analytical chemistry

Article

pubs.acs.org/ac

Direct Identification of Human Cellular MicroRNAs by Nanoflow Liquid Chromatography–High-Resolution Tandem Mass Spectrometry and Database Searching

Hiroshi Nakayama,*[,†,‡] Yoshio Yamauchi,[‡,§] Masato Taoka,[‡,§] and Toshiaki Isobe*[,‡,§]

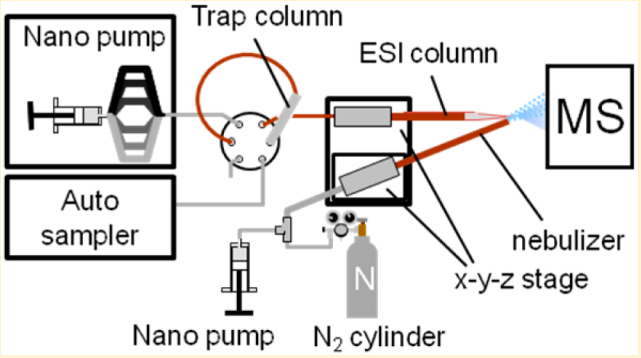
[†]Biomolecular Characterization Team, RIKEN Advanced Science Institute, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

[‡]Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Sanbancho 5, Chiyoda-ku, Tokyo 102-0075, Japan

[§]Department of Chemistry, Graduate School of Science and Engineering, Tokyo Metropolitan University, Minami-osawa 1-1, Hachioji-shi, Tokyo 192-0397, Japan

Supporting Information

ABSTRACT: MicroRNAs (miRNAs) are small noncoding RNAs that regulate gene networks and participate in many physiological and pathological pathways. To date, miRNAs have been characterized mostly by genetic technologies, which have the advantages of being very sensitive and using high-throughput instrumentation; however, these techniques cannot identify most post-transcriptional modifications of miRNAs that would affect their functions. Herein, we report an analytical system for the direct identification of miRNAs that incorporates nanoflow liquid chromatography–high-resolution tandem mass spectrometry and RNA-sequence database searching. By introducing a spray-assisting device that stabilizes negative nanoelectrospray ionization of RNAs and by searching an miRNA sequence database using the obtained tandem mass spectrometry data for the RNA mixture, we successfully identified femtomole quantities of human cellular miRNAs and their 3′-terminal variants. This is the first report of a fully automated, and thus objective, tandem mass spectrometry-based analytical system that can be used to identify miRNAs.



MicroRNAs (miRNA), a class of small noncoding RNAs, regulate gene networks by hybridizing to complementary mRNA sequences, which results in targeted mRNA degradation or translational repression. On the basis of the sequences contained in the current version of the miRNA database miRBase[1] (release 21, June 2014), the human genome encodes >1800 possible miRNAs, which can target ∼60% of human protein-coding transcripts.[2] Because human miRNAs are involved in many physiological pathways, their deregulation and/or dysfunction can lead to disease states, including cancer, metabolic diseases, and cardiovascular disorders.[3−5] Thus, various miRNAs are thought to be excellent targets for drug development and, possibly, valuable diagnostic markers.

The pathways for miRNA biogenesis are complex[6] and include post-transcriptional modifications (PTMs).[7−9] For example, plant miRNAs are 2′-*O*-methylated at the 3′-terminus ribose by the methyltransferase HEN1,[7] and miR-122, a mammalian liver-specific miRNA, is adenylated by the cytoplasmic poly(A) polymerase GLD-2 at its 3′-end.[9] These modifications contribute to the stabilization of those RNAs by protecting their 3′-ends from exonuclease activity and/or enzymatic uridylation.[10] Thus, PTMs play important roles in the structure and function of miRNAs, as has been reported for other noncoding RNAs.[11,12]

miRNAs have been characterized mostly with the use of highly sensitive and high-throughput genetic technologies, including deep sequencing by high-speed sequencers, microarrays, and real-time PCR after reverse transcription.[13,14] In fact, deep sequencing allows for the identification of new miRNAs and can assess expression levels by counting the number of sequence reads.[13,14] These technologies, however, have limited ability to characterize PTMs. Conversely, mass spectrometry (MS) is a suitable technology for analyzing RNA chemical structures and, thereby, complements the genetic approach.[7,9] In particular, high-resolution electrospray ionization (ESI)-tandem mass spectrometry (MS/MS) that incorporates collision-induced dissociation, electron-detachment dissociation, or infrared multiphoton dissociation[15−21] has been used successfully for sequencing of miRNAs,[20,21] small interfering RNAs,[15,19] and heavily modified tRNA.[18] However, to date, MS-based characterization of RNA, in general, has been

Received: November 24, 2014
Accepted: February 7, 2015
Published: February 8, 2015

DOI: 10.1021/ac504378s
Anal. Chem. 2015, 87, 2884−2891

much less sensitive than the available proteomic techniques that incorporate nanoflow liquid chromatography (nLC), which drastically increases the sensitivity of mass spectral analyses.[22] Notably, prior to this report, nLC had hardly been applied to RNA characterization. Because RNA is a hydrophilic molecule, it elutes from a reversed-phase LC column at a low organic solvent concentration, which has a relatively high surface tension and thereby tends to interfere with stable electrospray ionization under the nanoflow condition.

MS-based characterization of miRNA has also suffered from the absence of an algorithm that allows for the automated sequencing/identification of miRNA using MS/MS data. Two programs, SOS and COMPAS, are available for assisting in *de novo* sequencing of RNA.[23,24] SOS expects an interactive operation for helping manual sequencing. COMPAS can perform automated *de novo* sequencing of only short nucleotide sequences (≤12 nucleotides) using MS/MS data.[24] We have developed the RNA identification program Ariadne that can search DNA/RNA sequence databases for sequences matching those found in MS/MS data of RNase-digested RNA.[25,26] Although the utility of Ariadne for the analysis of various RNAs from biological sources[25−30] has been proven, the program could not be applied to miRNA analysis because RNase digestion of miRNAs generates only a few small oligonucleotides, which are insufficient to uniquely identify an miRNA in a database.

Herein, we present an nLC−MS/MS-based analytical system applicable to the characterization of miRNAs. We stabilized the nanospray of our nLC−MS/MS system by introducing a spray-assisting device and enabled Ariadne to identify miRNAs by incorporating an intact-RNA search function. These improvements allowed us to fully automated and unbiased identification of a femtomole quantity of human cellular miRNAs and their variants.

EXPERIMENTAL SECTION

Chemicals. Triethylammonium acetate (2 M) was purchased from Glen Research Corporation (Sterling, VA, USA). A Dynamarker RNA Low II kit (RNAs 20−500 nucleotides in length) was from Biodynamics Laboratory (Tokyo, Japan). HPLC-grade acetonitrile and water were obtained from Wako Pure Chemical Industries (Osaka, Japan).

Synthetic miRNAs. Synthetic *Mus musculus* miRNAs, mmu-miR-335 (5′-HO-UCAAGAGCAAUAACGAAAAAUGU-OH-3′), mmu-miR-295 (5′-HO-AAAGUGCUACUACUUUUGAGUCU-OH-3′), and mmu-miR-92a (5′-HO-UAUUGCACUUGUCCCGGCCUG-OH-3′), and *Arabidopsis thaliana* miRNA, ath-miR-173 (5′-p-UUCGCUUGCAGAGAGAAAUCACm-OH-3′; Cm: 2′-*O*-methylated cytidine), were obtained from JBioS (Saitama, Japan).

Preparation of miRNAs from HeLa Cells. A small RNA fraction was prepared from HeLa cells (8×10^7 cells) using mirVana miRNA Isolation kit reagents (Life Technologies, Austin, TX, USA). The resulting fraction (100 μg of RNA) was further purified by reversed phase-LC through a stainless steel column (2 mm i.d. × 100 mm length) containing macro-porous polystyrene/divinylbenzene resin (PLRP-S 300 Å, 3 μm, Polymer Laboratories Ltd., UK).[29] A high-pressure binary gradient LC system LC10ADvp (Shimadzu, Kyoto, Japan) connected to a UV detector (SPD-20A) equipped with a semimicro flow cell (5 μL volume, 5 mm path length) was used. The column was held at 60 °C in a column oven during the chromatography. The HeLa-cell small RNAs (~300 ng) were separated manually upon elution and correspond to RNAs 18−30 nucleotides in length based on the elution positions of the Dynamarker Low II kit standards.

An nLC−MS/MS System Equipped with a Spray-Assisting Device. nLC−MS/MS was performed as described[28] with the following modifications. The LC system consisted of a direct nanoflow pump (LC Assist, Tokyo, Japan), a ReNCon gradient device (LC Assist), an autoinjector (SIL-10ADvp Shimadzu, Kyoto, Japan) that feeds samples into a trap column (MonoCap trap C18 column, 0.2 mm i.d. × 30 mm length; GL science, Tokyo, Japan) connected to a Valco 6-port, 2-position valve, a frit-less electrospray column (150 μm i.d. × 50 mm length; packed with Develosil C30-UG3 resin, Nomura Chemical, Nagoya, Japan), and a spray-assisting device as described below (Figure 1).

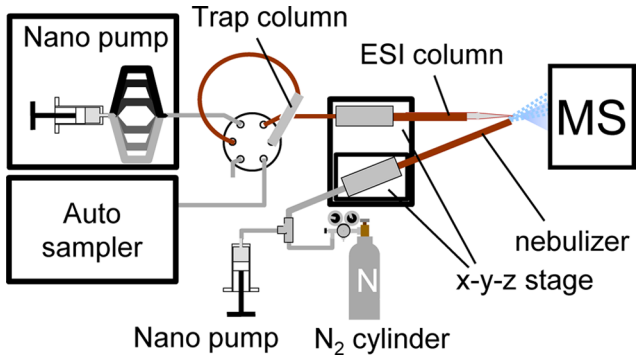


Figure 1. Schematic of the nLC−MS/MS system.

nLC was performed at a flow rate of 100 nL/min using a 30 min linear gradient (2−14% (v/v) acetonitrile in a solution containing 10% (v/v) methanol and 10 mM aqueous triethylammonium acetate, pH 7.0). We used this elution solvent instead of a conventional hexafluoro-2-propanol/triethylamine/methanol system,[31] because the latter exhibited insufficient separation of the synthetic miRNAs. The spray-assisting device consisted of a blunt-end fused silica capillary nozzle (50 μm i.d/375 μm o.d. × 50 mm length) that was placed on a *x-y-z* positioner, a second nano pump, and an N_2 gas cylinder, which were all connected by PEEK tubes and a T-connector (Figure 1). The tip of the spray-assisting nozzle was oriented to be 1 mm behind the ESI column tip by the *x-y-z* positioner. During nLC−MS/MS, acetonitrile was delivered at a flow rate of 300 nL/min by the nano pump and N_2 gas was delivered at 0.1 MPa (the N_2 flow rate was ~20 mL/min). Acetonitrile and N_2 were mixed at the T-connector, and the resulting mist was sprayed out of the tip of the assisting nozzle. The RNA sample was introduced by the autoinjector into the trap column, and after washing, the two-position valve was switched to connect the trap and ESI columns. The nLC eluate was sprayed online at −1.3 kV with the aid of the spray-assisting device to a Q-Exactive mass spectrometer, a hybrid quadrupole, and an Orbitrap spectrometer (Thermo Fisher Scientific, San Jose, CA), operating in the negative polarity. The spectrometer was operated in a data-dependent mode so that it was automatically switched between MS and MS/MS acquisition. Survey full-scan mass spectra (m/z 1500−1990) were acquired with a mass resolution of 35 000 at m/z 400. For the 3 most intense mass peaks in a survey scan with an intensity that was greater than 25 000 counts/s, each peak was isolated within a 3-m/z window for fragmentation. The mass peaks were

DOI: 10.1021/ac504378s
Anal. Chem. 2015, 87, 2884−2891

fragmented by higher-energy collisional dissociation with a normalized collision energy of 15. To retain mass resolution and to increase spectral quality, 3 "micro"-scans of MS/MS measurements, each of which fragmented up to 50 000 ions/s, with a mass resolution of 17 500 at m/z 400 were accumulated for a single MS/MS spectrum. The fixed starting mass value for MS/MS was 220.

DB Searching with Ariadne. The nLC−MS/MS raw data were processed using the peak-picking program, SpiceCmd (Mitsui Knowledge Inc., Tokyo, Japan), which we used to determine the mass and charge state of the monoisotopic ions in the mass or tandem mass spectra. SpiceCmd also deconvoluted these peaks and merged the data from the tandem mass spectra having the same precursor ion associated with a predetermined chromatographic time range (e.g., 1 min). The resulting peak list was written in a file in Mascot generic format (mgf) by an in-house Perl script, and the file was then used to search the miRNA sequence DB derived from miRBase[1] by Ariadne[25] with certain modifications. Ariadne was originally developed for the bottom-up identification of relatively long RNAs using the MS/MS spectral data of their RNase digests. For top-down analysis of intact RNAs, we added the Enzyme parameter "no-enzyme" so that Ariadne does not "cleave" RNA sequences in a DB but considers possible combinations of 5′ and 3′ termini when calculating mass values.

Ariadne initially calculates mass values for sequences retrieved from the database and filters them by comparison with the experimentally determined precursor mass value. It then calculates and compares mass values of *in silico*-derived product ions with those of the monoisotopic peaks extracted from the tandem mass spectrum. Product ions considered by Ariadne are the 5′-terminus-containing a and c ions, the 3′-terminus-containing w and y ions, multiply fragmented internal ions, and ions that had lost ≤2 neutral or charged bases from the molecular ion. Search tolerances were as follows: precursor mass tolerance, ± 5 ppm; MS/MS tolerance, ± 50 ppm.

miRNA sequence database were prepared as follows. *M. musculus* and *A. thaliana* miRNA sequences were extracted from the fasta file containing mature miRNA sequences (mature.fa) downloaded from the miRBase, release 21. New fasta files were created by the extracted sequences and were used as the miRNA databases (1193 and 325 entries for *M. musculus* and *A. thaliana* miRNAs, respectively) for Ariadne searches. For human miRNAs, the sequences with 5′ and/or 3′variants were generated from mature miRNA sequences (mature.fa) and another fasta file containing precursor sequences (hairpin.fa) downloaded from the miRBase, with the aid of Perl scripts developed in-house. Initially, miRNAs labeled "Homo sapiens" were extracted (1881 precursor and 2588 mature sequences). Then, their 5′ and 3′neighbor(s) were identified by comparing the sequence(s) with the corresponding precursor sequence(s). For this study, two neighbor residues of both termini were identified. Next, variant sequences were generated by adding a neighboring residue to and/or deleting the terminal residue from a mature miRNA sequence. Finally, the variant sequences were added into the mature sequence database, and multiplicated sequences in the merged database were unified. The resulting database containing 55 176 human miRNA sequences was used for Ariadne searches.

RESULTS

Spray-Assisting Device for Stabilizing the Nano Electrospray Ionization of RNA. A major drawback of the use of nLC−MS(/MS) for the identification of RNA sequences has been the instability of nano spray ionization when the mass spectrometer is run in the negative mode. The conventional coaxial setting for supplying sheath gas and/or liquid to stabilize the nano spray was difficult to implement because our ESI column was too short (50 mm) and fragile to set up properly into a nebulizer tube. Thus, we designed a spray-assisting device (Figure 1) and assessed its performance by evaluating the correlation between the spray conditions and the signal stabilities for the four test miRNAs (see Experimental Section). In the absence of the spray-assisting device, a droplet of eluate grew at the tip of the ESI source (data not shown). When the droplet was removed by manually wiping or blowing it off and when the concentration of the organic solvent was sufficiently large (e.g., ≥20% (v/v) acetonitrile/10% (v/v) methanol or ≥30% (v/v) methanol), continuous spraying occurred thereafter.

However, when the organic solvent concentration was not sufficient to promote spraying, only a new droplet formed. Conversely, when the spray-assisting device was operational, the spray was stable regardless of the organic solvent concentration of the chromatographic eluate, and the ion signals were always present during the analysis. It should be noted that our spray assisting device helped to assist ESI of miRNAs during the entire run, because the separation of miRNA was completed under the solvent condition containing <30% organic solvent (14% acetonitrile and 10% methanol). We examined two organic solvents, acetonitrile and methanol, as spray-assisting solvent with essentially the same results, while the addition of only N_2 gas without the organic solvent stabilized the spray less effectively and reproducibly. Thus, we anticipated that our spray-assisting device generates mists of acetonitrile/methanol and blows the mists continuously to merge with the chromatographic eluent and this would help to lower the surface tension of the eluent and facilitates its vaporization. We also noted that, if the organic solute content was sufficiently large, the addition of the device had little effect on the stability and intensity of the mass signals, suggesting that the spray-assisting device is useful only under low organic solvent concentration conditions.

We also assessed if the device affected the separation of the test miRNAs. The four miRNAs in a mixture were completely separated when the standard 30 min linear nLC-elution gradient, 2−14% (v/v) acetonitrile in 10% (v/v) methanol/10 mM triethylammonium acetate (pH 7.0), was used (Figure 2A). The full peak width at half the maximum height was 10−15 s for each peak. These peak widths are comparable to those found for conventional nLC−MS proteomic analyses.[22] Thus, the spray-assisting device stabilizes the nanoelectrospray ionization without broadening the chromatographic peaks and thereby allows for the automated and routine operation of the nLC−MS system.

Subfemtomole Detection of miRNAs and a Three-Order Dynamic Range in Their Mass Signals Using the nLC−MS System. To evaluate the sensitivity of our nLC−MS/MS system, we employed ath-miR-173 as a test miRNA. We could detect 300 amol of ath-miR-173 with a signal-to-noise ratio >10 (Figure 2B). Ions with four to six negative charges were detected, with the quadruply deprotonated anion at m/z 1778.000 being dominant. In addition, we detected minor ath-miR-173 anions noncovalently bound to triethylamine, which was used as the nLC ion-pair, and to diethylamine, which probably was a decomposition product

DOI: 10.1021/ac504378s
Anal. Chem. 2015, 87, 2884−2891

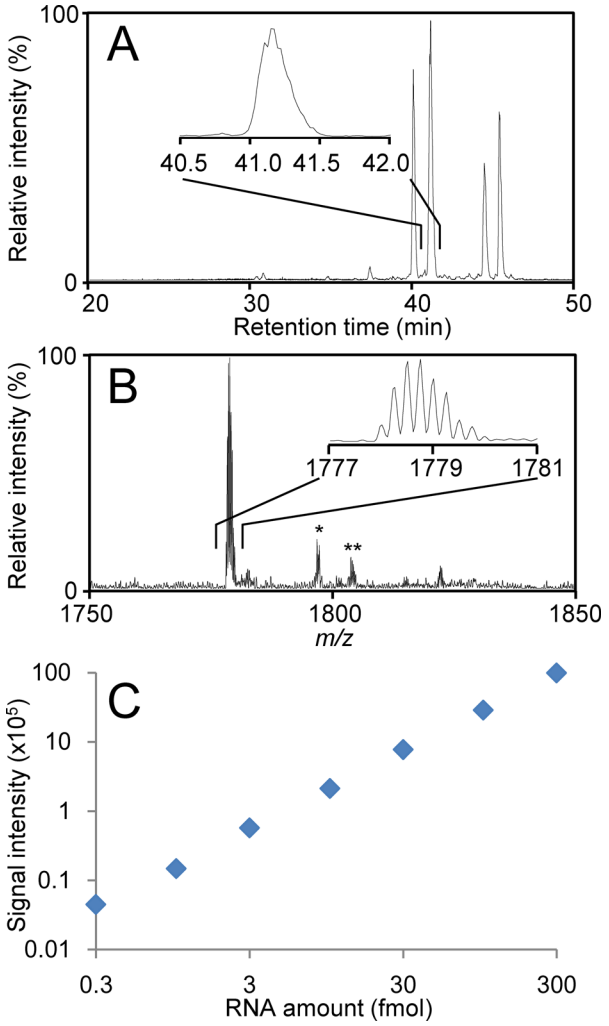


Figure 2. nLC−MS/MS separation of a mixture of the test miRNAs. (A) Base peak mass chromatogram of the mixture of the four test miRNAs. The sample amount was 100 fmol for each miRNA. The full peak widths at half-maximum height are 10, 14, 12, and 11 s for mmu-miR-92a, ath-miR-173, mmu-miR-295, and mmu-miR-335, respectively. The inset shows an expanded region around the ath-miR-173 peak. (B) The mass spectrum of the quadruply charged anion of ath-miR-173 (m/z 1778.000; sample amount, 300 amol). No sodium and potassium adduct ions are seen in the spectrum. The mass differences between the deprotonated anion and the peaks that are indicated by * and ** correspond to the molecular weights of diethylamine and triethylamine, respectively. The inset shows the attained isotopic resolution. (C) Standard curve for ath-miR-173 quantification. Double logarithmic plot of the injected amount (0.3, 1.0, 3.0, 10, 30, 100, and 300 fmol) vs the height of the monoisotopic-ion peak of the quadruply charged anion of ath-miR-173. The observed values of this measurement are given in Table S1, Supporting Information. The square of the Pearson product moment correlation coefficient (R^2) is 0.998.

of triethylamine. Notably, virtually no sodium or other metal adducts were detected, which if present, would have decreased the spectral sensitivity. The relationship between the injected amounts of ath-miR-173 and the intensities of its MS signal was linear within the range of 300 amol to 300 fmol (R^2 = 0.998; Figure 2C). Similar detection limits and linearities were seen for the other three miRNAs; namely, the R^2 values for mmu-miR-92a, mmu-miR-295, and mmu-miR335 between 300 amol and 300 fmol were 0.986, 0.993, and 0.997, respectively. Thus, our system is able to detect miRNAs semiquantitatively at subfemtomole to subpicomole levels, which is a three-order dynamic-signal range.

High-Resolution MS/MS for *de novo* miRNA Sequencing. The MS/MS spectrum obtained using our spray-assisting device coupled with the nLC−MS/MS system allowed us to completely sequence 100 fmol of each of the four miRNAs. The miRNAs were individually fragmented in the higher-energy collisional dissociation cell, and the resulting product ions were analyzed with the high-resolution Orbitrap spectrometer of the Q-Exactive instrument. Figure 3A shows the tandem mass

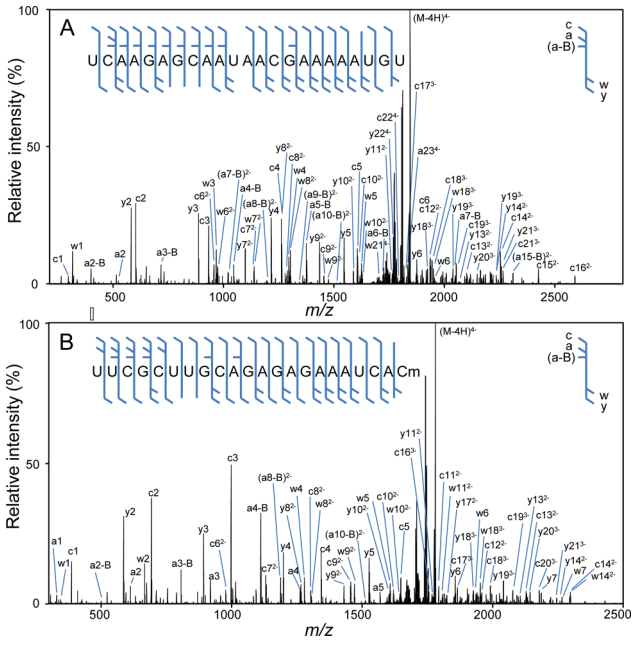


Figure 3. *de novo* sequencing of test miRNAs using high-resolution tandem mass spectra. Nearly complete sequences of (A) mmu-miR-335 and (B) ath-miR-173 were validated manually by assignment of 5′-nucleotide-containing a, a-B, and c ions and 3′-nucleotide-containing w and y ions. The w1 ion in the spectrum shown in panel B provides direct evidence for a 3′-terminus methylcytidine.

spectrum of the quadruply deprotonated anion of mmu-miR-335 at m/z 1850.519. The high-resolution spectrum facilitated to correctly determine monoisotopic peaks and their charge states of product ions that were generated by fragmentation of the multiply charged precursor. Manual inspection of the mmu-miR-335 tandem mass spectrum reveals that the 5′-nucleotide-containing a, a-B, and c ions and the 3′-nucleotide-containing w and y ions almost completely cover its sequence (Figure 3A). Similarly, manual sequencing of the miR-92a and miR-295 MS/MS spectra using the a, a-B, c, w, and y ions almost completely covers their sequences (data not shown).

In addition to its use as a sequencing tool, our nLC−MS/MS system allows for the identification of a modified nucleotide and its position within an miRNA. Figure 3B shows the tandem mass spectrum of the quadruply deprotonated ath-miR173 anion (starting amount, 100 fmol).

The peaks in that spectrum was assigned to a, a-B, c, w, and y ions and allowed for the identification of a methyl group attached to the 3′-terminus cytidine. Although the precise site of methylation within the cytidine molecule, i.e., at the 2′-*O*, 3′-*O*, 3-*N*, or 5-*C* position, was not determined by the data-dependent MS/MS measurement used in this study, it will be

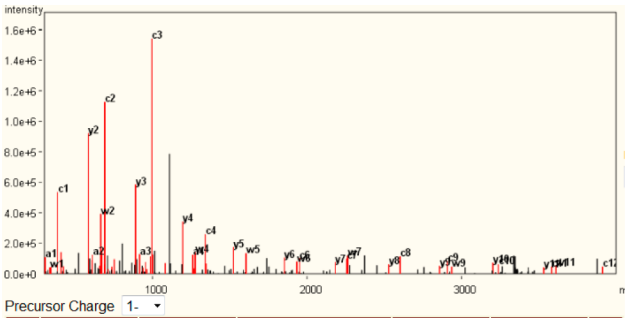
DOI: 10.1021/ac504378s
Anal. Chem. 2015, 87, 2884−2891

determined by an additional experiment such as in-source fragmentation and subsequent MS/MS or MS3.[20]

Automated Identification of miRNAs by Database Searching Using Tandem Mass Spectral Data. As an alternative approach to manually sequencing miRNAs using their tandem mass spectra, we examined fully automated identification of miRNAs by a database search. The same data of mmu-miR-335 as shown in Figure 3A was automatically processed by the SpiceCmd software. The program correctly picked the mass and charge state of the precursor quadruply deprotonated anion at m/z 1850.519 and also correctly detected relatively intense singly, doubly, and triply deprotonated product ions. However, the program failed to identify the relatively weak doubly and triply deprotonated anions that had larger m/z values than did the precursor and other relatively large quadruply deprotonated product ions that were manually assigned as the loss of one, two, or three base(s) from the molecular anion, probably because the corresponding peaks were largely overlapped (Table S2, Supporting Information). These peak lists, which included more than 30 incorrect assignments among 165 total peaks, were used for an Ariadne search of the *Mus musculus* miRNA database. Interestingly, Ariadne assigned most correctly deconvoluted product ions as a, c, w, and y (Table S2, Supporting Information) and identified the sample RNA as mmu-miR-335 using the sequence information contained in the database, even though the peak list contained ∼20% incorrect mass information. In principle, Ariadne selects the product ion peaks from a MS/MS spectrum within a subdivided 300-m/z zone and evaluates them in the descending order of their intensities. Thus, the erroneous detection of most of the base losses from the molecular ion would have an affect only on the assignment of product ions within a single zone and had little effect on the final result of RNA identification. Ariadne also correctly identified mmu-miR-92a and mmu-miR-295 by searching the *M. musculus* miRNA database using their tandem mass spectral data.

We then assessed the minimum amount of an miRNA needed for its identification when searching an miRNA database with Ariadne. Only 10 fmol of mmu-miR-335 was needed for a successful Ariadne search. We next characterized the peaks present in the tandem mass spectra of 10, 30, and 100 fmol samples of mmu-miR-335 (Figure S1 and Tables S3−S6, Supporting Information). The ions generated containing sequences within 5 to 7 residues of the 5′- or 3′-termini usually were found; however, ions from the middle of the mmu-miR-335 sequence diminished with decreasing quantities of starting material. Therefore, *de novo* sequencing when using the spectra of the 10 and 30 fmol samples provided two separate sequences, one from each terminus, which did not provide an unambiguous identification on a BLAST search. Conversely, a database search can employ effectively a precursor mass value to narrow the possibility of partial sequences in the middle of the miRNA molecule, which could not be obtained by the assignments of product ions.

We also assessed if Ariadne could detect a modified nucleotide in an miRNA. To do so, we used the tandem mass spectral data of ath-miR-173 (Figure 3B) with the Ariadne search parameter "max number of modifications" set to "1". This setting directs the program to associate the query mass with a calculated mass of an miRNA that has 0 or 1 modification. Ariadne correctly identified the test compound as ath-miR-173, which contains a methylated cytidine at the 3′-terminus (Figure 4). Thus, the results presented herein show



#	a	c	seq.	w	y	#
5'			P			5'
1	305.017	384.984	U		7034.990	22
2	611.043	691.009	U	6808.931	6728.965	21
3	916.084	996.050	C	6502.906	6422.939	20
4	1261.132	1341.098	G	6197.864	6117.898	19
5	1566.173	1646.139	C	5852.817	5772.851	18
6	1872.198	1952.164	U	5547.776	5467.809	17
7	2178.223	2258.190	U	5241.750	5161.784	16
8	2523.271	2603.237	G	4935.725	4855.759	15
9	2828.312	2908.278	C	4590.678	4510.711	14
10	3157.365	3237.331	A	4285.636	4205.670	13
11	3502.412	3582.378	G	3956.584	3876.617	12
12	3831.465	3911.431	A	3611.536	3531.570	11
13	4176.512	4256.478	G	3282.484	3202.517	10
14	4505.565	4585.531	A	2937.436	2857.470	9
15	4850.612	4930.578	G	2608.384	2528.417	8
16	5179.665	5259.631	A	2263.336	2183.370	7
17	5508.717	5588.683	A	1934.284	1854.318	6
18	5837.770	5917.736	A	1605.231	1525.265	5
19	6143.795	6223.761	U	1276.179	1196.212	4
20	6448.836	6528.802	C	970.153	890.187	3
21	6777.889	6857.855	A	665.112	585.146	2
22	7096.946	7176.912	C+methylation	336.060	256.093	1
3'			OH			3'

Figure 4. Identification of ath-miR-173 and its methylation site by database searching with Ariadne. The same data as shown in Figure 3B was used for automated database searching by Ariadne. The raw data file was initially processed by SpiceCmd to extract monoisotopic peaks. The resulting peak list was searched against the *A. thaliana* miRNA database with the Ariadne search parameter "max number of modifications" set to "1". Full sequence coverage of ath-miR-173 was obtained using the a, c, w, and y ions, and the methylation site was determined to be at the 3′-terminus cytidine.

that Ariadne is useful for the automated identification of an miRNA and two types of its PTMs (see the Discussion) when given high-resolution tandem mass spectral data and an miRNA database.

Identification of Human Cellular miRNAs. To examine the utility of our nLC−MS/MS-Ariadne system in the identification of miRNAs from a biological source, we prepared a small-RNA fraction from HeLa cells (∼8×10^7 cells), which had been cultivated in eight 10 cm diameter dishes. The resulting preparation (100 μg RNA) was separated by reverse-phase LC as described in the Experimental Section. We collected fractions 25−28 that eluted before the large fractions containing tRNAs for further study (Figure 5A). Half of each fraction was subjected to nLC−MS/MS, with the base peak mass chromatograms showing many possible RNA ions (m/z 1500−1990, z = −4 and −5) eluting between 35 and 43 min (Figure 5B).

Manual inspection of the tandem mass spectra identified a number of cellular miRNAs. For example, the tandem mass spectrum of the quadruply deprotonated ion at m/z 1769.219 (Figure 5B) has the sequence 5′-PO_4-UAGCUUAUCAGA-CUGAUGUU[AG or GA]-OH-3′ (Figure 5C). By performing a BLAST search against the nr database, this sequence was identified as human miR-21. Similarly, manual *de novo*

DOI: 10.1021/ac504378s
Anal. Chem. 2015, 87, 2884−2891

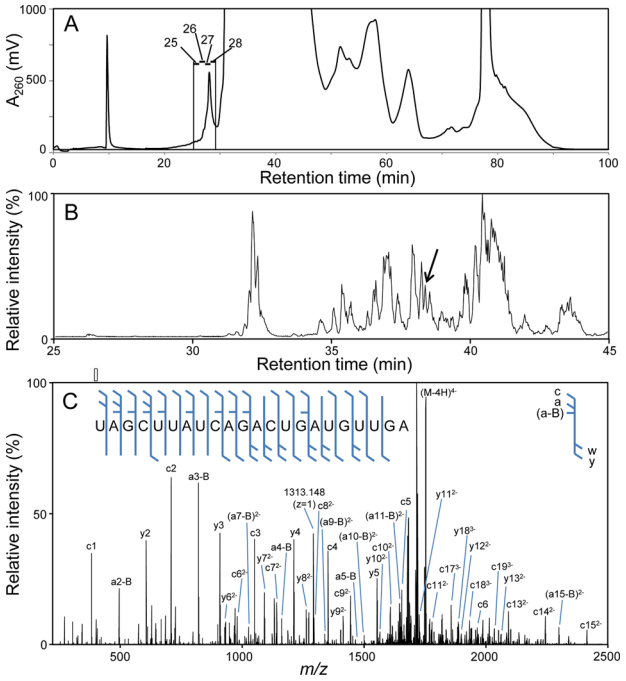


Figure 5. Direct identification of a human cellular miRNA. (A) Separation of the small-RNA fraction from HeLa cells by reverse-phase LC through a polystyrene/divinylbenzene column as described in the Experimental Section. (B) Base peak mass chromatogram of fraction 27. The three most intense peaks were subjected to higher-energy collisional dissociation MS/MS. (C) The high-resolution tandem mass spectrum of the quadruply deprotonated anion at m/z 1769.219 (the peak indicated by an arrow in panel B). The complete sequence was determined manually using the 5′-nucleotide-containing a, a-B, and c ions and 3′-nucleotide-containing w and y ions. Assignments are shown in the figure, and the sequence was identified as that of miR-21 by a BLAST search against the nr database.

sequencing of the nLC−MS/MS data with the use of the a, a-B, c, w, and y ions, identified let-7a, the canonical miR-23a, and its 3′ singly adenylated form. In addition, mass peaks at m/z 1816.01 and 1515.94 (z = −5 and −4, respectively) were identified as fragments of tRNA-Pro and 5S rRNA (data not shown), which seem to have contaminated the sample. Other MS/MS data, however, provided only partial sequences and thus could not be used to confidentially identify RNA sequence(s) by manual *de novo* sequencing and BLAST searching of a database.

We also evaluated the ability of Ariadne to identify miRNA found in the MS/MS data. Because previous deep sequencing studies determined the variable termini of many miRNAs[13] as did our manual *de novo* sequencing of miR-23a described above, we added these 5′- and/or 3′-processing variants of the human miRNAs into the human mature miRNA sequence database and used this modified database (55 176 entries) for an Ariadne search. Finally, the analysis successfully identified 19 miRNAs and their terminal variants, including the four miRNAs that were found by *de novo* sequencing and most of the miRNAs that could only be partially sequenced (Table 1). Interestingly, we found five 3′-terminus variants including those for miR-21 and miR-23a, whereas no 5′-terminus variants were detected. All identifications made by Ariadne were confirmed by visual examination of the spectra. Notably, those sequences that were identified as being tRNA and rRNA fragments by manual sequencing were not assigned by Ariadne to an miRNA in the database, indicating the reliability of Ariadne-based automated identification.

DISCUSSION

We present herein the utility of an automated MS-based system for the characterization of miRNAs. The complete system consists of an nLC system coupled with a high-resolution

Table 1. Identified miRNAs from HeLa Cells[a]

miRNA	sequence	m/z	MW (obs)	MW (calc)	Δm (ppm)	score	intensity
hsa-let-7a-5p	UGAGGUAGUAGGUUGUAUAGUU	1793.465	7177.890	7177.880	1.4	41.3	8.04×10^5
hsa-let-7b-5p	UGAGGUAGUAGGUUGUGUGGUU	1801.457	7209.860	7209.870	−1.4	23.3	1.40×10^5
hsa-let-7d-5p	AGAGGUAGUAGGUUGCAUAGUU	1798.972	7199.921	7199.924	−0.4	21.7	2.36×10^5
hsa-let-7f-5p	UGAGGUAGUAGAUUGUAUAGUU	1789.463	7161.882	7161.885	−0.4	29.9	2.20×10^5
hsa-let-7i-5p	UGAGGUAGUAGUUUGUGCUGUU	1781.703	7130.842	7130.842	0.0	21.0	3.07×10^5
hsa-miR-15b-5p	UAGCAGCACAUCAUGGUUUACA	1764.726	7062.934	7062.926	1.1	34.9	8.72×10^5
hsa-miR-16-5p	UAGCAGCACGUAAAUAUUGGCG	1784.480	7141.950	7141.955	−0.7	27.5	6.74×10^5
hsa-miR-21-5p	UAGCUUAUCAGACUGAUGUUGA	1769.219	7080.909	7080.889	2.8	40.0	1.63×10^6
hsa-miR-21-5p +3p:C	UAGCUUAUCAGACUGAUGUUGAC	1845.474	7385.927	7385.930	−0.4	44.1	1.71×10^6
hsa-miR-23a-3p	AUCACAUUGCCAGGGAUUUCC	1676.454	6709.848	6709.862	−2.1	44.6	4.34×10^5
hsa-miR-23a-3p +3p:A	AUCACAUUGCCAGGGAUUUCCA	1758.721	7038.914	7038.915	−0.1	34.2	1.50×10^6
hsa-miR-23a-3p -3p:C	AUCACAUUGCCAGGGAUUUC	1600.194	6404.809	6404.821	−1.9	44.6	2.54×10^5
hsa-miR-24-3p	UGGCUCAGUUCAGCAGGAACAG	1788.232	7156.960	7156.965	−0.7	29.6	1.32×10^6
hsa-miR-25-3p	CAUUGCACUUGUCUCGGUCUGA	1751.199	7008.827	7008.855	−4.0	12.1	7.46×10^5
hsa-miR-27a-3p -3p:C	UUCACAGUGGCUAAGUUCCG	1604.197	6420.821	6420.816	0.8	15.0	1.29×10^5
hsa-miR-27b-3p	UUCACAGUGGCUAAGUUCUGC	1680.698	6726.825	6726.841	−2.4	25.5	2.17×10^5
hsa-miR-30b-5p	UGUAAACAUCCUACACUCAGCU	1744.717	6982.900	6982.914	−2.0	18.0	3.96×10^5
hsa-miR-30c-5p +3p:U	UGUAAACAUCCUACACUCUCAGCU	1897.488	7593.982	7593.981	0.1	19.0	4.18×10^5
hsa-miR-98-5p	UGAGGUAGUAAGUUGUAUUGUU	1783.704	7138.848	7138.858	−1.4	25.5	1.10×10^5
hsa-miR-106b-5p	UAAAGUGCUGACAGUGCAGAU	1708.224	6836.928	6836.913	2.2	34.1	4.21×10^5
hsa-miR-125b-5p	UCCCUGAGACCCUAACUUGUGA	1752.720	7014.913	7014.904	1.3	25.9	5.92×10^5
hsa-miR-365a-3p	UAAUGCCCCUAAAAAUCCUUAU	1740.971	6967.915	6967.903	1.7	62.7	9.51×10^5

[a]Each miRNA was identified as a phosphorylated 5′-terminus and hydroxylated 3′-terminus based on the MS/MS spectrum of quadruply deprotonated precursor anion. The underlined miRNAs were also able to be identified by manual *de novo* sequencing.

DOI: 10.1021/ac504378s
Anal. Chem. 2015, 87, 2884−2891

tandem mass spectrometer and data-processing software used in conjunction with an miRNA database. In addition, a spray-assisting device (Figure 1) helps to spray a femtomole quantity of the miRNA eluted from the nLC system into the mass spectrometer. Thus, the system allowed direct MS/MS-based sequencing of RNAs of 20 nucleotides and longer (Figure 3) and fully automated identification of miRNAs of ∼10 fmol quantities by searching the miRNA DB. Furthermore, the data-processing software, Ariadne, can identify a modified nucleotide in the miRNA of interest, as demonstrated using methylated ath-miR-173 as the test compound (Figure 4). Currently, Ariadne can report two different PTMs in RNA molecule: methylated four canonical nucleotides and dihydrouridine. This condition is sufficient for the bottom-up identification of tRNAs, one of the most heavily modified RNAs;[25] however, the program could be adapted to more complex, multiple modifications by further refinement in the algorithm.

We found that our system has a sensitivity ∼100 times greater than do conventional infusion- or microLC−MS-based analytical systems that require greater than picomole amounts of RNA for complete sequencing.[15−21] We increased the sensitivity by incorporating an nLC system and a state-of-art high-resolution mass spectrometer. In addition, Ariadne-based automated searching of a database using the MS/MS data increased the sensitivity by ∼10-fold compared with manual *de novo* sequencing of a tandem mass spectrum. We believe that the increase in sensitivity arises as a result of the ability of Ariadne to compare the MS/MS data with the mass values of the full-length and fragments calculated from RNA sequences in the database, rather than the nucleotide sequence determined by *de novo* sequencing.

To our knowledge, this is the first report that describes direct MS/MS-based identification of cellular miRNAs. We were able to characterize dozens of HeLa cell miRNAs and their 3′-variants (Table 1). The miRNAs identified in this study include let-7, miR-16, and miR-21. According to a quantitative Northern blot study,[32] these miRNAs are present at a concentration of 1.3×10^3, 1.2×10^3, and 1.2×10^4 molecules per cell, respectively, suggesting that our system can identify an miRNA with a cellular abundance of $>10^3$ molecules per cell, and therefore, in principle, a single copy of an miRNA could be identified by processing $\sim 5 \times 10^{10}$ HeLa cells (∼5000, 10 cm diameter culture dishes). However, the limit of detection of our system is ∼300 amol, or 1.8×10^8 molecules (300×10^{-18} [mol] × 6.0×10^{23} [molecule/mol]) with the limit of identification being 10 fmol, or 6.0×10^9 molecules. Therefore, if we start with 4×10^7 HeLa cells, as was done in this study, we should be able to detect miRNA at a concentration of ∼5 molecules per cell and to unambiguously identify an miRNA if it is present at 1.5×10^2 molecules per cell. These calculations emphasize the great potential of MS-based RNA analysis; however, we actually identified only some of the miRNAs present in the HeLa cell, suggesting that further sample preparation and MS measurement improvements are required that would reduce interference from complex bio-oriented materials.

In conclusion, our system provides an improved tool for miRNA research that complements conventional genetic approaches. In particular, the identification protocol presented herein should be useful for miRNA analysis, including characterization of PTMs of miRNAs, and metabolic pathways involving miRNAs and other small RNAs, such as heavily modified pharmaceutical RNAs. Although our system is still limited in the types of studies applicable to RNAs present in low abundance, we expect that rapid advances in this and related fields will soon improve current MS-based technologies.

■ ASSOCIATED CONTENT

Supporting Information

Detailed assignments of product ions of mmu-miR-335. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*Tel: +81 48 462 1419. Fax: +81 48 462 4704. E-mail: knife@riken.jp (H.N.).

*Tel: +81 426 77 5667. Fax: +81 426 77 2525. E-mail: isobe-toshiaki@tmu.ac.jp (T.I.).

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. N. Takahashi of Tokyo University of Agriculture and Technology for valuable discussions, Mr. Y. Yokoi of Mitsui Knowledge Industry Co, Ltd. for advice concerning the appropriate parameter settings for SpiceCmd, and Ms. T. Shiina of RIKEN for assisting with data analysis. The work was supported by a grant from the Japan Science and Technology Agency for Core Research for Evolutional Science and Technology [ID 13415564].

■ REFERENCES

(1) Griffiths-Jones, S.; Grocock, R. J.; van Dongen, S.; Bateman, A.; Enright, A. J. *Nucleic Acids Res.* **2006**, *34*, D140−D144.

(2) Friedman, R. C.; Farh, K. K.-H.; Burge, C. B.; Bartel, D. P. *Genome Res.* **2009**, *19*, 92−105.

(3) Iorio, M. V.; Croce, C. M. *EMBO Mol. Med.* **2012**, *4*, 143−159.

(4) Chen, H.; Lan, H.; Roukos, D. H.; Cho, W. C. *J. Endocrinol.* **2014**, *222*, R1−R10.

(5) Thum, T. *EMBO Mol. Med.* **2012**, *4*, 3−14.

(6) Winter, J.; Jung, S.; Keller, S.; Gregory, R. I.; Diederichs, S. *Nat. Cell Biol.* **2009**, *11*, 228−234.

(7) Yu, B.; Yang, Z.; Li, J.; Minakhina, S.; Yang, M.; Padgett, R. W.; Steward, R.; Chen, X. *Science* **2005**, *307*, 932−935.

(8) Horwich, M. D.; Li, C.; Matranga, C.; Vagin, V.; Farley, G.; Wang, P.; Zamore, P. D. *Curr. Biol.* **2007**, *17*, 1265−1272.

(9) Katoh, T.; Sakaguchi, Y.; Miyauchi, K.; Suzuki, T.; Kashiwabara, S.-I.; Baba, T.; Suzuki, T. *Genes Dev.* **2009**, *23*, 433−438.

(10) Ji, L.; Chen, X. *Cell Res.* **2012**, *22*, 624−636.

(11) El Yacoubi, B.; Bailly, M.; de Crécy-Lagard, V. *Annu. Rev. Genet.* **2012**, *46*, 69−95.

(12) Decatur, W. A.; Fournier, M. J. *Trends Biochem. Sci.* **2002**, *27*, 344−351.

(13) Morin, R. D.; O'Connor, M. D.; Griffith, M.; Kuchenbauer, F.; Delaney, A.; Prabhu, A.-L.; Zhao, Y.; McDonald, H.; Zeng, T.; Hirst, M.; Eaves, C. J.; Marra, M. A. *Genome Res.* **2008**, *18*, 610−621.

(14) Jima, D. D.; Zhang, J.; Jacobs, C.; Richards, K. L.; Dunphy, C. H.; Choi, W. W. L.; Au, W. Y.; Srivastava, G.; Czader, M. B.; Rizzieri, D. A.; Lagoo, A. S.; Lugar, P. L.; Mann, K. P.; Flowers, C. R.; Bernal-Mizrachi, L.; Naresh, K. N.; Evens, A. M.; Gordon, L. I.; Luftig, M.; Friedman, D. R.; Weinberg, J. B.; Thompson, M. A.; Gill, J. I.; Liu, Q.; How, T.; Grubor, V.; Gao, Y.; Patel, A.; Wu, H.; Zhu, J.; Blobe, G. C.;

DOI: 10.1021/ac504378s
Anal. Chem. 2015, 87, 2884−2891

Lipsky, P. E.; Chadburn, A.; Dave, S. S.; Malignancies, H. *Blood* **2010**, *116*, e118−e127.
(15) Huang, T.; Liu, J.; Liang, X.; Hodges, B. D. M.; Mcluckey, S. A. *Anal. Chem.* **2008**, *80*, 8501−8508.
(16) Taucher, M.; Rieder, U.; Breuker, K. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 278−285.
(17) Taucher, M.; Breuker, K. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 918−929.
(18) Taucher, M.; Breuker, K. *Angew. Chem., Int. Ed. Engl.* **2012**, *51*, 11289−11292.
(19) Gardner, M. W.; Li, N.; Ellington, A. D.; Brodbelt, J. S. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 580−591.
(20) Izumi, Y.; Takimura, S.; Yamaguchi, S.; Iida, J.; Bamba, T.; Fukusaki, E. *J. Biosci. Bioeng.* **2012**, *113*, 412−419.
(21) Kullolli, M.; Knouf, E.; Arampatzidou, M.; Tewari, M.; Pitteri, S. J. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 80−87.
(22) Natsume, T.; Yamauchi, Y.; Nakayama, H.; Shinkawa, T.; Yanagida, M.; Takahashi, N.; Isobe, T. *Anal. Chem.* **2002**, *74*, 4725−4733.
(23) Rozenski, J.; McCloskey, J. A. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 200−203.
(24) Oberacher, H.; Mayr, B. M.; Huber, C. G. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 32−42.
(25) Nakayama, H.; Akiyama, M.; Taoka, M.; Yamauchi, Y.; Nobe, Y.; Ishikawa, H.; Takahashi, N.; Isobe, T. *Nucleic Acids Res.* **2009**, *37*, No. e47.
(26) Nakayama, H.; Takahashi, N.; Isobe, T. *Mass Spectrom. Rev.* **2011**, *30*, 1000−1012.
(27) Ishikawa, H.; Nobe, Y.; Izumikawa, K.; Yoshikawa, H.; Miyazawa, N.; Terukina, G.; Kurokawa, N.; Taoka, M.; Yamauchi, Y.; Nakayama, H.; Isobe, T.; Takahashi, N. *Nucleic Acids Res.* **2014**, *42*, 2708−2724.
(28) Taoka, M.; Yamauchi, Y.; Nobe, Y.; Masaki, S.; Nakayama, H.; Ishikawa, H.; Takahashi, N.; Isobe, T. *Nucleic Acids Res.* **2009**, *37*, No. e140.
(29) Yamauchi, Y.; Taoka, M.; Nobe, Y.; Izumikawa, K.; Takahashi, N.; Nakayama, H.; Isobe, T. *J. Chromatogr. A* **2013**, *1312*, 87−92.
(30) Taoka, M.; Ikumi, M.; Nakayama, H.; Masaki, S.; Matsuda, R.; Nobe, Y.; Yamauchi, Y.; Takeda, J.; Takahashi, N.; Isobe, T. *Anal. Chem.* **2010**, *82*, 7795−7803.
(31) Apffel, A.; Chakel, J. A.; Fischer, S.; Lichtenwalter, K.; Hancock, W. S. *Anal. Chem.* **1997**, *69*, 1320−1325.
(32) Lim, L. P.; Lau, N. C.; Weinstein, E. G.; Abdelhakim, A.; Yekta, S.; Rhoades, M. W.; Burge, C. B.; Bartel, D. P. *Genes Dev.* **2003**, *17*, 991−1008.

DOI: 10.1021/ac504378s
Anal. Chem. 2015, 87, 2884−2891