Published in final edited form as:

Anal Chim Acta. 2008 April 7; 612(2): 173–181.

Improving CMC-derivatization of pseudouridine in RNA for mass spectrometric detection

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

A protocol that utilizes matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) and N-cyclohexyl-N'-β-(4-methylmorpholinium)ethylcarbodiimide (CMC) derivatization to detect the post-transcriptionally modified nucleoside, pseudouridine, in RNA has been optimized for RNase digests. Because pseudouridine is mass-silent (i.e. the mass of pseudouridine is the same as the mass of uridine), after CMC derivatization and alkaline treatment, all pseudouridine residues exhibit a mass shift of 252 Da that allows its presence to be easily detected by mass spectrometry. This protocol is illustrated by the direct MALDI-MS identification of pseudouridines within *Escherichia coli* tRNA^{TyrII} starting from microgram amounts of sample. During this optimization study, it was discovered that the post-transcriptionally modified nucleoside 2-methylthio-N⁶-isopentenyladenosine, which is present in bacterial tRNAs, also retains a CMC unit after derivatization and incubation with base. Thus, care must be exercised when applying this MALDI-based CMC-derivatization approach for pseudouridine detection to samples containing transfer RNAs to minimize the misidentification of pseudouridine.

Keywords

CMCT derivatization; Endonucleases; RNase T1; Pseudouridine; 2-methylthio-N⁶-isopentenyladenosine; tRNA; MALDI-MS; RNA signature products

Introduction

The water-soluble carbodiimide, N-cyclohexyl-N'-β-(4-methylmorpholinium) ethylcarbodiimide (CMC) p-tosylate reacts with guanosine, thymidine and uridine-like components in DNA and RNA [1,2]. This single-strand specific reagent has served as a probe for unpaired and mismatched sites in DNA [3,4]. CMC-reacted DNA has also been shown to be inhibited in its ability to support transcription by RNA polymerase [5]. One of the most powerful applications of CMC derivatization has been its use as a sequencing tool for the detection of the posttranscriptionally modified nucleoside, pseudouridine (Ψ), in RNA [6-10]. Reaction with CMC at a pH of 8.5 modifies guanosine, uridines and the N-1 and N-3 positions of pseudouridine, but the CMC unit is retained selectively at higher pH (10.4) on the N-3 position of pseudouridine (N₃-CMC-Ψ). While N₁-CMC-Ψ and CMC-derivatized uridine and guanosine can release CMC under mild alkaline conditions, N₃-CMC-Ψ requires 7 M NH₄OH at 100 °C for 8 min for cleavage [1]. The method of Bakin and Ofengand exploits the

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

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

relative alkaline stability of N₃-CMC-Ψ as this derivatization product inhibits reverse transcriptase from reading the RNA sequence allowing one to determine these sites of modification [6,7]. While this method has been used with significant success for pseudouridine detection, it has been found that CMC-derivatized 4-thiouridine is also stable under mild alkaline conditions and that 5-methyluridine releases CMC slower than unmodified uridine [11].

In previous work, we applied the same principle of CMC derivatization with matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) to detect pseudouridines in two *Escherichia coli* tRNAs [9]. As in the method developed by Ofengand, pseudouridine is chemically modified using CMC p-tosylate and alkaline buffer. Thus, all pseudouridine residues except for 1-methyl-3-(amino-carboxypropyl) pseudouridine (m¹acp³Ψ) will exhibit a mass shift of 252.2 u due to the CMC unit. After HPLC purification of endonuclease digestion products, a comparison of the masses of the unreacted and CMC-reacted oligonucleotides will determine the number of pseudouridine residues in the oligonucleotide.

Although this approach was successfully demonstrated, its widespread implementation for pseudouridine identification by MALDI-MS has been hindered by incomplete derivatization, the requirement of purified CMC-modified oligonucleotides prior to analysis, and difficulties in co-crystallizing base-treated CMC-derivatized oligonucleotides with acidic matrices. Most significantly, the presence of unreacted oligonucleotides in the reaction mixture can hinder the identification of undermodified pseudouridines, and a failure to release CMC from guanosine and uridine residues during alkaline treatment can lead to a misinterpretation of those locations as pseudouridine containing sites. Ofengand and co-workers also advise this to be a major source of their error when they inadvertently reported an incorrect number of pseudouridines in *Haloarcula marismortui* RNA [10].

In related work, alternative chemical derivatizations of pseudouridine have been used prior to mass spectrometric detection [12,13]. As with the CMC-based approach, these derivatization strategies place a mass tag on pseudouridine that facilitates its detection by mass spectrometry. However, these reactions are also found to modify other uridine nucleosides, which potentially can be problematic when analyzing RNAs which have many naturally occurring uridine residues. Pseudouridine detection is not limited to methodologies involving chemical modification. Recent examples of non-derivatization approaches include LC/ESI-MS/MS for pseudouridine detection in mixtures of oligonucleotides [14] and the use of DNazymes and their cleavage efficiencies to determine the presence of pseudouridines [15]. In the case of DNazymes, the influence of pseudouridine on cleavage efficiencies of these synthetic ribonucleases is inconsistent and varies depending on pseudouridine location.

One goal of the current work was to improve the prior MALDI-MS protocol for N₃-CMC-Ψ detection so as to minimize any false positive identifications of pseudouridine. To do so required monitoring the CMC derivatization reaction. Previous efforts at determining the extent of CMC modification have utilized buoyant density methods [16], gel mobility assays [17], immunoassay based techniques [3,18] and primer extension methods [18]. These approaches have been found to be cumbersome, time-consuming and limited by size and structural conformation of the molecule under study. Moreover, most of these techniques were not found to be sensitive to CMC identification at the molecular level. In the present study, MALDI-MS was used to monitor the extent of CMC modification during the derivatization protocol, and that information allowed for the optimization of CMC derivatization conditions. As a result, the protocol presented here is built upon existing RNase mapping strategies [19,20], and permits the identification of pseudouridine in RNA samples starting with low microgram amounts of material.

In the process of optimizing this CMC-based MALDI-MS protocol for N₃-CMC-Ψ identification, it was discovered that the modified nucleoside 2-methylthio-N⁶-isopentenyladenosine (ms²i⁶A), which has only been found in bacterial tRNAs [21], can also undergo derivatization with CMC. The confirmation of this modified nucleoside as another potential interference for pseudouridine detection was also explored during this work. While an improved protocol for pseudouridine identification by CMC-derivatization and MALDI-MS analysis is presented, caution must be exercised when using this experimental strategy to identify pseudouridine in bacterial transfer RNAs due to potential interferences from the previously known 4-thiouridine and the newly identified 2-methylthio-N⁶-isopentenyladenosine.

Experimental

Materials

Poly d(T)_n oligomers and the heteropolymer, r(AUGCAUGC), were obtained from the University of Cincinnati DNA Core Facility (Cincinnati, OH, USA). *Escherichia coli* tRNA^{Tyr II}, CMC metho-*p*-toluenesulfonate, 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). Snake venom phosphodiesterase (SVP) and alkaline phosphatase were obtained from Worthington Biochemical Corp. (Lakewood, NJ). C18 Ziptips were obtained from Millipore Corporation (Billerica, MA, USA). C18 Sep-pak cartridges were purchased from Waters (Milford, Massachusetts, USA).

RNase Digestion

RNase T1 was precipitated from its original solution by the use of acetone. The resulting precipitate was purified by use of C18 Sep-Pak cartridges as previously described [22]. Approximately 5 µg of tRNA was digested with ca. 250 U of RNase T1 in 2 µL of 50 mM Tris-HCl (pH 7) containing 1 mM EDTA at 37 °C for 1 - 2 h. The digest was speed-vac dried and further purified by the use of C18 ziptips prior to CMC derivatization.

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 at -20 °C for further use. 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 ziptips prior to CMC derivatization.

CMC derivatization and alkaline treatment

A stock CMC solution was prepared from 20 µg CMC metho-*p*-toluenesulfonate in 1 mL of a buffer solution containing 50 mM Tris-HCl, 4 mM EDTA and 7 M urea. The solution was vortexed and incubated in a water bath. Reaction mixtures containing 0.5 µg to 2 µg of dT₁₀ or r(AUGCAUGC) or 5 µg of RNase digested tRNA were incubated with the stock CMC solution. Incubation time, temperature, pH and the ratio of amounts of CMC to sample were varied for optimization. Prior to alkaline treatment, the CMC reacted sample was first purified by the use of ziptips and then allowed to react for 1-2 h in 50 mM NH₄HCO₃ (pH 10.4) at a temperature of 75 – 80 °C. The resulting solution was dried down and reconstituted in 2 µL of nanopure water for MALDI analysis.

HPLC Purification of RNase Digests

For the exonuclease sequencing studies only, HPLC fractionation of specific CMC-derivatized RNase digestion products was required. HPLC separations were performed on a Beckman System Gold HPLC with a flow rate of 1 mL min⁻¹ and UV detection at 254 nm. An EC 125/4 Nucleogen DEAE 60-7 column (Nest Group, MA, USA) was used. Buffer A was composed of 25 mM triethylammonium acetate, pH 6.4, and buffer B was composed of 1 M triethylammonium acetate, pH 6.8. Gradient elution from 0 to 99% B at 1 %B min⁻¹ was applied. Fractions were collected manually and then dried down and reconstituted in 10 µL of nanopure water.

Exonuclease Digestion by Snake Venom Phosphodiesterase

A 10 µL aliquot of the HPLC fractions shown by MALDI-MS to contain the endonuclease digestion product of interest was subjected to CMC derivatization. The CMC derivatized fraction was incubated for 1 h at 37 °C with 1 µL of a 0.5U/µL solution of alkaline phosphatase. This solution was further reacted with 3 – 5 µL of a 1U/µL solution of SVP, 5 µL of 100 mM ammonium acetate and 5 µL of nanopure water. The reaction was allowed to proceed at 37 °C for 2 h with 3-5 µL aliquots removed every 15 min and placed upon ice. An aliquot of 1 µL of the reaction sample was mixed with 1 µL of matrix for MALDI analysis.

Mass Spectrometry

Mass spectrometric analysis was done on 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 external calibration using dT₃ and dT₁₅ was used in all analyses. Internal calibration was performed using underivatized and unmodified RNase digestion products. The matrix solutions used were 300 mM THAP in acetonitrile and 250 mM DAHC in water. A two-layer sample spotting approach was found to be most effective. First, 0.5 µL 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

tRNA sequences were obtained from the tRNA sequence database [23]. Theoretical RNase digestion products of *E. coli* tRNA^{TyrII} were calculated using the Mongo Oligo Mass Calculator (<http://www-medstat.med.utah.edu/massspec/mongo.htm>). Signature digestion product lists were obtained from the RNAccess database (<http://bearcatms.uc.edu/rnaccess/>) [24]. MALDI peak lists were exported to Microsoft Excel for manipulation and analysis.

Results and Discussion

In our prior work, it was noted that the initial derivatization reaction with CMC did not result in the maximum number of CMC units being bonded to the starting RNA or oligoribonucleotides. This inefficiency was handled experimentally by using HPLC to isolate only the CMC-derivatized product of interest prior to mass spectral analysis. Here, the initial goal was to improve the protocol so that mixtures of oligonucleotides could be analyzed by MALDI-MS after reaction with CMC without resorting to HPLC isolation of CMC-derivatized oligonucleotides.

CMC derivatization and mass spectrometric analysis of model oligonucleotides

Ofengand has reported that the reaction of CMC with DNA or RNA is dependent on pH, time, temperature and the CMC/nucleobase ratio [11]. Initial optimization studies were conducted

using the oligodeoxynucleotide dT₁₀ (data not shown) and the oligoribonucleotide r(AUGCAUGC) (Figure 1). The homopolymer was selected to identify the maximum number of CMC units that can bond to an oligonucleotide before steric crowding becomes limiting. The heteropolymer, composed of the four major ribonucleosides, was selected to determine the effects of a mixed base sequence on the efficiency of the derivatization reaction. Previous studies have shown that adenine and cytosine do not react with CMC [1,4], while guanine, thymine and uracil are susceptible to CMC derivatization. Although the reactivity of the ribonucleosides guanosine and uridine with CMC has already been evaluated in prior studies [18], in the present investigation reaction conditions that could generate a completely derivatized product (e.g., AUGCAUGC + 4 CMC units) were sought to minimize under-derivatization of oligoribonucleotide fragments.

The pH, time, temperature and CMC/nucleobase ratio were adjusted until appropriate derivatization conditions were obtained for both model oligonucleotides (e.g., Figure 1). The mass spectra reveal a series of ions separated by increments of 252 Da with each ion reflecting the number of CMC units bonded to the appropriate oligonucleotide. Although a single reaction product containing the maximum number of CMC units was not obtained for either model oligonucleotide, these data represent the greatest extent of CMC bonding to the starting material that could be obtained in these studies without leading to thermal degradation of the original oligonucleotide.

The derivatization of the starting material to the extent shown in Figure 1 was enabled by making significant changes to prior reaction conditions [9]. These changes included altering the CMC to nucleobase ratios and increasing the reaction period. The optimal CMC/nucleobase mole ratio was found to be 8×10^3 :1. Also, the number of CMC units bonded to an oligonucleotide was monitored using MALDI-MS while increasing the reaction time, and more CMC units were found to bond to an oligonucleotide as the reaction time increased (data not shown). The present results suggest that a reaction period from 12-24 h is required to bond the maximum number of CMC units to an oligonucleotide. Studies were also conducted regarding the effects of temperature and pH on the derivatization (data not shown). As the number of CMC units bonded to the oligonucleotide was found to decrease at increasing temperatures and pH, this avenue was not explored further. Based on these initial investigations, the following conditions for CMC derivatization of short, single-stranded oligonucleotide sequences were obtained: CMC to nucleobase mole ratio of 8×10^3 starting with microgram amounts of oligonucleotide using a 50 mM Tris, 4 mM EDTA and 7 M urea, pH 8.3 buffer and incubation for 12 – 24 hours at 37 °C.

The necessary second reaction step for CMC derivatization of pseudouridine is the incubation of the CMC-modified oligonucleotide under alkaline conditions. Under appropriate alkaline conditions, the CMC group is displaced from all nucleosides except pseudouridine and thioridines [11]. Non-ideal incubation conditions would result in unmodified uridines or guanosines in RNA retaining CMC, thus leading to false positives for pseudouridine during later analysis by MALDI-MS. As before, various experimental parameters including incubation period, pH and temperature were investigated to identify appropriate conditions for releasing CMC without degrading the oligonucleotide. The optimal conditions for release of CMC from short, single-stranded, CMC-reacted oligonucleotides were found to be an incubation time of 1 – 2 h at 75-80 °C in a 50 mM NH₄HCO₃ buffer, pH 10.4.

The CMC-reacted oligonucleotide analyzed in Figure 1a was incubated at pH 10.4 using these optimized conditions and then analyzed by MALDI-MS. As noted in Figure 1b, the detected oligonucleotide molecular ion clearly indicates that the CMC moiety is displaced from the initial derivatized oligonucleotide. The incubation temperature of this step in the protocol was found to be critical in the removal of CMC adducts from r(AUGCAUGC). Although thermal

degradation of an oligonucleotide could be a source of concern at the elevated temperatures (75–80 °C) used here, no evidence of degraded products were detected in the mass spectral data.

Analysis of CMC-Derivatized RNase T1 Digestion Products of *E. coli* tRNA^{TyrII}

The conditions identified from studies of simple oligonucleotides were then used as a starting point for modifying the MALDI-MS protocol for pseudouridine identification in RNA samples. The previously published protocol reacted intact RNAs with CMC p-tosylate followed by endonuclease digestion [9]. However, CMC p-tosylate is known to exhibit a higher reactivity towards single-stranded DNA/RNA while native, double-stranded oligonucleotides are unable to react with CMC p-tosylate [18]. Therefore, in this modified protocol, intact RNAs are first digested by an endonuclease (e.g., RNase T1) to generate a mixture of single-stranded oligoribonucleotides prior to reaction with CMC p-tosylate. The model tRNA chosen in this work, *E. coli* tRNA^{TyrII}, is completely sequenced and is known to contain two pseudouridine residues, thus providing a reasonable system on which to optimize the derivatization chemistry and analysis conditions.

Figure 2a is a representative mass spectrum arising from the RNase T1 digestion of tRNA^{TyrII} that has not been reacted with CMC p-tosylate. Because matrix background can be a source of interference for compounds below 700 Da, Table 1 is limited to *E. coli* tRNA^{TyrII} RNase T1 digestion products with molecular weights greater than 900 Da. Except for [s⁴U]CCCCp at *m/z* 1905 and the 3'-terminal fragment, AAUCCUCCCCCACCACCA-(OH), the complete RNase T1 map of tRNA^{TyrII} could be detected including the two pseudouridine-containing fragments, [m⁵U]ΨCGp (*m/z* 1293) and ACUQUA[ms²-i⁶A]AΨCUGp (*m/z* 4097) (modified nucleosides are identified in Table 1).

A number of other ions were also detected that did not correspond to missed cleavages during the RNase T1 digestion. Several could be assigned as contaminating *E. coli* tRNAs through identification of their signature digestion products [24]. The contaminating tRNAs were found to be tRNA^{Trp} (signature ions: *m/z* 1638, 1888, 2500, 2553 and 3944), tRNA^{Ser} (signature ions: *m/z* 1349, 2289 and 2402) and tRNA^{His} (signature ion: *m/z* 2425). It is apparent that this commercial *E. coli* tRNA^{TyrII} sample also contains tRNA^{Trp}, tRNA^{Ser} and tRNA^{His}, at a minimum, which is not surprising given the difficulty of purifying individual tRNAs [24]. These interfering tRNAs do not affect the ability to examine and optimize CMC derivatization conditions.

This pseudouridine identification protocol involves derivatization of RNase T1 digestion products with CMC and then releasing CMC from all sites except thiouridines and N-3 of pseudouridine, if present, by mild alkaline treatment. The resulting mixture is then analyzed by MALDI-MS to identify pseudouridine residues in RNase T1 digestion products by the characteristic 252 u mass shift from the unreacted RNase T1 digestion products (i.e., Figure 2a). Because this mixture contains a number of components including oligonucleotides, excess CMC, buffers and base, the quality of the MALDI data can be adversely affected. Experimental studies focused on sample purification, MALDI matrices and sample/matrix preparation were conducted to identify conditions most appropriate for the MALDI analysis of CMC-derivatized RNase digestion products.

Sample purification is important for successful MALDI-MS analysis as the CMC derivatization is pH and buffer sensitive, so great care must be taken to remove any residual effects from the preceding reactions in the multi-step derivatization. Spectral quality of CMC derivatized samples also is diminished by solution components such as salts, buffers and excess CMC. To obtain high quality MALDI data, these substances need to be removed from the sample prior to analysis. Several approaches for sample purification were investigated including the use of

centrifuge filters, dialysis membranes and C18 Ziptips. Among these options, C18 Ziptips were found to be most effective at sample purification including removal of salts, excess CMC and other residual solvents in the reaction mixture.

Successful CMC derivatization and MALDI analysis were found to require the following sample purification steps. First, the initial RNase digest was ziptipped to remove the enzyme, and the digestion products were then reacted with CMC p-tosylate as described. After CMC derivatization, the reaction mixture was ziptipped a second time to remove any unreacted CMC p-tosylate in the solution in addition to the buffer components before incubating the digestion products with base. After alkaline treatment, the mixture volume was reduced by evaporation, the sample was reconstituted in water and then spotted with matrix for MALDI analysis using the two-layer sample spotting approach. Although sample loss is a source of concern in the use of Ziptips for purification, it was determined that ~5 µg of starting material (i.e., tRNA subjected to RNase digestion) was sufficient to allow for this CMC derivatization strategy with successful MALDI-MS identification of CMC-derivatized pseudouridine residues.

In addition to determining appropriate sample purification and spotting procedures, the MALDI matrices THAP and 3-hydroxypicolinic acid (3-HPA) were examined to identify the more appropriate matrix for CMC-derivatized oligonucleotides. After examining these matrices in both polarities and with the time-of-flight operating in both linear and reflectron mode, it was found that THAP in negative polarity and reflectron mode provides the best data quality for CMC-derivatized oligonucleotides. The stability of CMC-derivatized oligonucleotides to reflectron mode TOF-MS allowed for high resolution and high mass measurement accuracy (using internal calibration) to facilitate RNase mapping of the digest products. These instrumental conditions were also sufficient to obtain representative spectra of RNase digestion products from tRNA amounts compatible with the CMC derivatization conditions.

Figure 2b is a representative mass spectrum arising from the RNase T1 digestion of tRNA^{TyrII} after reaction with CMC and alkaline treatment. Optimal reaction conditions utilized for the CMC modification of RNase T1 fragments of tRNA^{TyrII} were found to be very similar to those conditions used for the CMC reactions with dT₁₀ and r(AUGCAUGC). Table 1 includes those ions appearing in Figure 2b which are integer multiples of 252 Da from any ions detected in Figure 2a as those represent CMC-derivatized digestion products stable to alkaline treatment. As expected, the two known pseudouridine containing RNase T1 fragments of tRNA^{TyrII} were found to increase in *m/z* due to the CMC derivatization reaction. The ion corresponding to m⁵UΨCGp + 1 CMC unit (*m/z* 1544) was detected, with no evidence found here of an additional CMC unit due to the presence of 5-methyluridine (m⁵U) illustrating the alkaline treatment conditions were appropriate to release CMC from 5-methyluridine [11].

The second pseudouridine containing RNase T1 fragment, ACUQUA[ms²i⁶A]AΨCUGp, was detected as both the 2',3'-cyclic phosphate (>p) and linear phosphate species with one CMC unit (*m/z* 4332 and 4350, respectively) as well as a second ion corresponding to the presence of an additional CMC unit (*m/z* 4601). The attachment of an additional CMC suggests that a second nucleoside may also retain CMC under alkaline conditions. It has already been demonstrated in this work that CMC is not retained on unmodified guanosines and uridines under these reaction conditions, and previous studies have shown that CMC does not react with the unmodified nucleobases, adenine and cytosine. It is unlikely the second CMC unit remains on pseudouridine as the [m⁵U]ΨCGp digestion product in the same reaction showed no evidence of two CMC units. The reaction of CMC p-tosylate with modified bases such as queuosine and 2-methylthio-*N*⁶-isopentenyladenosine has not been studied. Thus, it is possible that a second CMC unit has reacted with either the queuosine (Q) located at position 35 on the RNA sequence or the 2-methyl-*N*⁶-isopentenyladenosine (ms²i⁶A) at position 38.

As illustrated in the preceding data, this present protocol is built upon existing RNase mapping strategies [19,20], and improved conditions for CMC derivatization of oligonucleotides were obtained by first digesting the RNA with an RNase prior to CMC derivatization to ensure that the RNA secondary structure does not limit the derivatization process. Starting from approximately five micrograms of tRNA (~200 pmol), an RNase map of underivatized product (the control) and the subsequent analysis of CMC-derivatized product is possible. Although sample losses reduce the signal-to-noise ratio in the final mass spectrum of the CMC-derivatized product, sufficient information is obtained to identify those RNase digestion products which increase in m/z by an integer multiple of 252, allowing one to either confirm or reveal the presence of pseudouridine in the original RNA sample. However, as noted above, these data do suggest that 4-thiouridine and pseudouridine are not the only post-transcriptionally modified nucleosides that can retain CMC after alkaline treatment. To further clarify if queuosine or 2-methylthio- N^6 -isopentenyladenosine also retain CMC after alkaline treatment (thus potentially interfering with the identification of pseudouridine), additional studies were conducted.

Analysis of CMC Derivatized RNase A Digestion Products of *E. coli* tRNA^{TyrII}

To obtain more information on the selectivity of CMC derivatization of specific post-transcriptional modifications such as 2-methylthio- N^6 -isopentenyladenosine and queuosine, tRNA^{TyrII} was digested with RNase A. RNase A selectively cleaves RNAs after pyrimidine residues and has been found to cleave after 5-methylcytidine, 5-methyluridine, dihydrouridine, pseudouridine and 4-thiouridine in tRNA [12]. Thus, for the tRNA^{TyrII} sample, the 2-methylthio- N^6 -isopentenyladenosine (ms^2i^6A) at position 38 should be detected in the RNase A digestion product (37)-A[ms^2i^6A]AΨ-(39) while the queuosine (Q) residue at position 35 will not be detected because it would be present in a dimer (5'-QU-3') whose m/z value is too low to accurately characterize with these MALDI conditions. If one CMC unit remains on the RNase A fragment (37)-A[ms^2i^6A]AΨ-(39) after derivatization and alkaline incubation, then 2-methylthio- N^6 -isopentenyladenosine can be ruled out as a site of derivatization with CMC, while if two CMC units are detected, then it is most likely that this modified nucleoside also retains a CMC unit under alkaline conditions.

Using the same approach described above for RNase T1, tRNA^{TyrII} was digested with RNase A before subjecting the sample to CMC derivatization. Figure 3a is a representative mass spectrum arising from the RNase A digestion of tRNA^{TyrII} that has not been reacted with CMC. The detected m/z values and their tRNA^{TyrII} sequence assignments are listed in Table 2. RNase A mediated signature digestion products for contaminating tRNAs were also detected. These were observed at m/z values of 1751.9 (Trp), 2014.8 (Gly I), 1982.8 (Ser II) and 2690.6 (Ser V). This information confirms the identity of the contaminating tRNAs detected during RNase T1 analysis and adds tRNA^{GlyI} to the list of contaminating tRNAs.

Figure 3b presents the mass spectrum of the RNase A digestion products after CMC derivatization and alkaline incubation. Table 2 includes those ions appearing in Figure 3b which are integer multiples of 252 u from any ions detected in Figure 3a. Upon CMC derivatization and alkaline incubation, the 2-methylthio- N^6 -isopentenyladenosine containing RNase A digestion product, A[ms^2i^6A]AΨp (m/z 1424.0), was found to bond to a maximum of 2 CMC units as noted by the ion at m/z 1926.9 in Figure 3b. Such results support the conclusion that 2-methylthio- N^6 -isopentenyladenosine is another post-transcriptionally modified nucleoside that retains a CMC unit after derivatization and alkaline incubation, and these findings are consistent with the results obtained from the corresponding RNase T1 digest analysis.

A further examination of the data obtained in Figure 3a reveals the tRNA^{Trp} signature digestion product A[ms^2i^6A]AACp at m/z 1751.9. As this RNase A product does not contain any guanosines, uridines or pseudouridines, it would not be expected to undergo reaction with CMC

unless 2-methylthio-*N*⁶-isopentenyladenosine is indeed another nucleoside amenable to derivatization by CMC. As found in Figure 3b, an ion at *m/z* 2002.8 is detected, which corresponds to a single CMC unit bonded to A[ms²i⁶A]AACp. Similar data was also found in the RNase T1 digestion analysis (Table 1). Table 3 summarizes the data obtained from the CMC derivatization of all fragments known to contain 2-methylthio-*N*⁶-isopentenyladenosine in both RNase digests. These data consistently report 2-methylthio-*N*⁶-isopentenyladenosine as a modified nucleoside that can retain CMC after alkaline treatment.

Exonuclease Digestion of Selected RNase Digestion Products

For additional verification that the CMC unit resides on 2-methylthio-*N*⁶-isopentenyladenosine, exonuclease digestion using snake venom phosphodiesterase (SVP) of CMC-derivatized 2-methylthio-*N*⁶-isopentenyladenosine containing RNase digestion products was performed. SVP sequentially hydrolyzes oligonucleotides in the 3' to 5' direction. Sequence information is obtained from the generation of mass ladders and subsequent measurement of the mass differences between the consecutive mass ladder ions [20].

The RNase A signature digestion product ion of tRNA^{Trp}, A[ms²i⁶A]AACp, at *m/z* 1752 was selected for exonuclease sequencing. First, the RNase A digestion product containing A[ms²i⁶A]AACp was fractionated by HPLC and then subjected to CMC derivatization. After derivatization, alkaline phosphatase was added to the reaction mixture for removal of the 3'-phosphate group followed by the addition of SVP for sequential cleavage from the 3'-terminus of the oligonucleotide. Figure 4 is the MALDI mass spectra of the CMC-derivatized A[ms²i⁶A]AACp, and co-eluting RNase products, after sequencing.

Complete sequencing of this oligonucleotide by SVP was not possible; however the information obtained is consistent with that found from the RNase mapping experiments. Sequencing of this oligonucleotide does not proceed past 5'-A[ms²i⁶A]A-3', for either the underivatized or CMC derivatized oligonucleotide. Sequence analysis does reveal that the CMC adduct must be present on one of these three nucleosides and it was not present on the 3'-terminal cytosine or adenosine residues. As it has already been established that adenosine does not react with CMC p-tosylate [1,4], the data obtained from SVP digestion supports the conclusion that 2-methylthio-*N*⁶-isopentenyladenosine is the location of CMC derivatization. It is possible that 2-methylthio-*N*⁶-isopentenyladenosine causes a stalling of SVP digestion as a previous report has shown that steric interference plays a significant role in stalling SVP digestion [25]. Enzyme stalling by this bulky modification would explain why both the underivatized and CMC-derivatized digestion products yielded similar sequence information. Thus, in addition to 4-thiouridine [11], it appears that 2-methylthio-*N*⁶-isopentenyladenosine is another modified nucleoside that can retain the CMC unit even after exposure to alkaline conditions.

Conclusions

An optimized CMC derivatization approach has been successfully implemented for the detection of pseudouridines in RNA. These derivatization conditions should also prove to be amenable with pseudouridine detection via the CMC-reverse transcriptase approach. These optimized conditions are compatible with MALDI-MS analysis of CMC-derivatized RNase digests, and do not require HPLC isolation of CMC-derivatized RNase digestion products for identification of pseudouridine. During the course of this optimization study, it was also determined that the post-transcriptional modification 2-methylthio-*N*⁶-isopentenyladenosine also retains a CMC unit even after incubation of CMC-derivatized oligoribonucleotides with base. As 2-methylthio-*N*⁶-isopentenyladenosine is currently only known to be present in bacterial tRNAs, care must be exercised when utilizing this approach for pseudouridine identification of such tRNA samples. Future efforts are focused towards the application of this

improved methodology for pseudouridine detection and screening from complex mixtures of tRNAs.

Acknowledgements

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

References

1. Ho NWY, Gilham PT. *Biochemistry* 1971;10:3651. [PubMed: 4328867]
2. Naylor R, Ho NWY, Gilham PT. *J Am Chem Soc* 1965;87:4209. [PubMed: 4284810]
3. Wani AA, Sullivan JK, Lebowitz J. *Nucleic Acids Res* 1989;17:9957. [PubMed: 2602147]
4. Lebowitz J, Chaudhuri AK, Gonenne A, Kitos G. *Nucleic Acids Res* 1977;4:1695. [PubMed: 896472]
5. Flashner MS, Katopes MA, Lebowitz J. *Nucleic Acids Res* 1977;4:1713. [PubMed: 331264]
6. Bakin A, Ofengand J. *Biochemistry* 1993;32:9754. [PubMed: 8373778]
7. Bakin A, Lane BG, Ofengand J. *Biochemistry* 1994;33:13475. [PubMed: 7947756]
8. Hansen MA, Kirpekar F, Ritterbushch W, Vester B. *RNA* 2002;3:202. [PubMed: 11911366]
9. Patteson KG, Rodicio LP, Limbach PA. *Nucleic Acids Res* 2001;29:e49. [PubMed: 11353094]
10. Campo DL, Recinos C, Yanez G, Pomerantz SC, Guymon R, Crain PF, McCloskey JA, Ofengand J. *RNA* 2005;11:210. [PubMed: 15659360]
11. Ofengand J, Campo DM, Kaya Y. *Methods* 2001;25:365. [PubMed: 11860291]
12. Mengel-Jorgensen J, Kirpekar F. *Nucleic Acids Res* 2002;30:e135. [PubMed: 12466567]
13. Emmerechts G, Herdewijn P, Rozenski J. *J Chrom B* 2005;825:233.
14. Pomerantz SC, McCloskey JA. *Anal Chem* 2005;77:4687. [PubMed: 16053277]
15. Buchhaupt M, Peifer C, Entian K. *Anal Biochem* 2007;361:102. [PubMed: 17181990]
16. Hale P, Woodward RS, Lebowitz J. *J Biol Chem* 1983;258:7828. [PubMed: 6305957]
17. Novack DF, Casna NJ, Fischer SG, Ford JP. *Proc Natl Acad Sci USA* 1986;83:586. [PubMed: 3003741]
18. Wani AA, Yamasaki EF. *Biochim Biophys Acta* 1991;1088:259. [PubMed: 2001400]
19. Kirpekar F, Douthwaite S, Roepstorff P. *RNA* 2000;6:296. [PubMed: 10688367]
20. Limbach PA. *Mass Spectrom Rev* 1996;15:297.
21. Limbach PA, Crain PF, McCloskey JA. *Nucleic Acids Res* 1994;22:2183. [PubMed: 7518580]
22. Meng Z, Limbach PA. *Anal Chem* 2005;77:1891. [PubMed: 15762601]
23. Sprinzl M, Vassilenko KS. *Nucleic Acids Res* 2005;33:D139. [PubMed: 15608164]
24. Hossain M, Limbach PA. *RNA* 2007;13
25. Driscoll MD, Sathya G, Saidi LF, DeMott MS, Hilf R, Bambara RF. *Molecular Endocrinology* 1999;13:958. [PubMed: 10379894]

Figure 1.

(a) MALDI mass spectrum of $r(\text{AUGCAUGC})$ after derivatization with CMC p-tosylate in 50 mM Tris, 4 mM EDTA, 7M urea (pH 8.3) overnight at 37 °C. The number of CMC adducts are noted on each mass spectrum. (b) MALDI mass spectrum of $r(\text{AUGCAUGC})$ after incubating the sample in Figure 1a in 50 mM NH_4OH (pH 10.4) for 60 min at 80 °C. These incubation conditions allow for the complete release of CMC without degrading the oligonucleotide.

Figure 2.

(a) MALDI mass spectrum of *E. coli* tRNA^{TyrII} digested with RNase T1. (b) MALDI mass spectrum of the CMC-derivatized RNase T1 digest of *E. coli* tRNA^{TyrII} after alkaline treatment. Table 1 lists the RNase T1 digestion products including those which retain CMC after alkaline treatment along with tRNA signature digestion products identified from contaminating *E. coli* tRNAs in the sample.

Figure 3.

(a) MALDI mass spectrum of *E. coli* tRNA^{TyrII} digested with RNase A. (b) MALDI mass spectrum of the CMC-derivatized RNase A digest of *E. coli* tRNA^{TyrII} after alkaline treatment. Table 2 lists the RNase A digestion products including those which retain CMC after alkaline treatment along with tRNA signature digestion products identified from contaminating *E. coli* tRNAs in the sample.

Figure 4. MALDI mass spectrum of the products arising from a 2 h snake venom phosphodiesterase digestion of the RNase A digestion product, A[ms²i⁶A]AACp, from *E. coli* tRNA^{TyrII} after CMC derivatization and alkaline treatment. Underivatized A[ms²i⁶A]AACp is also present in this sample.

Table 1

RNase T1 digestion products of *E. coli* tRNA^{TyrII} arising from data in Figure 2. Sequences listed are those that map to the known sequence of tRNA^{TyrII} or tRNA signature products [24] for contaminating *E. coli* tRNAs. '>p' represent 2'-3' cyclic phosphate digestion products. Modified nucleosides are defined below.

tRNA^{TyrII} sequence: 5'pGGUGGGG[s⁴U]UCCCGAGC[Gm]GCCAAAGGGAGCAGACUQUA[ms²i⁶A]AΨCUGCCG UCAACAGACUUCGAAGG[m⁵U]ΨCGAAUCCUUCUUUUUACCAACCA(OH)-3'

tRNA ^{TyrII} RNase T1 fragment	Calc <i>m/z</i>	(Fig. 2a) Exp <i>m/z</i>	(Fig. 2b) After Reaction with CMC (n = # CMC units)
CCGp	972.2	972.3	
CAGp	996.2	996.2	
AAG>p	1002.2	1002.2	
AAGp	1020.2	1020.2	
C[Gm]Gp	1026.2	1026.2	
m ³ UΨCG>p	1275.1	1275.0	
m ³ UΨCGp	1293.1	1293.0	1544.3 (n = 1)
ACUUCG>p	1895.2	1894.6	
ACUUCGp	1913.2	1913.0	2165.7 (n = 1)
UCACAGp	1936.3	1935.9	
CCAAAGp	1959.3	1959.2	
ACUQUA[ms ² i ⁶ A]AΨCUG>p	4079.6	4080.8	4332.6 (n = 1)
ACUQUA[ms ² i ⁶ A]AΨCUGp	4097.6	4097.6	4350.9 (n = 1); 4601.0 (n = 2)
RNase T1 Signature Ions (tRNA)	Cal <i>m/z</i>	(Fig. 2a) Exp <i>m/z</i>	(Fig. 2b) After Reaction with CMC (n=#CMC units)
AAA _{Gp} (Ser)	1349.2	1349.1	
U[m ⁷ G]UUGp (Trp)	1639.2	1638.6	
CCCCCGp (Trp)	1887.3	1887.5	
UCAAAG _{Gp} (Ser)	2289.3	2289.3	
A[ms ² i ⁶ A]AAACCGp (Ser)	2402.3	2402.3	2653.8 (n = 1)
CCACCCCA(OH) (His)	2425.4	2425.2	
UCUCUCCGp (Trp)	2500.3	2500.4	
UUCAADDGp (Trp)	2553.3	2553.2	2805.5 (n = 1)
U[Cm]UCCA[ms ² i ⁶ A]AAACCGp (Trp)	3943.6	3943.6	4195.7 (n = 1)

Modified Nucleosides

Ψ: pseudouridine

Cm: 2'-*O*-methylcytidine

D: Dihydrouridine

Gm: 2'-*O*-methylguanosine

m⁷G: 7-methylguanosine

I: inosine

ms²i⁶A: 2-methylthio-*N*⁶-isopentenyladenosine

Q: queuosine

m⁵U: 5-methyluridine

s⁴U: 4-thiouridine

Table 2

RNase A digestion products of *E. coli* tRNA^{TyrII} arising from data in Figure 3. Sequences listed are those that map to the known sequence of tRNA^{TyrII} or tRNA signature products [24] for contaminating *E. coli* tRNAs. '>p' represent 2'-3' cyclic phosphate digestion products. Modified nucleosides are defined in Table 1.

tRNA ^{TyrII} RNase A fragment	Calc <i>m/z</i>	(Fig. 3a) Exp <i>m/z</i>	(Fig. 3b) After Reaction with CMC (n = # CMC units)
GGU>p	995.1	995.1	
G[Gm]C>p	1008.1	1007.9	
G[Gm]Cp	1026.1	1026.1	
GAAU>p	1308.1	1308.0	
AGACp	1325.2	1324.9	
A[ms ^{2,6} A]AΨ>p	1406.0	1406.0	1657.6 (n = 1); 1909.9 (n = 2)
A[ms ^{2,6} A]AΨp	1424.0	1424.0	1675.6 (n = 1); 1926.9 (n = 2)
GGGG[s ⁴ U]p	1719.0	1718.8	minor peak detected at 1970.9
GAAGGm ⁵ Up	2030.2	2029.8	
AAAGGGAGCp	3018.4	3018.4	
RNase A Signature Ions (tRNA)	Calc <i>m/z</i>	(Fig. 3a) Exp <i>m/z</i>	(Fig. 3b) After Reaction with CMC (n=#CMC units)
A[ms ^{2,6} A]AACp (Trp)	1752.3	1751.9	2002.8 (n = 1)
AAAAGCp (Ser)	1983.3	1982.8	
GAGAGCp (Gly)	2015.3	2014.8	
GAAAGGUp (Ser)	2690.3	2690.6	

Table 3

Summary of RNase digestion products containing ms²i⁶A. “n” depicts the number of CMC units bonded to the oligonucleotide. Modified nucleosides are defined in Table 1.

Endonuclease	ms ² i ⁶ A Sequence and tRNA	Control <i>m/z</i>	CMC-derived <i>m/z</i>
RNase A	A[m ² i ⁶ A]A~p; tRNA ^{TyrII}	1424.0	1675.6 (n=1) 1926.9 (n=2)
RNase T1	ACUQUA[ms ² i ⁶ A]A~CUGp; tRNA ^{TyrII}	4097.6	4350.9 (n = 1) 4601.0 (n = 2)
RNase A	A[ms ² i ⁶ A]AACp; tRNA ^{Trp}	1751.9	2002.8 (n=1)
RNase T1	U[Cm]UCCA[ms ² i ⁶ A]AACCGp; tRNA ^{Trp}	3943.6	4195.6 (n=1)
RNase A	A[ms ² i ⁶ A]AACp; tRNA ^{Ser}	1751.9	2002.8 (n=1)
RNase T1	A[ms ² i ⁶ A]AACCGp; tRNA ^{Ser}	2402.3	2653.8 (n=1)