A Role for the MS Analysis of Nucleic Acids in the Post-Genomics Age

Daniele Fabris
Department of Chemistry and Biochemistry, University of Maryland Baltimore County, Baltimore, Maryland, USA

The advances of mass spectrometry in the analysis of nucleic acids have tracked very closely the exciting developments of instrumentation and ancillary technologies, which have taken place over the years. However, their diffusion in the broader life sciences community has been and will be linked to the ever evolving focus of biomedical research and its changing demands. Before the completion of the Human Genome Project, great emphasis was placed on sequencing technologies that could help accomplish this project of exceptional scale. After the publication of the human genome, the emphasis switched toward techniques dedicated to the exploration of sequences not coding for actual protein products, which amount to the vast majority of transcribed elements. The broad range of capabilities offered by mass spectrometry is rapidly advancing this platform to the forefront of the technologies employed for the structure-function investigation of these noncoding elements. Increasing focus on the characterization of functional assemblies and their specific interactions has prompted a re-evaluation of what has been traditionally construed as nucleic acid analysis by mass spectrometry. Inspired by the accelerating expansion of the broader field of nucleic acid research, new applications to fundamental biological studies and drug discovery will help redefine the evolving role of MS-analysis of nucleic acids in the post-genomics age. (J Am Soc Mass Spectrom 2010, 21, 1–13) © 2010 Published by Elsevier Inc. on behalf of American Society for Mass Spectrometry

The development of technologies for the analysis of nucleic acids by mass spectrometry (MS) has followed step-by-step the evolution of the field, with the introduction of electrospray ionization (ESI) [1, 2] and matrix assisted laser desorption ionization (MALDI) [3, 4] representing watershed events that clearly delimit two different eras. Before the advent of these ionization techniques, the realm of application consisted mainly of the analysis of fundamental units and components of limited size, such as nucleosides (Figure 1) [5, 6], nucleotides [7, 8], and small oligonucleotides [9–11], which often required derivatization of their highly polar functional groups to increase volatility and to improve analytical performance. Overcoming inherent technical challenges, this era established MS as an irreplaceable tool for the characterization of rare nucleotides produced by normal post-transcriptional processes [12–14] and aberrant chemical damage [15–17]. Efforts aimed at elucidating the building blocks of these biopolymers culminated in the creation of a comprehensive database of RNA modifications (available at http://library.med.utah.edu/RNAmods/) [18–20], which today constitutes an invaluable resource for the broader life sciences community. Proceeding forward in this type of pursuit, MS-technologies continue to support the identification of DNA adducts produced by alkylating agents, carcinogens, and antitumor drugs [21–24].

Making Headways in Biomedical Research

The introduction of ESI and MALDI made progressively larger nucleic acids accessible to MS analysis (Figure 2) [25–28], which ushered in a period of rapid expansion into different areas of biomedical research. The challenges posed by the larger analytes were answered by the development of new strategies and experimental methods (reviewed in references [29–33]), which sought to take full advantage of favorable characteristics exhibited by these ionization techniques, such as their intrinsic low-energy and ability to handle highly polar species. In the early 1990s, the rapid advance of these technologies was largely fueled by the demand for analytical platforms capable of affording reliable and cost-effective DNA sequencing, which was associated with the inception of the Human Genome Project [34, 35]. Different approaches were devised to obtain unambiguous sequence information through the gas-phase dissociation of selected precursor ions in MS/MS experiments [36–38], or the molecular mass determination of ladder products obtained in solution by terminating chain elongation [39–41], or inducing...
partial degradation of target sequences (Figure 3) [42–44] (reviewed in references [45–47]). The expectations were that high-throughput MS sequencing would play a significant role in this project of unprecedented scale by handling a fair share of the ~3 billion base pair genome. Additionally, the new approaches were expected to make MS the platform of choice for applications in medical diagnostics and for securing the information necessary to implement personalized medicine. By the mid-1990s, however, it was clear that such approaches could not overtake the concomitant advances made by sequencing strategies based on electrophoresis [48–50], which were capable of handling longer DNA segments, afforded greater multiplexing capabilities, and were significantly less expensive [51]. In 1997, a well-publicized editorial conceded that competitive MS technologies would not be ready to contribute before the then-loomiing completion of the Human Genome Project, but predicted that a bright future was still in store for genotyping and diagnostic applications [52]. This prediction was realized by the introduction of technologies for the rapid characterization of single nucleotide polymorphisms [53–55] and for the detection/identification of infectious pathogens [56–58]. At the time of this editorial, however, the broader MS community had been already shifting its attention toward the burgeoning field of MS-based proteomics [59–61].

Confronting New Challenges in the Post-Genomics Era

The completion of the Human Genome Project has brought to light surprising facts and posed new challenges [62–64]. Its results have led to conservative

![Figure 1](image-url)
estimates that less than 1.5% of the entire genome may be coding for actual proteins, in spite of the observation that at least 70% of its sequence is transcribed into RNA [65, 66]. Once upon a time dismissed as “selfish” or “junk DNA” [67, 68], the intragenic (i.e., intronic) and intergenic sequences representing the vast majority of the transcribed pool are now undergoing extensive re-evaluation. Indeed, the last decade has witnessed resurgent interest in the elucidation of the functions of noncoding elements (ncRNA) [69–71]. Although new classes of functional RNA are being discovered on a regular basis, their total amount still does not account for the entire transcribed pool [72]. These discoveries have keenly pointed out that sequence information alone is not sufficient to reveal the function of all classes of ncRNAs and to explain their mechanism of action. This conclusion is substantiated, for example, by the investigation of the activity of microRNAs (miRNAs) and riboswitches, which are receiving a great deal of attention for their participation in complex gene-regulation networks. In the case of miRNAs [73, 74], their ~22-nucleotide sequence enables specific recognition of mRNA targets in the cell [75], but the factors driving the proper selection of cognate proteins to produce active RNA-protein complexes (miRNPs) or determining whether a certain miRNP may induce down- or up-regulation of protein production remain to be elucidated [76]. In the case of riboswitches [77, 78], expression of downstream gene(s) is modulated by conformational changes induced by the interaction of a specific cellular component (e.g., metabolite, metal ion, etc.) with the 3D structure of the sensing domain, which cannot be predicted or explained according to the simple rules of base-pairing recognition [79, 80]. These examples clearly highlight the need for technologies capable of providing not only the higher-order structure of non-coding elements, but also unambiguous information about the identity of cognate ligands, the nature of their interactions, and the effects of binding on structure and dynamics. Beyond its well-developed sequencing capabilities, MS possesses unique strengths that could advance this analytical platform to the forefront of the technologies employed for the discovery and structure-function elucidation of noncoding nucleic acids.

The potential of MS-technologies for large scale analysis of cellular RNAs has been demonstrated by the implementation of fingerprinting strategies based on endonuclease digestion of total tRNA extracts, followed by MALDI-MS detection of signature products (Figure 4) [81, 82]. Analogous to approaches employed in quantitative proteomics, 18O end-labeling by nuclease digestion can be employed to evaluate RNA production and observe possible variations in the amount of post-transcriptionally modified nucleotides [83]. The development of these types of approaches will be expected to benefit the growing field of experimental RNomics by complementing current technologies based on microarray techniques and cDNA-based “deep sequencing,” which can provide comprehensive views of entire cellular transcriptomes [84, 85]. In fact, the dependence of established technologies on amplification/sequencing and electrophoretic analysis makes them ill-suited for the detection of possible post-transcriptional modifications, which is likely the cause of our insufficient understanding of the extent and biological significance of ncRNA modifications. MS-based strategies for the investigation of post-transcriptional modifications on a full transcriptome scale, analogous to those available for the identification of post-translational modifications of proteins on a full proteome scale, could provide missing information on the biological activities of noncoding elements and take their study in unexpected directions.

The most significant contributions to experimental RNomics, however, could arguably arise from the availability of complementary ionization/desorption techniques, which have pushed MS closer to becoming a...
universal analytical platform applicable to samples of virtually any type and origin. This uncanny flexibility could be harnessed to achieve the characterization of any ligand co-purified with a target nucleic acid of interest, which could represent the constitutive elements of a putative functional assembly. To this effect, affinity chromatography and related “pull-down” strategies have been routinely employed to recognize cognate proteins interacting with specific RNA [86] and DNA [87] sequences (reviewed in references [88 –90]). Similar strategies could be readily adapted to identify any type of biomolecule or small molecule ligand that may be capable of binding a certain construct with sufficient affinity, taking advantage of a growing number of metabolome databases for correct assignments [91–94]. This type of scheme could represent an excellent complement to microarray technologies involving the immobilization of specific RNA sequences covering the desired portion of the transcribed pool. Reversing the terms of this experimental design, the ligand of interest could be immobilized to enable the isolation of any nucleic acid structure with specific binding activity, thus leading to the identification of cognate sequences that may be involved with the selected ligand in biologically active assemblies. High-throughput schemes could be implemented to perform systematic screening of sequences capable of binding series of metabolites involved in the same or parallel pathways, or representative compounds from selected classes of xenobiotics. The potential in this direction is still largely untapped.

Elucidating Structure-Function Relationships

Mass spectrometry offers additional capabilities that make it an invaluable platform for the study of the structure-function relationships of these types of nu-

Figure 4. MALDI-MS spectra obtained from total E. coli tRNA after digestion with either RNase T1 (a) and (b), or RNase A (c). These experiments demonstrated that the detection of unique digestion products, or signatures, can provide the identification of individual isoacceptors present in the tRNA pool of a certain cell. Reproduced from reference [81], with permission.
cleic acids. The activity of most noncoding elements known to date is associated with their ability to establish specific interactions with other cellular components. Therefore, their elucidation places a premium on the possibility of observing intact functional assemblies and probing their salient structural features. The advent of soft ionization techniques and the advances of analyzers design have made this possibility a reality. Indeed, relatively short duplexes formed by complementary deoxyoligonucleotides were among the first noncovalent complexes detected by ESI without unwelcome dissociation [95–97]. Although this desirable outcome can be obtained also by MALDI with a judicious selection of matrix, additives, and other experimental conditions [98–101], ESI remains the ionization technique of choice for the detection of nucleic acid noncovalent assemblies (reviewed in references [102–105]).

Direct ESI-MS analysis of nucleic acid complexes with other nucleic acids [106–110], proteins [111–116], and small molecule ligands [117–120] can reveal their exact composition and stoichiometry from the observed molecular mass, dispensing with the typical curve-fitting of bulk data required by spectroscopic and calorimetric methods. Unlike these techniques, MS is capable of resolving any free/bound species at equilibrium in solution, even when such species possess very similar spectroscopic characteristics. With proper experimental design and data treatment, their respective signal intensities can be employed to obtain relative [121, 122] and absolute [123–128] dissociation constants (K_d’s) in solution, matching those afforded by established methods [129]. In this direction, competitive binding experiments in which multiple ligands are mixed simultaneously with the substrate of interest have proven very effective in providing relative scales of binding affinities based on the aspect ratio and distribution of the detected complexes [130–133]. Reversing the terms of the competition scheme, a ligand of interest can be also added to multiple substrates simultaneously, for example, to directly compare the effects of subtle variations in substrate structure on binding mode (Figure 5) [134]. In more rigorous fashion, quantitative determinations of K_d can be accomplished by following titration schemes in which the amount of ligand is increased stepwise and the abundances of free/bound species in solution are determined after each addition [125–128]. Possible dissociative effects induced by adding a certain ligand to a preformed complex can be also assessed to obtain quantitative determinations of inhibition constants (K_i’s) or ligand concentrations inducing 50% dissociation of the initial complex (IC_50’s) [135]. It is expected that these approaches will become increasingly popular as desirable features, such as limited sample consumption and exemption from radio- or fluorescent-labeling, will become better known within the broader life sciences community.

In addition to information attainable by taking
“snapshots” of species at equilibrium in solution, mass spectrometry can provide further insights into the nature of the interactions involved in functional nucleic acid assemblies by studying them in the gas-phase. The dissociation of these species has been activated directly in the ESI source [136–138], or after ion selection in MS/MS experiments [139–141], to assess their stability and to glean valuable details about determinant structural features [142–145]. For example, the dissociation of duplex structures in the absence/presence of ligands has been investigated to demonstrate the preservation of Watson-Crick pairing and stacking interactions in the gas phase and to elucidate the effects of selected binders on the stability of these interactions [142–146]. Conversely, analogous experiments have been completed to study the effects of ligand structure on binding mode [110, 136, 147]. A great deal of effort has been dedicated over the years to understand the fundamentals, to conciliate gas-phase observations with solution behaviors, and to enable predictions of experimental outcome from structural information and vice versa.

Finally, mass spectrometry can be also employed to elucidate higher-order structures of nucleic acids and their functional assemblies by following a variety of solution and gas-phase approaches. Chemical footprinting and bifunctional crosslinking have been combined with MS analysis to reveal the position of base-pairing interactions, which define elements of secondary structure of nucleic acids, or to identify long-range contacts between discrete domains, which outline tertiary structures and determine the global fold of the larger species [148–151]. The spatial constraints obtained by these approaches constitute valid input for performing molecular modeling by established computational methods, thus enabling the creation of all-atom 3D models based on actual experimental data (Figure 6) [152, 153]. Crosslinking strategies can also help identify the contacts between nucleic acids and cognate proteins [154–157] and reveal the organization of multi-subunit functional complexes [158–160]. The possible application of hydrogen/deuterium exchange (HDX) to characterize the contact interfaces between assembled units has been hampered by the fast exchange rate exhibited in solution by the nucleic acid components [161, 162]. However, this approach has been successfully employed to highlight the protein surfaces in contact with them and to study the dynamics associated with the specific interactions [163–165]. The implementation of HDX in the gas phase has been demonstrated for nucleotides [166, 167], unstructured oligonucleotides [168], and relatively small structured constructs [169, 170], but its application to the larger assemblies has yet to be reported.

Gas-phase approaches have been developed to map ligand sites and to gain precious insights into the architecture of large multi-subunit assemblies with nucleic acid components. Under suitable activation conditions, the binding of selected ligands to RNA substrates can prevent underlying nucleotides from undergoing the typical backbone fragmentation that produces the characteristic ion series employed in MS/MS sequencing. This observation has promoted strategies for achieving the characterization of specific binding sites onto target nucleic acid structures, which are revealed by recognizable gaps in the detected ion series [134, 171, 172]. Conversely, the same protection effects can be employed to screen libraries of small molecule ligands for their ability to bind to desired structural motifs, which is evaluated from their power to induce site-directed inhibition of nucleic acid fragmentation [123, 173]. In the absence of covalent fragmentation, the order by which bound subunits dissociate as discrete products may reveal their spatial situation within the complex of interest. Indeed, submitting intact ribosomes and their constitutive subunits to collisional activation induced the progressive release of protein units according to their affinity for each other and for the RNA components, but also according to their placement relative to the surface of the initial precursor ion [138, 174]. In similar fashion, the dissociation of RNA-RNA pairing interactions was employed to interrogate the conformational state of assemblies folding into alternative isomeric forms [175]. In this example, the pattern afforded by collisional activa-
tion clearly differentiated dimeric complexes capable of assuming either a loop-loop kissing or an extended-duplex conformation in the presence of their cognate viral protein (Figure 7), which provided new insights into the mechanism of structure remodeling mediated by the chaperone protein. The conformational study of progressively larger assemblies will be expected to take full advantage of recent exciting developments in ion mobility mass spectrometry, which has been already shown to be capable of affording direct information about nucleic acid structures in the gas phase [169, 176].

The Road Ahead

The unquestionable benefits offered by MS technologies for nucleic acid analysis come with intrinsic challenges posed by the very nature of these biomolecules. Some of these challenges are general and apply to any analytical platform, such as those concerning the integrity of RNA samples, which require utilizing RNase-free solvents, plastic ware, and tools to stave off rapid degradation. Others are specific to the MS platform, such as those originating from the presence of phosphate groups in the biopolymer structure, which affect the selection of sample preparation and mode of analysis. Indeed, the negative charges afforded by these functional groups confer nucleic acids a strong tendency to form unwanted cation adducts that can deteriorate the attainable resolution and signal to noise ratio. Over the years, alternative strategies have been devised to replace metal cations with the more volatile ammonium [42, 177], which employ ion-exchange [178], reversed-phase high-performance liquid chromatography [179], metal chelation [180, 181], ethanol precipitation [177], ultrafiltration, and microdialysis [182–184]. Although separation-based desalting is very effective for these types of biomolecules, it may also result in the unwanted dissociation of their noncovalent assemblies, either through excessive reduction of the solution ionic strength, or through direct perturbation of binding equilibria. In these cases, a viable alternative consist of performing ion-ion reactions with specific metal chelators, which can be accomplished in the gas phase after the desorption process is complete and equilibrium considerations are no longer a concern [185].

The presence of phosphate groups makes negative ion mode the preferred mode of operation for the MS analysis of nucleic acids. However, the sensitivity limits achieved in this polarity appear to be consistently worse than those afforded in the positive ion mode by the same instrumentation/technique. While this practical observation is well-known in the MS community, its causes are still not understood. Addressing this discrepancy will be critical for the direct analysis of cellular...

Figure 7. Collision-induced dissociation (CID) mass spectra obtained from a 3:2 protein-RNA assembly folded in either extended-duplex or loop-loop kissing conformation, (a) and (b), respectively. Proteins are represented by circles, RNA strands by triangles. Squares mark the remaining precursor ions. The extended-duplex complex displayed sequential loss of protein units, whereas the loop-loop conformer underwent dissociation of interstrand pairings to provide 1:1 protein-RNA products. Adapted with permission from reference [175].
analyses without additional amplification procedures. In the same direction, gas-phase activation techniques, such as electron capture dissociation (ECD) [186] and electron-transfer dissociation (ETD) [187], have been extensively investigated for sequence analysis of proteins and peptides in cationic form. However, the exploration of equivalent methods for the activation of anionic precursors and their application to nucleic acids are still at a very early stage [188].

Clearly, the multifaceted capabilities summarized here provide a strong rationale for a broader application of mass spectrometry to the structure-function investigation of nucleic acids. If the ever increasing demands of the post-genomics age extend beyond the need for mere sequence information, a new definition of what can be considered as MS analysis of nucleic acids should be formulated, which should include also the analysis of their complexes with other biomolecules, physiologic ligands, and xenobiotics. In many cases, the physical-chemical characteristics and structural stability of these functional assemblies are still largely determined by the nucleic acid components, even when such components amount to only a small fraction of the entire complex under investigation. In the case of protein-nucleic acid conjugates, for example, a high density of negatively-charged phosphates may affect the overall charge balance and may become a critical factor in the selection of polarity and mode of operation. A higher propensity to form cation adducts may command the selection of polarity and mode of operation. A higher charge balance and may become a critical factor in the entire complex under investigation. In the case of proteins and peptides in cationic form. However, the more commonly associated with analytes of purely nucleic acid nature. For these reasons, while the protein moiety could still be perceived as the critical component according to strict functional considerations, analytical considerations could make the nucleic acid moiety the focus of attention, due to the more stringent requirements imposed on the MS analysis. Often, what we call “substrate” or “ligand” becomes a question of perspective.

Now, if the range of information accessible by MS-based technologies is so astonishingly wide, then why isn’t their application more popular in the broader life sciences community? Part of the answer may be found in the intrinsic technical challenges posed by these types of experiments, which may require know-how that has yet to spread beyond a few dedicated laboratories. However, the progressive refinement of these technologies, the streamlining of experimental strategies, the complementary capabilities and advantages over established methodologies, as well as the nearly capillary diffusion of MS instrumentation, should contribute to a considerable expansion in coming years. Part of the answer may be found in the actual size of the public engaged in this field of research, which is still a small portion of the broader life sciences community and includes only a handful of commercial enterprises with a continued stake in nucleic acids. After all, the proportion of nucleic acid-targeted drugs in the market is miniscule in comparison with that of protein-targeted therapeutics. However, with less than 1.5% of the human genome coding for proteins, the potential for discovering viable targets among the remaining transcribed elements is too large to be ignored. MS-technologies will be expected to play a very significant role in the exploration of this extensive druggable space, supporting traditional tasks in drug discovery, as well as new strategies for the identification of possible targets and therapeutic candidates. When antisense and gene silencing technologies will overcome the last remaining hurdles that hamper their widespread introduction in human therapy, MS will be uniquely placed to accomplish the operations required by manufacturing and regulatory monitoring. Combined with new applications to fundamental biological studies, the activities spurred by the changing demands of a rapidly developing field will help redefine the evolving role of MS analysis of nucleic acids in the post-genomics age.

Acknowledgments

This research was funded by the National Institutes of Health (R01-GM43208 and R21-NS063871) and the National Science Foundation (CHE-0439067).

References

102. Loo, J. A. Studying Noncovalent Protein Complexes by Electrospray
Beck, J. L.; Colgrave, M. L.; Ralph, S. F.; Shell, M. E. Mass Spectrometry.
Hofstadler, S. A.; Griffey, R. H. Analysis of Noncovalent Complexes of
Goodlett, D. R.; Camp, D. G. II; Hardin, C. C.; Corregan, M.; Smith,
Wunschel, D. S.; Fox, K. F.; Fox, A.; Bruce, J. E.; Muddiman, D. C.;
Smidt, R. D. Direct Measurement of Oligonucleotide Binding Stoichiometry of
93(14), 7022–7027.

103. Cheng, X.; Morin, P. E.; Harms, A. C.; Bruce, J. E.; Ben-David, Y.; Smith,

104. David, W. M.; Brodbelt, J. S.; Kerwin, S. M.; Thomas, P. W. Investigation

105. David, W. M.; Brodbelt, J. S.; Kerwin, S. M.; Thomas, P. W. Investigation


108. Naylor, J. A.; Moreland, D. W.; Sannes-Lowery, K. A.; Sharmeen, L.; Truong,


