Published in final edited form as:

Rapid Commun Mass Spectrom. 2008 December ; 22(23): 3727–3734. doi:10.1002/rcm.3789.

MALDI-MS SCREENING FOR PSEUDOURIDINE IN MIXTURES OF SMALL RNAS BY CHEMICAL DERIVATIZATION, RNASE DIGESTION AND SIGNATURE PRODUCTS

Anita Durairaj and Patrick A. Limbach*

Rieveschl Laboratories for Mass Spectrometry, Department of Chemistry, PO Box 210172, University of Cincinnati, Cincinnati, OH 45221-0172

Abstract

We have developed a method to screen for pseudouridines in complex mixtures of small RNAs using Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-MS). First, the unfractionated crude mixture of tRNAs is digested to completion with an endoribonuclease, such as RNase T1, and the digestion products are examined using MALDI-MS. Individual RNAs are identified by their signature digestion products, which arise through the detection of unique mass values after nuclease digestion. Next, the endonuclease digest is derivatized using N-cyclohexyl-N'-(2-morpholinoethyl)-carbodiimide metho-*p*-toluenesulfonate (CMCT), which selectively modifies all pseudouridine, thiouridine and 2-methylthio-6-isopentenyladenosine nucleosides. MALDI-MS determination of the CMCT-derivatized endonuclease digest reveals the presence of pseudouridine through a 252 Da mass increase over the underivatized digest. Proof-of-concept experiments were conducted using a mixture of *Escherichia coli* transfer RNAs and endoribonucleases T1 and A. More than 80% of the expected pseudouridines from this mixture were detected using this screening approach, even on a unfractionated sample of tRNAs. This approach should be particularly useful in the identification of putative pseudouridine synthases through detection of their target RNAs and can provide insight into specific small RNAs that may contain pseudouridine.

Keywords

CMC derivatization; Endonucleases; RNase T1; RNase A; Pseudouridine; 4-thiouridine; tRNA; MALDI-MS; transfer RNA; non-coding RNA

Pseudouridine (Ψ), the 5-ribosyl isomer of uridine, is the most common post-transcriptionally modified nucleoside in RNA. Pseudouridine is present in all three phylogenetic domains of life and it is found in both large and small non-coding RNAs (ncRNAs).^{1–3} Pseudouridine is found to be particularly abundant in smaller non-coding RNAs such as small nuclear RNAs (snRNAs) and transfer RNAs (tRNAs).^{1,4} In snRNAs, pseudouridines are located within regions that are involved in snRNA-snRNA interactions and may play a role in strengthening these intermolecular associations.⁴ Additionally, their presence in specific regions of snRNAs also suggests a role in the functional assembly of small nuclear ribonucleoproteins (snRNPs) and the spliceosome.⁵

Among tRNAs, the conservation of pseudouridine at specific positions across all three domains of life is more obvious. For example, the pseudouridine at position 55 (Ψ -55) is universally conserved in almost all tRNAs of known sequence.^{6,7} Pseudouridines in

*To whom correspondence should be addressed. Phone (513) 556-1871, Fax (513) 556-9239, Pat.Limbach@uc.edu.

conserved positions in tRNA exhibit a stabilizing influence on RNA structure by virtue of their extra hydrogen bonding capability and their ability to enhance base stacking.^{8–10} Pseudouridines may also play an important role in the efficiency of the translational process.¹¹ One focus of tRNA research is geared towards understanding the significance of pseudouridine and other modified nucleosides in mitochondrial tRNA function and mitochondrial diseases. Recent studies demonstrate the association of loss of tRNA pseudouridylation at specific sites with functional defects of mitochondrial tRNAs and pathogenesis in humans.^{4,12} Additionally, lack of pseudouridine in tRNAs has been found to be linked to deficiencies in the expression of certain genes.¹³

An essential aspect for clarifying the functional and molecular basis for pseudouridine in RNA is its structural characterization and sequence location. This requires the fractionation and purification of individual RNAs prior to sequencing. Although a number of RNA fractionation techniques have been reported,¹⁴ the isolation and purification of small RNAs, such as tRNAs, from a complex cellular pool remains a challenging process.¹⁵ In particular, the isolation of mitochondrial tRNAs in amounts sufficient for structural characterization is difficult and laborious. Moreover, a different set of challenges arises upon sequencing pseudouridine because it is an isomer of uridine.

The classical approach for identifying pseudouridine in small RNAs such as tRNAs and snRNAs involves radioactive labeling, ribonuclease digestion and two-dimensional thin-layer chromatography.^{5,13,16} That approach was replaced by the Bakin and Ofengand approach for pseudouridine mapping in RNA,¹⁷ which is based on chemical modification of pseudouridine by N-cyclohexyl-N'-(2-morpholinoethyl)-carbodiimide metho-*p*-toluenesulfonate (CMCT) and reverse transcription. This popular and effective method is most successful with larger ncRNAs such as ribosomal RNAs. A modified CMCT-reverse transcriptase approach for tRNAs has been reported, although this requires the use of polyadenylation to provide complete sequence coverage of the 3'-end of the molecule.¹⁰

More recently, mass spectrometry has begun to play an important role in the identification and localization of pseudouridine.^{18–24} Unlike conventional biochemical techniques, mass spectrometry offers a fast, sensitive and direct approach to analyze RNA.²⁵ Previously we described the application of CMCT derivatization with mass spectrometry-based detection.²⁰ CMCT derivatization incorporates a 252 Da mass tag onto pseudouridine, facilitating its detection by mass spectrometry. More recently, we improved this approach and incorporated pseudouridine detection into a general RNase mapping strategy.²⁶ However, those prior reports have focused on the identification of pseudouridine within isolated and purified single RNA species. As the RNase mapping approach is extendable to mixtures of tRNAs,²⁷ we sought to determine if the CMCT derivatization strategy could also be extended to mixtures of tRNAs.

In the present study, we have developed a matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) method to screen for pseudouridines in a mixture of unfractionated tRNAs. Knowledge of the constituent tRNAs in the mixture is provided by detection of signature endonuclease digestion products.²⁷ The simultaneous detection of pseudouridine would be accomplished by a CMCT-derivatization/MALDI-MS based approach. Because signature digestion products are unique *m/z* values associated with a particular tRNA and pseudouridine is an isomer of uridine (*i.e.*, pseudouridine does not have a unique mass value), in principle these two approaches should be complementary. For sequence placement of the detected pseudouridine, an RNase mapping strategy with multiple endonucleases should enable the mapping of pseudouridines to the known sequences of those tRNAs that were identified by their signature digestion products. Proof-

of-concept studies were conducted using an unfractionated mixture of *Escherichia coli* tRNAs known to contain pseudouridine.

EXPERIMENTAL

Materials

Escherichia coli tRNAs, N-cyclohexyl-(2-morpholmoethyl)-carbodiimide metho-*p*-toluenesulfonate (CMCT), Tris-HCl, urea, EDTA, ammonium bicarbonate, diammonium hydrogen citrate (DAHC) and 2,4,6-trihydroxyacetophenone (THAP) were obtained from Sigma-Aldrich (St. Louis, MO, USA) and used without further purification. RNases A and T1 were purchased from Roche Molecular Biochemicals (Indianapolis, IN, USA). C18 Zip-Tips were obtained from Millipore Corporation (Billerica, MA, USA). Sep-Pak C18 cartridges were purchased from Waters (Milford, Massachusetts, USA).

RNase Digestion

RNase T1 was precipitated from its original suspension by use of acetone. The precipitate was resuspended and then eluted in 1 mL of 75% aqueous acetonitrile from a Sep-Pak C18 cartridge. The solution was taken to dryness and resuspended in RNase T1 buffer, 50 mM Tris-HCl (pH 7) containing 1 mM EDTA. Approximately 5 µg of tRNA was digested with ca. 250 U of RNase T1 in 2 µL of RNase T1 buffer at 37 °C for 1–2 h.

The digest was speed-vac dried and reconstituted in 10 µL of C18 Zip-Tip equilibration solution (0.1 M TEAA, pH 7). The C18 Zip-Tip was equilibrated for binding by washing three times with 10 µL of equilibration solution. The oligonucleotide sample was bound to the C18 Zip-Tip by aspirating and dispensing the sample in 5 to 10 cycles. This step was followed by washing with 0.1 M TEAA, pH 7 and 100% H₂O. For the final elution step, 1–10 µL of the elution solution (50% aqueous acetonitrile) was dispensed into an empty vial. The eluent was then aspirated and dispensed through the C18 Zip-Tip at least three times without introducing air. The eluent can be speed-vac dried or directly spotted with matrix for MALDI analysis.

For the RNase A digestion, approximately 1 mg of the stock RNase A was dissolved in 1 mL of 50 mM Tris-HCl (pH 7) and 1 mM EDTA by boiling for 20 min. After the solution had cooled, the RNase A was divided into aliquots and stored for further use at –20 °C. Approximately 5 µg of tRNA was digested with 3 µL of the stock RNase A solution (0.01 U RNase A/µg) at 37 °C for 2–4 h. The digest was then speed-vac dried and purified by C18 Zip-Tips, as described above, prior to CMCT derivatization.

CMCT derivatization and alkaline hydrolysis

RNase A or RNase T1 digests were incubated in 3–5 µL of a stock CMCT solution, prepared from 20 mg CMCT in 1 mL of a buffer containing 50 mM Tris-HCl, 4 mM EDTA and 7 M urea. The reaction solution was vortexed and incubated in a water bath for 12–24 h at 37 °C. For the alkaline hydrolysis reaction, which is necessary to remove CMCT from all reacted nucleosides except pseudouridine, thiouridines, and 2-methylthio-6-isopentenyladenosine, the CMCT-reacted RNA was first purified by the use of C18 Zip-Tips and then allowed to react for 1–2 h in 50 mM NH₄HCO₃ (pH 10.4) at 75–80 °C (Supplemental Material Scheme S1). The resulting solution was speed-vac dried and reconstituted in 2 µL of nanopure water for MALDI analysis.

Mass Spectrometry

Mass spectrometric analysis was performed using a Bruker Reflex IV MALDI-TOF instrument (Bruker Daltonics, Billerica, MA) equipped with a nitrogen laser ($\lambda = 337$ nm).

MALDI spectra were obtained in negative ion and reflectron mode. A two point calibration using dT₃ and dT₁₅ was used in all analyses. The matrix solutions used were 300 mM THAP in acetonitrile and 250 mM DAHC in water. A two layer sample spotting approach was used. First, 0.5 μ L of the THAP matrix was spotted onto the MALDI plate and allowed to dry, then two microliters of THAP and DAHC combined in a 1:1 ratio were mixed with two microliters of sample. Approximately one microliter of this mixture was spotted on top of the previously dried THAP on the MALDI plate.

Data Analysis

Genomic tRNA sequences were obtained from the tRNA sequence database.²⁸ Theoretical endonuclease digestions of *E. coli* tRNAs were done using “Mongo Oligo Mass Calculator” (<http://library.med.utah.edu/masspec/mongo.htm>). The assigned base compositions for the experimentally observed mass values were obtained using “Oligo Composition Calculator” (<http://library.med.utah.edu/masspec/compo.htm>) with mass accuracies of $\pm 0.05\%$. Signature digestion product lists were obtained from the RNAccess web site (<http://bearcatms.uc.edu/rnaccess/>). MALDI peak lists were exported to Microsoft Excel for manipulation and analysis.

RESULTS AND DISCUSSION

Experimental Scheme

The experimental scheme developed here involves two separate MALDI-MS measurements of the endonuclease digest from the initial mixture of RNAs (Figure 1). In one measurement, the endonuclease digest is directly characterized by MALDI-MS to both identify constituent RNAs through their signature digestion products and to generate the control data set for comparison to the CMCT-derivatized sample. In the other measurement, the endonuclease digest first was derivatized following the CMCT protocol described in the Experimental section. That sample was then analyzed by MALDI-MS to manually identify mass shifts of 252 Da, which are indicative of CMCT derivatization. The base compositions of all ions undergoing a 252 Da shift were then identified on the basis of the RNase specificity and accurate mass measurement.²⁹ These base compositions were then matched against the genomic RNA sequences of the RNAs identified by signature product ions to place pseudouridine within existing RNA sequences. Where feasible, information from multiple endonucleases can be used to confirm RNA sequence placement of pseudouridine.

Previous studies have demonstrated the utility of CMCT derivatization and MALDI-MS for the identification of pseudouridine in specific, isolated tRNAs.^{20,26} For the present study, a commercial sample of *E. coli* tRNAs was obtained. CMCT derivatization experiments were conducted to determine whether or not the CMCT derivatization/MALDI-MS approach could be applied to mixtures of tRNAs, and to determine whether pseudouridines could be identified and localized to specific tRNAs present in the commercial sample.

tRNA Identification

To identify the tRNAs present in the sample, the signature digestion approach was used.²⁷ The enzymatic digestion of an individual tRNA by an RNase (e.g., RNase T1) will generate a number of specific endonuclease digestion products. A comparison of an organism's complete complement of tRNA RNase digestion products yields a set of unique or “signature” digestion product(s) that ultimately enable the detection of individual tRNAs from a total tRNA pool.

For example, *E. coli* tRNA^{Cys} has four RNase T1 signature digestion products (m/z 1614, 1793, 2685 and 2914) and three RNase A signature digestion products (m/z 1303, 1649 and

1959). The mass spectrometric detection of any one or more of these signature products from a mixture of *E. coli* tRNAs will confirm the presence of *E. coli* tRNA^{Cys}. When RNase T1 and A are used in combination, all *E. coli* tRNA families generate signature products that can be detected by MALDI-MS as well as all isoaccepting tRNAs except AlaIII, GluII, GluIII and TyrI.

As the first step in this experimental scheme, the sample was digested with either endonuclease RNase A or RNase T1. Figure 2 presents representative mass spectral data obtained from the analysis of the RNase A (Fig. 2a) and RNase T1 (Fig. 2b) digested tRNA sample. Table 1 lists all of the identified signature digestion products, which were arrived at by comparing experimental peak lists (provided as supplemental material in Supplemental Tables S1 (RNase A) and S2 (RNase T1)) with the *E. coli* tRNA signature digestion product lists available from the RNAccess database.²⁷

Table 1 summarizes all the *E. coli* tRNAs in the mixture identified by their RNase A and T1 signature digestion products. The tRNA families identified in this manner included: Arg, Gln, Glu, Gly, His, Leu, Met, Phe, Ser, Thr, Trp, Tyr and Val. These results are consistent with the tRNAs stated to be present by the vendor, which included Ala, Arg, Asp, Glu, Gly, His, Ile, Leu, Lys, Phe, Ser, Thr, Tyr and Val. *E. coli* tRNAs Ala, Asp, Ile and Lys each have several possible signature digestion products, which have been identified previously.²⁷ Thus, it is possible that these tRNAs are present but below the limit of detection (~ 50 ng) for the signature digestion approach. For all subsequent analyses, only those tRNAs confirmed to be present through identification of their signature digestion products were analyzed for pseudouridine.

RNase Digestion, CMCT Derivatization and MALDI-MS Determination of *E. coli* tRNAs

In a separate step, the RNase digests were reacted with CMCT as described in the Experimental section. CMCT derivatization followed by incubation in base leads to the selective modification of pseudouridines, 4-thiouridine and 2-methylthio-*N*⁶-isopentenyladenosine.²⁶ Thus, after this reaction, RNase digestion products whose mass increases by a multiple of 252 u should contain one or more of these post-transcriptionally modified nucleosides.

To identify derivatized RNase digestion products, a portion of each reaction mixture was analyzed by MALDI-MS (Figure 3). Fig. 3a contains representative results from the reaction of CMCT with the RNase A digest, and Fig. 3b presents results from the reaction of CMCT with the RNase T1 digest. In addition, Supplemental Tables S3 (RNase A) and S4 (RNase T1) present the ions arising from the mass spectral data.

RNase digestion products that contain pseudouridine should be derivatized by CMCT resulting in a 252 Da increase. To identify potential pseudouridine containing RNase digestion products, the *m/z* values between unreacted and CMCT-reacted RNase digestion products (e.g., Figure 2a and Figure 3a) were compared using an in-house developed macro for Microsoft Excel. Any pairs differing by 252 could represent CMCT-derivatized RNase digestion products. For RNase A, nine *m/z* pairs (1325/1577; 1341/1593; 1356/1607; 1406/1657; 1424/1675; 1687/1938; 2081/2332; 2098/2350; 2362/2614) differing by 252 were identified from the experimental data. For RNase T1, five *m/z* pairs (1293/1545; 1302/1554; 2265/2517; 2402/2653; 3319/3572) differing by 252 were identified. In addition, for the RNase A digestion, two *m/z* pairs (1424/1926; 2098/2601) were identified which differed by 504 (i.e., two CMC adducts) and one *m/z* pair of similar mass difference (3319/3822) was found for the RNase T1 digestion.

Identification of RNase Digestion Products Undergoing Derivatization with CMCT

The next step is to identify the base composition of ions undergoing derivatization with CMCT and mapping those base compositions against the sequences of the tRNAs identified through their signature digestion products (Table 1). Base compositions were identified as described in the Experimental section through the combination of enzyme specificity and accurate mass measurement. If unmodified nucleotides were unable to match the measured m/z values, then the modified nucleosides listed in Table 1 were added to the search parameters. The modifications included are those known to be common in bacterial transfer RNAs as well as those present in the detected signature digestion products of specific tRNAs.

For example, after CMCT derivatization of the RNase A digest, an ion at m/z 2614 was detected, which is 252 Da higher than the ion at m/z 2362 in the unreacted RNase A digest. The only base composition that matches this m/z under the search constraints of one uridine residue, one terminal phosphate and mass measurement accuracy of 0.05% is A_2G_4Up . Likewise, the RNase A digestion ion at m/z 2098, which is 252 Da less than the ion at m/z 2350 found after CMCT derivatization, did not match any unmodified base compositions subject to the same constraints listed above. While no base composition matched when including the possibility of a methylation or thiolation, the incorporation of a 114 Da modification, which would be indicative of the post-translationally modified nucleoside 2-methylthio- N^6 -isopentenyladenosine, yields the matching base composition of $A_4GUp + 114$ Da. The entries listed in Table 2, which provides the base compositions matching the m/z values of RNase digestion products that increased by 252 or 504 u after reaction with CMCT, were all determined in a similar manner.

Placing Pseudouridines within Specific tRNAs

To then locate particular pseudouridines, the base compositions provided in Table 2 were searched manually against the tRNA sequences of those tRNAs identified by their signature digestion products (Table 1). For example, the base composition of the RNase A digestion product mentioned above, A_2G_4Up , which was found to be derivatized by CMCT, was searched against the tRNA sequences for the tRNA families noted in Table 1. The only matching sequence consistent with RNase A digestion was found to be 5'-(C)AGGGGAU(U)-3', which matches the sequence of *E. coli* tRNA Phe. Thus, A_2G_4Up is considered to be an RNase A digestion product arising from *E. coli* tRNA Phe, which was identified through signature digestion products as being a constituent tRNA in the original sample.

The specific tRNAs that have RNase digestion products which match particular base compositions are listed in Table 2. As noted in this table, several base compositions, such as AG_2Up and $CU_2Gp +$ one methyl group, are found in a significant number of tRNAs. However, there are several RNase digestion products which underwent a 252 or 504 Da mass increase after CMCT derivatization that only map to one or several tRNAs. Examples include A_3C_2UGp , which only maps onto tRNA Gln II as 5'-(58)AAUCCAG(64)-3', and A_2G_4Up and $[ms^2i^6A]A_3GUp$, which only map onto tRNA Phe as 5'-(26)AGGGGAU(32)-3' and 5'-(34)GAA $[ms^2i^6A]AU(39)$ -3', respectively. In those cases, this approach can be used to identify particular sites within specific tRNAs that contain a pseudouridine and confirm the presence of 2-methylthio- N^6 -isopentenyladenosine.

The advantage of using signature digestion products to first identify constituent tRNAs within the mixture is evident by the number of tRNAs that contain matching base compositions that were not detected in the original mixture. For example, the RNase A fragment at 1424 was found to increase by 252 and 504 Da after CMC derivatization. The

matching base composition for this m/z value required the inclusion of the modified nucleoside 2-methylthio- N^6 -isopentenyladenosine to yield $[ms^2i^6A]A_2Up$. Without additional constraints, five isoaccepting tRNAs have matching RNase A base compositions: Cys, Leu I, Phe, Tyr I and Tyr II. Two of these, tRNA Cys and Leu I, can be excluded due to the absence of any signature digestion products for those tRNAs. It should also be noted that the matching base composition contains two nucleosides that could be derivatized by CMCT under our experimental conditions, 2-methylthio- N^6 -isopentenyladenosine and uridine (or more accurately, pseudouridine). Thus, there is consistency in the base compositions identified using mass and RNase specificity along with CMCT reactivity within the data in Table 2.

As noted in Table 1, 13 tRNA families represented by up to 24 isoaccepting tRNAs were detected in the commercial mixture of *E. coli* tRNAs. To investigate how efficient the experimental protocol is at identifying pseudouridines within a tRNA mixture, the RNA sequences of the 24 possible isoaccepting tRNAs were obtained. Table 3 lists the known pseudouridines and their sequence placement for these 24 isoaccepting tRNAs. Because several RNase digestion products are not unique to any one isoaccepting tRNA (Table 2), we cannot verify whether redundant RNase digestion products, such as the commonly conserved T Ψ CGp RNase T1 digestion product, were detected from each particular isoaccepting tRNA. However, as noted in Table 3, more than 80% of the expected pseudouridines amenable to MALDI-MS analysis were identified, even starting from an unfractionated mixture.

In arriving at the assignments in Table 3, it was assumed that any base composition from Table 2 that contained only one possible nucleoside for CMCT derivatization (i.e., pseudouridine, thiouridine, or 2-methylthio- N^6 -isopentenyladenosine) was uniquely modified by CMCT. For those base compositions that could contain more than one nucleoside for CMCT derivatization, no specific CMCT assignment is possible in the absence of further sequencing of the CMCT derivatized RNase digestion products. However, it is noteworthy that the three RNase pairs that differed by two CMC adducts (RNase A: 1424/1926; 2098/2601 and RNase T1: 3319/3822) did match tRNA sequences that contained both pseudouridine and 2-methylthio- N^6 -isopentenyladenosine, suggesting that this approach can identify tRNAs containing more than one type of CMCT-derivatized nucleoside.

Obtaining sequence coverage with RNase T1 was especially problematic as analysis of the control RNase T1 digest in Figure 1b revealed that the majority of predicted RNase T1 digestion product sequences with m/z values greater than 4000 were missing and the quality of the mass spectral data was lower than obtained from the RNase A digestion. The most likely reason for the absence of higher molecular weight species is the peak suppression effect³⁰ which is common in MALDI-MS of complex tRNA digests.²⁷ Previous work shows that it is possible to obtain coverage of higher molecular weight sequences from a complex tRNA mixture, although signal intensities are found to be significantly lower.²⁷ The poorer mass spectral quality of the RNase T1 digestion products is most likely due to contamination arising from the enzymatic digestion, as prior MALDI-MS results from RNase T1 digests were of similar quality to that found for RNase A.²⁷ However, the data reported here are sufficient to identify constituent tRNAs and RNase T1 digestion products undergoing derivatization by CMCT.

Limitations

Although this experimental approach was found to reveal more than 80% of the pseudouridines expected to be present in this well-characterized sample, there are several limitations to this approach that arose during this study. Clearly, the lack of sequence

information limits the utility of this experimental scheme to the putative placement of pseudouridine within specific oligonucleotide sequences. Only those sequences with one potential site for pseudouridine or those pseudouridines identified by both endonucleases can be considered as rigorous localizations of pseudouridine. Additional experiments utilizing endonucleases such as RNase TA or RNase U2 could improve sequence coverage and lead to greater sequence overlap between digest fragments detected by this approach. Alternatively, the use of MS/MS, via MALDI-TOF/TOF or Q-TOF instruments, to sequence RNase digestion products that were derivatized by CMCT would provide more accurate information on the sequence location of particular pseudouridines.

Another limitation arises from the redundancy of pseudouridine in common sequence motifs of tRNAs. As mentioned above, Ψ -55 in the T-stem loop is universally conserved in bacterial tRNAs, and determining the specific isoacceptor tRNA that contains this pseudouridine would require prior isolation of the isoacceptor. In those cases, the requirement that isoaccepting tRNAs be identified through their signature digestion product(s) will limit the potential isoaccepting tRNAs present in the sample, but cannot differentiate which isoaccepting tRNA has its T-stem loop derivatized by CMCT.

More general limitations in this method arise from the requirement for CMCT derivatization for pseudouridine detection and signature digestion products for tRNA identification. Previously, we examined the experimental conditions to minimize under-derivatization of pseudouridine.²⁶ As noted in that work, under-derivatization would lead to false negatives – failing to identify pseudouridine when it actually is present. Another experimental concern would be false positives – oligonucleotide ions that differ by 252 Da which do not contain pseudouridine, thiouridine, or 2-methylthio- N^6 -isopentenyladenosine. Although false positives have not previously been found by us to be a significant concern, as mentioned above, the use of multiple endonucleases and/or tandem mass spectrometry provides for more rigorous sequence and tRNA assignments. As noted previously,²⁷ one limitation associated with the Signature Digestion Product approach for RNA identification is that RNA sequences are not known for many organisms. Use of this approach for organisms not in the RNA Access database is not recommended for this reason, although the CMCT derivatization scheme can still be used to identify the presence of pseudouridine in RNA mixtures.

CONCLUSION

It has been demonstrated that the CMCT derivatization/mass spectrometry based approach can enable the detection of pseudouridine in complex samples of small RNAs. The approach described here did not require prior HPLC purification or fractionation for determination of the mixture of tRNAs. While this approach would not be suitable for the de novo localization and sequence placement of pseudouridine in unknown RNAs, it was found to be reasonably effective at reporting the majority of pseudouridines and, when combined with knowledge of RNA identity and gene sequence, can be used to localize pseudouridines to particular RNA(s) in the mixture.

There are several applications where this screening approach for pseudouridine detection may be of interest. Because this approach is a relatively quick route to characterize pseudouridine in mixtures of RNAs, comparative studies focused on pseudouridine abundance or distribution would be enhanced through the screening of mixtures. For example, Ψ -15 in the D stem-loop is modified by a single pseudouridine synthase, *truD*, in bacteria. Putative homologs in other phylogenetic domains could be examined by screening pseudouridines in wild type and viable deletion mutants to determine whether the deleted gene is indeed the postulated pseudouridine synthase. In those cases, rigorous purification of

a particular substrate is not required for initial studies. Alternatively, this screening approach should be particularly applicable to initial examination of mixtures of small RNAs whose pseudouridine status is undetermined. Results obtained after CMCT screening could be used to limit the specific substrates that require more rigorous examination and sequence confirmation. In this fashion, the developed screening approach would be most useful in pseudouridine discovery studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support of this work was provided by the National Institutes of Health (GM58843) and the University of Cincinnati.

REFERENCES

1. Lane BG, Ofengand J, Gray MW. *Biochim.* 1995; 77:7.
2. Ofengand J. *FEBS Lett.* 2002; 514:17. [PubMed: 11904174]
3. Helm M. *Nucleic Acids Res.* 2006; 34:721. [PubMed: 16452298]
4. Patton J, Bykhovskaya Y, Mengesha E, Bertolotto C, Fischel-Ghodsian N. *J Biol Chem.* 2005; 280:19823. [PubMed: 15772074]
5. Zhao X, Yu Y. *RNA.* 2004; 10:681. [PubMed: 15037777]
6. Charette M, Gray WM. *Life.* 2000; 49:341. [PubMed: 10902565]
7. Grosshans H, Lecointe F, Grosjean H, Hurt E, Simos G. *J Biol Chem.* 2001; 276:46333. [PubMed: 11571299]
8. Davis D. *Nucleic Acids Res.* 1995; 23:5020. [PubMed: 8559660]
9. Yarian C, Basti M, Cain R, Ansari G, Guenther R, Sochacka E, Czerwinska G, Malkiewicz A, Agris P. *Nucleic Acids Res.* 1999; 27:3543. [PubMed: 10446245]
10. Ofengand J, Campo DM, Kaya Y. *Methods.* 2001; 25:365. [PubMed: 11860291]
11. Lecointe F, Namy O, Hatin I, Simos G, Rousset J, Grosjean H. *J Biol Chem.* 2002; 277:30445. [PubMed: 12058040]
12. Zeharia A, Fischel-Ghodsian N, Casas K, Bykhovskaya Y, Tamari H, Lev D, Mimouni M, Lerman-Sagie T. *Child Neurology.* 2005; 20:449.
13. Urbonavicius J, Durand JM, Bjork GR. *J. Bacteriol.* 2002; 184:5348. [PubMed: 12218021]
14. RajBhandary U, Köhrer C. *Bioscience.* 2006; 31:439.
15. Cayama E, Yopez A, Rotondo F, Bandeira E, Ferreras A, Triana-Alonso F. *Nucleic Acids Res.* 2000; 28:e64. [PubMed: 10871387]
16. Matsugi J, Murao K, Ishikura H. *Nucleic Acids Res.* 1992; 20:3514. [PubMed: 1630926]
17. Bakin A, Ofengand J. *Biochemistry.* 1993; 32:9754. [PubMed: 8373778]
18. Mengel-Jorgensen J, Jensen SS, Rasmussen A, Poehlsgaard J, Iverson JLL, Kirpekar F. *J. Biol. Chem.* 2006; 281:22108. [PubMed: 16731530]
19. Mengel-Jorgensen J, Kirpekar F. *Nucleic Acids Res.* 2002; 30:e135. [PubMed: 12466567]
20. Patteson KG, Rodicio LP, Limbach PA. *Nucleic Acids Res.* 2001; 29:e49. [PubMed: 11353094]
21. Hansen MA, Kirpekar F, Ritterbushch W, Vester B. *RNA.* 2002; 3:202. [PubMed: 11911366]
22. Kirpekar F, Douthwaite S, Roepstorff P. *RNA.* 2000; 6:296. [PubMed: 10688367]
23. Pomerantz SC, McCloskey JA. *Anal. Chem.* 2005; 77:4687. [PubMed: 16053277]
24. Durairaj A, Limbach P. *Anal Chim Acta.* 2008; 623:117. [PubMed: 18620915]
25. Thomas B, Akoulitchiev A. *Trends Biochem Sci.* 2006; 31:173. [PubMed: 16483781]
26. Durairaj A, Limbach P. *Anal Chim Acta.* 2008; 612:173. [PubMed: 18358863]
27. Hossain M, Limbach PA. *RNA.* 2007; 13:295. [PubMed: 17194720]

28. Sprinzl M, Vassilenko KS. Nucleic Acids Res. 2005; 33:D139. [PubMed: 15608164]
29. Pomerantz SC, Kowalak JA, McCloskey JA. J. Am. Soc. Mass Spectrom. 1993; 4:204.
30. Knochenmuss R. Anal Chem. 2003; 75:2199. [PubMed: 12918956]

Figure 1.
Experimental scheme for the screening of pseudouridines in mixtures of small RNAs.

Figure 2. MALDI mass spectra of *E. coli* tRNA mixture digested using (a) RNase A and (b) RNase T1. Detected m/z values are listed in Supplemental Tables S1 (RNase A) and S2 (RNase T1).

Figure 3. MALDI mass spectra of *E. coli* tRNA mixture digested using (a) RNase A and (b) RNase T1 after derivatization with CMCT. Detected m/z values are listed in Supplemental Tables S3 (RNase A) and S4 (RNase T1).

Table 1

Summary of all tRNAs in mixture detected by their experimentally obtained signature digestion products from RNase A (Fig. 2a) and T1 (Fig. 2b) mediated cleavage.

Isoaccepting tRNAs	RNase signature ion(s)	Sequence
Arg III	2160.9	[t ⁶ A]AGGGCp
Arg V	2046.0	G[m ¹ G]AGGCp
Gln II	1952.3; 2265.5	[m ² A]UACCGp; AAUCCAGp
Glu I	1960.2	UAACAGp
Glu II/III	1960.2; 3182.1	UAACAGp; AAUCCCCUAGp
Gly I	2016.2	GAGAGCp
Gly II	1671.9	AAGGUp
His	3512.2	AAUCCCAUAGp
Leu II/III	1996.3	AADD[Gm]Gp
Leu III	2571.5	CCCAAUAGp
Met	2242.4	AAUCCCGp
Phe	3319.8	AA[ms ² i ⁶ A]AUCCCGp
Ser I	2402.5	A[ms ² i ⁶ A]AACCGp
Ser II	1984.0	AAAAGCp
Ser III	2998.5	DGAAGGAGCp
Ser IV	1349.8; 2402.5	AAAGp; A[ms ² i ⁶ A]AACCGp
Ser V	1685.9; 2690.8	AGAGTp; GAAAGGGUp
Ser III/V	1349.8	AAAGp
Thr I/II	1011.4	AGTp
Trp	1752.9	A[ms ² i ⁶ A]AACp
Tyr I/II	3019.4	AAAGGGAGCp
Val III	4001.2	UCAUCACCCACCA

Modifications List

Gm: 2'-O-methylguanosine

m¹G: 1-methylguanosine

m²A: 2-methyladenosine

t⁶A: N⁶-threonylcarbamoyladenine

T (m⁵U): ribothymidine (5-methyluridine)

ms²i⁶A: 2-methylthio-N⁶-isopentenyladenosine

s⁴U: 4-thiouridine

yW: wybutosine

Ψ: pseudouridine

Table 2

Base compositions of RNase digestion products that increased by 252 or 504 u after reaction with CMCT. Base compositions were determined using a mass accuracy of 0.05% or better. If the experimental m/z did not match a base composition with no modifications, the known modification masses of tRNA modified nucleosides were then used to identify probable base compositions.

RNase	Exp m/z	Base compositions	Matching tRNAs (in Table 1)	Matching tRNAs (not in Table 1)
A	1341.7	AG ₂ Cp; AG ₂ Up	GlnII, GlyII, His, LeuII, LeuIII, Phe, SerI, SerII, SerIII, SerIV, Trp	ArgI, ArgII, ArgIV, Asn, Asp, Cys, IleII, LeuI, ValI, ValII
A	1424.7	A ₃ Up + 114 (ms^2i^6)	Phe, TyrI, TyrII	Cys, LeuI
A	1687.7	AG ₃ Up	LeuIII, Phe SerI, SerII, SerIII, SerV, ThrI, ThrII	IleI, IleII, LeuI
A	2081.3	A ₅ Up + 114 (ms^2i^6)	no match	no match
A	2098.4	A ₄ GUp + 114 (ms^2i^6)	Phe	no match
A	2362.4	A ₂ G ₄ Up	Phe	no match
T1	1293.6	CU ₂ Gp + 14 (methyl)	ArgIII, ArgV, GlnII, GluI, GluII, GluIII, GlyI, GlyII, His, Met, Phe, ArgIII, ArgV, SerII, SerIII, SerIV, SerV, ThrI, ThrII, Trp, TyrI, TyrII	AlaI, AlaII, AlaIII, ArgI, ArgII, ArgIV, Asn, Asp, Cys, GlnI, GlyIII, Lys, ValI, ValII, ValIII, Sec6
T1	1302.8	AU ₂ Gp	ArgIII, His, Phe, LeuII, SerII, ThrII	AlaI, ArgIV, Cys, Lys, ValII
T1	2265.5	A ₃ C ₂ UGp	GlnII	no match
T1	2402.5	A ₄ C ₂ Gp + 114 (ms^2i^6)	SerI, SerV, Trp	no match
T1	3319.8	A ₄ C ₄ UGp + 114 (ms^2i^6); A ₄ C ₃ U ₂ Gp + 114 (ms^2i^6)	Phe	no match

Table 3

Predicted Ψ -fragments from endonuclease digested *E. coli* tRNAs. Endonucleases include RNase A and RNase T1. tRNA sequences were obtained from the tRNA database.²⁸ Pseudouridines are highlighted for emphasis only. Sequences are only shown for RNase digestion products greater than m/z 1000 due to MALDI-MS background interferences present at lower masses. N.D. - not detected in this study.

tRNA	Ψ -sequence	Base compositions detected in this work
Arg III	Ψ – 55	T Ψ CGp (RNase T1)
Arg V	Ψ – 55	T Ψ CGp (RNase T1)
Gln II	Ψ – 38 Ψ – 55	[m ² A] Ψ ACCGp (RNase T1); G[m ² A] Ψ p (RNase A) T Ψ CGp (RNase T1)
Glu I	Ψ – 13 Ψ – 55	UC Ψ AGp (RNase T1) T Ψ CGp (RNase T1)
Glu II	Ψ – 13 Ψ – 55	UC Ψ AGp (RNase T1) T Ψ CGp (RNase T1)
Glu III	Ψ – 13 Ψ – 55	UC Ψ AGp (RNase T1) T Ψ CGp (RNase T1)
Gly I	Ψ – 55	T Ψ CGp (RNase T1)
Gly II	Ψ – 55	T Ψ CGp (RNase T1)
His	Ψ – 38 Ψ – 39 Ψ – 55	G[m ² A] Ψ p (RNase A) N.D. T Ψ CGp (RNase T1)
Leu II	Ψ – 38 Ψ – 40 Ψ – 55	AGG Ψ p (RNase A) Ψ UAGp (RNase T1) T Ψ CGp (RNase T1)
Leu III	Ψ – 38 Ψ – 55	GAGG Ψ p (RNase A) T Ψ CGp (RNase T1)
Met	Ψ – 39 Ψ – 55	N.D. T Ψ CGp (RNase T1)
Phe	Ψ – 32 Ψ – 39 Ψ – 55	A Ψ UGp (RNase T1); AGGGGA Ψ p (RNase A) AA[ms ² i ⁶ A]A Ψ CCCCGp (RNase T1); GAA[ms ² i ⁶ A]A Ψ p (RNase A) T Ψ CGp (RNase T1)
Ser I	Ψ – 55	T Ψ CGp (RNase T1)
Ser II	Ψ – 55	T Ψ CGp (RNase T1)
Ser III	Ψ – 40 Ψ – 55	GGAAAG Ψ p (RNase A) T Ψ CGp (RNase T1)
Ser IV	Ψ – 40 Ψ – 55	GGAAAG Ψ p (RNase A) T Ψ CGp (RNase T1)
Ser V	Ψ – 55	T Ψ CGp (RNase T1)
Thr I	Ψ – 55	T Ψ CGp (RNase T1)
Thr II	Ψ – 55	T Ψ CGp (RNase T1)
Trp	Ψ – 55	T Ψ CGp (RNase T1)
Tyr I	Ψ – 39 Ψ – 55	A[ms ² i ⁶ A]A Ψ p (RNase A) T Ψ CGp (RNase T1)
Tyr II	Ψ – 39 Ψ – 55	A[ms ² i ⁶ A]A Ψ p (RNase A) T Ψ CGp (RNase T1)
Val III	Ψ – 55	T Ψ CGp (RNase T1)