

# Mass Spectrometry Analysis of Intact Proteins from Crude Samples

Shay Vimer, Gili Ben-Nissan, and Michal Sharon\*



Cite This: *Anal. Chem.* 2020, 92, 12741–12749



Read Online

ACCESS |

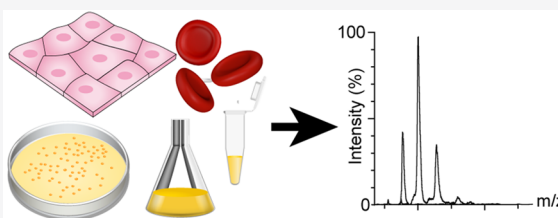


Metrics & More



Article Recommendations

**ABSTRACT:** Analysis of intact proteins by native mass spectrometry has emerged as a powerful tool for obtaining insight into subunit diversity, post-translational modifications, stoichiometry, structural arrangement, stability, and overall architecture. Typically, such an analysis is performed following protein purification procedures, which are time consuming, costly, and labor intensive. As this technology continues to move forward, advances in sample handling and instrumentation have enabled the investigation of intact proteins *in situ* and in crude samples, offering rapid analysis and improved conservation of the biological context. This emerging field, which involves various ion source platforms such as matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) for both spatial imaging and solution-based analysis, is expected to impact many scientific fields, including biotechnology, pharmaceuticals, and clinical sciences. In this Perspective, we discuss the information that can be retrieved by such experiments as well as the current advantages and technical challenges associated with the different sampling strategies. Furthermore, we present future directions of these MS-based methods, including current limitations and efforts that should be made to make these approaches more accessible. Considering the vast progress we have witnessed in recent years, we anticipate that the advent of further innovations enabling minimal handling of MS samples will make this field more robust, user friendly, and widespread.



Mass spectrometry (MS) has become an indispensable tool for protein analysis.<sup>5,6</sup> This is exemplified by the large-scale quantitative analysis of complex cellular systems, including the investigation of clinical samples by proteomics.<sup>3,7–9</sup> In this approach, proteins are enzymatically digested prior to MS analysis, and it is the peptide fragments that are identified and quantified. An alternative strategy, which is the focus of this Perspective, is MS-based analysis of intact proteins, in which the full-length polypeptide is examined.<sup>10–13</sup> When native conditions are used for such an analysis, not only the intact protein is detected but also the three-dimensional structure of the protein and its noncovalent interactions with protein partners and/or biomolecules are preserved.<sup>10–12</sup> This mode of analysis has been mainly the regime of native MS electrospray ionization (ESI) experiments, which has yielded unprecedented insights into the array of coexisting protein isoforms, the crosstalk between post-translational modifications (PTMs) and the structure and function of protein complexes ranging from ribosomes through viral capsids to membrane protein complexes.

Top-down analysis of intact proteins allows the identification and PTM mapping of individual protein components in the protein complex, while native MS studies the upper levels of the protein complex organization.<sup>2,14,15</sup> Both approaches, however, usually call for substantial sample preparation. Current top-down protocols, for example, often include cleanup strategies like protein precipitation, molecular weight cutoff ultrafiltration, liquid chromatography (LC) separation approaches (such as reversed phase, size exclusion,

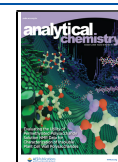
ion exchange and chromatofocusing), and capillary electrophoresis techniques (such as capillary zone electrophoresis and capillary isoelectric focusing).<sup>14</sup> Conversely, native MS studies are usually assayed with purified proteins,<sup>16</sup> but their purification procedures require pronounced time and labor investments and frequently involve prior molecular tagging strategies to enable affinity purification techniques and multiple biochemical isolation steps. Moreover, as most of the commonly used solutions for purification or storage of protein complexes contain buffers, salts, and solubilizing agents that are largely incompatible with MS, a buffer exchange step into a volatile buffer is necessary.<sup>16</sup>

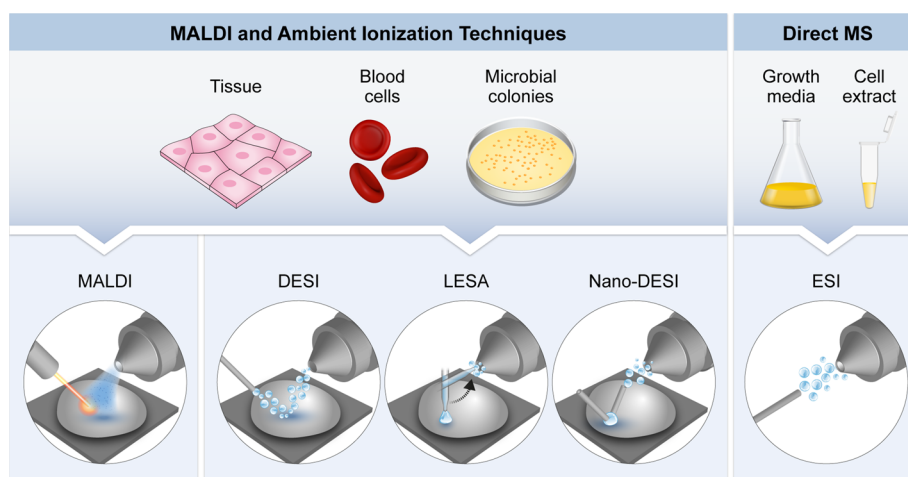
Therefore, methods for the rapid analysis of proteins that do not require multiple preparation steps, i.e., can be analyzed in their crude state, offer great advantages. Obviating the multistep sample preparation procedures and/or the need for protein purification will overcome this labor-intensive and costly process. There is also the advantage of better preserving the native state of the protein, thus maintaining PTMs and noncovalent associations of the protein with other biomolecules, such as ligands and cofactors, which are likely to be lost

Received: May 20, 2020

Accepted: September 8, 2020

Published: September 8, 2020





**Figure 1.** Scheme summarizing the various mass spectrometry ionization techniques for intact protein analysis from crude samples. *In situ* analysis of intact proteins is enabled by MS-based imaging approaches such as MALDI, DESI, LESA, and nano-DESI. These methods have successfully been used to probe intact proteins from tissue sections, dried blood spots, and microbial colonies. In contrast, the direct MS approach is based on an ESI-MS analysis of crude solution samples. It enables the assessment of cellular lysates of intracellular overexpressed proteins in bacterial systems or the growth medium of eukaryotic secretion systems. For further reading, several comprehensive reviews on the different ionization methods have been published.<sup>1–4</sup>

during lengthy procedures. Shortening the time lag between sample production and characterization will also be highly beneficial for screening programs of engineered constructs by enabling the rapid selection and ranking of lead protein candidates. Moreover, the ability to rapidly analyze proteins may also assist clinical diagnostics initiatives and pharmaceutical applications.

Traditionally, ESI-based methods have been used for the investigation of intact proteins. Indeed, there are emerging examples of harnessing the native MS method for the characterization of proteins from crude samples.<sup>17–21</sup> Recent studies have also put forward the power of *in situ* ambient surface sampling and matrix-assisted laser desorption ionization (MALDI) imaging MS methods for the analysis of proteins from surfaces, cells, and tissues with minimal sample handling (see reviews in refs 2 and 22–27). Although these *in situ* methods have been mainly harnessed for the investigation of metabolites and peptides, they are now being expanded toward protein analysis. In this Perspective, we highlight recent MS-based studies that take advantage of different sampling platforms to rapidly analyze intact proteins (Figure 1). We review the most relevant technical challenges associated with these approaches and present ideas about how to improve their performance and promote utilization in the future. We anticipate that the development of rapid protein analysis approaches will progress in the next decade, with increasing applications and usage.

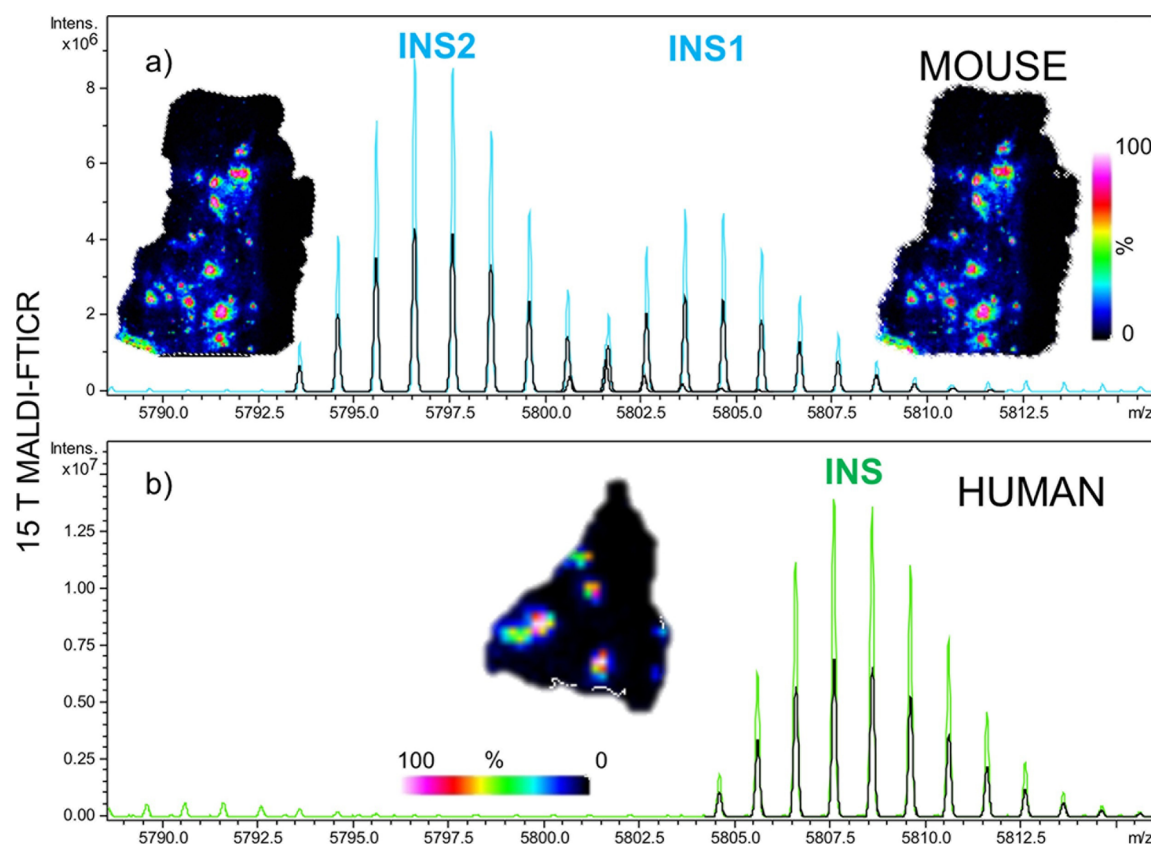
## ■ MALDI IMAGING MS

MALDI is a “soft” technique that can be used for the ionization of large biomolecules.<sup>28,29</sup> The method uses UV laser irradiation to photovolatilize samples that are cocrystallized with an energy-absorbent organic matrix (Figure 1). The activation of the sample–matrix mixture gives rise to matrix evaporation, which carries the sample with it into the mass spectrometer.<sup>30</sup> While the exact mechanism of ion generation is not fully understood, MALDI typically generates singly charged ions with small populations of doubly and triply charged ions.<sup>31</sup> As a result of its ability to analyze samples

directly from surfaces, MALDI-MS can be combined with histology in a process known as imaging mass spectrometry to obtain the spatial distribution of proteins directly from tissue sections.<sup>32</sup>

MALDI imaging of intact proteins directly from mammalian tissues started in the early 2000s. In one of the first examples, the spatial distribution of intact thymosin  $\beta$ .4 (~5 kDa) and S100 calcium binding protein A4 (S100A4) (10 kDa) was mapped.<sup>33</sup> More recent examples involve determining the expression, localization, and PTM profile of intact lens crystalline proteins ( $\alpha$ ,  $\beta$ , and  $\gamma$ ).<sup>34,35</sup> The method has also been extensively used for the study of intact histones in MCF7 breast cancer cells, which have indicated that histone H3 (15.2 kDa) and H4 (11.3 kDa) are highly abundant in tumor regions.<sup>36</sup> MALDI examination of mouse brains has revealed the existence of multiple histone H1 variants and their PTMs (20–25 kDa mass range).<sup>37</sup> This study also enabled mapping of the distribution of H1 variants in the brain, providing information not accessible by other methods, given the limited availability of suitable antibodies.

Identifying tissue- or tumor-specific proteoforms is challenging, especially if the mass difference between variants is small, as in the case of different point mutations, oxidation (16 Da), disulfide bond formation (2 Da), or amide formation (1 Da).<sup>38</sup> The resolving power of time-of-flight (TOF) mass analyzers may not be sufficient in such cases.<sup>39</sup> However, coupling a MALDI source with high-resolution mass analyzers such as Fourier transform ion cyclotron resonance (FTICR)<sup>40</sup> and Orbitrap instruments<sup>41</sup> that have high resolving powers and mass accuracy, can overcome this limitation and resolve highly similar proteoforms, including proteins with isobaric isotopic distributions. A recent MALDI-FTICR high-resolution imaging analysis of glioblastoma mouse brain tissue has identified a large number of intact proteins in the range of 4–15 kDa, including heavily modified histones. The imaging resolved many of the PTMs of histones H2, H3, and H4, whose acetylation and methylation status was dependent on whether tumor or healthy regions were analyzed.<sup>40</sup> Similarly, in another high-resolution MALDI-FTICR imaging study, sections of



**Figure 2.** MALDI-FTICR MS imaging of intact insulin from pancreatic tissues. Spectra and images of a mouse (A) and human (B) pancreas indicating the localization of the two mouse insulin isoforms INS1 and INS2 and the human protein INS, respectively, within the islets of Langerhans. The theoretical mass distribution is highlighted in blue (for mouse) and green (for human). Adapted with permission from ref 43. Copyright (2019) Elsevier.

mouse kidneys infected with *Staphylococcus aureus* were analyzed.<sup>42</sup> Twelve modified variants were identified for S100A8 (10.1 kDa), a protein subunit of the heterodimer calprotectin, with one form of the oxidized protein found to localize in the center of infectious foci, where staphylococcal microcolonies reside.<sup>42</sup> The high mass and spatial resolution of MALDI-FTICR imaging was used recently to determine the localization of the human insulin protein (INS) and to distinguish between the two mouse insulin proteins INS1 and INS2 within the pancreatic islets of Langerhans (Figure 2), demonstrating this method's ability to resolve highly similar proteins.<sup>43</sup>

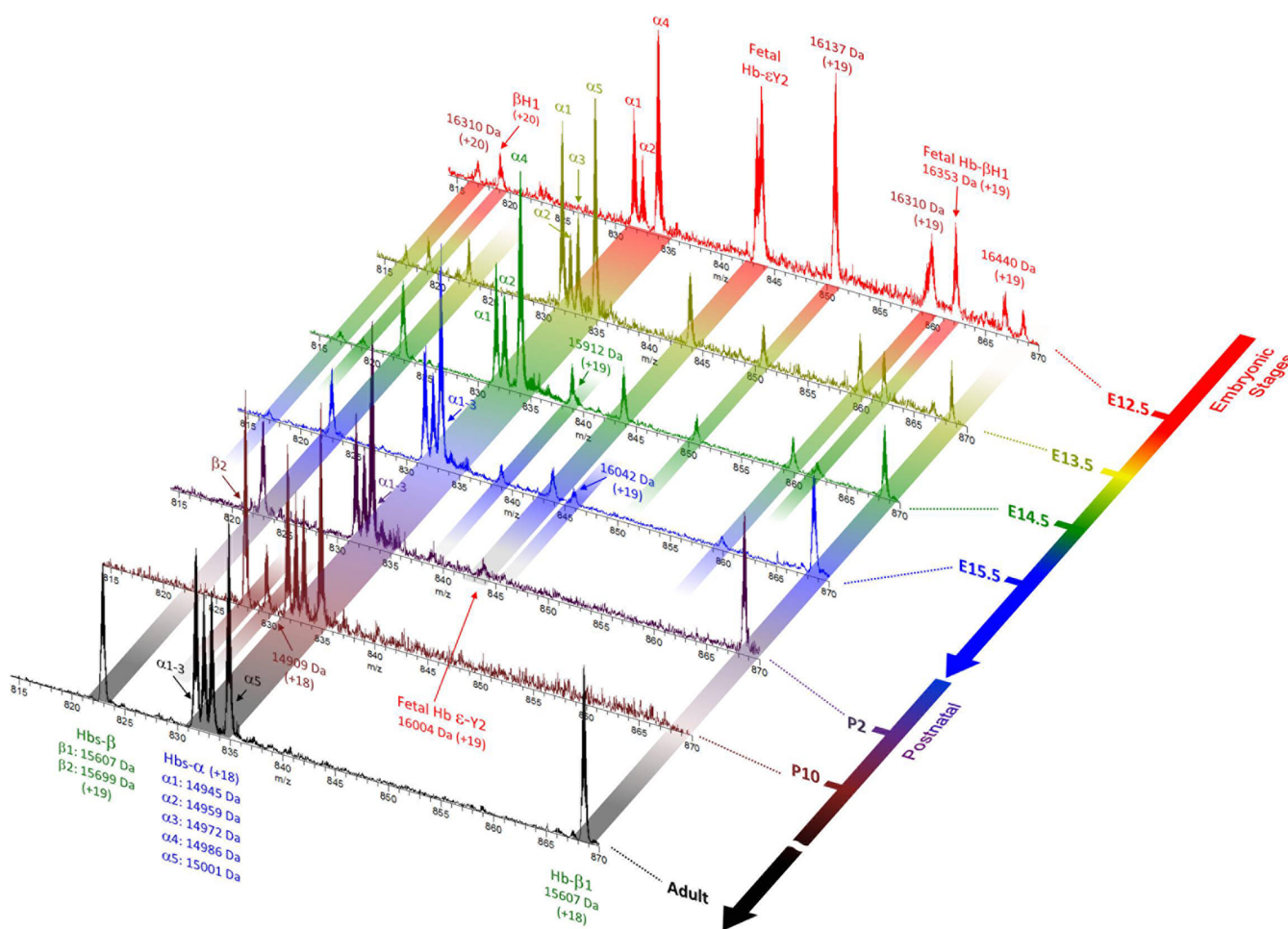
The benefit of MALDI imaging is that intact proteins are analyzed directly from tissues, without the isolation and purification process. Another significant advantage is the spatial molecular landscape it provides. In recent years, this aspect has been rapidly progressing with the development of instruments equipped with low-diameter laser beams, suitable for high spatial resolution imaging.<sup>44,45</sup> There is no doubt that the ability to resolve the spatial localization of multiple proteins within a single section of pathological tissue can enable the detection of disease candidates and improve our understanding of pathophysiology. Nevertheless, the method still requires extensive tissue preparation techniques.<sup>46</sup> Tissues must be pretreated with a matrix, and although multiple matrix options and coating methods are available, the commonly used matrices are acidic and contain organic solvents, which tend to denature the proteins of interest.<sup>47,48</sup> Thus, while intact proteins can be ionized in the presence of detergents and high

salt concentrations,<sup>48</sup> the preservation of protein–protein or protein–ligand interactions is challenging.<sup>25,49</sup>

Top-down strategies provide an efficient approach for the protein identification of MALDI imaging data, wherein intact protein measurements are followed by on-tissue MS/MS analysis. However, the typical low-charge ions generated by MALDI not only yield high  $m/z$  values, which may limit the detection of the intact protein, but also restrict the fragmentation efficiency.<sup>50,51</sup> Current attempts to overcome this hurdle focus on supercharging matrices, which are expected to increase the charge state of ions generated from tissue.<sup>52,53</sup> Another inherent limitation of MALDI imaging is that it is mainly suitable for small proteins, up to 25 kDa.<sup>24</sup> However, given the recent developments in the field, it is expected that in the future the method will expand toward the detection and imaging of higher-mass proteins. Progress in this direction is already being made by the use of the matrix ferulic acid, which remarkably increases signal acquisition in the mass range from 20 to 150 kDa.<sup>54</sup> An additional promising direction involves a new detector that enables proteins up to 110 kDa to be detected directly from the tissue and proteins of up to 70 kDa to be spatially resolved.<sup>55</sup>

## ■ AMBIENT SURFACE MS

Ambient surface sampling techniques are performed with minimal sample handling at atmospheric pressure and make use of ESI.<sup>2</sup> Specifically, the analysis of intact protein involves various liquid extraction techniques, in which target molecules are removed from the sample surface and extracted into a



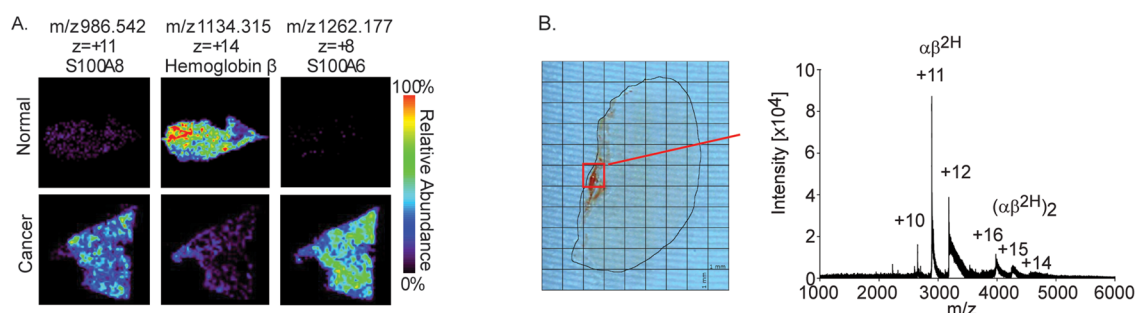
**Figure 3.** Mouse hemoglobin expression during development is characterized by coupling nano-DESI and microscopy. The data show the variety of hemoglobin subtypes and isoforms over time, performed directly on spinal cords tissue sections at different stages of mouse development. High expression levels of fetal isoforms are detected during early embryonic stages (E12.5–E15.5). During the developmental transition from fetal to adult, the hemoglobin forms gradually change: the fetal isoforms decline in number over time and disappear altogether at the postnatal stage, at which time the adult variant becomes the dominant form. Adapted with permission from ref 63. Copyright (2013) National Academy of Sciences.

solvent before the ionization process.<sup>56</sup> These liquid extraction approaches can be categorized into three main classes (Figure 1): (1) Spray-based techniques, such as desorption electrospray ionization mass spectrometry (DESI-MS), in which charged droplets are sprayed directly onto a sample of interest, while extracting the analytes from the surface and producing gaseous ions that are directed into the mass spectrometer.<sup>57</sup> (2) Direct liquid extraction techniques, such as liquid extraction surface analysis (LESA), in which the sample is extracted from the surface by dispensing a solvent on its surface from a pipet tip. This creates a liquid junction between the tip and the sample surface, allowing analytes to dissolve and reaspire with a conductive pipet tip prior to ESI-MS.<sup>2</sup> (3) Flow-based techniques, such as nanospray desorption electrospray ionization (nano-DESI), in which a solvent bridge is created between two capillaries and the surface of interest. The main difference between nano-DESI and LESA is the configuration of the capillaries. In LESA, only one capillary is responsible for sample extraction and ionization, while in nano-DESI two capillaries are used, a primary capillary which directs the solvent to the surface, where the sample material is dissolved, and a second capillary, which collects the sample and directs it into the nano-ESI mass spectrometer.<sup>58</sup> The liquid

junction created in nano-DESI is typically smaller than that of LESA, due to the low flow rate of the liquid onto the surface.<sup>59</sup>

Studies using liquid extraction surface sampling techniques for *in situ* analysis of intact proteins are starting to emerge. Hemoglobin, due to its high abundance in blood, represents one of the main examined systems. Using a DESI FTICR platform, measurements from hemolysate cells have allowed the identification of the intact hemoglobin  $\alpha$  and  $\beta$  chains ( $\sim 16$  kDa) as well as the released heme group (616 Da).<sup>60</sup> Another ambient surface MS study, this time involving a LESA-based analysis of intact proteins from dried blood spots, demonstrated this method's ability to identify multiple hemoglobin variants that lead to sickle cell disease and thalassemia in a rapid ( $<10$  min) and unambiguous manner.<sup>61,62</sup> Nano-DESI, in combination with light microscopy, has too been employed to investigate hemoglobin, specifically, the variation in its expression during mice development from fetal to adult stages (Figure 3).<sup>63</sup>

While the above examples were all conducted under denaturing conditions using high concentrations of organic solvents, more recent native MS studies have opted to employ ammonium acetate-based solvents.<sup>22,64</sup> In these studies, LESA-MS detected the intact tetrameric hemoglobin complex ( $\sim 64$



**Figure 4.** Intact-protein analysis by ambient surface MS. (A) DESI-MS images of normal and cancerous breast tissue samples reveal the different expression levels of intact S100 proteins and hemoglobin. The detected upregulation of the S100 proteins in breast cancer tissues is in accordance with previous records. Adapted with permission from ref 70 Copyright (2018) American Chemical Society. (B) LESA-MS analysis under native conditions of a mouse kidney section. (Left) Microscope image of a mouse kidney section overlaid with a pixel grid. (Right) Mass spectrum of an intact hemoglobin heterotetramer (+16) and heterodimer (+11) sampled directly from the tissue section labeled by a red square in the left panel, which marks the position of the renal pelvis, which is rich in blood vessels. Adapted with permission from ref 71. Copyright (2020) American Chemical Society.

kDa) bound to its heme group in dried blood spots<sup>64</sup> and in thin tissue sections of mouse liver.<sup>22</sup>

High-resolution instruments enable resolving isotopic patterns of proteins from tissue sections, which are then combined with spatial mapping. This has been exemplified by nano-DESI imaging of lymphoma and healthy thymus sections, in which truncated forms of proteins, such as thymosin  $\beta$ -4 and ubiquitin, were largely found in the tumor, particularly in the thymus region infiltrated by cancerous cells.<sup>65</sup> Notably, such truncations are highly likely to escape detection by the immuno-based techniques commonly used in biochemical studies, highlighting the benefits of this MS-based method. Similarly, a top-down LESA-MS analysis of healthy and diseased liver tissues was able to unambiguously distinguish between two different variants of a protein biomarker, the liver fatty acid binding protein (FABP1). The two highly similar forms of FABP1 differ only by one amino acid substitution, Thr  $\rightarrow$  Ala, with the Ala variant displaying increased risk of developing nonalcoholic fatty liver disease.<sup>66</sup> Another benefit of this approach, in comparison to proteomic analysis or immunohistochemistry, derives from the ability to define the assembly state of the analyzed protein complex. This has been shown in a LESA-based analysis of kidney tissues, which identified the presence of a 42 kDa homotrimer,<sup>23</sup> whose stoichiometry was confirmed by MS/MS.

The ambient surface techniques have also been expanded to the extraction of periplasmic and cytosolic proteins directly from living bacterial colonies, offering the means to analyze bacterial growth, communication, and response to external factors.<sup>67</sup> In one study using the *E. coli* K-12 laboratory model strain, a LESA extraction pipet tip was maneuvered into contact with the colony. Protein identification using top-down MS revealed the identity of six different proteins, among them DNA-binding and stress-resistant proteins, ranging in mass between 6.5 and 15.5 kDa.<sup>67</sup>

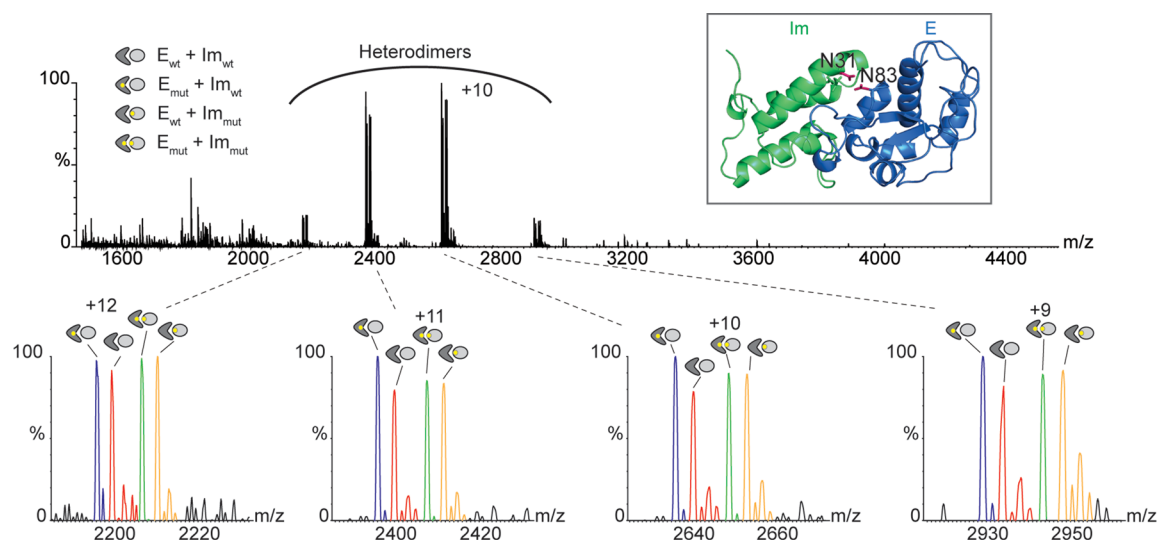
Unlike bacterial colonies, the analysis of yeast colonies is not possible using contact LESA because of their thick cell walls. Therefore, a combination of LESA-MS and electroporation was explored to lyse yeast colonies directly from growth agar media and release intact proteins for LESA-MS analysis.<sup>68</sup> Electroporation of different yeast species, such as *Saccharomyces cerevisiae*, *Candida glabrata*, and *Cryptococcus neoformans*, enabled the identification of different intact proteins, ranging in mass from 5 to 15 kDa. Two of the *S. cerevisiae*

proteins and 20 of the *C. glabrata* and *C. neoformans* proteins were detected, for the first time, in their intact form.

To reduce the overlap of charge states and increase the protein identification capacity, *in situ* ambient surface approaches were recently integrated with ion mobility separation. Significant improvement in the signal-to-noise ratio was achieved by the separation of highly complex overlapping signals of liver tissues through DESI-MS imaging via coupling the measurement with traveling wave ion mobility mass spectrometry (TWIMS). In this research, the intact mass and spatial distribution of the hemoglobin  $\alpha$  and  $\beta$  chains ( $\sim$ 16 kDa), a fatty acid binding protein (14.3 kDa), and a 10 kDa heat shock protein (CH10) could be detected.<sup>69</sup> In a similar effort to increase sensitivity so as to detect proteins that would have otherwise remained buried in the noise, DESI-MS was coupled to a high field asymmetric waveform ion mobility spectrometer (FAIMS), which allowed the detection of up to 16 proteoforms in mouse kidney, mouse brain, and human ovarian samples. The different expression levels of members of the S100 protein family were imaged directly in healthy and cancer breast tissue samples (Figure 4A).<sup>70</sup> Moreover, when such experiments are performed using native-like conditions, the collision cross sections of folded proteins can be profiled directly from tissues, as demonstrated by LESA-TWIMS for  $\beta$ -thymosin (4.9 kDa), ubiquitin (8.5 kDa), and tetrameric hemoglobin (64 kDa) (Figure 4B).<sup>71</sup> This type of analysis opens the way to studying protein–protein interactions or protein misfolding in disease tissues.

Taken together, ambient surface analysis techniques benefit from the fact that sample preparation or disruption procedures are not required, enabling preserving a more biologically relevant environment compared to MALDI imaging methods.<sup>2</sup> Moreover, the multiple charge states generated by electrospray ionization enables top-down analysis by a range of techniques, such as collision-induced dissociation (CID), electron capture dissociation (ECD), or electron transfer dissociation (ETD).<sup>2</sup> Another major advantage is the integration of ion mobility separation with the imaging workflows to analyze folded proteins and protein complexes in a spatially defined manner. Nevertheless, the method is still limited by the ionization efficiency of protein assemblies, restricting the analysis of low-abundance and large species.

Incomplete desolvation and the formation of protein–protein and protein–contaminant clusters have been shown to



**Figure 5.** Determining the strengths of intermolecular pairwise interactions between residues at the interfaces of interacting proteins by direct MS. A wild type colicin endonuclease (E) and a wild type immunity (Im) protein and two respective mutants (mut), in which the two interface residues (N83 and N31) were mutated to alanine, were overexpressed in the same bacterial cells. The crude lysate contained the four different complexes constituting the WT and mutated proteins. The pairwise interaction energy was calculated from a single high-resolution native mass spectrum directly from the crude lysate by measuring the intensities of the complexes formed by the two WT proteins (red peaks), the complex of each WT protein with a mutant protein (blue and orange peaks), and the complex of the two mutant proteins (green peaks). Yellow dots in the dimer cartoons represent the mutated interface residues. Inset shows the structures of the E-Im dimer. The mutated residues are labeled. Adapted with permission from refs 3 and 21. Copyright (2020) Springer Nature and Copyright (2018) American Chemical Society, respectively.

be responsible for the mass-dependent loss of sensitivity in large protein analyses.<sup>72</sup> Recent progress in mitigating these limitations has, however, been achieved by using solvent additives<sup>73</sup> and improving platform setups.<sup>74,75</sup> These technical advances suggest that it may become possible to analyze the spatial distribution of less-abundant intact protein complexes in tissues, aiding disease understanding and therapeutic avenues.

## ■ DIRECT MS

Direct MS is a new approach for the analysis of recombinant proteins under native conditions that does not require prior protein purification while allowing the rapid analysis of intact, overproduced proteins from crude samples, with minimal sample handling.<sup>76</sup> In this method, overexpressing bacterial cells are directly lysed in a volatile native MS compatible solution, and only the clearance of cell debris and insoluble materials is required before measurement.<sup>18</sup> In a similar manner, secreted proteins from eukaryotic hosts can be directly sprayed into the mass spectrometer from a precleared culture medium, following buffer exchange into a native MS-compatible solution.<sup>17</sup> Thus, analysis is based on the ESI-MS measurement of samples in solution, unlike the surface-based methods described above (Figure 1). The method can be carried out on multiple mass spectrometers, such as Orbitrap and quadrupole time-of-flight (QTOF)-based mass spectroscopy platforms, that enable intact protein detection.<sup>17–19</sup>

The key requirement for direct MS analysis is the overproduction of the target protein, such that it will outperform the cellular levels of the endogenous proteins.<sup>17–19</sup> In its basis, the method relies on a drawback of MS, its limited dynamic range, wherein low-abundance proteins are masked by higher-abundance ones.<sup>77</sup> This inherent property is turned into an advantage that enables biased detection of the highly produced target protein, disregarding the lower-abundant

endogenous proteins. Thus, signal suppression enables overcoming the need for prior protein purification.

Recombinant proteins produced in *E. coli*, the most commonly used bacterial expression system,<sup>78,79</sup> can accumulate several milligrams of protein per milliliter of lysate, making them an excellent system for direct MS analysis.<sup>18</sup> Production quantities are usually lower in eukaryotic hosts, challenging the ability to detect proteins that are overexpressed intracellularly. Secretion systems, however, overcome this bottleneck by transferring the recombinant protein from the cell to the growth medium, making it the dominant protein in the culture despite the presence of the background endogenous proteins in the medium.<sup>80</sup> Therefore, for direct MS, the use of protein-free growth media is recommended in order to eliminate ion suppression and superposition of signals due to a high background of medium proteins, particularly when expression levels are low.<sup>17</sup>

The direct MS method has been applied to a broad array of proteins and protein complexes, ranging from 20 kDa monomers to 185 kDa protein complexes.<sup>17,18</sup> The approach is particularly useful for characterizing engineered proteins, as it provides rapid assessment of their quality along with structural input for iterative redesign and optimization.<sup>17,19</sup> This aspect has been shown in the analysis of designed antibodies, demonstrating not only the proper folding, assembly state, and glycan modification of the generated antibodies but also their improved expression levels and stability compared to the wild type forms.<sup>20</sup> In another example, the method was utilized for a “quick and dirty” determination of proteins’ interactions and function.<sup>19</sup> RAB1A, a Ras-related protein, was suspected to bind the 20S proteasome and inhibit its proteolytic activity. MS/MS experiments using a cell lysate expressing RAB1A in the presence of the 20S proteasome confirmed the binding of the two, while time course analysis of the levels of a 20S

proteasome substrate in this lysate confirmed that RAB1A coordinates the proteasome activity.<sup>19</sup>

Recently, direct MS was adjusted so as to measure pairwise interaction strengths from crude lysates (Figure 5).<sup>21</sup> The approach is based on the double mutant-cycle method, wherein the two target residues are mutated both separately and in combination, usually to alanine, and the energetic effects of the mutations are determined.<sup>81</sup> An earlier study has shown that pairwise interaction energies can be determined from a single mass spectrum.<sup>82</sup> On the basis of this understanding, four different proteins, consisting of the two wild type (wt) and two mutated (mut) variants, were coexpressed in the same bacterial cells.<sup>21</sup> The spectrum measured from the crude lysate showed the four coexisting complexes (wt/wt, wt/mut, mut/wt, and mut/mut), from which the coupling energy was calculated, obviating the need to purify each of the four proteins and determine individual binding constants for each of the generated dimers.

Unlike the *in situ* surface techniques described above, the direct MS method, although less restricted by the protein or protein complex size, is constrained to the analysis of overproduced recombinant targets rather than endogenous proteins. Consequently, while it provides in-depth analysis of PTMs and profiles of coexisting variants,<sup>17,19</sup> it should be noted that the relative abundance of the different modified forms of the target protein may not reflect the true biological distribution. Moreover, the method is limited to the use of expression systems that enable the target proteins to become the dominant component in the sample. Therefore, current applications using eukaryotic expression systems are limited to the analysis of recombinant secreted proteins rather than those produced intracellularly. This limitation may be overcome by the development of improved expression plasmids that can scale up the yield. Another way to overcome the relative abundance limitation is to simplify existing fractionation methods employed in the native MS analysis of complex protein lysates.<sup>83–85</sup> Currently, however, these strategies require a significant amount of sample handling, which is not only labor intensive but may also result in the loss of important noncovalent associations. The recent development of charge detection MS,<sup>86,87</sup> which allows detailed analyses of high mass and heterogeneous samples, is also expected to advance direct characterization of endogenous proteins from crude lysates. Alternatively, data processing techniques may be adapted from the *in-cell* NMR method, in which background signals are significantly reduced by subtracting from the original spectrum a spectrum acquired on a control sample of cells transfected with an empty vector, where protein expression did not occur.<sup>88</sup>

We expect that further extension of this method will involve its progression to high-throughput analysis. Recent progress toward this direction has been made by the development of an online buffer exchange method for clarifying cell lysates.<sup>89</sup> Similarly, improvements in cell lysis techniques, like microwell arrays,<sup>90</sup> high-throughput electromechanical lysis,<sup>91</sup> or pressure-based methods for tissue lysis,<sup>92</sup> may facilitate throughput analysis. Another expected future direction involves coupling to microfluidic systems, which will enable the continued analysis of cultures grown in a bioreactor and the capture in real time of the state of protein production.

## CONCLUDING REMARKS

Herein, we propose that the MS analysis of intact proteins from crude samples, with minimal perturbations to their structure, modifications, and/or assembly, opens up multiple opportunities for rapid characterization and high-throughput screening. Specifically, the approaches described above, although fundamentally differing from each other in terms of the characteristics of the sample (surface analysis vs solution), all benefit from the relative simplicity and feasibility across different MS platforms. They also benefit from the minimal amount of sample required, the absence of labeling, the reduced time gap between sample production and characterization, and the in-depth information afforded by MS.

We anticipate, for example, that the direct MS method that involves solution analysis will facilitate protein engineering and successful mass production of recombinant proteins for both industrial and pharmaceutical applications, achieving batch-to-batch consistency and maximum productivity at a reasonable cost. The screening of surfaces, as exemplified by the ambient ionization methods, is foreseen to be translated into benchtop analytical devices for on-site tests, such as intraoperative tumor margin assessment and environmental, forensics, and defense applications.<sup>27</sup> Moreover, given the promising results of MALDI and ambient MS methods, these methods are expected to be highly beneficial for the rapid identification of microorganisms, analysis of biofluids, and diagnostic studies.<sup>26</sup> Thus, a critical aspect in the coming years will be to keep evolving these approaches into a robust and user-friendly set of technologies.

Despite the diverse potential applications of the MS techniques for protein analyses with minimal handling, they are mostly used in certain specialized laboratories, and their translation into everyday life practice is hindered due to multiple challenges. For instance, all the described methods, regardless of the applied ionization method, will benefit from the development of robust, easy-to-use, cost-effective devices that can be employed routinely by nonexperts. Automation of platforms and their integration with robotic laboratories are also expected to expand the outreach of the techniques. There is also a need for improving the ionization efficiency, overcoming signal suppression issues, and developing high-throughput capabilities. Considering that, irrespective of the method by which the protein sample is introduced into the mass spectrometer, fragmentation of the intact protein must be used for identification (as CID, ECD, and ETD), improvements in these capabilities along with automated assignment tools will advance *de novo* sequence analysis and unambiguous interpretation. Moreover, as the advanced methods outlined in this Perspective are often adjusted to one analytical aspect, combining different types of techniques may bring new advances. Finally, as the field grows to maturity, standardization of the methodologies for clinical, industrial, and pharmaceutical applications also will be required.

## AUTHOR INFORMATION

### Corresponding Author

Michal Sharon – Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel; [orcid.org/0000-0003-3933-0595](https://orcid.org/0000-0003-3933-0595); Phone: (+972)-8-934-3947; Email: [michal.sharon@weizmann.ac.il](mailto:michal.sharon@weizmann.ac.il); Fax: (+972)-8-934-6010

## Authors

Shay Vimer – Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel

Gili Ben-Nissan – Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.analchem.0c02162>

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We are grateful for the support of the Israel Science Foundation (ISF) Grant 300/17. M.S. is the incumbent of the Aharon and Ephraim Katzir Memorial Professorial Chair.

## REFERENCES

- Hale, O. J.; Cooper, H. J. *Biochem. Soc. Trans.* **2020**, *48*, 317–326.
- Kocurek, K. I.; Griffiths, R. L.; Cooper, H. J. *J. Mass Spectrom.* **2018**, *53*, 565–578.
- Vimer, S.; Ben-Nissan, G.; Sharon, M. *Nat. Protoc.* **2020**, *15*, 236–265.
- Baker, T. C.; Han, J.; Borchers, C. H. *Curr. Opin. Biotechnol.* **2017**, *43*, 62–69.
- Domon, B.; Aebersold, R. *Science* **2006**, *312*, 212–217.
- Mann, M.; Hendrickson, R. C.; Pandey, A. *Annu. Rev. Biochem.* **2001**, *70*, 437–473.
- Doll, S.; Gnad, F.; Mann, M. *Proteomics: Clin. Appl.* **2019**, *13*, No. 1800113.
- Uzozie, A. C.; Aebersold, R. *J. Proteomics* **2018**, *189*, 1–10.
- Mallick, P.; Kuster, B. *Nat. Biotechnol.* **2010**, *28*, 695–709.
- Leney, A. C.; Heck, A. J. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 5–13.
- Sharon, M. *Science* **2013**, *340*, 1059–1060.
- Mehmood, S.; Allison, T. M.; Robinson, C. V. *Annu. Rev. Phys. Chem.* **2015**, *66*, 453–474.
- Toby, T. K.; Fornelli, L.; Kelleher, N. L. *Annu. Rev. Anal. Chem.* **2016**, *9*, 499–519.
- Donnelly, D. P.; Rawlins, C. M.; DeHart, C. J.; Fornelli, L.; Schachner, L. F.; Lin, Z.; Lippens, J. L.; Aluri, K. C.; Sarin, R.; Chen, B.; Lantz, C.; Jung, W.; Johnson, K. R.; Koller, A.; Wolff, J. J.; Campuzano, I. D. G.; Auclair, J. R.; Ivanov, A. R.; Whitelegge, J. P.; Pasa-Tolic, L.; et al. *Nat. Methods* **2019**, *16*, 587–594.
- Benesch, J. L.; Ruotolo, B. T.; Simmons, D. A.; Robinson, C. V. *Chem. Rev.* **2007**, *107*, 3544–3567.
- Hernandez, H.; Robinson, C. V. *Nat. Protoc.* **2007**, *2*, 715–726.
- Ben-Nissan, G.; Vimer, S.; Warszawski, S.; Katz, A.; Yona, M.; Unger, T.; Peleg, Y.; Morgenstern, D.; Cohen-Dvashi, H.; Diskin, R.; Fleishman, S. J.; Sharon, M. *Commun. Biol.* **2018**, *1*, 213.
- Gan, J.; Ben-Nissan, G.; Arkind, G.; Tarnavsky, M.; Trudeau, D.; Noda Garcia, L.; Tawfik, D. S.; Sharon, M. *Anal. Chem.* **2017**, *89*, 4398–4404.
- Vimer, S.; Ben-Nissan, G.; Morgenstern, D.; Kumar-Deshmukh, F.; Polkinghorn, C.; Quintyn, R. S.; Vasil'ev, Y. V.; Beckman, J. S.; Elad, N.; Wysocki, V. H.; Sharon, M. *ACS Cent. Sci.* **2020**, *6*, 573–588.
- Warszawski, S.; Borenstein Katz, A.; Lipsh, R.; Khmel'nitsky, L.; Ben-Nissan, G.; Javitt, G.; Dym, O.; Unger, T.; Knop, O.; Albeck, S.; Diskin, R.; Fass, D.; Sharon, M.; Fleishman, S. J. *PLoS Comput. Biol.* **2019**, *15*, No. e1007207.
- Cveticanin, J.; Netzer, R.; Arkind, G.; Fleishman, S. J.; Horovitz, A.; Sharon, M. *Anal. Chem.* **2018**, *90*, 10090–10094.
- Griffiths, R. L.; Cooper, H. J. *Anal. Chem.* **2016**, *88*, 606–609.
- Griffiths, R. L.; Konijnberg, A.; Viner, R.; Cooper, H. J. *Anal. Chem.* **2019**, *91*, 6962–6966.
- Aichler, M.; Walch, A. *Lab. Invest.* **2015**, *95*, 422–431.
- Chen, F.; Gulbakan, B.; Weidmann, S.; Fagerer, S. R.; Ibanez, A. J.; Zenobi, R. *Mass Spectrom. Rev.* **2016**, *35*, 48–70.
- Dudley, E. *Adv. Exp. Med. Biol.* **2019**, *1140*, 27–43.
- Kuo, T. H.; Dutkiewicz, E. P.; Pei, J.; Hsu, C. C. *Anal. Chem.* **2020**, *92*, 2353–2363.
- Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, *60*, 2299–2301.
- Karas, M.; Bachmann, D.; Bahr, U.; Hillenkamp, F. *Int. J. Mass Spectrom. Ion Processes* **1987**, *78*, 53–68.
- Zaluzec, E. J.; Gage, D. A.; Watson, J. T. *Protein Expression Purif.* **1995**, *6*, 109–123.
- Zenobi, R.; Knochenmuss, R. *Mass Spectrom. Rev.* **1998**, *17*, 337–366.
- Ryan, D. J.; Spraggins, J. M.; Caprioli, R. M. *Curr. Opin. Chem. Biol.* **2019**, *48*, 64–72.
- Stoeckli, M.; Chaurand, P.; Hallahan, D. E.; Caprioli, R. M. *Nat. Med.* **2001**, *7*, 493–496.
- Anderson, D. M.; Nye-Wood, M. G.; Rose, K. L.; Donaldson, P. J.; Grey, A. C.; Schey, K. L. *J. Mass Spectrom.* **2020**, *55*, No. e4473.
- Grey, A. C.; Schey, K. L. *Mol. Vis* **2008**, *14*, 171–179.
- Djidja, M.-C.; Francese, S.; Loadman, P. M.; Sutton, C. W.; Scriven, P.; Claude, E.; Snel, M. F.; Franck, J.; Salzter, M.; Clench, M. R. *Proteomics* **2009**, *9*, 2750–2763.
- Lahiri, S.; Sun, N.; Solis-Mezarino, V.; Fedisch, A.; Ninkovic, J.; Feuchtinger, A.; Götz, M.; Walch, A.; Imhof, A. *Proteomics* **2016**, *16*, 437–447.
- Tong, H.; Bell, D.; Tabei, K.; Siegel, M. M. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1174–1187.
- Norris, J. L.; Cornett, D. S.; Mobley, J. A.; Andersson, M.; Seeley, E. H.; Chaurand, P.; Caprioli, R. M. *Int. J. Mass Spectrom.* **2007**, *260*, 212–221.
- Dilillo, M.; Ait-Belkacem, R.; Esteve, C.; Pellegrini, D.; Nicolardi, S.; Costa, M.; Vannini, E.; Graaf, E. L. d.; Caleo, M.; McDonnell, L. A. *Sci. Rep.* **2017**, *7*, 603.
- Hecht, E. S.; Scigelova, M.; Eliuk, S.; Makarov, A. *Fundamentals and Advances of Orbitrap Mass Spectrometry*; Meyers, R.A., Ed.; Wiley, 2019.
- Spraggins, J. M.; Rizzo, D. G.; Moore, J. L.; Rose, K. L.; Hammer, N. D.; Skaar, E. P.; Caprioli, R. M. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 974–985.
- Piga, I.; Heijs, B.; Nicolardi, S.; Giusti, L.; Marselli, L.; Marchetti, P.; Mazzoni, M. R.; Lucacchini, A.; McDonnell, L. A. *Int. J. Mass Spectrom.* **2019**, *437*, 10–16.
- Smith, A.; L'Imperio, V.; Denti, V.; Mazza, M.; Ivanova, M.; Stella, M.; Piga, I.; Chinello, C.; Ajello, E.; Pieruzzi, F.; Pagni, F.; Magni, F. *Proteomics: Clin. Appl.* **2019**, *13*, No. 1800016.
- Norris, J. L.; Caprioli, R. M. *Chem. Rev.* **2013**, *113*, 2309–2342.
- Neagu, A. N. *Adv. Exp. Med. Biol.* **2019**, *1140*, 55–98.
- Morisasa, M.; Sato, T.; Kimura, K.; Mori, T.; Goto-Inoue, N. *Foods* **2019**, *8*, 633.
- Madler, S.; Boeri Erba, E.; Zenobi, R. *Top. Curr. Chem.* **2012**, *331*, 1–36.
- Beaufour, M.; Ginguene, D.; Le Meur, R.; Castaing, B.; Cadene, M. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 1981–1994.
- Spraggins, J. M.; Rizzo, D. G.; Moore, J. L.; Rose, K. L.; Hammer, N. D.; Skaar, E. P.; Caprioli, R. M. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 974–985.
- Flangea, C.; Schiopu, C.; Capitan, F.; Mosoarca, C.; Manea, M.; Sisu, E.; Zamfir, A. D. *Central European Journal of Chemistry* **2013**, *11*, 25–34.
- Inutan, E. D.; Richards, A. L.; Wager-Miller, J.; Mackie, K.; McEwen, C. N.; Trimpin, S. *Mol. Cell. Proteomics* **2011**, *10*, M110.000760.
- Trimpin, S.; Inutan, E. D.; Herath, T. N.; McEwen, C. N. *Mol. Cell. Proteomics* **2010**, *9*, 362–367.
- Mainini, V.; Bovo, G.; Chinello, C.; Gianazza, E.; Grasso, M.; Cattoretti, G.; Magni, F. *Mol. Biosyst.* **2013**, *9*, 1101–1107.
- van Remoortere, A.; van Zeijl, R. J.; van den Oever, N.; Franck, J.; Longuespee, R.; Wisztorski, M.; Salzter, M.; Deelder, A. M.;



- Fournier, I.; McDonnell, L. A. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1922–1929.
- (56) Laskin, J.; Lanekoff, I. *Anal. Chem.* **2016**, *88*, 52–73.
- (57) Takáts, Z.; Wiseman, J. M.; Gologan, B.; Cooks, R. G. *Science* **2004**, *306*, 471–473.
- (58) Roach, P. J.; Laskin, J.; Laskin, A. *Analyst* **2010**, *135*, 2233–2236.
- (59) Addie, R. D.; Balluff, B.; Bovee, J. V.; Morreau, H.; McDonnell, L. A. *Anal. Chem.* **2015**, *87*, 6426–6433.
- (60) Takats, Z.; Kobliha, V.; Sevcik, K.; Novak, P.; Kruppa, G.; Lemr, K.; Havlicek, V. *J. Mass Spectrom.* **2008**, *43*, 196–203.
- (61) Edwards, R. L.; Creese, A. J.; Baumert, M.; Griffiths, P.; Bunch, J.; Cooper, H. J. *Anal. Chem.* **2011**, *83*, 2265–2270.
- (62) Edwards, R. L.; Griffiths, P.; Bunch, J.; Cooper, H. J. *Proteomics* **2014**, *14*, 1232–1238.
- (63) Hsu, C.-C.; White, N. M.; Hayashi, M.; Lin, E. C.; Poon, T.; Banerjee, I.; Chen, J.; Pfaff, S. L.; Macagno, E. R.; Dorrestein, P. C. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 14855–14860.
- (64) Martin, N. J.; Griffiths, R. L.; Edwards, R. L.; Cooper, H. J. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1320–1327.
- (65) Hsu, C.-C.; Chou, P.-T.; Zare, R. N. *Anal. Chem.* **2015**, *87*, 11171–11175.
- (66) Sarsby, J.; Martin, N. J.; Lalor, P. F.; Bunch, J.; Cooper, H. J. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1953–1961.
- (67) Randall, E. C.; Bunch, J.; Cooper, H. J. *Anal. Chem.* **2014**, *86*, 10504–10510.
- (68) Kocurek, K. I.; Havlikova, J.; Buchan, E.; Tanner, A.; May, R. C.; Cooper, H. J. *Anal. Chem.* **2020**, *92*, 2605–2611.
- (69) Towers, M. W.; Karancsi, T.; Jones, E. A.; Pringle, S. D.; Claude, E. J. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 2456–2466.
- (70) Garza, K. Y.; Feider, C. L.; Klein, D. R.; Rosenberg, J. A.; Brodbelt, J. S.; Eberlin, L. S. *Anal. Chem.* **2018**, *90*, 7785–7789.
- (71) Hale, O. J.; Sisley, E. K.; Griffiths, R. L.; Styles, I. B.; Cooper, H. J. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 873–879.
- (72) Douglass, K. A.; Venter, A. R. *J. Mass Spectrom.* **2013**, *48*, 553–560.
- (73) Honarvar, E.; Venter, A. R. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1109–1117.
- (74) Ambrose, S.; Housden, N. G.; Gupta, K.; Fan, J.; White, P.; Yen, H. Y.; Marcoux, J.; Kleanthous, C.; Hopper, J. T. S.; Robinson, C. V. *Angew. Chem., Int. Ed.* **2017**, *56*, 14463–14468.
- (75) Hale, O. J.; Illes-Toth, E.; Mize, T.; Cooper, H. J. *Anal. Chem.* **2020**, *92*, 6811.
- (76) Vimer, S.; Ben-Nissan, G.; Sharon, M. *Nat. Protoc.* **2020**, *15*, 236–265.
- (77) Bakalarski, C. E.; Elias, J. E.; Villen, J.; Haas, W.; Gerber, S. A.; Everley, P. A.; Gygi, S. P. *J. Proteome Res.* **2008**, *7*, 4756–4765.
- (78) Karyolaimos, A.; Ampah-Korsah, H.; Hillenaar, T.; Mestre Borrás, A.; Dolata, K. M.; Sievers, S.; Riedel, K.; Daniels, R.; de Gier, J. W. *Front. Microbiol.* **2019**, *10*, 1511.
- (79) Hochkoeppler, A. *Biotechnol. Lett.* **2013**, *35*, 1971–1981.
- (80) Dumont, J.; Ewart, D.; Mei, B.; Estes, S.; Kshirsagar, R. *Crit. Rev. Biotechnol.* **2016**, *36*, 1110–1122.
- (81) Hidalgo, P.; MacKinnon, R. *Science* **1995**, *268*, 307–310.
- (82) Sokolovski, M.; Cveticanin, J.; Hayoun, D.; Korobko, I.; Sharon, M.; Horovitz, A. *Nat. Commun.* **2017**, *8*, 212.
- (83) Shen, X.; Kou, Q.; Guo, R.; Yang, Z.; Chen, D.; Liu, X.; Hong, H.; Sun, L. *Anal. Chem.* **2018**, *90*, 10095–10099.
- (84) Belov, A. M.; Viner, R.; Santos, M. R.; Horn, D. M.; Bern, M.; Karger, B. L.; Ivanov, A. R. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 2614–2634.
- (85) Skinner, O. S.; Haverland, N. A.; Fornelli, L.; Melani, R. D.; Do Vale, L. H. F.; Seckler, H. S.; Doubleday, P. F.; Schachner, L. F.; Srzentic, K.; Kelleher, N. L.; Compton, P. D. *Nat. Chem. Biol.* **2018**, *14*, 36–41.
- (86) Keifer, D. Z.; Pierson, E. E.; Jarrold, M. F. *Analyst* **2017**, *142*, 1654–1671.
- (87) Kafader, J. O.; Melani, R. D.; Durbin, K. R.; Ikwuagwu, B.; Early, B. P.; Fellers, R. T.; Beu, S. C.; Zabrouskov, V.; Makarov, A. A.; Maze, J. T.; Shinholt, D. L.; Yip, P. F.; Tullman-Ercek, D.; Senko, M. W.; Compton, P. D.; Kelleher, N. L. *Nat. Methods* **2020**, *17*, 391–394.
- (88) Barbieri, L.; Luchinat, E.; Banci, L. *Nat. Protoc.* **2016**, *11*, 1101–1111.
- (89) VanAernum, Z. L.; Busch, F.; Jones, B. J.; Jia, M.; Chen, Z.; Boyken, S. E.; Sahasrabudhe, A.; Baker, D.; Wysocki, V. H. *Nat. Protoc.* **2020**, *15*, 1132–1157.
- (90) Dura, B.; Choi, J.-Y.; Zhang, K.; Damsky, W.; Thakral, D.; Bosenberg, M.; Craft, J.; Fan, R. *Nucleic Acids Res.* **2019**, *47*, No. e16.
- (91) Kim, M.; Wu, L.; Kim, B.; Hung, D. T.; Han, J. *Anal. Chem.* **2018**, *90*, 872–880.
- (92) Lucas, N.; Robinson, A. B.; Marcker Espersen, M.; Mahboob, S.; Xavier, D.; Xue, J.; Balleine, R. L.; deFazio, A.; Hains, P. G.; Robinson, P. J. *J. Proteome Res.* **2018**, *18*, 399–405.