Published online 15 December 2015

Nucleic Acids Research, 2016, Vol. 44, No. 6 ***e59***
doi: 10.1093/nar/gkv1462

A mass spectrometry-based method for direct determination of pseudouridine in RNA

Yoshio Yamauchi[1,2], Yuko Nobe[1,2], Keiichi Izumikawa[2,3], Daisuke Higo[4], Yoko Yamagishi[4], Nobuhiro Takahashi[2,3], Hiroshi Nakayama[2,5], Toshiaki Isobe[1,2,*] and Masato Taoka[1,2,*]

[1]Department of Chemistry, Graduate School of Science and Engineering, Tokyo Metropolitan University, Minami-osawa 1-1, Hachioji-shi, Tokyo 192-0397, Japan, [2]Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency (JST), Sanbancho 5, Chiyoda-ku, Tokyo 102-0075, Japan, [3]Department of Biotechnology, United Graduate School of Agriculture, Tokyo University of Agriculture and Technology, Saiwai-cho 3-5-8, Fuchu-shi, Tokyo 183-8509, Japan, [4]Thermo Fisher Scientific, 3-9 Moriya-cho, Kanagawa-ku, Yokohama-shi, Kanagawa 221-0022, Japan and [5]Biomolecular Characterization Unit, RIKEN Center for Sustainable Resource Science, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

Received October 7, 2015; Revised November 17, 2015; Accepted November 27, 2015

ABSTRACT

Pseudouridine (5-ribosyluracil, Ψ) is the only 'mass-silent' nucleoside produced by post-transcriptional RNA modification. We describe here a novel mass spectrometry (MS)-based method for direct determination of Ψ in RNA. The method assigns a Ψ-containing nucleolytic RNA fragment by an accurate measurement of a signature doubly dehydrated nucleoside anion ($[C_9H_7N_2O_4]^{1-}$, *m/z* 207.04) produced by collision-induced dissociation MS, and it determines the Ψ-containing nucleotide sequence by pseudo-MS^3, i.e. in-source fragmentation followed by MS^2. By applying this method, we identified all of the known Ψs in the canonical human spliceosomal snRNAs and, unexpectedly, found two previously unknown Ψs in the U5 and U6 snRNAs. Because the method allows direct determination of Ψ in a subpicomole quantity of RNA, it will serve as a useful tool for the structure/function studies of a wide variety of non-coding RNAs.

INTRODUCTION

Pseudouridine (Ψ) was among the first RNA post-transcriptional modification (PTM) to be described (1). This modification is catalyzed by pseudouridine synthases (2), and the resulting uridine isomer, Ψ, is highly conserved among RNAs in all domains of life (3) and is the most frequently found modified nucleoside in cellular RNAs, including rRNA, tRNA, mRNA, small nuclear RNA (snRNA) and small nucleolar RNA (4–8). Pseudouridylation reduces the conformational flexibility and/or induces structural changes in cellular RNA (9,10) that thereby affects RNA function. Ψ is dynamically regulated by environmental cues, providing a potential mechanism for the regulation of RNA fate (5,6,11). In addition, pseudouridylation appears to mediate nonsense-to-sense codon conversion in mRNA and thus may introduce post-transcriptional genetic recoding and diversify the proteome (12).

Ψ can be distinguished from uridine by the method introduced by Bakin and Ofengand in 1993 (13), which utilizes primer extension by reverse transcription of RNA with gene-specific primers after Ψ-specific chemical labeling to interrupt the extension reaction. This method, coupled with a next-generation sequencing technique, recently demonstrated several hundred Ψ sites in non-coding RNAs and mRNAs (5–8). This approach was innovative with respect to its high-sensitivity and -throughput, and it has provided new insight into the biological importance of pseudouridylation of mRNA. However, the method has several principal limitations, such as that the surrounding modified nucleotides frequently interfere with the Ψ identification and that the 3′ end of RNA cannot provide the primer binding sites for reverse transcriptase (13). Indeed, previous applications of this method failed to detect, for example, some Ψs in U5 snRNA and all Ψs in U2 snRNA (5).

Direct chemical analysis of RNA PTMs, on the other hand, has been performed extensively by mass spectrometry (MS) (14). Because pseudouridylation is a mass-silent modification, however, conventional MS-based Ψ analysis relies on RNase mapping after Ψ-directed chemical labeling of RNA, such as by cyanoethylation with acrylonitrile (15–18), or on the metabolic pre-labeling of RNA with 5,6-D_2-uracil (19) so that when the deuterium is exchanged with solvent hydrogen, the resulting mass shift dis-

*To whom correspondence should be addressed. Tel: +81 426 77 2543; Fax: +81 426 77 2525; Email: mango@tmu.ac.jp
Correspondence may also be addressed to Toshiaki Isobe. Tel: +81 426 77 5667; Fax: +81 426 77 2525; Email: isobe-toshiaki@tmu.ac.jp

tinguishes 5,6-D_2-uridine from 6-D-Ψ. However, the results of the chemical labeling are frequently complex, especially for RNA fragments containing multiple U/Ψs, with the non-specific reaction of acrylonitrile with uridine, and the metabolic RNA labeling has limited applicability to mutant cells deficient in the pathway for *de novo* synthesis of uridine monophosphate from aspartic acid. As an alternative to the approaches described above, Pomerantz and McCloskey reported a direct MS-based method to determine Ψ in RNA that utilizes a Ψ-characteristic fragmentation pattern produced by collision-induced dissociation (CID)-MS (20). Namely, they found that CID of Ψ-containing RNA produce a doubly dehydrated nucleoside anion at m/z 207 due to a Ψ-characteristic stable glycosidic C-C bond and its product ion produced by retro-Diels-Alder reaction at m/z 164, and by monitoring this fragmentation pattern, they identified a number of Ψs in archaeal and bacterial rRNAs (21,22).

We present here a method for direct MS-based sequencing of Ψ-containing RNA. In this study, we show that the m/z 207 ion detected by Pomeratz and McCloskey is the mixture of a Ψ-characteristic m/z 207.04 ion and a 207.01 ion produced from 2′-O-methylated nucleoside–containing RNA and that a current state-of-art mass spectrometer can discriminate those ions. Thus, the method described here depends on the accurate measurement of the CID-derived m/z 207.04 ion as a Ψ-specific signature anion and determines the Ψ-containing RNA sequence directly by pseudo-MS^3 analysis, i.e. in-source fragmentation followed by MS/MS (MS^2) analysis, for identification and sequencing of the Ψ-containing RNA. We show that this method allows determination of all Ψs in the canonical human spliceosomal snRNAs, including two previously unknown Ψs in the U5 and U6 snRNAs.

MATERIALS AND METHODS

Reagents

Standard laboratory chemicals were obtained from Wako Pure Chemical Industries (Osaka, Japan). RNase T1 and A were purchased from Worthington (Lakewood, NJ, USA) and further purified by reversed-phase liquid chromatography (LC) before use. Triethylammonium acetate buffer (TEAA, pH 7.0) was purchased from Glen Research (Sterling, VA, USA). The chemically synthesized RNAs AAUUGp and AAΨUGp and a guide oligonucleotide for RNase H digestion of a U5 snRNA fragment (UAUGCGAUCUGAA(GAGA)AACCAGA, in which the GAGA sequence in parentheses represents deoxyribonucleotides and all others are 2′-O-methyl ribonucleotides), were purchased from JBioS (Saitama, Japan).

Purification and RNase digestion of snRNA

Human U snRNAs were extracted from the nuclear fraction of 293T cells (CRL-3216, ATCC, Manassas, VA, USA) and purified by denaturing ion-paired reversed-phase LC on a macro-porous polystyrene-divinylbenzene resin (23). The purified RNAs (∼10 ng) were digested with 2 ng of RNase T1 or A at 37°C for 60 min. The digests were diluted 5-fold with 100 mM TEAA (pH 7.0) and immediately subjected to nanoflow LC–MS^2 or stored frozen at −20°C until use.

Sequence-specific RNase H cleavage was performed using a guide oligonucleotide to target the RNA site (18). The human U5 snRNA (∼300 fmol) was denatured at 65°C for 10 min and digested with 15 U RNase H (Takara Bio Inc, Shiga, Japan) at 42°C for 1 h in the presence of the guide oligonucleotide (2 pmol) complementary to the cleavage site in 100 μl of 40 mM Tris–HCl (pH 7.7), 4 mM $MgCl_2$ and 1 mM DTT. The resulting fragments were digested further with RNase T1 and subjected directly to LC-MS (described below).

Direct nanoflow LC-MS and MS^2 analysis of RNA fragments

Nucleolytic RNA fragments were analyzed on a direct nanoflow LC-MS system as described (24,25). Briefly, the RNase digests were first concentrated on a trap column (MonoCap C18, 200 μm i.d. × 40 mm; GL Sciences, Tokyo, Japan) connected to a sample injection valve in 100 mM TEAA (pH 7.0) containing 0.1 mM diammonium phosphate. The samples were then injected onto a reversed-phase tip column (Develosil C30 UG, 150 μm i.d. × 60 mm, 3 μm particle size; Nomura Chemical Co. Ltd, Aichi, Japan) equilibrated with solvent A (10 mM TEAA in water:methanol, 9:1). Samples were then eluted at 100 nl/min with a 35-min linear gradient to 35% solvent B (10 mM TEAA:acetonitrile, 60:40). The column was subsequently washed with 70% B for 10 min and re-equilibrated with buffer A for the next analysis.

The LC eluate was sprayed online at −1.3 kV with the aid of a spray-assisting device (25) to a Q Exactive mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) in negative ion mode. The spectrometer was operated in a data-dependent mode to automatically switch between MS and MS^2 acquisition. Survey full-scan mass spectra (from m/z 500 to 2000) were acquired at a mass resolution of 35 000. Up to five of the most intense mass peaks, which exceeded 100 000 counts/s (60 ms maximum injection time), were isolated with a 3 m/z window for fragmentation with data dependent acquisition mode. Precursors were fragmented by higher-energy CID with normalized collision energy of 20%. To retain mass resolution and to increase spectral quality, three micro-scans of MS^2 measurements were accumulated for a single MS^2 spectrum. A fixed starting mass value for MS^2 was set at m/z 100. A mass resolution of 17 500 at m/z 200 for MS^2 was set to detect the Ψ-characteristic signature ion at m/z 207.04.

Determination of Ψ-containing RNA sequence by pseudo-MS^3 and full MS^3

In-source fragmentation followed by MS^2 (pseudo-MS^3) analysis was performed using a Q Exactive mass spectrometer to determine the Ψ-containing RNA sequence. The analysis consisted of four steps: (i) in-source fragmentation was performed by switching the source CID voltage from 0 to 70 or 80 eV (the voltage was set to produce a-, c-, w- and y- series ions in the spectrum) during the passage of a chromatographic peak, (ii) the series ions were detected with a

single micro-scan set at a mass resolution of 35 000 at m/z 200 in the Orbitrap, (iii) the a-, c-, w- or y-series ions in the inclusion mass list exceeding intensity of 4000 counts/s were isolated by the quadrupole with the isolation window of 3 m/z units and (iv) each isolated ion was dissociated with a normalized collision energy of 40% up to 1 000 000 ions/s or the maximum injection time of 250 ms and then detected by the Orbitrap. The inclusion mass list was prepared by calculating m/z values of the series ion from the candidate Ψ-containing RNA sequence. Fixed start masses for the in-source fragmentation and pseudo-MS^3 were set at m/z 300 and 100, respectively.

The full-MS^3 analysis of Ψ-containing RNA fragments was performed on an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific). Survey full-scan mass spectra (from m/z 500 to 1800) were acquired in a negative ion mode at a mass resolution of 60 000 at m/z 400. The ions exceeding the intensity of 200 000 counts/s (maximum injection time 100 ms) were isolated with a 2 m/z window, and the precursor ions were fragmented by CID with normalized collision energy of 25%. To retain mass resolution and to increase spectral quality, two micro-scans of MS^2 measurements, each fragmented up to 50 000 ions/s (maximum injection time 200 ms) with a mass resolution of 15 000 at m/z 200, were accumulated for a single MS^2 spectrum. The fixed starting mass value for MS^2 was set at m/z 600. The full-MS^3 scan mass spectra were obtained by higher-energy CID with normalized collision energy of 40%. Two micro-scans of full-MS^3 measurements, each fragmented up to 200 000 ions/s (maximum injection time 250 ms) with a mass resolution of 15 000 at m/z 200, were accumulated for a single MS^3 spectrum. The fixed starting mass value for full-MS^3 was set at m/z 100.

Cyanoethylation of Ψ-containing RNA with acrylonitrile

RNA samples were digested before the derivatization reaction. The cyanoethylation was performed as described by Mengel-Jorgensen *et al*. (16) with minor modifications (18). Briefly, RNase–digested RNA (<1 pmol in 5 μl) was added with 1 μl acrylonitrile and 40% ethanol/1.2 M TEAA, pH 8.6, (34 μl) and incubated at 70°C for 120 min. After cyanoethylation, the reaction mixture was diluted with 10 volumes of water for LC-MS analysis.

RESULTS

Strategy of the method

The strategy of the method reported here is outlined in Figure 1A. Because conventional MS has difficulty discriminating between the Ψ-diagnostic ion at m/z 207.04 $[C_9H_7N_2O_4]^{1-}$ and the O-methylated dehydrophosphoribose at m/z 207.01 $[C_6H_8O_6P]^{1-}$ produced from 2′-O-mehtylated nucleoside–containing RNA (Figure 1B), previous studies utilized a set of ions to identify a Ψ-containing RNA (the dehydrated nucleoside anion at m/z 207 and its product ion at m/z 164) and estimated the position of Ψ based on the ambiguous trend of a-series ions of the fragment (20). We predicted that the higher resolution and more accurate mass measurement of pseudo-MS^3 could discriminate the ions of m/z 207.04 and 207.01 and simplify sequencing of Ψ-containing RNA. In the procedure shown in Figure 1A, a sample RNA purified from a biological source is digested with RNase T1 and the resulting fragments are analyzed by direct nanoflow LC-MS. From the MS^2 spectra, RNA fragments carrying mass-altering PTMs are identified using the Ariadne software (27), and then Ψ-containing RNA fragments are assigned by the signature ion at m/z 207.04 in the MS^2 spectrum. When RNase T1 digestion of a sample RNA produced multiple fragments having the same U/Ψ–containing sequence that could hardly be located within the final sequence, we performed RNase A digestion or/and systematically segmented the sample RNA with RNase H to avoid the production of such fragments. Subsequently to the 1st LC-MS^2 analysis, we analyzed the data and prepared manually an inclusion mass list of the m/z values of U/Ψ-containing product ion (a-, c-, w- and y-series) estimated from the identified sequence. The second-round in-source fragmentation LC-MS (second run in Figure 1A) detects the Ψ-containing product ions utilizing the inclusion list as a reference and further fragmented by pseudo-MS^3, and the position of Ψ in the RNA sequence is determined by inspection of the resulting m/z 207.04 ion as a signature of the Ψ-containing fragment.

Analysis of a Ψ-containing synthetic RNA

We first evaluated our method by analyzing a mixture of chemically synthesized short RNAs (AAUUGp and AAΨUGp). The resulting extracted ion chromatogram (EIC) and the product ion chromatogram extracted from the MS^2 spectra at m/z 207.04, which we termed here the extracted product ion chromatogram (EPIC), are shown in Figure 2. The RNAs were separated by reversed-phase LC and the MS^2 fragment ion at m/z 207.04 was detected only in the chromatogram from AAΨUGp, demonstrating that the method clearly distinguished the Ψ-containing RNA from the corresponding U-containing RNA. In the MS^2 spectrum, we detected many other product ions with m/z < 230, which were composed of bases, ribose, phosphoric acid and their dehydrated ions (Figure 3). However, only the Ψ-containing RNA generated doubly dehydrated ions of Ψ at m/z 207.04 (Figure 3A and B) and, as expected, this Ψ-characteristic signature ion was clearly discriminated from the O-methylated dehydrophosphoribose ion at m/z 207.01 under the conditions employed (Figure 3B and C). The signal of this signature ion was usually less intense than the sequence-determining a-, c-, w- and y-series ions but was still detectable under the conditions necessary for Ariadne to identify the sequence from the MS^2 spectrum. In the second-round LC-MS, the $[AAΨUGp]^{2-}$ precursor ion at m/z 815.6 of the synthetic RNA was fragmented in-source (Figure 4A) and the resulting c3, c4, y2 and y3 ions were further fragmented to obtain the series of pseudo-MS^3 spectra. The signature ion at m/z 207.04 was found in the MS^3 spectra of c3, c4 and y3 but was not detectable in the MS^3 spectrum of the y2 ion (Figure 4B and C), indicating that Ψ is present at the third position from the 5′ terminus of this nucleotide. We estimated that the detection limit of the Ψ-specific signature ion at m/z 207.04 in the first-round LC-MS analysis was ∼10 fmol (Supplementary Figure S1) and

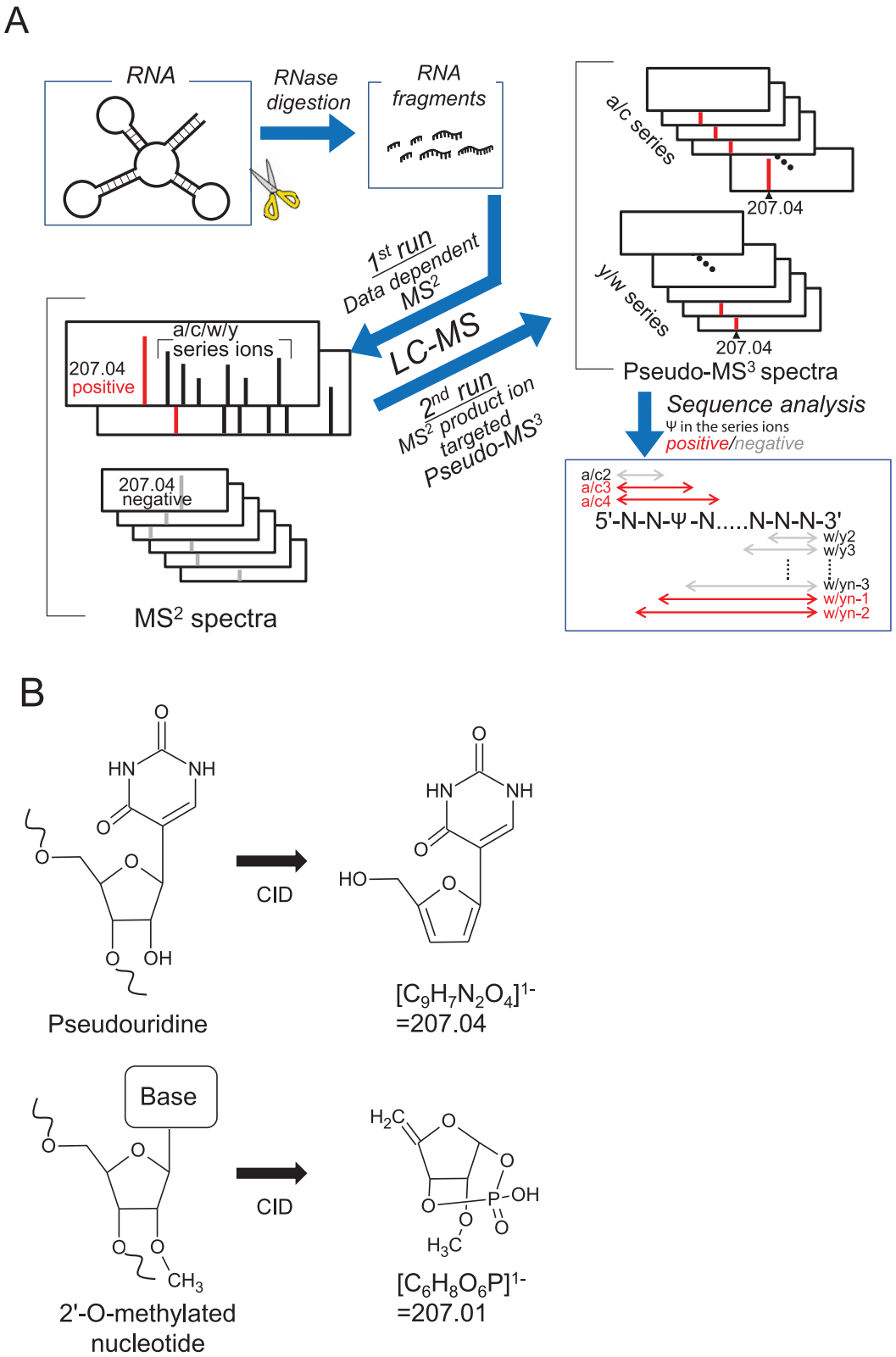

Figure 1. Strategy for MS-based determination of Ψ-containing RNA sequence. (**A**) Step-by-step overview of the method and (**B**) chemical structures of the Ψ-characteristic signature ion $[C_9H_7N_2O_4]^{1-}$ at m/z 207.04 and the 2′-O-methylated dehydrophosphoribose ion $[C_6H_8O_6P]^{1-}$ at m/z 207.01 derived from 2′-O-methylated nucleotides. The structure of the product ions are inferred from the references (20,26). Sites of deprotonation are not known, and all structures are drawn as neutral species. Note that the high-resolution and accurate mass spectrometer used in this study discriminates between these ions produced during the CID-MS analysis of RNA.

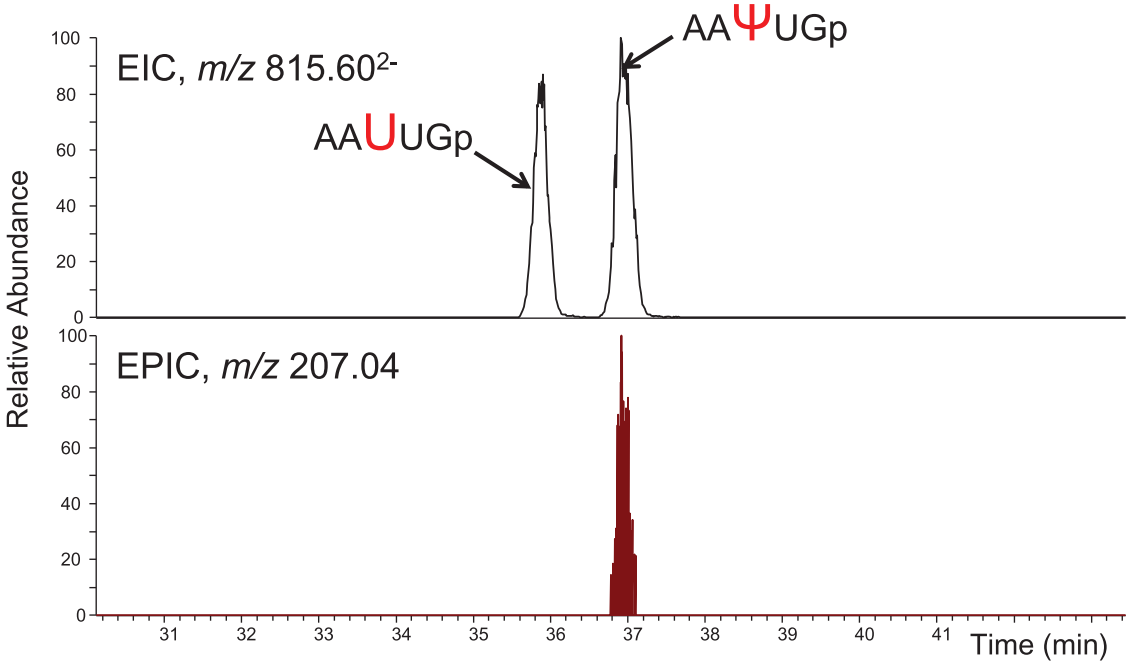

Figure 2. Selective detection of Ψ-containing synthetic RNA by the signature ion at *m/z* 207.04. Upper panel: an MS of EIC for the synthetic RNAs. Lower panel: an MS^2 of EPIC for the Ψ-characteristic signature ion. A mixture of synthetic RNAs (50 fmol each) was analyzed by LC-MS. The sequences of the synthetic RNAs are indicated with the *m/z* values for extraction (mass tolerance, 15 ppm). Note that monitoring the reaction at *m/z* 207.04 showed the presence of Ψs in the synthetic RNA.

that ∼50 fmol of RNA is required for the pseudo-MS^3–based sequencing of Ψ-containing fragments.

Analysis of human canonical spliceosomal RNAs

We next evaluated our method by analyzing the human spliceosomal snRNAs. The spliceosome is a highly complex and dynamic RNA–protein complex that assembles on each pre-mRNA to be spliced (28). The canonical spliceosome is made up of five snRNAs (U1, U2, U4, U5 and U6) and many associated proteins, and it catalyzes ∼99% of the pre-mRNA splicing reactions in human cells (29–32). The chemical structures of these snRNAs have been studied in detail and, according to the literature (33–39), contain a total of 32 methylated nucleotides and 25 Ψs.

We purified these five snRNAs from a nuclear extract of cultivated human 293T cells, digested them with RNase T1 and analyzed the resulting fragments by LC-MS^2. By analyzing the MS^2 spectra with Ariadne, we could assign all fragments covering the total sequence of each snRNA except for free guanosine monophosphate released by RNase T1 cleavage (Supplementary Table S1). All mass values estimated for each fragment and the determined sequences were compatible with the reported chemical structures, including the 18 fragments carrying mass-altering PTMs (base- and/or 2′-O-methylated nucleosides) reported for human U1, U2, U4, U5 and U6 snRNAs (38,39).

To find the mass-silent PTM Ψ, we focused on the signal of *m/z* 207.04 in the MS^2 spectra. According to the literature (38,39), the human snRNAs contain 2 (U1), 14 (U2), 3 (U4), 3 (U5), and 3 (U6) Ψs, and therefore, based on the cleavage specificity, RNase T1 should produce the following number of Ψ-containing fragments from each snRNA: 1 (U1), 6 (U2), 2 (U4), 2 (U5) and 3 (U6) (38,39). In the experimental LC-MS^2 spectrum derived from the RNase T1 digest of each purified snRNA, we actually found 1 (U1), 6 (U2), 2 (U4), 3 (U5) and 4 (U6) Ψ-containing fragments (Figures 5 and 6; see also Supplementary Figures S2–S4 and Supplementary Table S1). This suggested that U1, U2 and U4 snRNAs have the reported number of Ψs, whereas U5 and U6 snRNAs may each have an additional Ψ that has not been reported.

Identification of $^9\Psi$ in human U6 snRNA

To confirm that the U5/U6 snRNAs contain previously unknown Ψs, we first analyzed the U6 snRNA. According to a previous report (37), U6 snRNA is 106 bases long, and Ψs are located at positions 31, 40 and 86. Four RNase T1 fragments contained the Ψ-characteristic signature ion at *m/z* 207.04, and we assigned the positions of the Ψs in three of the sequences as ^{17}CACAUAUACUAAAAΨU^{33}Gp, ^{39}AΨAC(mA)^{44}Gp and ^{81}CAAAUΨC^{88}Gp (Figure 5 and Supplementary Figures S5–S7). The positions of these Ψs in the U6 snRNA were confirmed by detecting the *m/z* 207.04 signature ion in the LC-MS^2 spectrum of the RNase A digests and pseudo-MS^3 spectra of the RNase T1 digests (Supplementary Figures S5 and S7 and Supplementary Table S2). On the other hand, the remaining Ψ-containing fragment appeared to have a tentative sequence ^{8}C(U/Ψ)(U/Ψ)C^{12}Gp. Thus, we performed pseudo-MS^3 analysis and determined the sequence of this fragment as ^{8}CΨUC^{12}Gp, with Ψ located at position 9 from the 5′ terminus (Figure 7). Examination by conventional MS-based

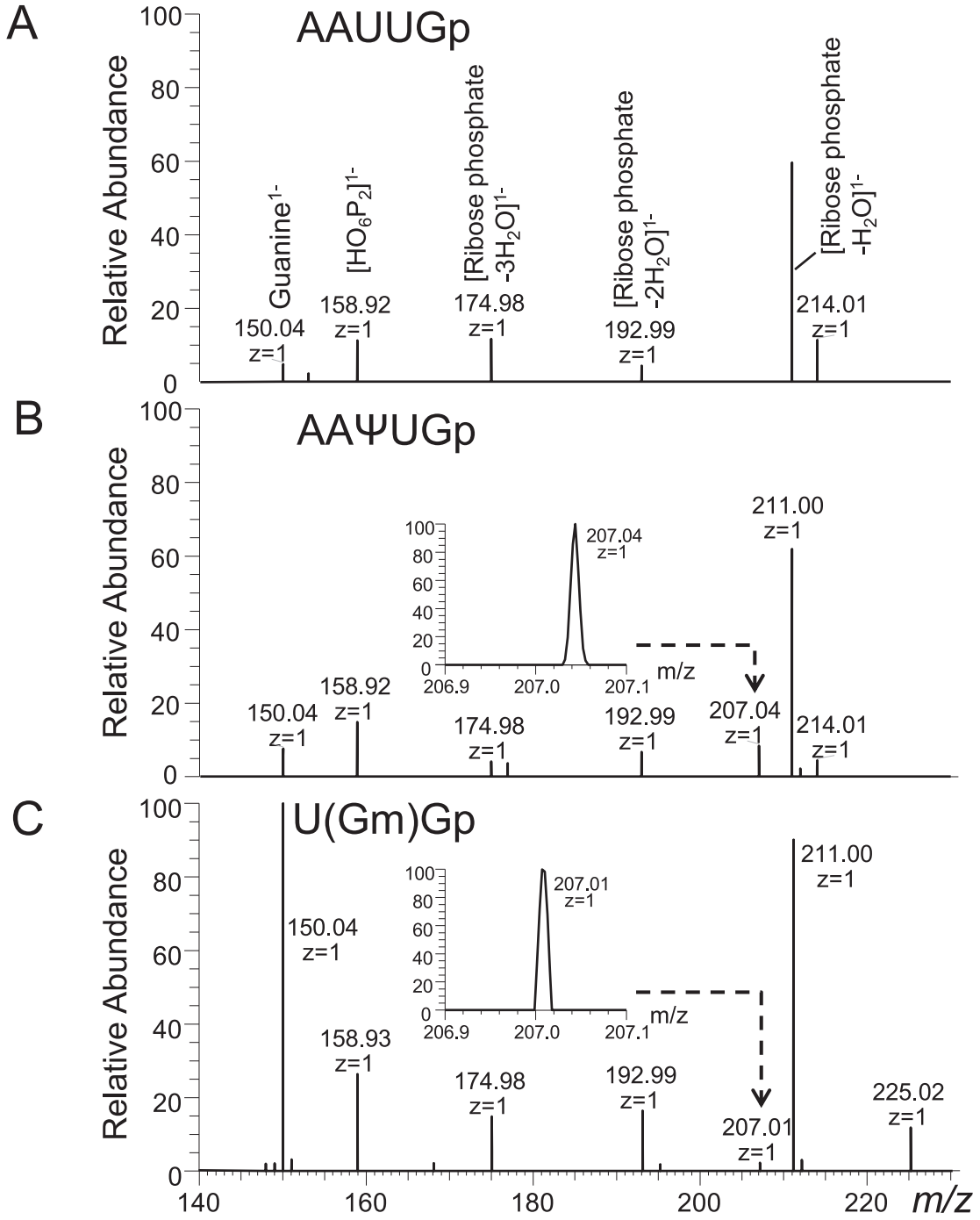

Figure 3. Low m/z region in the MS^2 spectra of (**A**) AAUUGp, (**B**) AAΨUGp and (**C**) U(Gm)Gp with the fragment ions assigned. The Ψ-characteristic signature ion at m/z 207.04 was detected only in the spectrum for AAΨUGp (inset in B) and is clearly discriminated from the m/z 207.01 ion derived from 2′-O-methylated RNA (inset in C).

analysis after Ψ-directed cyanoethylation of this fragment also supported this sequence (Supplementary Figure S8), demonstrating that our method allows direct *de novo* sequencing of Ψ-containing RNA at sub-picomole to femtomole quantities.

Identification of ^{11}Ψ in human U5 snRNA

U5 snRNA is 117 bases long, and Ψs are located at positions 43, 46 and 53 (36). Our preparation of U5 snRNA contained the structural isomers U5A and U5B, which differ in nucleotides at seven positions (Supplementary Table S3). Nevertheless, the LC-MS^2 analysis of this preparation resulted in four RNase T1 fragments containing the Ψ-characteristic signa-

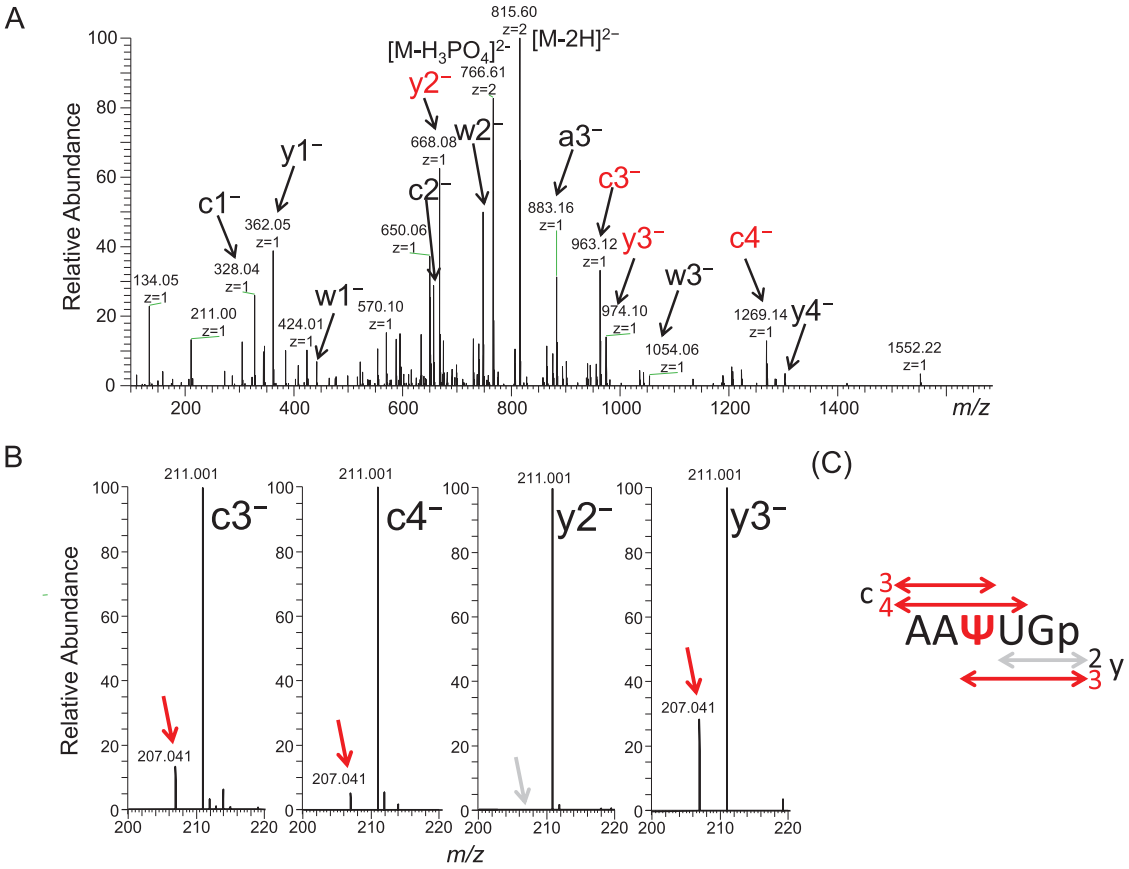

Figure 4. Determination of Ψ in a synthetic RNA by pseudo-MS^3. (**A**) In-source fragmentation spectrum of AAΨUGp. The major sequencing ions are indicated, with those containing potential Ψ shown in red. M, precursor molecule. (**B**) Pseudo-MS^3 spectra produced by CID of the c- and y-series ions in (A). Regions of the spectra near the *m/z* 207.04 ion are shown. Note that the MS^3 fragments produced from c3-, c4- and y3-series ions contained the *m/z* 207.04 signature ion for Ψ (red arrows), whereas the fragment from the y2$^-$ ion did not. (**C**) Sequence of the Ψ-containing synthetic RNA determined by pseudo-MS^3.

ture ion at *m/z* 207.04 (Figure 6). The three known Ψs were recovered in the two fragments having the sequences ^{25}CAUAAAUCUUUC(Gm)CCU(Um)UΨA(Cm)ΨAAA^{50}Gp and ^{51}AUΨUCC^{57}Gp, as described previously. The remaining Ψ-containing fragment, on the other hand, had a tentative sequence ^{10}(U/Ψ)(U/Ψ)(U/Ψ)C(U/Ψ)C(U/Ψ)(U/Ψ)CA^{20}Gp. Determination of the Ψ position in this 11-nucleotide fragment was relatively difficult because it contained six potential Ψs. First, the analysis of a series of sequence ions resulting from the pseudo-MS^3 analysis of this fragment showed the sequence ^{10}(U/Ψ)(U/Ψ)(U/Ψ)C(U/Ψ)CUUCA^{20}Gp, in which Ψ might be at the first to fifth position from the 5′ terminus (Supplementary Figure S9). Next, this fragment was cleaved with RNase H to prepare a shorter 5′ fragment, ^{10}(U/Ψ)(U/Ψ)(U/Ψ)C^{14}(U/Ψ), which was subjected to pseudo-MS^3 analysis. The inspection of a series of sequence ions in this spectrum clearly indicated that the sequence of this RNase H fragment was ^{10}(U/Ψ)ΨUC^{14}U (Supplementary Figure S10). Finally, we performed LC-MS^2 analysis of an RNase A digest of the purified U5 snRNA and identified the fragment ^{8}GG^{10}Up with no ^{8}GG^{10}Ψp (Supplementary Figure S11). Thus, the sequence of the Ψ-containing fragment derived from U5 snRNA was finally determined to be ^{10}UΨUCUCUUCA^{20}Gp, with a single Ψ located at position 11 from the 5′ terminus. This sequence, containing ^{11}Ψ, was confirmed by the Ψ-directed cyanoethylation of this fragment followed by the LC-MS analysis (Supplementary Figure S12). Notably, we found that the pseudouridylation at this position appeared to be partial, with Ψ occupying this position in only ∼40% of fragments as estimated from the EIC (Figure 6).

DISCUSSION

We present a method for direct MS-based characterization of a 'mass-silent' RNA PTM, namely pseudouridylation. This method utilizes an *m/z* 207.04 anion produced by CID-mediated cleavage leaving a Ψ-specific C–C bond as a Ψ-characteristic signature and determines a Ψ-containing oligonucleotide sequence directly by RNase mapping coupled with pseudo-MS^3 analysis. This method is applicable by various types of mass spectrometer currently available, *e.g.* quadrupole-Orbitrap, ion trap-Orbitrap, quadrupole-time of flight and Fourier transform ion cyclotron resonance mass spectrometer, which can perform the data dependent MS^2, can discriminate *m/z* 207.04 and 207.01, and can perform pseudo-MS^3 or full-MS^3. Because the procedure requires no previous RNA labeling and can determine

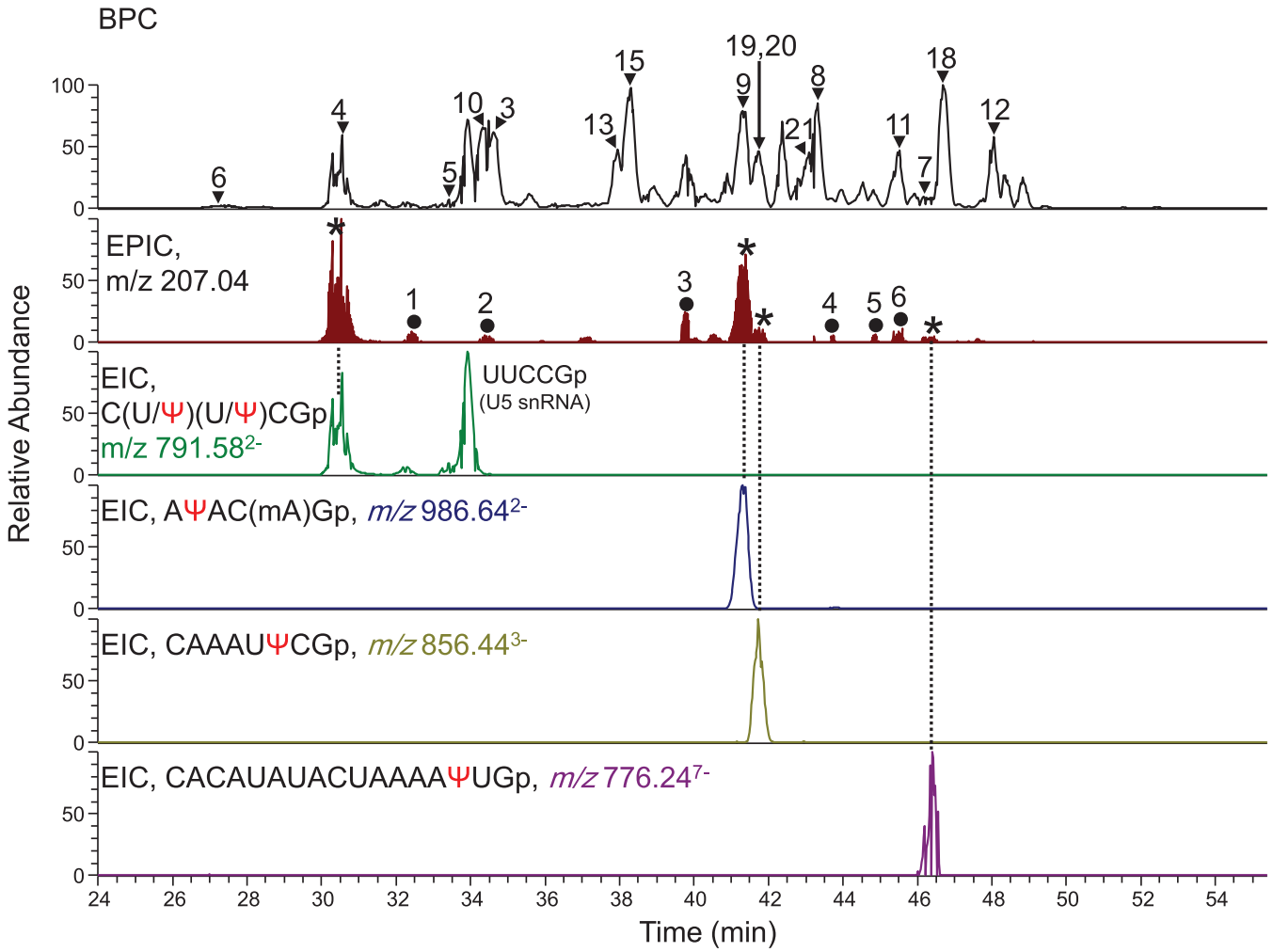

Figure 5. Identification of four Ψ-containing fragments in human U6 snRNA. The RNase T1 digest of human U6 snRNA (100 fmol) was analyzed by LC-MS and MS^2. The base peak chromatogram (BPC) of MS, EPIC of MS^2 and four EICs of MS are shown. The major peaks in the BPC, indicated by the arrow or arrowheads with peak numbers, were assigned to the fragments of U6 snRNA (see Supplementary Table S1). The peaks in the EPIC indicated by asterisks were assigned to the Ψ-containing fragments of U6 snRNA by Ariadne, whereas the peaks with numbered closed circles indicate fragments from contaminating RNAs in our U6 snRNA preparation and the U6 snRNA fragment with the 2′,3′-cyclic-phosphate terminus assigned as follows: 1, CΨUCG > p (U6 snRNA); 2, TΨCGp (tRNA); 3, AUΨUCCGp (U5 snRNA); 4, A(Cm)UAAAGp (U5 snRNA); 5, CAUAAAUCUUUC(Gm)CCUUU(Um)A(Cm)ΨAAAGp (U5 snRNA); 6, UAUAAAUCUUUC(Gm)CCUUU(Um)A(Cm)ΨAAAGp (U5 snRNA). The oligoribonucleotide with a sequence UUCCGp detected in the EIC with *m/z* 791.58 was derived from U5 snRNA contaminated in our U6 snRNA preparation. Note that contaminated RNA fragments can easily be distinguished those derived from a sample RNA by comparing the identified sequence with its genomic sequence using Ariadne. The peaks of the Ψ-containing fragments of U6 snRNA reappear in the EIC. The sequence of each RNase T1 fragment and its *m/z* value (±15 ppm mass tolerance) for extraction are indicated on each EIC. Corresponding peaks on the EPIC and EIC are linked with dotted lines.

the sequence from femtomole to sub-picomole quantities of RNA, it will serve as a useful tool for structure/function studies of pseudouridylation. We note that further development of software to generate automatically the inclusion list for targeted LC-MS^3 from the data obtained by 1st LC-MS^2 analysis will be important to expand the utility of this method and to make it more general approach for MS-based Ψ analysis.

Separate from the pseudo-MS^3–dependent determination of Ψ-containing RNA sequences reported here, we also examined the ability of conventional MS^3 analysis to determine the Ψ-containing sequences. In this approach, a sample RNA was digested with RNase T1 and the resulting digest subjected directly to LC-MS^3 analysis. Application of this method to human U6 snRNA, for example, allowed us to identify the two Ψ-containing fragments ^{8}CΨUC^{12}Gp and ^{81}CAAAUΨC^{88}Gp by monitoring the signature ion at *m/z* 207.04 in the full-MS^3 spectra obtained from the fragmentation of c- and y-series ions, whereas we could detect, but could not sequence, the other two Ψ-containing fragments in U6 snRNA, owing to low signal intensity (Supplementary Figures S13 and S14). We estimated that conventional MS^3 analysis was ∼10-fold less sensitive than pseudo-MS^3 analysis under the conditions reported here, probably because the signal-dependent automatic switching from MS^2 to MS^3 measurement might consume the acquisition time for separated sample ions. Thus, further development of MS instrumentation or/and sample handling techniques

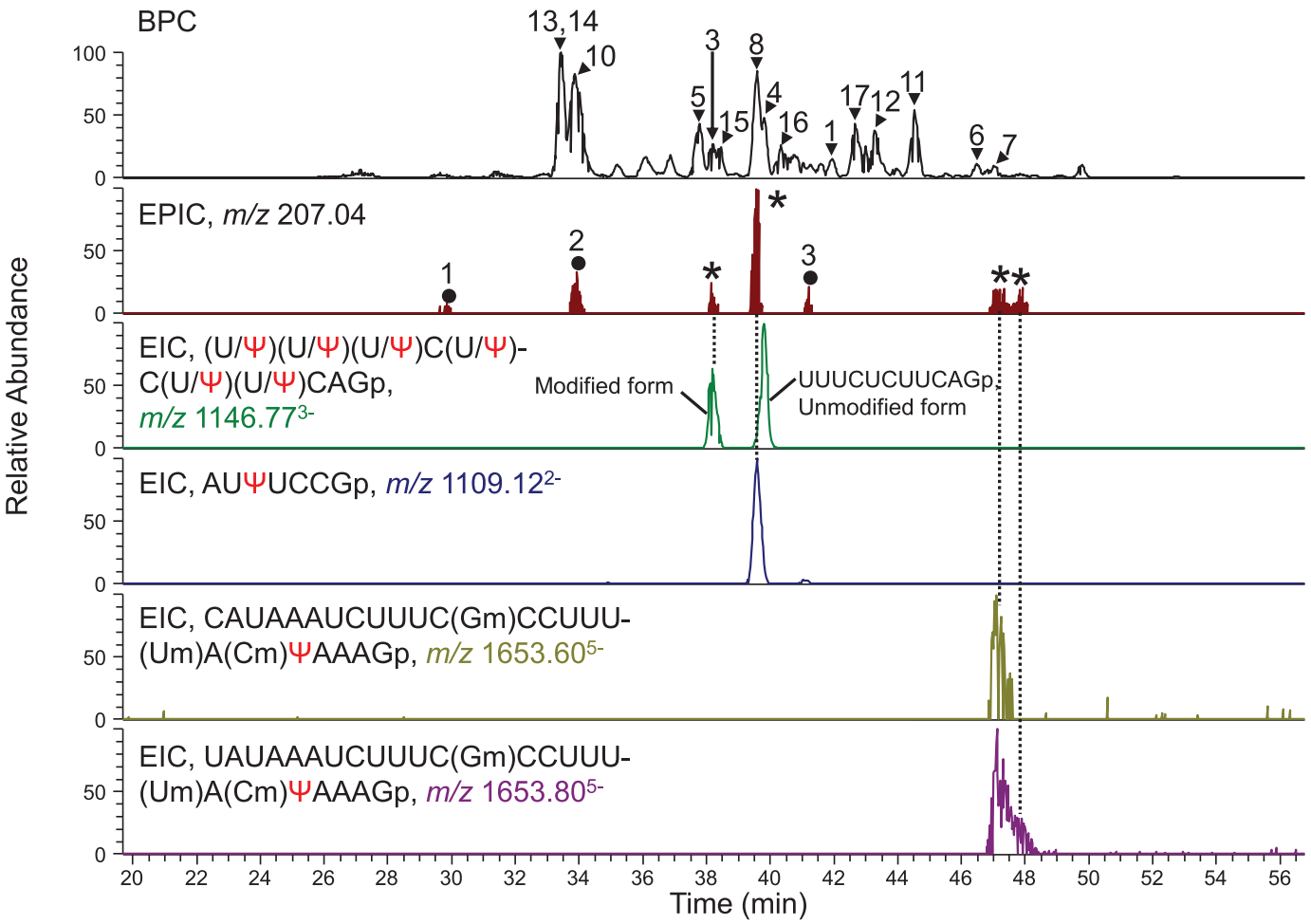

Figure 6. Identification of four Ψ-containing fragments in human U5 snRNA. The RNase T1 digest of human U5 snRNA (100 fmol mixture of U5A and U5B) was analyzed by LC-MS and MS^2. The BPC of MS, EPIC of MS^2 and four EICs of MS are shown. The major peaks in the BPC, indicated by the arrow or arrowheads with peak numbers, were assigned to the fragments of U5A or U5B snRNA (see Supplementary Table S1). The major peaks in EPIC indicated by asterisks were assigned to the Ψ-containing fragments of U5A/U5B snRNA. The peaks with numbered closed circles in the EPIC were fragments of contaminating RNA in our U5A/B preparation assigned as follows: 1, CΨUCGp (U6 snRNA); 2, TΨCGp (tRNA); 3, ΨAC(mA)Gp (U6 snRNA). The peaks of Ψ-containing fragments of U5A or U5B snRNA reappear in the EIC. The sequence of each RNase T1 fragment and its *m/z* value (±15 ppm mass tolerance) are indicated. Corresponding peaks on EPIC and EIC are linked with dotted lines.

will be necessary to improve the sensitivity of the MS^3 mode to make it applicable to minute amounts of RNA.

Using our method, we identified two previously unknown Ψs—one in human U5 snRNA and one in U6 snRNA. The U5, U6 and U4 snRNAs are the RNA components of the tri-snRNP (small nuclear ribonucleoprotein) spliceosomal complex, an extremely large (>1.5 MDa) intermediate snRNP containing more than 30 proteins, which participates in the later steps of pre-mRNA splicing (40). Thus, the loop 1 structure, consisting of the 11-nucleotide sequence GCCUUUUAC, is highly conserved among U5 snRNAs from various species (41). In humans, this loop contains four PTM sites and has the sequence ^{37}(Gm)CCU(Um)UΨA^{45}(Cm) (Supplementary Table S2 and Supplementary Figure S14). It has also been shown that loop 1 contributes to the stabilization of the interaction mediated by p220 (human ortholog of yeast Prp8) between the exon and U5 snRNA during pre-mRNA splicing (39,42,43). The ^{11}Ψ identified in this study is located within the stem of loop 1. Given that this stem structure appears to associate with the N-terminal region of Prp8 and Snu114 in the yeast spliceosomal complex (44), ^{11}Ψ may contribute to the stabilization of interactions of the human complex with p220 and SNU114.

Including ^{9}Ψ identified in this study, human U6 snRNA contains four Ψs and two base- and eight 2′-O-methylated nucleotides. In the secondary structure, more than half of those modified nucleotides, including two Ψs, are located at the binding surface for U4 snRNA and may stabilize the base-pairing interaction (Supplementary Figure S15). The unwinding of this U4/U6 duplex is ATP-dependent and provides a proofreading step to ensure proper positioning of the tri-snRNP on the spliceosome (45), and thus PTMs in the U4/U6 interface may regulate the rate of the proper splicing by affecting the thermodynamic stability of the base-pairing interactions (39). The ^{9}Ψ in U6 snRNA found in this study, on the other hand, is located within the 5′ stem–loop, which is not part of the U4/U6 interface. This

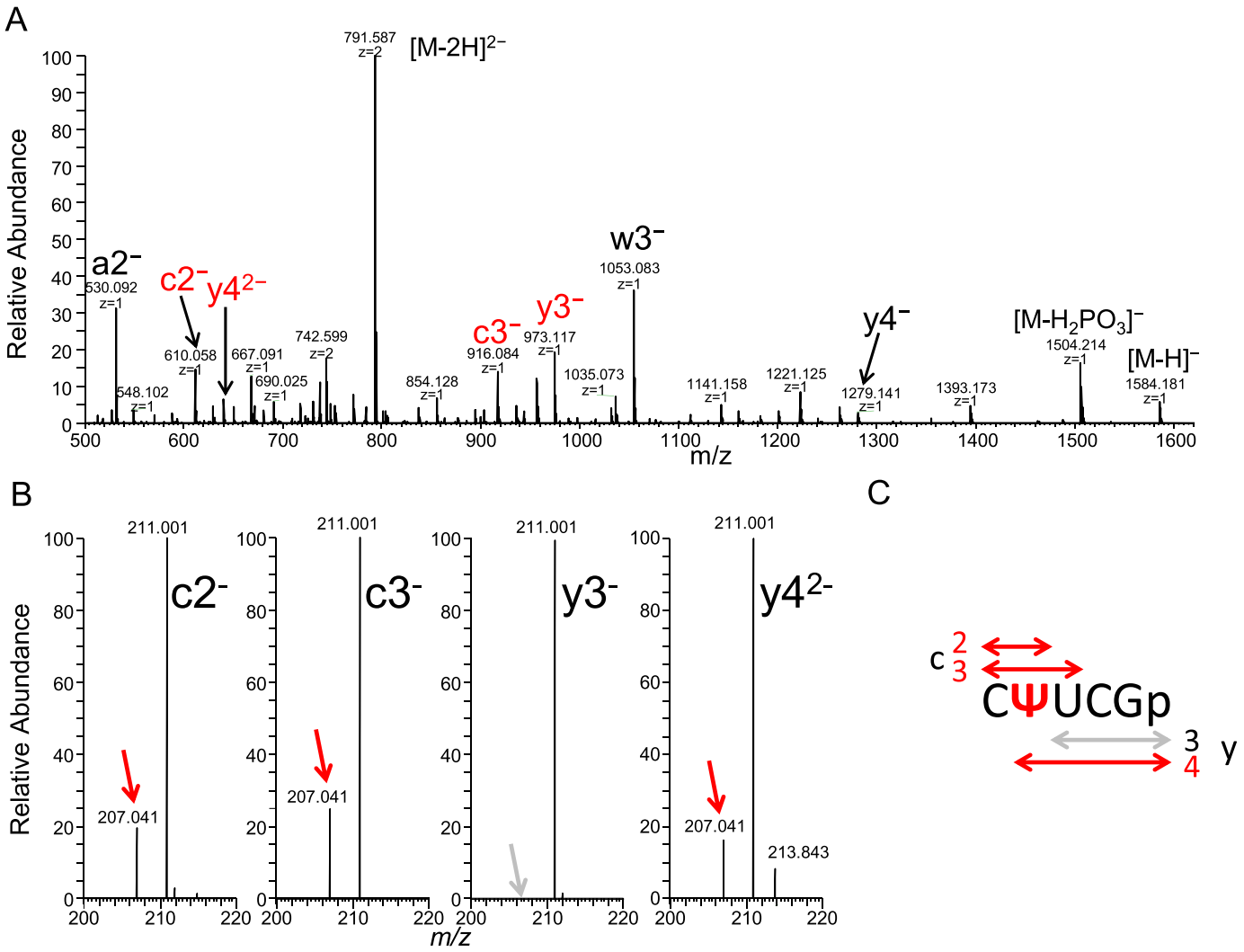

Figure 7. Determination of $^{9}\Psi$ in human U6 snRNA by pseudo-MS^3. (**A**) In-source fragmentation spectrum of CΨUCGp derived from U6 snRNA. The major sequencing ions are indicated in the figure, with those containing potential Ψ shown in red. M, precursor molecule. (**B**) Pseudo-MS^3 spectra produced by CID of the c- and y-series ions in (A). Regions of the spectra near the *m/z* 207.04 ion are shown. Note that the MS^3 fragments produced from the c2$^-$, c3$^-$ and y4$^-$ series ions contained the *m/z* 207.04 signature ion for Ψ (red arrows), whereas the fragment from the y3$^-$ ion did not. (**C**) Sequence of the Ψ-containing U6 snRNA fragment determined by pseudo-MS^3.

loop structure is conserved among U6 snRNAs from various species (46) and is tightly associated with protein factors in the yeast spliceosomal complex (47). However, the role of $^{9}\Psi$ in this interaction is not clear and requires further investigation.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Ms Chiharu Fujita for technical assistance.

FUNDING

Japan Science and Technology Agency for Core Research for Evolutional Science and Technology [ID 13415564]. Funding for open access charge: Core Research for Evolutional Science and Technology (CREST); Japan Science and Technology Agency (JST).
Conflict of interest statement. None declared.

REFERENCES

1. Cohn,E.W. and Volkin,E. (1951) Nucleoside-5′-phosphates from ribonucleic acid. *Nature*, **167**, 483–484.
2. Ofengand,J., Malhotra,A., Remme,J., Gutgsell,N.S., Del Campo,M., Jean-Charles,S., Peil,L. and Kaya,Y. (2001) Pseudouridines and pseudouridine synthases of the ribosome. *Cold Spring Harb. Symp. Quant. Biol.*, **66**, 147–159.
3. Ge,J. and Yu,Y.T. (2013) RNA pseudouridylation: new insights into an old modification. *Trends Biochem. Sci.*, **38**, 210–218.
4. Machnicka,M.A., Milanowska,K., Osman Oglou,O., Purta,E., Kurkowska,M., Olchowik,A., Januszewski,W., Kalinowski,S., Dunin-Horkawicz,S., Rother,K.M. *et al.* (2013) MODOMICS: a database of RNA modification pathways–2013 update. *Nucleic Acids Res.*, **41**, D262–D267.

5. Schwartz,S., Bernstein,D.A., Mumbach,M.R., Jovanovic,M., Herbst,R.H., Leon-Ricardo,B.X., Engreitz,J.M., Guttman,M., Satija,R., Lander,E.S. *et al.* (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell*, **159**, 148–162.
6. Carlile,T.M., Rojas-Duran,M.F., Zinshteyn,B., Shin,H., Bartoli,K.M. and Gilbert,W.V. (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature*, **515**, 143–146.
7. Lovejoy,A.F., Riordan,D.P. and Brown,P.O. (2014) Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. *PLoS One*, **9**, e110799.
8. Li,X., Zhu,P., Ma,S., Song,J., Bai,J., Sun,F. and Yi,C. (2015) Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.*, **11**, 592–597.
9. Charette,M. and Gray,M.W. (2000) Pseudouridine in RNA: what, where, how, and why. *IUBMB Life*, **49**, 341–351.
10. Uliel,S., Liang,X.H., Unger,R. and Michaeli,S. (2004) Small nucleolar RNAs that guide modification in trypanosomatids: repertoire, targets, genome organisation, and unique functions. *Int. J. Parasitol.*, **34**, 445–454.
11. Wu,G., Xiao,M., Yang,C. and Yu,Y.T. (2011) U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. *EMBO J.*, **30**, 79–89.
12. Karijolich,J. and Yu,Y.T. (2011) Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature*, **474**, 395–398.
13. Bakin,A. and Ofengand,J. (1993) Four newly located pseudouridylate residues in Escherichia coli 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry*, **32**, 9754–9762.
14. Durairaj,A. and Limbach,P.A. (2008) Mass spectrometry of the fifth nucleoside: a review of the identification of pseudouridine in nucleic acids. *Anal. Chim. Acta*, **623**, 117–125.
15. Patteson,K.G., Rodicio,L.P. and Limbach,P.A. (2001) Identification of the mass-silent post-transcriptionally modified nucleoside pseudouridine in RNA by matrix-assisted laser desorption/ionization mass spectrometry. *Nucleic Acids Res.*, **29**, e49.
16. Mengel-Jorgensen,J. and Kirpekar,F. (2002) Detection of pseudouridine and other modifications in tRNA by cyanoethylation and MALDI mass spectrometry. *Nucleic Acids Res.*, **30**, e135.
17. Suzuki,T. (2014) A complete landscape of post-transcriptional modifications in mammalian mitochondrial tRNAs. *Nucleic Acids Res.*, **42**, 7346–7357.
18. Taoka,M., Nobe,Y., Hori,M., Takeuchi,A., Masaki,S., Yamauchi,Y., Nakayama,H., Takahashi,N. and Isobe,T. (2015) A mass spectrometry-based method for comprehensive quantitative determination of post-transcriptional RNA modifications: the complete chemical structure of Schizosaccharomyces pombe ribosomal RNAs. *Nucleic Acids Res.*, **43**, e115.
19. Popova,A.M. and Williamson,J.R. (2014) Quantitative analysis of rRNA modifications using stable isotope labeling and mass spectrometry. *J. Am. Chem. Soc.*, **136**, 2058–2069.
20. Pomerantz,S.C. and McCloskey,J.A. (2005) Detection of the common RNA nucleoside pseudouridine in mixtures of oligonucleotides by mass spectrometry. *Anal. Chem.*, **77**, 4687–4697.
21. Del Campo,M., Recinos,C., Yanez,G., Pomerantz,S.C., Guymon,R., Crain,P.F., McCloskey,J.A. and Ofengand,J. (2005) Number, position, and significance of the pseudouridines in the large subunit ribosomal RNA of Haloarcula marismortui and Deinococcus radiodurans. *RNA*, **11**, 210–219.
22. Guymon,R., Pomerantz,S.C., Crain,P.F. and McCloskey,J.A. (2006) Influence of phylogeny on posttranscriptional modification of rRNA in thermophilic prokaryotes: the complete modification map of 16S rRNA of Thermus thermophilus. *Biochemistry*, **45**, 4888–4899.
23. Yamauchi,Y., Taoka,M., Nobe,Y., Izumikawa,K., Takahashi,N., Nakayama,H. and Isobe,T. (2013) Denaturing reversed phase liquid chromatographic separation of non-coding ribonucleic acids on macro-porous polystyrene-divinylbenzene resins. *J. Chromatogr. A*, **1312**, 87–92.
24. Taoka,M., Yamauchi,Y., Nobe,Y., Masaki,S., Nakayama,H., Ishikawa,H., Takahashi,N. and Isobe,T. (2009) An analytical platform for mass spectrometry-based identification and chemical analysis of RNA in ribonucleoprotein complexes. *Nucleic Acids Res.*, **37**, e140.
25. Nakayama,H., Yamauchi,Y., Taoka,M. and Isobe,T. (2015) Direct identification of human cellular microRNAs by nanoflow liquid chromatography-high-resolution tandem mass spectrometry and database searching. *Anal. Chem.*, **87**, 2884–2891.
26. Qiu,F. and McCloskey,J.A. (1999) Selective detection of ribose-methylated nucleotides in RNA by a mass spectrometry-based method. *Nucleic Acids Res.*, **27**, e20.
27. Nakayama,H., Akiyama,M., Taoka,M., Yamauchi,Y., Nobe,Y., Ishikawa,H., Takahashi,N. and Isobe,T. (2009) Ariadne: a database search engine for identification and chemical analysis of RNA using tandem mass spectrometry data. *Nucleic Acids Res.*, **37**, e47.
28. Wahl,M.C., Will,C.L. and Luhrmann,R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell*, **136**, 701–718.
29. Jurica,M.S. and Moore,M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell*, **12**, 5–14.
30. Roca,X., Akerman,M., Gaus,H., Berdeja,A., Bennett,C.F. and Krainer,A.R. (2012) Widespread recognition of 5' splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes Dev.*, **26**, 1098–1109.
31. Hartmuth,K., Urlaub,H., Vornlocher,H.P., Will,C.L., Gentzel,M., Wilm,M. and Luhrmann,R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 16719–16724.
32. Zhou,Z., Licklider,L.J., Gygi,S.P. and Reed,R. (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature*, **419**, 182–185.
33. Branlant,C., Krol,A., Ebel,J.P., Lazar,E., Gallinaro,H., Jacob,M., Sri-Widada,J. and Jeanteur,P. (1980) Nucleotide sequences of nuclear U1A RNAs from chicken, rat and man. *Nucleic Acids Res.*, **8**, 4143–4154.
34. Shibata,H., Ro-Choi,T.S., Reddy,R., Choi,Y.C., Henning,D. and Busch,H. (1975) The primary nucleotide sequence of nuclear U-2 ribonucleic acid. The 5'-terminal portion of the molecule. *J. Biol. Chem.*, **250**, 3909–3920.
35. Krol,A., Branlant,C., Lazar,E., Gallinaro,H. and Jacob,M. (1981) Primary and secondary structures of chicken, rat and man nuclear U4 RNAs. Homologies with U1 and U5 RNAs. *Nucleic Acids Res.*, **9**, 2699–2716.
36. Krol,A., Gallinaro,H., Lazar,E., Jacob,M. and Branlant,C. (1981) The nuclear 5S RNAs from chicken, rat and man. U5 RNAs are encoded by multiple genes. *Nucleic Acids Res.*, **9**, 769–787.
37. Epstein,P., Reddy,R., Henning,D. and Busch,H. (1980) The nucleotide sequence of nuclear U6 (4.7 S) RNA. *J. Biol. Chem.*, **255**, 8901–8906.
38. Reddy,R. (1985) Compilation of small RNA sequences. *Nucleic Acids Res.*, **13**(Suppl), r155–r163.
39. Karijolich,J. and Yu,Y.T. (2010) Spliceosomal snRNA modifications and their function. *RNA Biol.*, **7**, 192–204.
40. Will,C.L. and Luhrmann,R. (2011) Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.*, **3**, 1–23.
41. Frank,D.N., Roiha,H. and Guthrie,C. (1994) Architecture of the U5 small nuclear RNA. *Mol. Cell. Biol.*, **14**, 2180–2190.
42. Umen,J.G. and Guthrie,C. (1995) A novel role for a U5 snRNP protein in 3' splice site selection. *Genes Dev.*, **9**, 855–868.
43. Hinz,M., Moore,M.J. and Bindereif,A. (1996) Domain analysis of human U5 RNA. Cap trimethylation, protein binding, and spliceosome assembly. *J. Biol. Chem.*, **271**, 19001–19007.
44. Nguyen,T.H., Galej,W.P., Bai,X.C., Savva,C.G., Newman,A.J., Scheres,S.H. and Nagai,K. (2015) The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature*, **523**, 47–52.
45. Raghunathan,P.L. and Guthrie,C. (1998) RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr. Biol.*, **8**, 847–855.
46. Butcher,S.E. and Brow,D.A. (2005) Towards understanding the catalytic core structure of the spliceosome. *Biochem. Soc. Trans*, **33**, 447–449.
47. Karaduman,R., Fabrizio,P., Hartmuth,K., Urlaub,H. and Luhrmann,R. (2006) RNA structure and RNA-protein interactions in purified yeast U6 snRNPs. *J. Mol. Biol.*, **356**, 1248–1262.