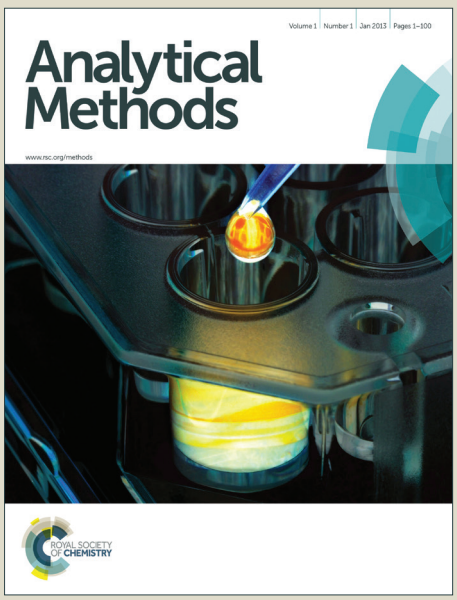
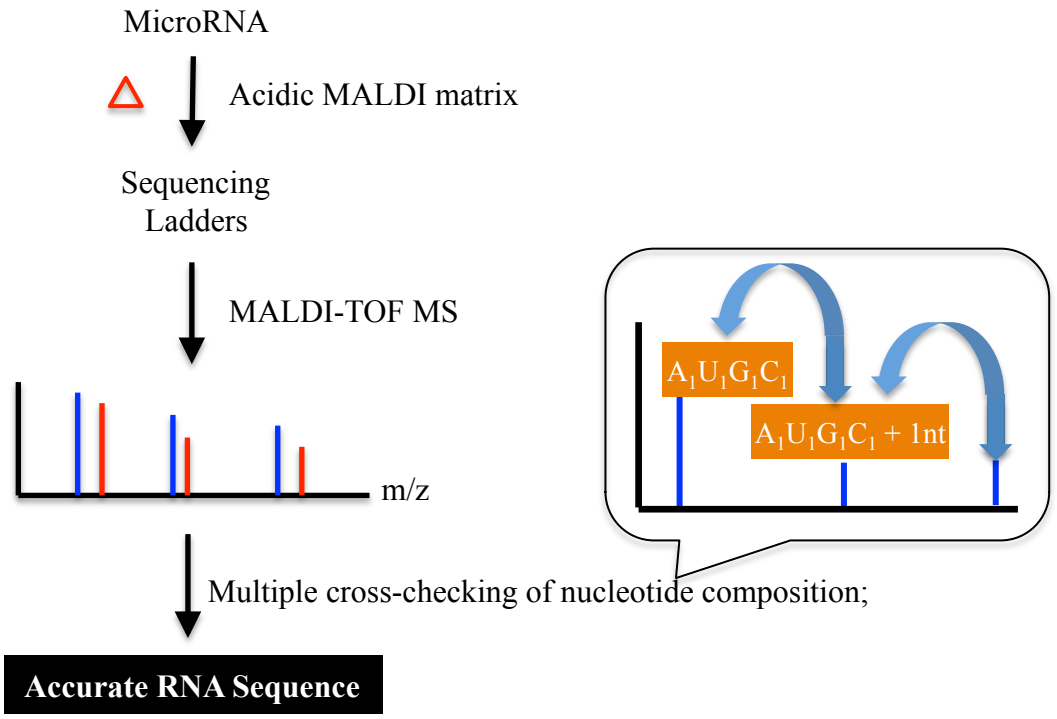
Analytical Methods

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods



Bottom-Up Mass Spectrometric Sequencing of MicroRNA with 100% Sequence Coverage and Sequence Accuracy

Dickson M. Wambua[1], Masaaki Ubukata[2], John Dane[2], Robert B. Cody[2], Norman H. L. Chiu[1,3]**

[1]Department of Chemistry and Biochemistry, University of North Carolina at Greensboro, 301 McIver Street, Greensboro, NC 27412, USA.

[2]JEOL USA, Inc. 11 Dearborn Road, Peabody, MA 01960, USA.

[3]Department of Nanoscience, Joint School of Nanoscience and Nanoengineering, Greensboro, NC 27401, USA.

* To whom correspondence should be addressed. Both authors have made equal contributions. Email and Phone #: prof.chiu@gmail.com, 336-334-3126; cody@jeol.com, 978-535-5900.

ABSTRACT

The increasing interests in microRNA (miRNA) as diagnostic biomarkers or potential drug targets have raised the demand for more accurate miRNA detection. One way to improve the accuracy is by using mass spectrometry (MS) to measure miRNA directly. Matrix-assisted laser desorption/ionization (MALDI) MS stands apart from other MS techniques is due to the fact that MALDI matrix is required for sample preparation. In this study, by exploiting the acidity of MALDI matrix and its mixing with miRNA prior to MS measurements, a simple method to generate RNA sequencing ladders is developed. The method utilizes MALDI matrix to hydrolyze RNA at high temperature. The resulting sequencing ladders are ready to be measured without any desalting. By using MALDI SpiralTOF MS, monoisotopic mass of each RNA fragment was measured. The RNA sequence was determined by sequentially comparing nucleotide compositions that are calculated from measured monoisotopic masses. The use of nucleotide compositions to assist the spectral interpretation has the advantages on distinguishing the complementary sequencing ladders, and allows the nucleotide identity at each position to be crosschecked multiple times. Together with the analysis of both complementary sequencing ladders, 100% sequence coverage and sequence accuracy were achieved in a blinded study.

INTRODUCTION

MicroRNA (miRNA), which is encoded by eukaryotic nuclear DNA, is a group of small RNA. Following nuclear and cytoplasmic processing, each mature miRNA has a unique RNA

sequence. The size of the majority (82 %) of human mature miRNA is equal or shorter than 22 nucleotides (Supplemental Figure S1). With the relative small size, the sequence similarity among different miRNAs is expected to be high, especially those belong to the same family of miRNA.[1] Despite of these characteristics, by forming RNA-induced silencing complexes (RISC) with cellular proteins, miRNA can regulate approximately 60 % of gene expression activities. Thus, miRNA have already been associated with many diseases.[2] The use of miRNA as potential targets for the development of novel drugs, and as diagnostic biomarkers for human diseases both require the detection of specific miRNA to be accurate and cost effective.[3,4]

In the human body, there are more than 2,000 miRNA. To identify a known miRNA or discover a new miRNA, similar to the detection of other types of RNA, the RNA sequence has to be determined. Traditionally, the Sanger-based sequencing method is the gold standard for determining the RNA sequence of a new miRNA. However, the method is labor intensive, time consuming and requires the use of multiple reagents. For the detection of known miRNA, nucleic acid probe with a complementary sequence to the target miRNA sequence has been extensively used in many existing methods.[1,5,6] Depending upon the purity of miRNA samples as well as the conditions in which the binding between the nucleic acid probe and target miRNA is carried out, non-specific binding can often occur and potentially leads to false positive results. Furthermore, in many cases, the binding between a specific miRNA and its corresponding nucleic acid probe does not generate any reporting signal, thus a reporting label is required. By measuring the reporting label that has been attached to either target miRNA or a detection nucleic acid probe, the target's RNA sequence can be indirectly determined.

Over the past decade, several alternative methods for sequencing nucleic acids have been developed. Amongst the emerging methods, MS based methods stand out to be more accurate.

This is because the intrinsic molecular mass of nucleic acids are directly measured versus the indirect measurement of a signal generated from a reporting label that has been attached to a probe or target nucleic acid. The most recent review article on mass spectrometry of RNA was written by Kirpekar and his associate.[7] In general, MS methods for sequencing nucleic acids can be divided into two groups, namely top-down and bottom-up method. In the top-down methods, the RNA molecule of interest is first ionized by one of the soft ionization techniques, for example electrospray ionization (ESI) has been commonly used. The resulting RNA ions are then selected and transmitted into a collision cell, where ions undergo a gas-phase collision-induced dissociation (CID). In comparison to DNA, the molecular ions of RNA are more stable, thus less CID fragment ions are usually produced.[8,9] By reducing the charges on the fragment ions via ion/ion reaction, McLuckey and co-workers had reduced the spectral overlap between multiply charged fragment ion peaks, and allowed roughly 60% of a selected transfer RNA molecule with 75 nucleotides to be successfully sequenced.[10] Breuker and her group had used heterocyclic amine additives in RNA sample to reduce the charges on precursor ESI ions, and extended the sequence coverage up to 61 nucleotides.[11] Recently, Pitteri and her associates reported mixtures of up to six different miRNAs can be quantitated with high accuracy by combining nanoLC-MS with the top-down approach.[12] Similar to other studies, their MS/MS experiments could not achieve complete sequence coverage. Despite the feasibility to extend the sequence coverage, the drawbacks of top-down MS methods include: (1) requirement for higher sample concentration and/or purity such that more RNA precursor ions can be generated to support the CID process; (2) each phosphodiester linkage in the RNA backbone can be cleaved in four different ways during the CID process, and the signals of resulting fragment ions may potentially overlap with other signals that have the same mass-to-charge ratios; (3) as described

above, extra chemicals or procedure are often required to achieve a long and complete read in RNA sequencing; and (4) computer programs for analyzing tandem MS spectra for RNA sequencing are less available than those programs for peptide sequencing. In the bottom-up MS methods, similar to the Sanger sequencing method, RNA sequencing ladders are generated prior to the MS measurements. Depending on the approach for generating the sequencing ladders, the bottom-up methods can be further divided into two groups. In the first case, various enzymatic activities have been used to digest RNA into sequencing ladder(s).[13,14] For example, 5' exonuclease removes a single nucleotide sequentially from the 5' end, and generates a 5'→3' sequencing ladder. To achieve a double read, both 5'→3' and 3'→5' sequencing ladders are required. This can be achieved by digesting a specific RNA with either 5' exonuclease or 3' exonuclease separately, thus two individual enzymatic digestion and subsequent MS measurements are required. By measuring the mass differences between adjacent RNA fragments within a sequencing ladder, the RNA sequence can be determined. Although the experimental set up including MS measurements for the bottom-up methods is relatively simple, the drawbacks of the enzymatic methods include (1) two separate digestions are required to achieve double read or a complete sequence coverage; (2) the exonuclease activity can be inhibited by RNA modifications; and (3) the desalting of the digested RNA fragments prior to MS measurements requires extra chemicals and procedure. Recently, Limbach and his co-workers has developed a comparative sequencing method named CARD, in which reference RNA and unknown RNA are labeled with different stable isotopes by using the activity of RNase T1.[15] For the bottom-up methods, an alternative approach for generating a RNA sequencing ladder is by using chemical hydrolysis. Bahr et al. had recently reported the use of trifluoroacetic acid to partially hydrolyze RNAs with 21 mer.[16] The resulting sequencing ladders were first

measured by using the conventional linear mode in MALDI-TOF MS. In order to achieve complete sequence coverage, Bahr et al. had to rely on using tandem MS to measure some of the RNA fragments with 2 or 3 mer. To enable *de novo* sequencing or resequencing of RNA, several attempts to develop methods that would adequately distinguish C from U or vice versa have been reported. These include differentiating RNA fragments bearing C or U by their peak heights,[17] combination of alkaline hydrolysis with base-specific enzymatic cleavages such as RNase T1, RNase U2 or RNase A,[18] and a combination of acidic hydrolysis with CID.[9] Overall, each of the above methods involves the use of multiple reagents and/or MS experiments.

Among various MS techniques, MALDI time-of-flight (TOF) MS requires the smallest sample size and provides the highest sample throughput as well as the largest mass range. In a standard method for preparing MALDI sample, a sample of interest is mixed with an excess amount of MALDI matrix. In this study, by taking advantage of the mixing between a miRNA sample and MALDI matrix prior to the MS measurements and the acidity of MALDI matrix, a unique yet simple and accurate method to generate RNA sequencing ladders is developed.[19] The method utilizes acidic MALDI matrix to hydrolyze a single phosphodiester bond in each miRNA molecule, thus generating two sets of complementary 5' and 3' sequencing ladder, which are ready to be measured by MALDI-TOF MS (Figure 1). Monoisotopic mass of each RNA fragment was accurately measured by using the newly developed MALDI SpiralTOF MS. To determine the RNA sequence, the nucleotide composition of each monoisotopic mass was calculated, and the nucleotide compositions of two adjacent peaks within the same sequencing ladder were compared. The sequential comparison of nucleotide compositions between two adjacent peaks also allows the RNA sequence to be crosschecked multiple times, thus improving

the accuracy. To evaluate the developed method, a blinded study for sequencing selected human miRNA was carried out.

MATERIAL AND METHODS

2,5-dihydroxybenzoic acid (DHB, pKa_1=3.01) was purchased from Acros Organics (Morris Plains, NJ, USA). 3-hydroxypicolinic acid (3-HPA, pKa_1=1.14), sinapinic acid (SA, pKa_1=4.53), ammonium acetate and ammonium citrate dibasic were acquired from Sigma Aldrich (St. Louis, MO, USA). Acetonitrile and methanol, both of HPLC grade were obtained from Fisher Scientific (Pittsburgh, PA, USA). Non-sterile 0.22 µm low protein binding Durapore (PVDF) syringe driven membrane filter units (25mm) were purchased from Millipore Corp. (Bedford, MA, USA). Three different miRNA were purchased from Integrated DNA technologies (San Diego, CA), which include hsa-miR-153 (5' p-UUGCAUAGUCACAAA AGUGAUC-OH 3'), hsa-miR-183-5p (5' p-UAUGGCACUGGUAGAAUUCACU-OH 3'), and mml-miR-124a (5' p-UUAAGGCACGCGGUGAAUGCCA-OH 3'). Water from a Barnstead NANOpure Diamond Water Purification System (Dubuque, Iowa, USA) was used for the reconstitution and dilution of RNA oligos.

Preparation of saturated MALDI matrices, and limited hydrolysis of miRNA

Three MALDI matrices, namely 3 HPA, SA and DHB, were separately dissolved in either 10% or 50% acetonitrile to saturation. Each MALDI matrix was centrifuged; and the saturated supernatant was drawn for the hydrolysis of miRNA. Saturated matrices were freshly prepared before use. For the limited hydrolysis of miRNA, 2 µL of 10 pmol/µL of miRNA was

mixed with 2 μL of saturated MALDI matrix. The mixture was incubated at 65 °C for 30 minutes, unless otherwise stated.

MALDI MS measurements

For all MALDI MS measurements, 3-HPA was used as the MALDI matrix unless otherwise stated. The 3-HPA matrix solution was prepared by dissolving 35.0 mg of 3-HPA and 8.80 mg of ammonium citrate dibasic in 1.0 mL of 10% acetonitrile. The matrix solution was vortexed for 1-2 minutes, filtered through 0.22 μm membrane filter, and stored at −20 °C. MALDI sample plate was cleaned with deionized water and methanol. When MALDI matrix was used, 0.3 μL of matrix was spotted on the plate and allowed to air dry before 0.3 μL of miRNA sample was added on top of the dried matrix and allowed to air dry.

In this study, two different MALDI mass spectrometers were used. During an initial study to optimize the conditions for carrying out the limited hydrolysis of miRNA, a conventional MALDI-TOF mass spectrometer (4700 Proteomics Analyzer, Applied Biosystems, Framingham, MA) was used. Each sample was measured by using the linear high mass positive mode. The Nd:YAG laser (200 Hz) was set at ~5,000 arbitrary units with the maximum setting at 7,900 arbitrary units. Molecular ions were extracted after 450 ns delay. The accelerating voltage was +20.0 kV and the grid voltage was +18.8 kV. The linear detector voltage was +2.0 kV. Each spectrum was acquired by accumulating ~3,000 shots with random edge-biased for positioning the laser spot. By using the Data Explorer Version 4.6 software, the resulting mass spectra were internally calibrated.

To achieve higher mass resolution and accuracy, a second MALDI-TOF mass spectrometer, namely MALDI SpiralTOF (*JMS-S3000* SpiralTOF™, JEOL Ltd., Tokyo, Japan), was used.[20-26] A 349 nm Nd-YLF laser (500 Hz) was used. Data from 250,000 laser shots were accumulated for each mass spectrum. The extraction delay (250 ns) and laser power were optimized to maximize the isotopic separation. JEOL *MS Tornado* software was used to process the SpiralTOF mass spectra. Nucleotide compositions were calculated by using the *Mass Mountaineer* software, which is available from http://mass-spec-software.com.

RESULTS AND DISCUSSION

Prior to MALDI MS measurements, the sample of interest is normally mixed and co-crystallized with a large molar excess of MALDI matrix compound, which protects the analyte by absorbing the laser energy during the MALDI process. In the case of positive ion mode, the MALDI matrix can also serve as a proton donor. Among the most commonly used MALDI matrices, three of them are acidic and shown to support RNA analysis. These include sinapinic acid (SA), 2,5-dihydroxybenzoic acid (DHB), and 3-hydroxypicolinic acid (3-HPA) (Figure 2a). The acidic nature of these MALDI matrices presents a unique opportunity for developing a new method for sequencing small RNA. In this study, for the first time, the use of acidic MALDI matrix to partially hydrolyze RNA prior to MALDI-TOF MS measurements was explored. To demonstrate the feasibility of this approach, a human mature miRNA called miR-153 with a known sequence (22 mer) was selected as a model, which has the same size as the majority of human mature miRNA (Supplemental Figure S1).

Generation of RNA Sequencing Ladders

It has been known for many years that RNA is susceptible to acidic hydrolysis mainly through the cleavage of its phosphodiester bonds.[27,28] As summarized in Figure 2b, the results from a limited acidic hydrolysis of RNA fragment are the production of two fragments, namely 5' and 3' fragment. In the case of the 3' fragment, the resulting 5' terminal has a hydroxyl group. For the 5' fragment, its 3' terminal consists of either a 2',3'-cyclic phosphate group or a linear phosphate group (Figure 2b). The cyclic phosphate group can be rapidly hydrolyzed and converted into a linear phosphate group. However, under an acidic condition, the non-bridging phosphoryl oxygen in the linear phosphate can be protonated. This will then be followed by a bond formation between the phosphorus and the hydroxyl group at the 2' position. With a proton transfer, the cyclic phosphate group is regenerated. Thus, both cyclic phosphate group and linear phosphate group can co-exist at the 3' end of 5' fragments, which were detected in this study. Similar to other chemical methods for RNA degradation, the cleavage of phosphodiester bond during the limited acidic hydrolysis of RNA has no sequence specificity. Hence, each phosphodiester bond within a small RNA fragment can be cleaved during the limited acidic hydrolysis. With sufficient amount of RNA material, the net outcome of limited acidic hydrolysis of miRNA is the production of two sets of complementary 5' and 3' sequencing ladders. As shown in the schematic diagram of the developed method (Figure 1), following the limited acidic hydrolysis, the complementary 5' and 3' sequencing ladders can be directly measured by using MALDI-TOF MS without the use of any additional reagent or experimental procedure.

Although the pKa value of MALDI matrices in aqueous solution are known (Figure 2a) and the yield of acidic RNA hydrolysis is also known to be pH dependent,[27] the yield from using

different acidic MALDI matrices to hydrolyze the selected miRNA model (miR-153) were compared. This is due to the fact that MALDI matrices are usually prepared in non-aqueous solutions, and the use of different non-aqueous solutions were reported, which may change the acidity of MALDI matrices. The results from using the three most common acidic MALDI matrices to hydrolyze miR-153 are shown in Figure 3a. Although the hydrolyzed RNA fragments in the MALDI matrix solutions can be directly spotted on a MALDI sample plate and measured, the overall signal-to-noise ratios in the MALDI spectra can be improved by adopting the thin-layer sample preparation method; and MALDI sample plates with pre-loaded MALDI matrix on their surfaces are commercially available. Among those matrices, the highest percentage of sequence coverage (74%) was achieved when 3-HPA was used to partially hydrolyze miR-153. The percentage of sequence coverage is calculated from a ratio between the number of ribonucleotide that have been experimentally determined to the total number of ribonucleotide in the selected miRNA model. Unless otherwise stated, the identity of each ribonucleotide was determined by adopting the traditional approach of using the mass difference between two adjacent peaks within a specific sequencing ladder. In Figure 3, the results were obtained by analyzing both sets of complementary sequencing ladder, which was equivalent to a double read. For MALDI mass spectrometric sequencing of RNA, the use of both sets of complementary sequencing ladder to deduce the RNA sequence is important. This is due to the fact that MALDI matrix ion and its cluster ions may interfere with the MS measurements at the lower mass region (≤1,000 m/z). Owing to this reason, RNA fragments with less than 4 nucleotides were not included in both sets of complementary sequencing ladder, thus lowering the sequence coverage. This, however, can be overcome by using both sets of complementary sequencing ladder to deduce the entire RNA sequence. Besides achieving high sequence coverage, the benefit of using

both sets of complementary sequencing ladder also allows most of the RNA sequence to be confirmed by crosschecking the results obtains from each individual set of sequencing ladder. To ensure the highest rate of hydrolysis was achieved, saturated MALDI matrix solutions in either 10% acetonitrile or 50% acetonitrile were used. Despite more 3-HPA could be dissolved in 50% acetonitrile, higher percentage of sequence coverage was achieved when 3-HPA was dissolved in 10% acetonitrile (Figure 3a). This is attributed mainly to the higher signal-to-noise ratios attainable by using 3-HPA in 10% acetonitrile as a MALDI matrix. To further optimize the sequence coverage, the effects of using 3-HPA in 10% acetonitrile to hydrolyze miR-153 at different incubation temperature and time were investigated. As shown in Figure 3b, the optimal incubation temperature and time were determined to be at 65 ^{0}C for 30 min. By using the optimal conditions, 89% of sequence coverage was achieved. This is equal to reading 20 out of 22 nucleotides in the entire miRNA molecule. With 89% sequence coverage, it is sufficient to distinguish the RNA sequence of a specific miRNA from all other miRNA that have been reported. In Figure 3b, the decrease in the percentage of sequence coverage at temperature higher than 65 ^{0}C was due to excessive hydrolysis of miR-153. This included the production of some internal RNA fragments, which resulted from cleaving more than one phosphodiester bond in a single miRNA molecule. In contrast, under the optimal incubation temperature and time, no internal RNA fragment was detected.

MALDI-TOF Mass Spectrometry of RNA Sequencing Ladders

Initially, the RNA sequencing ladders were measured using a conventional MALDI-TOF MS instrument (4700 Proteomics Analyzer MALDI-TOF MS). To attain the highest signals for

measuring the RNA sequencing ladders within a mass range between 1,000 m/z and ~8,000 m/z, the linear mode in the conventional MALDI-TOF MS was used. Similar to the mass spectrum in Figure 4, the peaks that corresponded to the 3' fragment ions often had higher signals than those corresponded to the 5' fragment ions. This is attributed to the differences in the functional group at the terminals. In the case of 3' fragments, both 5' and 3' terminals have a hydroxyl group. Whereas, the 5' and 3' terminal of the 5' fragments both contains a phosphate group (Figure 2b). The difference in peak intensities between the 3' and 5' fragments can be useful to distinguish the two sets of complementary sequencing ladders. Among the peaks that corresponded to the 5' fragment ions, the peaks that corresponded to ions containing the 2',3'-cyclic phosphate group at the 3' terminal often had higher signals than those corresponded to ions containing the linear phosphate group at the 3' terminal. This can be attributed to the protonation of the 2',3'-cyclic phosphate group as shown in Figure 2b, and the measurement of its positive ion was carried out. Although the 3' terminal of each 5' fragment can contain either a 2',3'-cyclic phosphate group or a linear phosphate group, the mass difference (17 Da) between the two phosphate groups is unique in comparison to the mass differences between the four ribonucleotides. However, the limited mass resolution in the high mass region of the conventional MALDI-TOF MS poses a challenge for analyzing the entire set of sequencing ladders with high accuracy.

In this study, with the linear mode in the conventional MALDI-TOF MS, the mass resolution at 7,000 m/z was about 1,000. This is far too low for the differentiation of the mass difference between uridine and cytidine, which has only 0.98 Da. To achieve higher mass resolution, the newly developed MALDI SpiralTOF MS was also used to analyze the RNA sequencing ladders of miR-153. In comparison to the conventional TOF mass analyzer, which has approximately 1 m flight distance, the SpiralTOF mass analyzer uses a unique ion optics to

extend the flight distance to 17 m while maintaining a high efficiency on ion transmission and keeping the divergence of ion beam at the detection plane to a minimum.[20-26] The results of using MALDI SpiralTOF MS to measure the RNA sequencing ladders of miR-153 are shown in Figure 4. The average signal-to-noise ratio in Figure 4 was higher than the results obtained by using the conventional MALDI-TOF MS. With the ability to identify additional members of the RNA sequencing ladders, the sequence coverage for resequencing miR-153 was increased from 89 % to 100 %. Both complementary RNA sequencing ladders were used, but no MS/MS measurement was required to resequence the entire miR-153 with 22 mer. The sequence coverage can be lower if the sample concentration is lowered. By using MALDI SpiralTOF MS, isotopic resolution was achieved in the entire mass range between 1,000 m/z and ~8,000 m/z (Figure 5). The isotopic pattern of the unhydrolyzed miR-153 molecular ions, which includes the number of isotopic peaks and their relative peak heights, matches well to the results of a computational simulation (Supplemental Figure S2). To the best of our knowledge, this represents the first report on using the SpiralTOF mass analyzer to achieve isotopic resolution starting from 1,000 m/z to ~8,000 m/z in a single mass spectrum. The mass resolution at 7,000 m/z was close to 20,000. With external calibration, the average mass accuracy for measuring miR-153 sequencing ladders was equal to 7.1 ppm (Table 1 and 2), which represents a significant improvement over the use of conventional MALDI-TOF MS in the linear mode.

Deduction of RNA Sequence Using Nucleotide Composition

With the ability to achieve monoisotopic mass measurements as shown in Figure 4, the RNA sequence can be determined by using the traditional approach of mass difference between

two adjacent peaks within a sequencing ladder.[29,30] However, there are challenges for using this traditional approach. In this study, both sets of complementary sequencing ladder co-exist in the same sample, and are measured together at the same time. Together with the 0.98 Da mass difference between uridine and cytidine, it poses difficulties to identify which is the correct peak in the mass spectrum that corresponds to a particular RNA fragment. To address these issues and achieve a higher accuracy on determining the RNA sequence, an alternative approach to deduce the RNA sequence is adopted. As shown in the schematic diagram of the developed method (Figure 1), the nucleotide composition of each RNA fragment is first calculated from its measured monoisotopic mass. Since the elemental composition is constrained by the four natural ribonucleotides and the function group at the 5' and 3' end of each RNA fragment, only a small number of nucleotide compositions are calculated within the error tolerance. As an example, let us consider the calculation of the nucleotide composition of a 3' fragment from miR-153, which had a measured monoisotopic mass of 4436.6755 m/z (Table 2). This is done with the *Mass Mountaineer* software by defining "pseudoelements" or "superatoms" consisting of the isotopic distributions for each of the ribonucleotide subunits. Superatoms designated "C", "U". "A", and "G" are defined from the calculated isotope distributions for $C_9H_{12}N_3O_7P$, $C_9H_{11}N_2O_8P$, $C_{10}H_{12}N_5O_6P$, and $C_{10}H_{12}O_7N_5P$ respectively. Additional superatoms "c", "u", "a" and "g" are defined from the calculated isotope abundances for the protonated ribnucleotides; one and only one of these must be present for any given nucleotide composition. It does not matter which one is used for the calculation as long as that nucleotide is present anywhere in the composition. An elemental composition calculation for the nucleotide superatom limits $u_1C_{0\text{-}25}U_{0\text{-}25}A_{0\text{-}25}G_{0\text{-}25}$ and an error tolerance of 10 ppm results in only one composition: $u_1C_3U_2A_6G_2$. The correct nucleotide composition of the selected 3' fragment is $U_3C_3A_6G_2$ with the elemental composition

$C_{134}H_{167}N_{55}O_{93}P_{13}$. Isotope matching provides additional confirmation of the correct composition. For comparison, a traditional elemental composition for the elements $C_{0\text{-}150}H_{0\text{-}175}N_{0\text{-}60}O_{0\text{-}100}P_{0\text{-}20}$ results in 830 possible compositions. The calculated nucleotide compositions were then ranked by comparing their corresponding monoisotopic masses and isotopic peak patterns to the results obtained from the MS measurements. Candidate nucleotide compositions for successive fragments that differed by more than a single nucleotide were rejected. After the top ranking nucleotide composition is assigned to a measured isotopic peak, the RNA sequence is determined by sequentially comparing the nucleotide composition of two adjacent peaks that correspond to a particular set of sequencing ladder. By using the difference between two nucleotide compositions instead of the difference between two measured monoisotopic masses, the possibility on misidentifying one of the peaks or both peaks in the mass spectrum is significantly lowered. This is because, as exemplify in Figure 1, the correct difference between two nucleotide compositions that correspond to a particular set of sequencing ladder should always be the identity of the terminal nucleotide in the longer RNA fragment. The advantage of using the approach of nucleotide composition also includes the ability to crosscheck the nucleotide identity at previous position(s) while performing the subsequent comparison of nucleotide compositions within the same set of sequencing ladder. If a mistake on the RNA sequence is identified during the crosschecking of nucleotide compositions, this approach does allow a correction on the RNA sequence to be made. Overall, after the entire RNA sequence is determined, the nucleotide identity at each position would have been crosschecked multiple times, which depend on the total number of RNA fragments that makes up the entire sequencing ladder and how many of them are detectable by the mass spectrometric measurement. Together with the ability to crosscheck the RNA sequences obtain from the complementary sequencing

ladder, the accuracy on the RNA sequence can be significantly higher than using the traditional approach of mass difference between two adjacent peaks. To the best of our knowledge, this represents the first report on using the difference in nucleotide compositions as described above to crosscheck the RNA sequence and ensure high accuracy is achieved.

In an attempt to evaluate the accuracy on the mathematical conversion of a specific monoisotopic mass to its corresponding nucleotide composition, the theoretical monoisotopic mass of all possible nucleotide compositions of a 5 mer RNA fragment were calculated. The results indicated that there is no identical monoisotopic mass and the smallest mass difference among all the possible nucleotide compositions for a 5 mer RNA fragment is 0.04 Da. With <10 ppm mass accuracy that could be achieved by MALDI SpiralTOF MS, the accuracy on converting a measured monoisotopic mass to its corresponding nucleotide composition in this study is expected to be relatively high, providing the RNA fragment does not have more than 5 nucleotides. As shown in Figure 4 (and Supplemental Figures S3 and S4) the smallest detectable RNA fragment in each sequencing ladder is never longer than 5 mer. Using the smallest detectable RNA fragment is the starting point, the conversion of other monoisotopic masses to their corresponding nucleotide compositions of any longer RNA fragments can become more accurate. This is because the number of possible nucleotide compositions is considerably decreased when the identities of more than one ribonucleotide are known.[31]

As shown in Figure 4, the signal-to-noise ratios across the sequencing ladders are not equal to each other. To ensure the correct assignment of nucleotide composition is achieved when the monoisotopic peak is too weak, unresolved or absent, an approach similar to the "averagine" method proposed by McLafferty and his associates for making an estimate on the monoisotopic mass of proteins is explored.[31] The sequences of all human miRNA show a nearly

equal distribution among the four ribonucleotides (23.20% C, 26.78% U, 23.52% A, and 26.50% G).[32,33] Based on these values, we define an average nucleotide, "averageotide", having the composition of $C_{9.5001}H_{11.7321}N_{3.7325}O_{7.0325}P$ and an average molecular mass of 321.6983 Da. This is closely related to the values defined by Zubarev and Marshall for "averabaseine", although our values are calculated specifically for human microRNAs.[34,35] As an example, by using the averageotide method, the monoisotopic mass of hsa-miR-328 precursor was calculated. The estimated monoisotopic mass of hsa-miR-328 precursor is 24,214.3 Da, which differs from the theoretical value of 24,214.2 Da by only 0.1 Da or 4 ppm. Hence, by using an estimated monoisotopic mass, it is also possible to calculate the nucleotide composition with high accuracy.

Through the collaboration between our laboratories, the developed method was used to sequence three different miRNA in a blinded study, in which no information on the size and the RNA sequence of each miRNA were available when the mass spectral data were analyzed. The three different miRNA were selected randomly and contain different percentage of UC content, thus posing different levels of challenges on the MS measurements as well as the data analysis. All the results are summarized in Table 3, and additional details are available in the Supplemental Figures S3-4 and Supplemental Tables S1-4. In all three cases, 100% sequence coverage as well as 100% sequence accuracy, i.e. measured RNA sequence matched 100% with the expected RNA sequence, were achieved.

CONCLUSIONS

This study represents an important advance on using mass spectrometry to sequence small RNA. Starting from the sample preparation prior to the MS measurements, this is the first report on using acidic MALDI matrix to generate RNA sequencing ladders. Not only the use of any additional chemical or reagent to generate the sequencing ladders is not required, the hydrolyzed RNA fragments can be readily measured without any purification or desalting. For analyzing biological samples in the future, target miRNA has to be isolated from the samples prior to the partial acidic hydrolysis. The isolation of target miRNA can be achieved by the approach of using a complementary DNA probe under the optimal binding conditions. If the biological samples contain a high salt content, the samples can be desalted by various methods, e.g. ZipTip purification, which are fully compatible with MALDI-TOF MS measurements. A second major advance is the ability to achieve accurate monoisotopic mass measurements within the mass range of ~1,000 - 7,000 m/z in a single experiment by using MALDI SpiralTOF MS. Among all three miRNAs, with external calibration, the average mass accuracy was 5.4 ppm. A third major advance is the unique approach of using the difference in nucleotide compositions to deduce the RNA sequence as well as crosschecking the RNA sequence multiple times. No MS/MS experiment was required to achieve 100% sequence coverage for sequencing 22 mer miRNA. By using the developed method, 100% sequence accuracy was achieved in a blinded study. With these results, we do believe the developed method can be useful for *de novo* sequencing of small RNA, or pinpointing the location of RNA modifications. If, the RNA modification creates an isobaric peak in the MALDI spectrum, the use of conventional tandem mass spectrometry is expected to provide sufficient resolving power to identify the RNA modification and allows accurate sequencing to be carried out. Overall, this study provides a simple and accurate method for sequencing miRNA and applicable to other small RNAs.

ACKNOWLEDGEMENTS

We wish to acknowledge the financial support from UNCG and the GK-12 program at NSF as well as the assistance given by Brandie Ehrmann and Vamsikrishna Kandhi.

REFERENCES

1. Dong H, et al. (2013) MicroRNA: Function, Detection, and Bioanalysis. *Chem Rev*, **113**, 6207-6233.
2. Calin, G. and C. Croce. (2006) MicroRNA-cancer connection: the beginning of a new tale. *Cancer Res*, **66(15)**, 7390-4.
3. Jiang, Q., et al. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res*, **37**, D98-104.
4. Park, N., et al. (2009) Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. *Clin Cancer Res*, **15(17)**, 5473-7.
5. Wark, A., H. Lee, and R. Corn. (2008) Multiplexed Detection Methods for Profiling microRNA Expression in Biological Samples. *Angew Chem Int Ed Engl*, **47(4)**, 644-52.
6. Hunt EA, Goulding AM, Deo SK. (2009) Direct detection and quantification of microRNAs. *Anal Biochem*, **387**, 1-12.
7. Giessing AMB, Kirpekar F. (2102) Mass spectrometry in the biology of RNA and its modifications. *Journal of Proteomics*, **75**, 3434-3449.
8. Kwon, Y., et al. (2001) DNA sequencing and genotyping by transcriptional synthesis of chain-terminated RNA ladders and MALDI-TOF mass spectrometry. *Nucleic Acids Res*, **29(3)**, E11.
9. Tang, W., L. Zhu, and L.M. Smith. (1997) Controlling DNA Fragmentation in MALDI-MS by Chemical Modification. *Anal Chem*, **69**(3), 302-12.
10. Huang TY, Liu J, McLuckey SA. (2009) Top-down tandem mass spectrometry of tRNA via ion trap collision-induced dissociation. *J Am Soc Mass Spectrom*, **21**, 890-898.
11. Taucher, M. and K. Breuker. (2010) Top-down mass spectrometry for sequencing of larger (up to 61 nt) RNA by CAD and EDD. *J Am Soc Mass Spectrom*, **21(6)**, 918-29.
12. Kullolli, M., Knouf, E., Arampatzidou, M., Tewari, M., Pitteri, S. J. (2014) Intact microRNA analysis using high resolution mass spectrometry. *J. Am. Soc. Mass Spectrom*., **25**, 80-87.
13. Gao, H; Liu, Y; Rumley, M; Yuan, H; Mao, B. (2009) Sequence confirmation of chemically modified RNAs using exonuclease digestion and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, **23(21)**, 3423-3430.

14. Spottke, B; Gross, J; Galla, H; Hillenkamp, F. (2004) Reverse Sanger sequencing of RNA by MALDI-TOF mass spectrometry after solid phase purification. *Nucleic Acids Res*, **32(12)**, e97/1-e97/8.

15. Li, S., Limbach, P.A. (2012) Method for comparative analysis of ribonucleic acids using isotope labeling and mass spectrometry. *Anal Chem*, **84**, 8607-8613.

16. Bahr, U., H. Aygün, and M. Karas. (2009) Sequencing of single and double stranded RNA oligonucleotides by acid hydrolysis and MALDI mass spectrometry. *Anal Chem*, **81(8)**, 3173-9.

17. Faulstich, K., et al. (1997) a sequencing method for RNA oligonucleotides based on mass spectrometry. *Anal Chem*, **69(21)**, 4349-53.

18. Hahner, S., et al. (1997) Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI) of endonuclease digests of RNA. *Nucleic Acids Res*, **25(10)**, 1957-64.

19. Wambua, D.M., Cody, R. B., Ubukata, M., Dane, J., Chiu, N.H.L. (2012) *De novo* Sequencing of Small RNA by Bottom-Up High Resolution MALDI SpiralTOF Mass Spectrometry. Abstract at ASMS Conference on Mass Spectrometry & Allied Topics.

20. Ishihara, M., M. Toyoda, and T. Matsuo. (2000) Perfect spacial and isochronous focusing ion optics for multi-turn time of flight mass spectrometer. *Int. J. Mass Spectrom.*, **197**, 179-189.

21. Toyoda, M. (2010) Development of multi-turn time-of-flight mass spectrometers and their applications. *Eur. J. Mass Spectrom.,* **16(3)**, 397-406.

22. Toyoda, M., et al. (2003) Multi-turn time-of-flight mass spectrometers with electrostatic sectors. *J. Mass Spectrom.*, **38**, 1125-1142.

23. Satoh, T., et al. (2011) Tandem time-of-flight mass spectrometer with high precursor ion selectivity employing spiral ion trajectory and improved offset parabolic reflectron. *J Am Soc Mass Spectrom*, **22(5)**, 797-803.

24. Satoh, T., T. Sato, and J. Tamura. (2007) Development of a high-performance MALDI-TOF mass spectrometer utilizing a spiral ion trajectory. *J Am Soc Mass Spectrom*, **18(7)**, 1318-1323.

25. Satoh, T., et al. (2005) The design and characteristic features of a new time-of-flight mass spectrometer with a spiral ion trajectory. *J Am Soc Mass Spectrom*, **16(12)**, 1969-1975.

26. Satoh, T., et al. (2006) A new spiral time-of-flight mass spectrometer for high mass analysis. *J. Mass Spectrom. Soc. Jpn*., **541(1)**, 11-17.

27. Oivanen, M., S. Kuusela, and H. Lönnberg. (1998) Kinetics and Mechanisms for the Cleavage and Isomerization of the Phosphodiester Bonds of RNA by Brønsted Acids and Bases. *Chem Rev*, **98(3)**, 961-990.

28. Perreault, D.M. and E.V. Anslyn. (1997) Unifying the Current Data on the Mechanism of Cleavage–Transesterification of RNA. *Angew Chem Int Ed Engl.*, **36**, 432-450.

29. Pieles, U., et al. (1993) Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: a powerful tool for the mass and sequence analysis of natural and modified oligonucleotides. *Nucleic Acids Res*, **21(14)**, 3191-6.

30. Pomerantz SC, Kowalak JA, McCloskey JA. (1993) Determination of oligonucleotide composition from mass spectrometrically measured molecular weight. *J Am Soc Mass Spectrom*, **4**, 204-209.

31. Senko, M.W., S.C. Beu, and F.W. McLafferty. (1995) Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J Am Soc Mass Spectrom*, **6(4)**, 229-233.

32. Griffiths-Jones S, Grocock R, van Dongen S, Bateman A, Enright A. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res,* **34**, D140-144.

33. Griffiths-Jones S, Saini H, van Dongen S, Enright A. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res*, **36**, D154-158.

34. Zubarev, R., Demirev, P. (1998) Isotope depletion of large biomolecules: Implications for molecular mass measurements. *J Am Soc Mass Spectrom, **9** **(2)***, 149-156.

35. Xiong, Y., Schroeder, K., Greenbaum, N. L., Hendrickson, C. L., Marshall, A. G. (2004) Improved Mass Analysis of Oligoribonucleotides by ^{13}C, ^{15}N Double Depletion and Electrospray Ionization FT-ICR Mass Spectrometry. *Anal Chem, **76** **(6)***, 1804-1809.

TABLE AND FIGURES LEGENDS

Figure 1. Schematic diagram of the developed method for achieving accurate RNA sequencing of microRNA. In the developed method, microRNA is partially hydrolyzed with the conventional acidic MALDI matrix. During the hydrolysis, each microRNA fragment is cut into two halves, thus generating a 5' fragment and a 3' fragment. Collectively, the 5' and 3' fragments are equivalent to two sets of opposite sequencing ladders (5'→3' and 3'→5'). Both 5' and 3' fragment ladders are measured simultaneouslyby using MALDI-TOF MS. For determining the RNA sequence, the nucleotide composition of each fragment is calculated from its monoisotopic mass. The RNA sequence is then determined by comparing the nucleotide composition of two adjacent peaks within the same sequencing ladder. As indicated in the callout, the comparison of nucleotide composition also serves as a unique approach for crosschecking the ribonucleotide identity at each position multiple times. Together with a double read, i.e. analyzing two opposite sequencing ladders, the RNA sequence of a specific microRNA is accurately determined.

Figure 2. (a) Molecular structure and pKa value of acidic MALDI matrices that were used in this study. 3-HPA = 3-hydroxypicolinic acid; DHB = 2,5-dihydroxybenzoic acid; and SA= sinapinic acid. All pKa values at 25^0C were calculated by using Advanced Chemistry Development (ACD/Labs) Software V11.02. (b) Mechanism for acidic hydrolysis of RNA fragment.

Figure 3. (a) Effects of using selected MALDI matrices to hydrolyze 20 pmol of miR-153 for 30 min at 60^oC. In all cases, saturated MALDI matrix solution was prepared in either 10% or 50% acetonitrile. (b) Optimization of incubation temperature and time on the hydrolysis of 20 pmol of miR-153 with saturated 3-HPA in 10% acetonitrile. All data were obtained by using 4700 Proteomic Analyzer from Applied Biosystem. The error bars represent the standard deviation of 3 replicates.

Figure 4. MALDI SpiralTOF mass spectrum of partially hydrolyzed miR-153. Positive ions were measured. Each 5' fragment resulting from the limited acidic hydrolysis of miR-153 is labeled with an asterisk and a numerical number that corresponds to how many ribonucleotides are absent from the 3' end of an unhydrolyzed miR-153. Each 3' fragment resulting from the limited acidic hydrolysis of miR-153 is labeled with a numerical number that corresponds to how many ribonucleotides are absent from the 5' end of an unhydrolyzed miR-153. The sequence coverage provided by each sequencing ladder is shown above the spectrum.

Figure 5. Isotopic resolution of molecular ions that correspond to each 3' fragment as shown in Figure 4, which resulted from the limited acidic hydrolysis of miR-153. Each subspectrum is labeled with a numerical number that corresponds to how many ribonucleotides are absent from the 5' end of an unhydrolyzed miR-153, and follows by the ribonucleotide identity at the 5' terminal of each 3' fragment. Both theoretical mass and observed monoisotopic mass of each 3' fragment of miR-153 are listed in Table 2.

Table 1. Accuracy on using MALDI SpiralTOF MS to measure the monoisotopic mass of each 5' fragment resulting from the limited acidic hydrolysis of miR-153. Each 5' fragment contains a phosphate group at the 5' terminal and a 2',3'-cyclic phosphate at the 3' terminal, except the 3' terminal of unhydrolyzed miR-153 fragment has a hydroxyl group. RMS = root mean square.

Table 2. Accuracy on using MALDI SpiralTOF MS to measure the monoisotopic mass of each 3' fragment resulting from the limited acidic hydrolysis of miR-153. Each 3' fragment contains a hydroxyl group at both 5' and 3' terminal, except the 5' terminal of unhydrolyzed miR-153 fragment has a phosphate group. RMS = root mean square.

Table 3. Summary of results obtained from the sequencing of selected miRNA.

Supplemental Figures and Tables

Figure S1. Size distribution of human mature microRNA. All data were obtained from miRBase.

Figure S2. Distribution of isotopic peaks that correspond to $[\text{miR-153} + \text{H}]^+$ ions. (a) Experimental data obtained by using MALDI SpiralTOF, and (b) theoretical simulation with resolving power of 20,000. The simulation was generated by using *msTornado Analysis* software.

Figure S3. MALDI SpiralTOF mass spectrum of partially hydrolyzed miR-183-5p. Positive ions were measured. Each 5' fragment resulting from the limited acidic hydrolysis of miR-183-5p is labeled with an asterisk and a numerical number that corresponds to how many ribonucleotides are absent from the 3' end of an unhydrolyzed miR-183-5p. Each 3' fragment resulting from the limited acidic hydrolysis of miR-183-5p is labeled with a numerical number that corresponds to how many ribonucleotides are absent from the 5' end of an unhydrolyzed miR-183-5p.

Figure S4. MALDI SpiralTOF mass spectrum of partially hydrolyzed miR-124a. Positive ions were measured. Each 5' fragment resulting from the limited acidic hydrolysis of miR-124a is labeled with an asterisk and a numerical number that corresponds to how many ribonucleotides are absent from the 3' end of an unhydrolyzed miR-124a. Each 3' fragment resulting from the

limited acidic hydrolysis of miR-124a is labeled with a numerical number that corresponds to how many ribonucleotides are absent from the 5' end of an unhydrolyzed miR-124a.

Table S1. Accuracy on using MALDI SpiralTOF MS to measure the monoisotopic mass of each 5' fragment resulting from the limited acidic hydrolysis of miR-183-5p. Each 5' fragment contains a phosphate group at the 5' terminal and a 2',3'-cyclic phosphate at the 3' terminal, except the 3' terminal of unhydrolyzed miR-183-5p fragment has a hydroxyl group. RMS = root mean square.

Table S2. Accuracy on using MALDI SpiralTOF MS to measure the monoisotopic mass of each 3' fragment resulting from the limited acidic hydrolysis of miR-183-5p. Each 3' fragment contains a hydroxyl group at both 5' and 3' terminal, except the 5' terminal of unhydrolyzed miR-183-5p fragment has a phosphate group. RMS = root mean square.

Table S3. Accuracy on using MALDI SpiralTOF MS to measure the monoisotopic mass of each 5' fragment resulting from the limited acidic hydrolysis of miR-124a. Each 5' fragment contains a phosphate group at the 5' terminal and a 2',3'-cyclic phosphate at the 3' terminal, except the 3' terminal of unhydrolyzed miR-124a fragment has a hydroxyl group. RMS = root mean square.

Table S4. Accuracy on using MALDI SpiralTOF MS to measure the monoisotopic mass of each 3' fragment resulting from the limited acidic hydrolysis of miR-124a. Each 3' fragment contains a hydroxyl group at both 5' and 3' terminal, except the 5' terminal of unhydrolyzed miR-124a fragment has a phosphate group. RMS = root mean square.

Figure 1

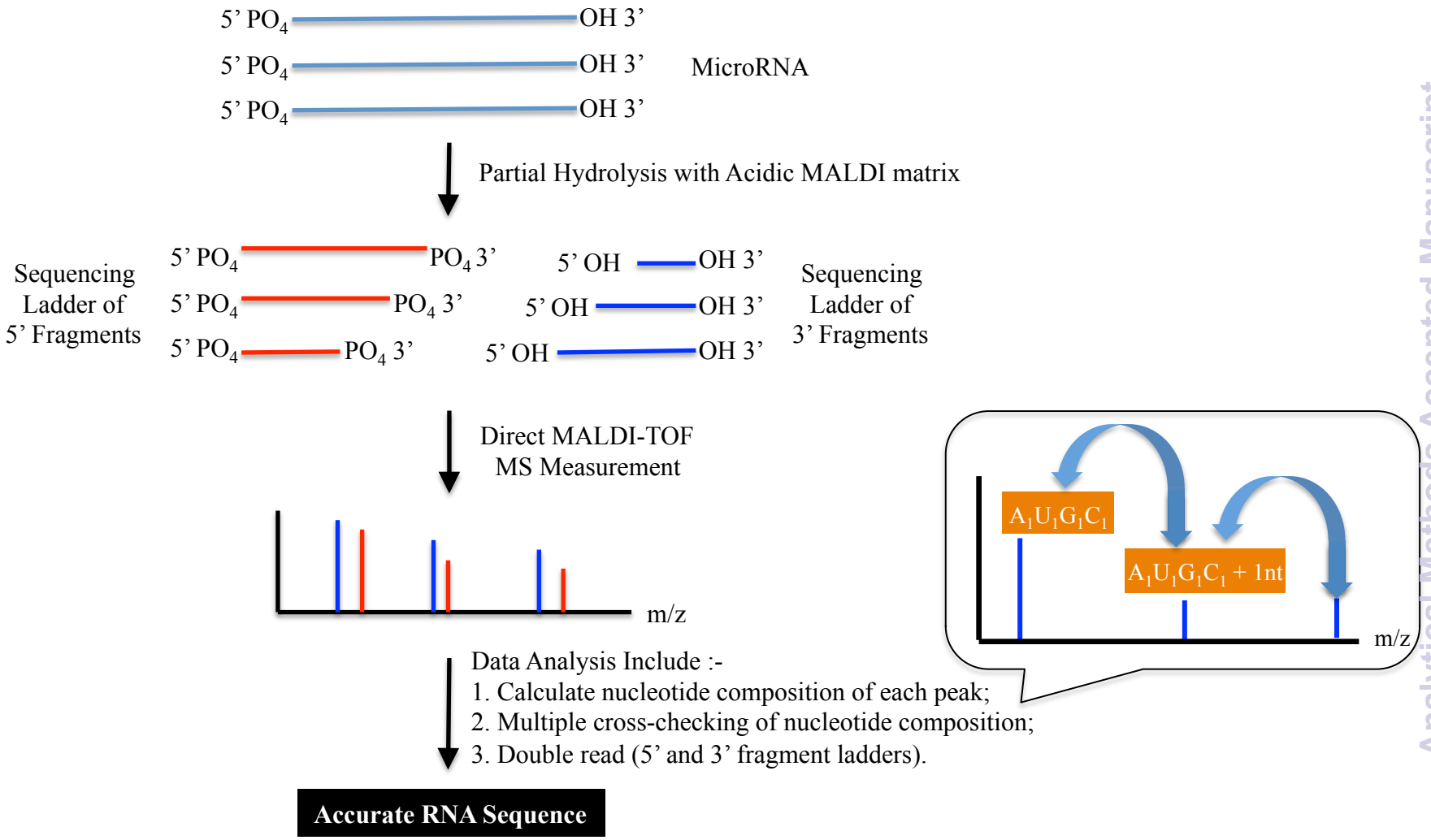
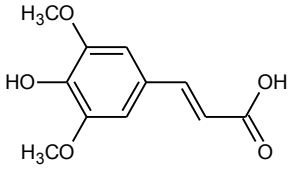
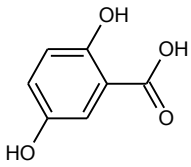


Figure 2

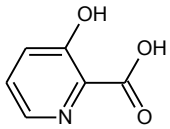
(a)



SA ($pKa_1 = 4.53$)



DHB ($pKa_1 = 3.01$)



3-HPA ($pKa_1 = 1.14$)

(b)

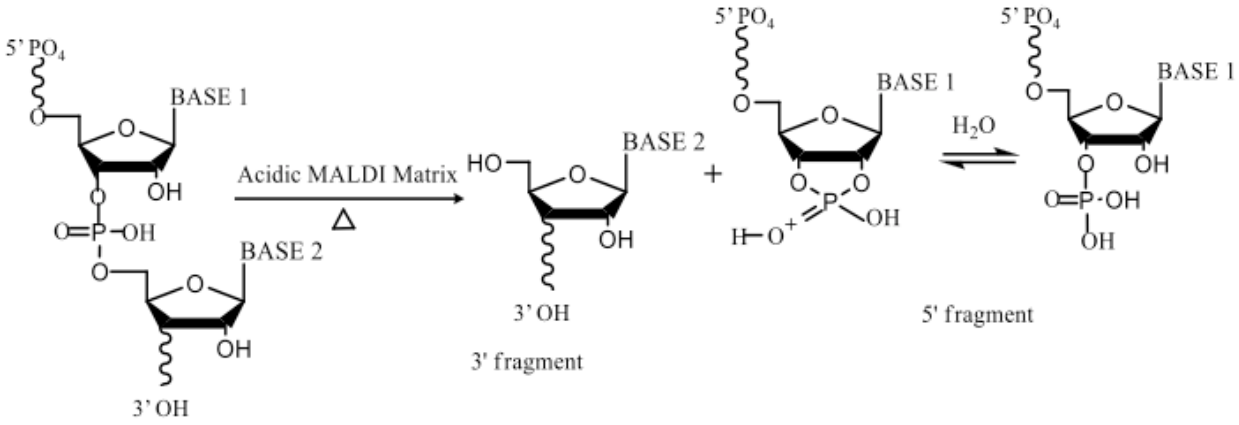
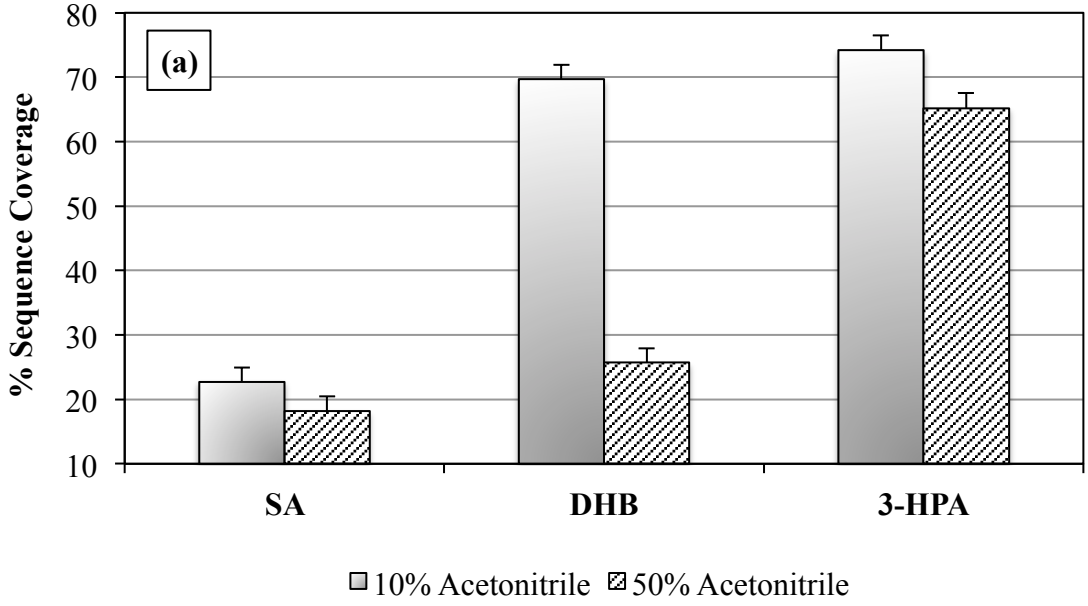


Figure 3



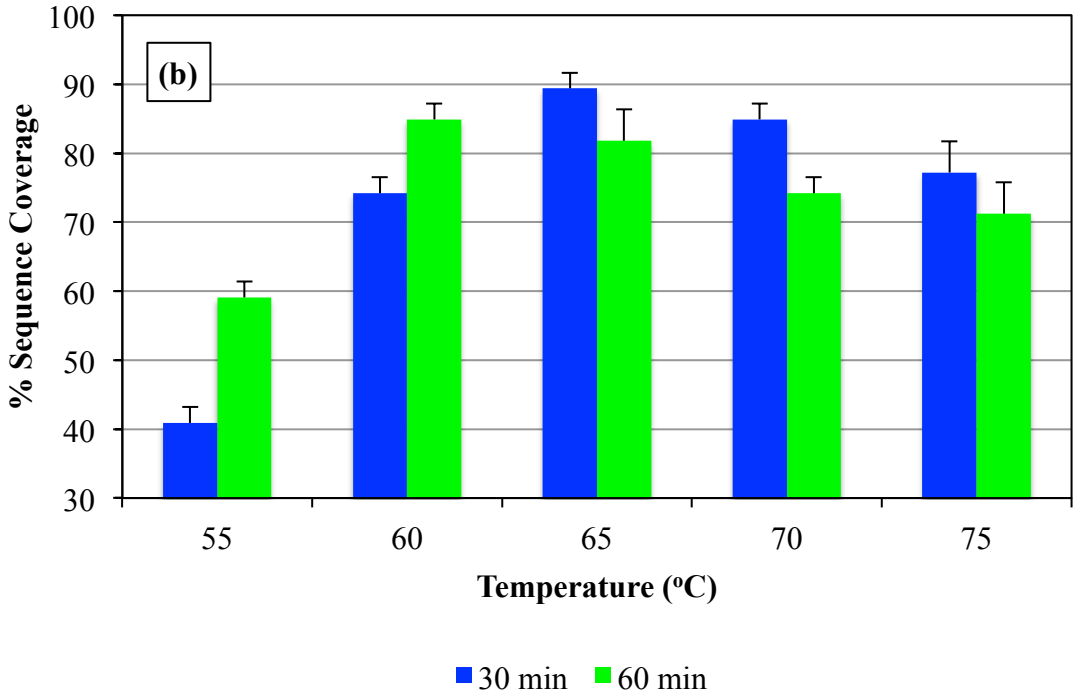


Figure 4

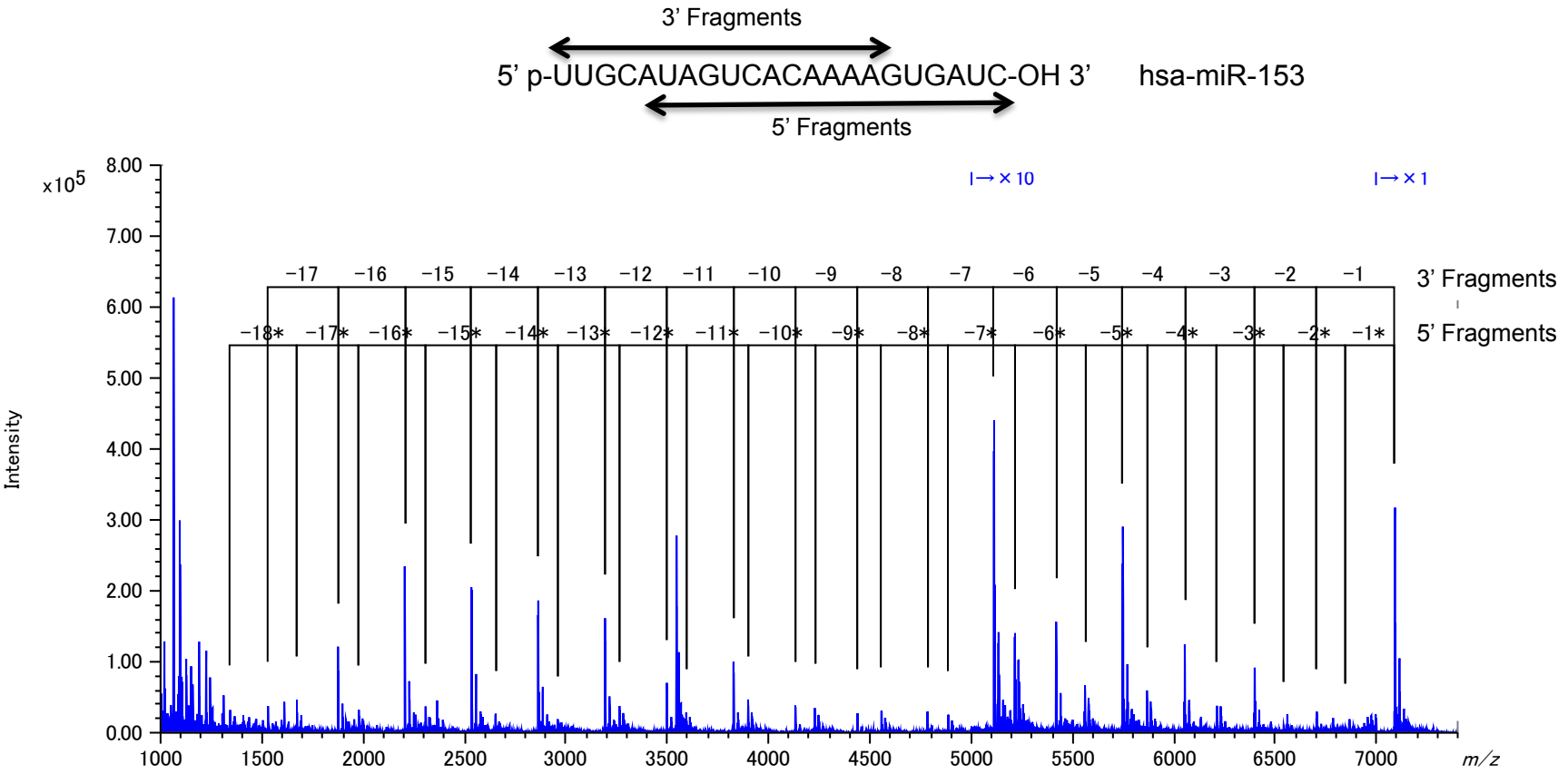


Figure 5

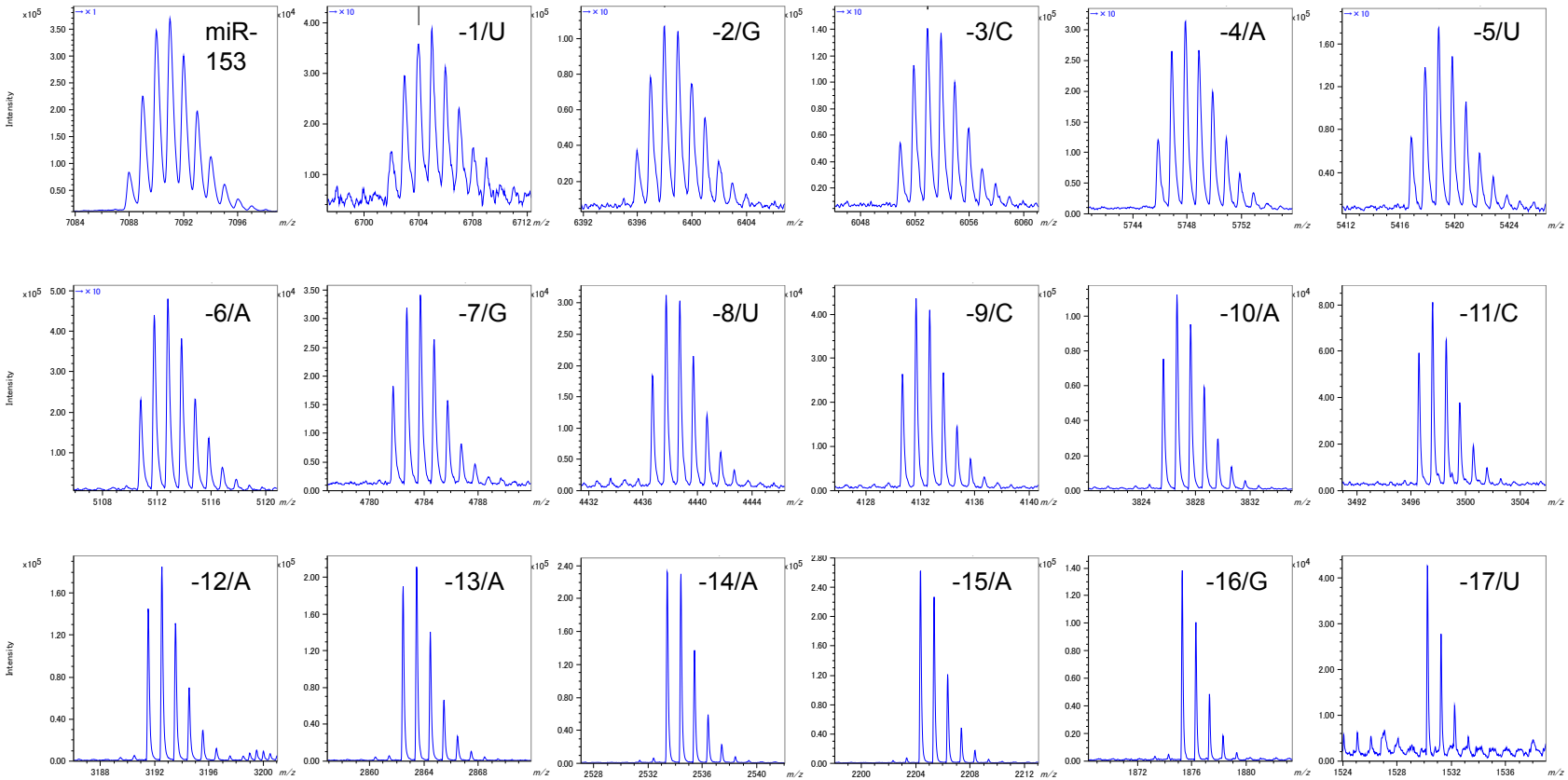


Table 1

5' Fragments of miR-153	Theoretical Mass (m/z)	Observed Mass (m/z)	Mass Difference (m/z)	Mass Accuracy (ppm)
pUUGCp'	1343.1129	1343.1204	-0.0075	-5.561707
pUUGCAp'	1672.1655	1672.1711	-0.0057	-3.378852
pUUGCAUp'	1978.1908	1978.2036	-0.0129	-6.495835
pUUGCAUAp'	2307.2433	2307.2535	-0.0102	-4.433863
pUUGCAUAGp'	2652.2907	2652.3023	-0.0116	-4.369808
pUUGCAUAGUp'	2958.3160	2958.3336	-0.0176	-5.945950
pUUGCAUAGUCp'	3263.3573	3263.3716	-0.0143	-4.385055
pUUGCAUAGUCAp'	3592.4098	3592.4445	-0.0347	-9.656471
pUUGCAUAGUCACp'	3897.4511	3897.4715	-0.0204	-5.234190
pUUGCAUAGUCACAp'	4226.5036	4226.5211	-0.0175	-4.135806
pUUGCAUAGUCACAAp'	4555.5561	4555.5786	-0.0225	-4.930243
pUUGCAUAGUCACAAAp'	4884.6087	4884.6317	-0.0230	-4.716857
pUUGCAUAGUCACAAAAp'	5213.6612	5213.6933	-0.0321	-6.160738
pUUGCAUAGUCACAAAAGp'	5558.7086	5558.7486	-0.0400	-7.194117
pUUGCAUAGUCACAAAAGUp'	5864.7339	5864.7926	-0.0587	-10.00557
pUUGCAUAGUCACAAAAGUGp'	6209.7814	6209.8190	-0.0377	-6.063015
pUUGCAUAGUCACAAAAGUGAp'	6538.8339	6538.9382	-0.1043	-15.95544
pUUGCAUAGUCACAAAAGUGAUp'	6844.8592	6844.9683	-0.1091	-15.94335
pUUGCAUAGUCACAAAAGUGAUC	7087.9447	7088.0505	-0.1058	-14.92816
				RMS = 8.3

Table 2

3' Fragments of miR-153	Theoretical Mass (m/z)	Observed Mass (m/z)	Mass Difference (m/z)	Mass Accuracy (ppm)
UGAUC	1530.2434	1530.2468	-0.0034	-2.254543
GUGAUC	1875.2908	1875.2924	-0.0016	-0.858534
AGUGAUC	2204.3433	2204.3494	-0.0061	-2.762728
AAGUGAUC	2533.3958	2533.4020	-0.0062	-2.435466
AAAGUGAUC	2862.4484	2862.4553	-0.0069	-2.427991
AAAAGUGAUC	3191.5009	3191.5091	-0.0082	-2.578724
CAAAAGUGAUC	3496.5422	3496.5500	-0.0078	-2.245075
ACAAAAGUGAUC	3825.5947	3825.6079	-0.0132	-3.458286
CACAAAAGUGAUC	4130.6360	4130.6508	-0.0148	-3.592667
UCACAAAAGUGAUC	4436.6613	4436.6755	-0.0142	-3.209621
GUCACAAAAGUGAUC	4781.7087	4781.7250	-0.0163	-3.408823
AGUCACAAAAGUGAUC	5110.7612	5110.7828	-0.0216	-4.222463
UAGUCACAAAAGUGAUC	5416.7865	5416.8136	-0.0271	-4.999274
AUAGUCACAAAAGUGAUC	5745.8390	5745.8657	-0.0267	-4.639879
CAUAGUCACAAAAGUGAUC	6050.8803	6050.9061	-0.0258	-4.260537
GCAUAGUCACAAAAGUGAUC	6395.9278	6395.9766	-0.0488	-7.636109
UGCAUAGUCACAAAAGUGAUC	6701.9531	6702.0371	-0.0840	-12.53963
pUUGCAUAGUCACAAAAGUGAUC	7087.9447	7088.0505	-0.1058	-14.92816
				RMS = 5.8

Table 3

microRNA	Size (nt)	UC Content[1] (%)	Sequence Crosscheck[2] (%)	Sequence Coverage[3] (%)	Ave. Mass Accuracy[4] (ppm)	Sequence Accuracy[5] (%)
miR-183-5p	22	50	80	100	3.0	100
miR-153	22	45	80	100	7.1	100
miR-124a	22	41	82	100	6.2	100

[1] Percentage of nucleotide that are either uridine (U) or cytidine (C) within a specific miRNA.
[2] Average percentage of sequence that can be crosschecked by comparing nucleotide composition obtained from a single read.
[3] Total sequence coverage achieved by double read.
[4] Average mass accuracy achieved by using MALDI SpiralTOF MS to measure both 5' and 3' fragments.
[5] Comparing RNA sequence obtained from a blinded study in this report to the reference RNA sequence in miRBase.