</td>
<td>[25, 29]</td>
</tr>
<tr>
<td>1970</td>
<td>Methacarn</td>
<td>Coagulation</td>
<td>RNA</td>
<td>[28, 30]</td>
</tr>
<tr>
<td>1970</td>
<td>Microwave irradiation(MW)</td>
<td>Stabilization</td>
<td>Protein</td>
<td>[14]</td>
</tr>
<tr>
<td>1992</td>
<td>AmeX</td>
<td>Dehydration</td>
<td>RNA, DNA</td>
<td>[32]</td>
</tr>
<tr>
<td>1994</td>
<td>Formalin + MW</td>
<td>Crosslink+ Stabilization</td>
<td>Protein</td>
<td>[17]</td>
</tr>
<tr>
<td>2000</td>
<td>Genipin</td>
<td>Crosslink</td>
<td>Amino acid</td>
<td>[22]</td>
</tr>
<tr>
<td>2001</td>
<td>HOPE</td>
<td>Dehydration</td>
<td>DNA, RNA</td>
<td>[34]</td>
</tr>
</tbody>
</table>