Published in final edited form as:

J Am Soc Mass Spectrom. 2014 July ; 25(7): 1136–1145. doi:10.1007/s13361-014-0911-2.

Structure-Specific Ribonucleases for MS-Based Elucidation of Higher-Order RNA Structure

Matteo Scalabrin, Yik Siu, Papa Nii Asare-Okai, Daniele Fabris

The RNA Institute, University at Albany-SUNY, Albany, NY 12222, USA

Abstract

Supported by high-throughput sequencing technologies, structure-specific nucleases are experiencing a renaissance as biochemical probes for genome-wide mapping of nucleic acid structure. This report explores the benefits and pitfalls of the application of Mung bean (Mb) and V1 nuclease, which attack specifically single- and double-stranded regions of nucleic acids, as possible structural probes to be employed in combination with MS detection. Both enzymes were found capable of operating in ammonium-based solutions that are preferred for high-resolution analysis by direct infusion electrospray ionization (ESI). Sequence analysis by tandem mass spectrometry (MS/MS) was performed to confirm mapping assignments and to resolve possible ambiguities arising from the concomitant formation of isobaric products with identical base composition and different sequences. The observed products grouped together into ladder-type series that facilitated their assignment to unique regions of the substrate, but revealed also a certain level of uncertainty in identifying the boundaries between paired and unpaired regions. Various experimental factors that are known to stabilize nucleic acid structure, such as higher ionic strength, presence of Mg(II), etc., increased the accuracy of cleavage information, but did not completely eliminate deviations from expected results. These observations suggest extreme caution in interpreting the results afforded by these types of reagents. Regardless of the analytical platform of choice, the results highlighted the need to repeat probing experiments under the most diverse possible conditions to recognize potential artifacts and to increase the level of confidence in the observed structural information.

Graphical Abstract

Correspondence to: Daniele Fabris; fabris@albany.edu.

Electronic supplementary material The online version of this article (doi:10.1007/s13361-014-0911-2) contains supplementary material, which is available to authorized users.

Keywords

RNA; Higher-order structure; Structure-specific nucleases

Introduction

Biomolecules in living systems operate by establishing direct functional interactions with other cellular components, which are driven by their physical-chemical properties and structural features. In the case of nucleic acids, the indissoluble relationship between structure and function cannot be more evident than in the activity of ribozymes and riboswitches, which is not defined by the information encoded in their sequences, but by their ability to interact with cognate ligands [1, 2]. The absence of a direct link between sequence and function for a vast array of non-protein coding sequences—the bulk of the human genome—represents a major obstacle to the elucidation of their biological significance, which is still far from being understood in spite of the unprecedented wealth of available genome information [3, 4]. The need for comprehensive structural information has increased the demands placed on traditional high-resolution techniques, which have to contend with intrinsic limitations associated with sample size, availability, purity, and crystallization properties. Driven by these demands, increasing effort has been dedicated to the development of computational methods for predicting higher-order structure directly from sequence [5]. The great advances made in this direction have increased the appeal of alternative approaches for securing spatial constraints that may be immediately employable in model-building operations [6, 7]. The ever-improving accuracy of structure prediction algorithms has placed a premium on the availability of actual experimental data for model refinement and validation.

The 3D structure of nucleic acids is defined by the hierarchic combination of secondary, tertiary, and quaternary elements, which are stabilized by specific base pairing interactions [8, 9]. Watson-Crick pairing of contiguous sequences results in the formation of duplex and stem loop domains that represent fundamental elements of secondary structure. Annealing of complementary sequences located on distal sections or separate strands can lead to the formation of tertiary and quaternary structures, respectively. In this context, the preeminent role played by pairing interactions justifies the emphasis placed on their characterization to complete the structural determination of nucleic acids. For this purpose, a variety of experimental approaches have been developed over the years to identify and locate base pairs, which rely on the different chemical reactivity exhibited by paired versus unpaired nucleotides [10–12]. For example, dimethylsulfate and similar footprinting agents are typically employed to alkylate functional groups on the pairing edge of nucleobases, which are left unprotected in single-stranded regions exposed to the solvent [10–12]. In contrast, N-methylisatoic anhydride and other flexibility probes provide indirect information on base-pairing by attacking the 2'-hydroxyl of ribose units that bear unpaired or otherwise unconstrained nucleobases [13, 14]. Employed in concert or individually, these approaches are capable of providing comprehensive maps of pairing interactions, which can be translated into valid secondary and tertiary structures.

In recent years, we have explored the application of mass spectrometry (MS) as a possible platform for the characterization of probing products [15–17]. Traditional detection schemes involve electrophoretic analysis of oligonucleotide ladders that are generated either by strand cleavage at the modification site [10–12], or probe-induced inhibition of enzymatic primer extension [12–14]. Replacing electrophoresis with the MS platform allowed us to dispense with probe-specific schemes and to pursue approaches that may be applicable across the board with virtually any type of reagent [18]. In particular, we have explored the virtues of bottom-up strategies in which probe application is followed by digestion with specific ribonucleases to enable mass mapping and sequencing by tandem mass spectrometry (MS/MS) [15–17]. We have tested possible top-down strategies in which the probed substrates are submitted directly to gas-phase activation with no prior hydrolytic procedure [19]. The excellent flexibility afforded by mass spectrometry allowed us to expand the repertoire of structural probes to include bifunctional crosslinking agents, which can offer direct information on the mutual arrangement of contiguous domains, but are unsuitable for electrophoretic applications [20, 21]. We demonstrated the validity of the spatial constraints provided by these approaches by completing the determination of actual 3D structures [22, 23]. The studies illustrated also the ability of advanced computational tools to utilize these types of experimental data and offered a glimpse of the potential of these alternative strategies for structural elucidation.

In this report, we have investigated the merits of combining the MS platform with structure-specific nucleases to identify base pairing interactions. The spectrum of enzymes capable of cleaving the phosphodiester linkage includes nucleases that attack specific types of nucleotides in solvent-accessible regions, restriction enzymes that recognize well-defined sequence motifs, and nucleases that bind to distinct structural features with no preferred nucleotide or recognition sequence. For example, RNase A and T1 are nucleotide-specific enzymes that cleave the accessible 3'-end of C/U and G, respectively [24, 25]. In contrast, the activity of Mung bean (Mb) and V1 nuclease is directed towards single- or double-stranded regions, respectively, where they induce hydrolysis of any type of nucleotide [26–28]. In recent years, structure-specific nucleases have experienced a renaissance as essential ingredients of high-throughput approaches for mapping secondary structures at the genome-wide level [29]. These methods require converting the hydrolytic products into cDNA libraries or performing primer extension amplification to enable sequence characterization. The MS platform could represent a valid alternative for unknown sequences that stymie primer design, or shorter constructs that may not allow adequate priming. For this reason, we have explored possible strategies for combining structure-specific nucleases with MS detection and for employing them in concert with other structural probes. We investigated possible factors that may influence the specificity of enzymatic activity and their effects on the quality of the attained structural information.

Experimental

Sample Preparation and Probing Conditions

All RNA substrates were obtained from commercial sources (i.e., Integrated DNA Technologies, Coralville, IO, USA), or prepared in vitro by T7 polymerase transcription of

appropriate DNA templates. The template plasmid for human tRNA^{Lys3} was a generous gift by Dr. K. Musier-Forsyth, whereas that for 5'-UTR was prepared in-house. Sample purification was accomplished by gel electrophoresis, followed by electro-elution from manually excised bands. Desalting was performed by ultrafiltration against 150 mM ammonium acetate (pH adjusted to 7.0) in Millipore (North Bend, OH, USA) Centricon devices with 3 kDa MWCO. Mb and V1 nucleases were purchased from Promega (Madison, WI, USA) and Life Technologies (Carlsbad, CA, USA), respectively, and used without any further purification. RNase T1 was purchased from ThermoFisher Scientific (Waltham, MA, USA) and used without further purification. In typical probing reactions, a 50 μ L aliquot of 20 μ M solution of RNA was mixed with an appropriate amount of enzyme to achieve a final 50 pmol:100 units in the case of Mb and a 50 pmol:0.01 units for V1 nuclease, unless otherwise indicated. Each reaction mixture was subsequently incubated at 4°C overnight (~14 h), unless otherwise noted. A wide range of conditions was explored by varying the amount of Mb from 10 to 200 units and that of V1 from 0.001 to 0.1 units per 50 pmol substrate. The effects of incubation temperature were also explored by performing digestion at 4, 25, and 37°C. Each reaction was quenched either by completing ethanol precipitation or by performing immediate MS analysis. In general, ethanol precipitation was preferred when initial analysis indicated the need for a desalting step, which was dictated by the quality of the starting RNA sample. Samples were diluted prior to analysis to achieve a final 1–5 μ M concentration of total RNA in 150 mM ammonium acetate.

Mass Spectrometric Analysis

Samples were analyzed by direct infusion electrospray ionization (ESI) on ThermoFisher Scientific LTQ-Orbitrap Velos mass spectrometer. All analyses were performed in nanoflow mode by using quartz emitters produced in house by using a Sutter Instruments Co. (Novato, CA, USA) P2000 laser pipette puller. Up to 5 μ L samples were typically loaded onto each emitter by using a gel-loader pipette tip. A stainless steel wire was inserted through the back-end of the emitter to supply an ionizing voltage that ranged between 0.8 and 1.2 kV. Source temperature and desolvation conditions were adjusted by closely monitoring the incidence of ammonium adducts and water clusters, with typical source temperature of 200°C and 35 V desolvation voltage [30]. Tandem mass spectrometry (MS/MS) experiments involved isolating the precursor ion of interest in the LTQ element of the instrument, activating fragmentation in either the LTQ or the C-trap, and performing fragment detection in the Orbitrap. Duplex dissociation was typically achieved by using an activation voltage of 20 V, whereas strand fragmentation for sequencing purposes was achieved with a 50 V voltage. The instrument was calibrated by using a 0.5 mg/mL solution of CsI in water, which provided a typical <2 ppm mass accuracy. Data interpretation was accomplished with the aid of Mongo Oligo Calculator [31] and DataSnack 1.0 (developed in house).

Results and Discussion

Characteristic Features of Structure-Specific Nuclease Activity

Selected elements of RNA secondary structure were employed to investigate the determinants of the activity of structure-specific nucleases. The presence of discrete single- and double-stranded regions makes stemloop structures ideal targets for testing both Mb and

V1 nuclease. For this reason, a 20 mer construct replicating the stemloop 3 (SL3) domain of the HIV-1 packaging signal [15, 32] was used to monitor the products obtained under a variety of experimental conditions (see Experimental). Nuclease activity was initially tested in solutions in which typical biological buffers were replaced with MS-friendly ammonium equivalents, for the purpose of enabling direct infusion analysis of digest mixtures without separation procedures. As shown in Figure 1, abundant signals corresponding to hydrolytic products were readily detected by direct electrospray ionization (ESI) [33, 34], thus confirming the ability of these enzymes to operate in ammonium-based environments. Product identity was assigned by matching experimental masses with possible values calculated from sequence (Table 1). Initial interpretation corroborated by tandem mass spectrometry (MS/MS) [35, 36] experiments indicated that the observed species did not possess the terminal 3'-phosphate that is characteristic of fragments produced by RNase A and T1 digestion. These types of products were consistent with specific cleavage at the 3' side of the phosphodiester bond, which left the phosphate group attached to the 5'-end of the released moiety. In contrast, RNase A and T1 attack the 5' side of the phosphodiester bond and leave a terminal 3'-phosphate (in either cyclic or linear form), consistent with decidedly different mechanisms of action.

The ends of the products were carefully examined to identify possible preferences for specific types of nucleotides. The absence of any particular trend confirmed the lack of base specificity attributed to Mb and V1 cleavage [26–28]. Instead, numerous products appeared to form ladder series centered on the same region, but terminating with consecutive nucleotides in the sequence (e.g., n, n+1, n+2, etc.). In general, the average sizes of hydrolytic products were comparatively larger than those of fragments generated under identical conditions by base-specific nucleases, such as RNase T1 (Supplemental Material, Table 1S). This observation was not surprising in light of the high frequency of T1-susceptible G in typical RNA sequences (at least one every four bases). In general, T1 digestion tends to generate numerous small products with similar base compositions (e.g., di-, tri-, and tetra-nucleotides), which can be assigned to alternative regions of the substrate of interest. In contrast, Mb and V1 digestion tends to produce larger products with intrinsic redundancy afforded by the ladder series, which makes them more readily identifiable with unique regions, thus minimizing possible interpretation ambiguities.

The specificity for defined structural features was evaluated by mapping the position of the respective hydrolysis sites onto the substrate's secondary structure (Figure 1 insets). In agreement with the preference of Mb nuclease for single-stranded nucleotides, the cleavages were located exclusively in the loop region of the stemloop construct, whereas the double-stranded stem was unaffected (Figure 1a). Conversely, the V1 sites were concentrated in the middle of the stem region, where the helical structure was less likely to experience momentary strand dissociation or “breathing” (Figure 1b). Although the actual V1 fragments were unmistakably produced by hydrolysis of the double-stranded structure, their 5'- and 3'-ends did not necessarily correspond to complementary nucleotides (see for example U5:C14, A6:C14, C4:U15, etc. in Table 1), as expected from the possible attack of a well-defined base pair with concomitant hydrolysis of both phosphates located directly across in the opposing strands. Indeed, the observed products were not consistent with blunt-ended structures, but rather with overhangs of different sizes (Scheme 1). This cleavage pattern

eliminated possible models in which V1 nuclease included either a large catalytic site spanning the width of the duplex, or two contiguous sites with appropriate spacing.

The detection of V1 products corresponding to consecutive stretches of the construct's sequence, such as U5:C14 and U15:C20, or U5:U15 and A16:C20 (Table 1), suggested that they originated from independent cleavages of the single-stranded backbone. Further, the formation of ladder series containing the 3'-end of the construct, such as C4:C20, U5:C20, and A6:C20 (Table 1), was consistent with individual events leading to the scission of only one strand of the stem. Based on these observations, the profile of V1 activity resembles more closely that of a nicking enzyme than that of an actual restriction nuclease. In fact, although both classes operate on base-paired regions, the former induces single-strand hydrolysis, whereas the latter always cleaves both strands. The requirement for intact double-stranded structure may explain also the absence of a more homogeneous distribution of cleavage sites over the entire stem (Figure 1b inset). This distribution could be explained by the ability of the initial cleavage event to destabilize contiguous pairing interactions, thus possibly leading to strand dissociation and elimination of recognizable duplex structures.

Challenges Posed by the Analysis of Structure-Specific Products

The characterization of oligonucleotide mixtures obtained by enzymatic hydrolysis follows well-established strategies that rely on mass mapping and sequencing techniques (reviewed in reference [36, 37]). Owing to their uncommon features, however, the analysis of products obtained from structure-specific nucleases posed unique challenges that are not faced when pursuing typical base-specific digests. In fact, RNase A and T1 activity tends to provide distinct products that are not necessarily complementary to one another and, thus, are generally detected as individual single-stranded oligonucleotides. In contrast, numerous signals observed in the Mb digest corresponded to relatively stable complexes consisting of paired strands, which were readily assigned to the double-stranded region of SL3 (Figure 1a and Table 1). Their genesis was directly traced to the progressive elimination of loop nucleotides from the initial structure, with release of different stem sections as annealed hydrolysis products. Their effective detection as intact duplexes was ascribed to the non-denaturing digestion conditions and the relatively low energy afforded by the ESI desolvation process, which prevented the dissociation of complementary strands (see Experimental) [30].

The gas-phase activation of these types of complexes is known to result in alternative pathways that may lead to either duplex dissociation into individual strands, or partial unzipping followed by backbone fragmentation of now unpaired regions [38]. With the purpose of obtaining unambiguous sequence information, controlled strand separation was performed by gentle collisional activation in the LTQ region of the instrument (Figure 2a, see Section 2). Each individual strand was then mass isolated and submitted to further activation in the C-trap element to complete MS³ experiments. The results provided full sequence coverage that confirmed the identity of hydrolysis products (Figure 2b inset). In subsequent experiments, we investigated the possibility of performing ESI-MS analysis at progressively higher desolvation voltages (e.g., increasing this parameter from 35 to 80 V) to readily recognize putative duplexes from single-stranded species. This operation offered the

option of utilizing in-source dissociation to achieve duplex melting before gas-phase activation of individual strands, which makes this approach practical also on instruments devoid of MSⁿ capabilities.

The fact that nucleic acids consist of very few types of fundamental units increases the likelihood that typical RNA digests may contain oligonucleotide products with different sequences but identical base composition. Therefore, the concomitant formation of isobaric species cannot be readily recognized by mass mapping alone. A possible solution consists of repeating the hydrolysis with additional enzymes that possess different base specificity, in such a way as to obtain complementary maps covering the ambiguous sequences. When alternative enzymes are not available, however, product degeneracy may be unambiguously resolved by MS/MS approaches. For example, the signal detected at m/z 1104.81 in the V1 digest (Figure 1b) was initially assigned as the triply charged form of either U5:C14 or A6:U15 (Table 1). When the precursor ion was mass selected and activated in a MS/MS experiment, the observed fragments confirmed the simultaneous presence of both species in the sample (Figure 3a). Although their sequences were nearly identical, the presence of U at either the 3' - or 5' -end was sufficient to prevent significant overlap between the characteristic ion series, which afforded full sequence coverage for both possible species. MS/MS analysis was performed for comparison purposes on the individual 10 mer oligonucleotide that was obtained in separate form from a commercial source (Figure 3b). The characteristic ion series displayed by the individual species matched very closely those obtained from the isobaric mixture (Figure 3a), thus confirming that the simultaneous analysis of coexisting precursor ions did not interfere with the ability to achieve their positive identification. Considering that precursor ions with very similar sequences and identical charge state are likely to manifest comparable fragmentation efficiencies, the relative intensities of the respective series would be expected to reflect very closely the relative abundances of each sample in the isobaric mixture, thus providing the basis for correlating hydrolysis yields with structural context. The likelihood that isobaric products generated by different regions of an RNA substrate might also display the same sequence is significant for smaller products (e.g., up to ~3–4 mers), but decreases rapidly with size. Consequently, the challenge represented by product degeneracy is more likely to affect the analysis of RNase A/T1 than Mb/V1 digests that contain comparatively larger products.

Evaluation of Structure-Specific Nucleases as Actual Probes

The initial experiments on the individual stemloop structure provided a comprehensive overview of the possible outcomes expected from Mb and V1 digestion, and helped establish a possible range of conditions for their effective utilization. Based on this knowledge, we explored the potential of these nucleases as actual structural probes by tackling RNA substrates of increasing size and complexity. Human tRNA^{Lys3} was selected for its cloverleaf topology characterized by well-defined loop and stem regions, which offered ample opportunities for targeted nuclease attack [39]. The utilization of recombinant tRNA^{Lys3} was justified by the fact that this construct has been used in extensive functional studies as a valid surrogate for the natural version bearing post-transcriptional modifications [40, 41]. The observed digestion mixtures appeared to be more complex than those obtained from SL3 RNA. The number of observed products was noticeably larger, as expected from

the significant size difference between the constructs (i.e., 76 nt for tRNA^{Lys3} versus 20 nt for SL3). The increased sample complexity did not affect the ability of direct infusion ESI-MS to properly resolve the sample mixtures in the absence of front-end separation (Figure 4a and Supplementary Table 2S). As described above, initial mapping assignments were supported by extensive MS/MS analysis to minimize possible interpretation ambiguities (not shown).

The cleavage sites identified by the various products were marked onto the construct's secondary structure to examine their structural context. The pattern afforded by Mb nuclease after 5-h incubation (see Experimental) showed that hydrolytic activity was concentrated almost exclusively in single-stranded regions, such as loops and hinges (Figure 5a). Characteristic ladder series covered the double-stranded stems that were largely spared from hydrolysis. In contrast, V1 digestion produced cleavage sites localized in the double-stranded stems, with ladder series spanning the single-stranded regions (Figure 5c). In many instances, the position of susceptible sites did not appear to coincide exactly with the expected boundaries between paired and unpaired regions. Experimental conditions that are known to influence structure stability were explored to test whether this outcome was due to the confounding effects of local dynamics on enzyme recognition. For example, the concentration of ammonium acetate was increased in solution to boost ionic strength (e.g., from 150 to 300 mM); small amounts of Mg(II) compatible with MS analysis were added to the sample (e.g., up to 5 mM was reached when subsequent ethanol precipitation was completed); digestion reactions were performed at different temperatures (e.g., 4, 25, and 37°C); and incubation intervals were varied to gauge the susceptibility of dynamic structures (e.g., between 5 and 14 h/overnight). The results showed that the paired/unpaired boundaries tended to recede upon overnight incubation, consistent with the progressive cleavage of the more dynamic ends (Figure 5b). Addition of Mg(II) was also found to somewhat improve the accuracy of Mb activity, but did not completely eliminate deviations from the expected boundaries. These observations could be attributed to the overriding effects of enzymatic activity over environmental conditions, which may undermine the stability of duplex regions by digesting them in progressively smaller, less stable sections. Similar scenarios have been observed in the investigation of protein structures by limited proteolysis, in which initial attack may destabilize previously protected regions and expose them to hydrolysis [42, 43].

The possible virtues and pitfalls of these putative structural probes were finally evaluated by tackling the investigation of an unknown structure consisting of the 5'-untranslated region (5'-UTR) of the HIV-1 genome. This non-coding stretch of viral RNA contains discrete regulatory signals that mediate vital steps of viral replication, including reverse transcription, transactivation, splicing, genome recognition, dimerization, and packaging [44]. The structure of 5'-UTR has been the object of intense studies that employed different approaches and showed little consensus [44–48]. The prevailing views agree on the presence of a few conserved domains that may rearrange into alternative folds to conceal/expose specific sequences with different biological valence. For this reason, we decided to apply Mb and V1 nuclease to investigate the position of double-stranded regions that may enable the correct discrimination between local secondary structures and long-range tertiary interactions.

The digestion mixtures obtained from a 387-nt construct mimicking 5'-UTR provided complex spectra with large numbers of products (Figure 4b and Supplementary Table 3S). The vast majority of them were assigned by matching experimental masses with theoretical values calculated from sequence, and by performing extensive MS/MS analysis. The assignments allowed us to map the cleavage sites onto one of the alternative secondary structures proposed for 5'-UTR, which identified regions that had effectively withstood nuclease attack (Figure 6). Intact duplexes were detected in the Mb digest, which matched sections of the double-stranded stems of the transactivation response (TAR) element and poly-adenine (polyA) domain. The V1 digest included unique oligonucleotides that corresponded to loops and unpaired linker sequences between discrete domains. Probing experiments were repeated under different conditions to identify the regions most affected by local dynamics. Taken together, the results concurred in delineating the features of consensus domains, such as TAR, polyA, and the primer-binding site (PBS). In contrast, no definitive information was obtained for the various stem loops of the packaging signal (i.e., DIS, SL2, SL3, and AUG domains in Figure 6), even when local dynamics were minimized by elevated Mg(II) concentration and ionic strength (e.g., up to 5 and 300 mM, respectively). This region has been long suspected to undergo major structural rearrangements responsible for the alternative folds proposed for 5'-UTR. For this reason, it was not surprising that its pervasive plasticity resulted in nearly exhaustive digestion by either nuclease. Further experiments are currently underway in which ad hoc antisense oligonucleotides are employed to inhibit possible rearrangements, in such a way as to enable the separate probing of individual folds.

Conclusions

From the technology point of view, this study provided the opportunity to evaluate the possible combination of structure-specific nucleases with a MS-based detection platform. We demonstrated that these enzymes remained active in ammonium-based solutions that were amenable to direct infusion analysis. Both nucleases provided characteristic products that were readily identifiable on the basis of their masses and gas-phase dissociation properties. Possible ambiguities arising from the formation of hydrolytic products with identical base composition, but different sequences, were successfully solved by tandem mass spectrometry. The challenge represented by product degeneracy in complex RNA digests is typically overcome by performing separate experiments with enzymes that possess different target specificities. However, we have shown that MS/MS sequencing represents a viable alternative by virtue of the ability to achieve simultaneous activation of mixed precursor ions in the same experiment. Thanks to their different sequences, the isobaric precursors were found to provide overlapping but recognizable fragment ion series.

In general, Mb and V1 activity matched the reported specificity for unpaired and paired strands, respectively. The absence of preferred target nucleotides or specific recognition sequences supports their possible application to virtually any type of substrate. On one hand, the detection of characteristic ladder products facilitated the recognition of regions that were spared from cleavage. On the other, however, it also signaled an acute sensitivity of these enzymes towards the stability of targeted structures, which produced frequent inaccuracies in the identification of the boundaries between unpaired and paired regions. Possible

measures aimed at minimizing strand breathing and increasing overall fold stability helped reduce the incidence of inaccuracies without eliminating them. These observations prompted typical notes of caution that apply to any type of chemical/biochemical probing application, regardless of the selected detection platform. Caution recommends that experiments be repeated under widely different experimental conditions, including different probe to substrate ratios, and that the ensuing data be carefully compared to recognize the signs of possible structure distortion and other artifactual effects induced by probe activity. Conditional to these recommendations, the concerted application of Mb and V1 nuclease can provide valuable information on the distribution of base-paired structures that play a determinant role in defining the higher-order structure of RNA. In addition to their stand-alone application discussed here, future work will explore their possible utilization in the characterization of products of footprinting and crosslinking reactions, which will take advantage of their ability to produce larger fragments that are more readily identifiable.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors acknowledge support for this work by the RNA Institute of the University at Albany and by the National Institutes of Health (GM064328-12). The authors thank Dr. K. Musier-Forsyth for the generous gift of a plasmid encoding human tRNA^{Lys3}. Helpful discussions with Dr. J. B. Mangrum are gratefully acknowledged.

References

1. Doudna JA, Cech TR. The chemical repertoire of natural ribozymes. *Nature*. 2002; 418:222–228. [PubMed: 12110898]
2. Breaker RR. Prospects for riboswitch discovery and analysis. *Mol. Cell*. 2011; 43:867–879. [PubMed: 21925376]
3. Mattick JS, Makunin IV. Non-coding RNA. *Hum. Mol. Genet*. 2006; 15(Spec 1):R17–29. [PubMed: 16651366]
4. Taft RJ, Pang KC, Mercer TR, Dinger M, Mattick JS. Non-coding RNAs: regulators of disease. *J. Pathol*. 2010; 220:126–139. [PubMed: 19882673]
5. Cruz JA, Blanchet M-F, Boniecki M, Bujnicki JM, Chen S-J, Cao S, Das R, Ding F, Dokholyan NV, Flores SC, Huang L, Lavender CA, Lisi V, Major F, Mikolajczak K, Patel DJ, Phillips A, Puton T, Santalucia J, Sijenyi F, Hermann T, Rother K, Rother M, Serganov A, Skorupski M, Soltysinski T, Sripakdeevong P, Tuszyńska I, Weeks KM, Waldsich C, Wildauer M, Leontis NB, Westhof E. RNA-puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA*. 2012; 18:610–625. [PubMed: 22361291]
6. Weeks KM. Advances in RNA structure analysis by chemical probing. *Curr. Opin Struct. Biol*. 2010; 20:295–304. [PubMed: 20447823]
7. Fabris D, Yu ET. Elucidating the higher-order structure of biopolymers by structural probing and mass spectrometry: MS3D. *J. Mass Spectrom*. 2010; 45:841–60. [PubMed: 20648672]
8. Brenner SE. A tour of structural genomics. *Nat. Rev. Genet*. 2001; 2:801–809. [PubMed: 11584296]
9. Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY. Understanding the transcriptome through RNA structure. *Nat. Rev. Genet*. 2011; 12:641–655. [PubMed: 21850044]
10. Peattie DA, Gilbert W. Chemical probes for higher-order structure in RNA. *Proc. Natl. Acad. Sci. U. S. A*. 1980; 77:4679–4682. [PubMed: 6159633]
11. Kjems J, Egebjerg J. Modern methods for probing RNA structure. *Curr. Opin. Biotechnol*. 1998; 9:59–65. [PubMed: 9503589]

12. Brunel C, Romby P. Probing RNA structure and RNA–ligand complexes with chemical probes. *Methods Enzymol.* 2000; 318:3–21. [PubMed: 10889976]
13. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* 2005; 127:4223–4231. [PubMed: 15783204]
14. Wilkinson KA, Merino EJ, Weeks KM. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc.* 2006; 1:1610–1616. [PubMed: 17406453]
15. Yu E, Fabris D. Direct probing of RNA structures and RNA–protein interactions in the HIV-1 packaging signal by chemical modification and electrospray ionization Fourier transform mass spectrometry. *J. Mol. Biol.* 2003; 330:211–223. [PubMed: 12823962]
16. Yu ET, Fabris D. Toward multiplexing the application of solvent accessibility probes for the investigation of RNA three-dimensional structures by electrospray ionization-Fourier transform mass spectrometry. *Anal. Biochem.* 2004; 344:356–366.
17. Turner KB, Yi-Brunozzi HY, Brinson RG, Marino JP, Fabris D, Le Grice SF. SHAMS: combining chemical modification of RNA with mass spectrometry to examine polypurine tract-containing RNA/DNA hybrids. *RNA.* 2009; 15:1605–1613. [PubMed: 19535461]
18. Hawkins AE, Fabris D. RNA Structure Determination by Structural Probing and Mass Spectrometry: MS3D.. In: Leontis, N, Westhof, E, editors. *RNA Structure Analysis and Prediction.* Springer-Verlag; Berlin Heidelberg: 2012. 361–389.
19. Kellersberger KA, Yu E, Kruppa GH, Young MM, Fabris D. Top-down characterization of nucleic acids modified by structural probes using high-resolution tandem mass spectrometry and automated data interpretation. *Anal. Chem.* 2004; 76:2438–2445. [PubMed: 15117181]
20. Zhang Q, Yu ET, Kellersberger KA, Crosland E, Fabris D. Toward building a database of bifunctional probes for the MS3D investigation of nucleic acids structures. *J. Am. Soc. Mass Spectrom.* 2006; 17:1570–1581. [PubMed: 16875836]
21. Zhang Q, Crosland E, Fabris D. Nested Arg-specific bifunctional crosslinkers for MS-based structural analysis of proteins and protein assemblies. *Anal. Chim. Acta.* 2008; 627:117–128. [PubMed: 18790135]
22. Yu ET, Zhang Q, Fabris D. Untying the FIV frame-shifting pseudoknot structure by MS3D. *J. Mol. Biol.* 2005; 345:69–80. [PubMed: 15567411]
23. Yu ET, Hawkins AE, Eaton J, Fabris D. MS3D structural elucidation of the HIV-1 packaging signal. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105:12248–12253. [PubMed: 18713870]
24. Raines RT. Ribonuclease A. *Chem. Rev.* 1998; 98:1045–1066. [PubMed: 11848924]
25. Yoshida H. The ribonuclease T1 family. *Methods Enzymol.* 2001; 341:28–41. [PubMed: 11582784]
26. Wyatt JR, Walker GT. Deoxynucleotide-containing oligoribonucleotide duplexes: stability and susceptibility to RNase V1 and RNase H. *Nucleic Acids Res.* 1989; 17:7833–7842. [PubMed: 2552416]
27. Kedzierski W, Laskowski M Sr. Mung bean nuclease I. V. A tool for probing low melting regions in deoxyribonucleic acid. *J. Biol. Chem.* 1973; 248:1277–1280. [PubMed: 4631390]
28. Nilsen TW. RNA structure determination using nuclease digestion. *Cold Spring Harbor Protoc.* 2013:379–382.
29. Wan Y, Qu K, Ouyang Z, Chang HY. Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. *Nat. Protoc.* 2013; 8:849–869. [PubMed: 23558785]
30. Fabris D, Turner KB, Hagan NA. Electrospray Ionization-Mass Spectrometry for the Investigation of Protein–Nucleic Acid Interactions.. In: Banoub, JH, Limbach, PA, editors. *Mass Spectrometry of Nucleosides and Nucleic Acids.* CRC Press; 2009. 303–327.
31. Rozenski J. Mongo Oligo Mass Calculator, v2.06. 1999
32. Pappalardo L, Kerwood DJ, Pelczer I, Borer PN. Three-dimensional folding of an RNA hairpin required for packaging HIV-1. *J. Mol. Biol.* 1998; 282:801–818. [PubMed: 9743628]
33. Yamashita M, Fenn JB. Electrospray ion source. Another variation on the free-jet theme. *J. Phys. Chem.* 1984; 88:4671–4675.

34. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science*. 1989; 246:64–71. [PubMed: 2675315]
35. McLuckey SA, Habibi-Goudarzi S. Decompositions of multiply charged oligonucleotide anions. *J. Am. Chem. Soc.* 1993; 115:12085–12095.
36. Nordhoff E, Kirpekar F, Roepstorff P. Mass spectrometry of nucleic acids. *Mass Spectrom. Rev.* 1996; 15:67–138. [PubMed: 27082318]
37. Limbach PA. Indirect mass spectrometric methods for characterizing and sequencing oligonucleotides. *Mass Spectrom. Rev.* 1996; 15:297–336. [PubMed: 27082835]
38. Gabelica V, De Pauw E. Comparison of the collision-induced dissociation of duplex DNA at different collision regimes: evidence for a multistep dissociation mechanism. *J. Am. Soc. Mass Spectrom.* 2002; 13:91–98. [PubMed: 11777205]
39. Benas P, Bec G, Keith G, Marquet R, Ehresmann C, Ehresmann B, Dumas P. The crystal structure of HIV reverse-transcription primer tRNA(Lys,3) shows a canonical anticodon loop. *RNA*. 2000; 6:1347–1355. [PubMed: 11073212]
40. Hargittai MR, Musier-Forsyth K. Use of terbium as a probe of tRNA tertiary structure and folding. *RNA*. 2000; 6:1672–1680. [PubMed: 11105765]
41. Hargittai MR, Mangla AT, Gorelick RJ, Musier-Forsyth K. HIV-1 nucleocapsid protein zinc finger structures induce tRNA(Lys,3) structural changes but are not critical for primer/template annealing. *J. Mol. Biol.* 2001; 312:985–997. [PubMed: 11580244]
42. Zappacosta F, Pessi A, Bianchi E, Venturini S, Sollazzo M, Tramontano A, Marino G, Pucci P. Probing the tertiary structure of proteins by limited proteolysis and mass spectrometry: the case of Minibody. *Protein Sci.* 1996; 5:802–813. [PubMed: 8732752]
43. Serpa JJ, Parker CE, Petrotchenko EV, Han J, Pan J, Borchers CH. Mass spectrometry-based structural proteomics. *Eur. J. Mass Spectrom. (Chichester, Engl)*. 2012; 18:251–267.
44. Baudin F, Marquet R, Isel C, Darlix JL, Ehresmann B, Ehresmann C. Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains. *J. Mol. Biol.* 1993; 229:382–397. [PubMed: 8429553]
45. Damgaard CK, Dyr-Mikkelsen H, Kjems J. Mapping the RNA binding sites for human immunodeficiency virus type-1 gag and NC proteins within the complete HIV-1 and -2 untranslated leader regions. *Nucleic Acids Res.* 1998; 26:3667–3676. [PubMed: 9685481]
46. Abbink TE, Berkhout B. A novel long distance base-pairing interaction in human immunodeficiency virus type 1 RNA occludes the Gag start codon. *J. Biol. Chem.* 2003; 278:11601–11611. [PubMed: 12458192]
47. Paillart JC, Dettenhofer M, Yu XF, Ehresmann C, Ehresmann B, Marquet R. First snapshots of the HIV-1 RNA structure in infected cells and in virions. *J. Biol. Chem.* 2004; 279:48397–403. [PubMed: 15355993]
48. Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews DH, Giddings MC, Weeks KM. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.* 2008; 6:e96. [PubMed: 18447581]

Figure 1.

ESI-MS spectra of digestion mixtures obtained by treating SL3 RNA with (a) Mung bean and (b) V1 nuclease. The insets illustrate the location of the respective cleavage sites. Dashes are used to indicate duplex products

Scheme 1.

Putative mechanisms of double-stranded cleavage of SL3 RNA: (a) concerted attack of a well-defined base pair would be expected to produce exclusively species with blunt ends; (b) independent cleavages of the two strands would be expected to produce ladders with overhangs

Figure 2.

(a) MS/MS spectrum of the duplex product G1:G9-A11:C20 observed at m/z 1531.71 in Figure 1(a). (b) MS³ spectrum of the isolated G1:G9 component (m/z 1531.71→445.22→...)

Figure 3. MS/MS spectrum of (a) the species observed at m/z 1104.81 in Figure 1(b). (b) A triply charged ion with m/z 1104.81 obtained from individual A6:U15 oligonucleotide

Figure 4. Representative ESI-MS spectra of samples obtained by Mb digestion of **(a)** human tRNA^{Lys3} and **(b)** HIV-1 5'-UTR. Dashes are used to indicate duplex products

Figure 5.

Cleavage maps obtained by treating human tRNA^{Lys3} with (a) Mung bean nuclease for 5 h, (b) overnight, and (c) V1 nuclease overnight (see Supplementary Table 2S). Dashed lines mark cleavage sites. Highlighted in yellow are regions marginally affected by hydrolysis

Figure 6. Cleavage maps obtained by treating HIV-1 5'-UTR with (a) Mung bean nuclease for 5 h, and (b) V1 nuclease overnight (see Supplementary Table 3S). Dashed lines mark cleavage sites. Highlighted in yellow are regions marginally affected by hydrolysis

Table 1

Digestion Products Obtained by Treating SL3 RNA with Mung Bean and V1 Nuclease (see Experimental).
Monoisotopic Masses are Reported in Mass Units

Position	Sequence	Exp. mass	From seq.
Mung bean nuclease digestion of SL3 RNA			
G1:G9-A11:C20	5'-GGACUAGCG-3', 5'-pAGGCUAGUCC-3'	6130.87	6130.88
G1:G9-G12:C20	5'-GGACUAGCG-3', 5'-pGGCUAGUCC-3'	5801.82	5801.83
G1:G9	5'-GGACUAGCG-3'	2892.44	2892.45
A11:C20	5'-pAGGCUAGUCC-3'	3238.43	3238.43
G12:C20	5'-pGGCUAGUCC-3'	2909.38	2909.38
V1 nuclease digestion of SL3 RNA			
G1-C4	5'-GGAC-3'	1262.23	1262.23
G1-U5	5'-GGACU-3'	1568.26	1568.26
G1-C14	5'-GGACUAGCGGAGGC-3'	4561.68	4561.68
C4-U15	5'-pCUAGCGGAGGCU-3'	3928.52	3928.53
C4-C20	5'-pCUAGCGGAGGCUAGUCC-3'	5518.73	5518.74
U5-C14	5'-pUAGCGGAGGC-3'	3317.46	3317.46
U5-U15	5'-pUAGCGGAGGCU-3'	3623.49	3623.49
U5-A16	5'-pUAGCGGAGGCUA-3'	3952.54	3952.54
U5-C20	5'-pUAGCGGAGGCUAGUCC-3'	5213.69	5213.70
A6-C14	5'-pAGCGGAGGC-3'	3011.43	3011.44
A6-U15	5'-pAGCGGAGGCU-3'	3317.46	3317.46
A6-C20	5'-pAGCGGAGGCUAGUCC-3'	4907.66	4907.67
G7-C14	5'-pGCGGAGGC-3'	2682.38	2682.38
G7-U15	5'-pGCGGAGGCU-3'	2988.41	2988.41
G7-U18	5'-pGCGGAGGCUAGU-3'	3968.53	3968.54
C8-C14	5'-pCGGAGGC-3'	2337.33	2337.34
C8-U15	5'-pCGGAGGCU-3'	2643.36	2643.36
U15-C20	5'-pUAGUCC-3'	1914.24	1914.24
A16-C20	5'-pAGUCC-3'	1608.22	1608.22