446–451 *Nucleic Acids Research, 1998, Vol. 26, No. 2*

Sequencing RNA by a combination of exonuclease digestion and uridine specific chemical cleavage using MALDI-TOF

D. A. Tolson* and N. H. Nicholson

SmithKline Beecham Pharmaceuticals, New Frontiers Science Park (North), Coldharbour Road, The Pinnacles, Harlow, Essex CM19 5AD, UK

Received October 10, 1997; Revised and Accepted November 21, 1997

ABSTRACT

The determination of DNA sequences by partial exonuclease digestion followed by Matrix-Assisted Laser Desorption Time of Flight Mass Spectrometry (MALDI-TOF) is a well established method. When the same procedure is applied to RNA, difficulties arise due to the small (1 Da) mass difference between the nucleotides U and C, which makes unambiguous assignment difficult using a MALDI-TOF instrument. Here we report our experiences with sequence specific endonucleases and chemical methods followed by MALDI-TOF to resolve these sequence ambiguities. We have found chemical methods superior to endonucleases both in terms of correct specificity and extent of sequence coverage. This methodology can be used in combination with exonuclease digestion to rapidly assign RNA sequences.

INTRODUCTION

Relating of the function of RNA to its structure requires knowledge of the primary sequence. Current methods of RNA sequencing include the ‘Sanger’ chain termination, where the RNA is first transcribed into cDNA with reverse transcriptase (1), and polyacrylamide gel electrophoresis of chemical (2) or enzymatic (3,4) digests of radiolabelled RNA. These methods, which are inherently indirect, are prone to errors and are time consuming and labour intensive. Mass spectrometric techniques such as MALDI and Electrospray provide a more direct, specific and rapid means for the analysis of nucleic acids (5). Provided that sufficient mass accuracy and resolution can be obtained, digest fragments can be identified directly on the basis of their mass. Oligonucleotide sequencing by Exonuclease Digestion and MALDI-TOF has been reported (6,7). Unambiguous assignment of DNA fragments is possible due to the minimum 9 Da mass difference between the nucleotides dA and dT, which is readily resolved by the MALDI-TOF instrument. When the same procedure is applied to large (>5000 Da) RNA molecules, it is not possible to unambiguously assign the nucleotides U and C (1 Da mass difference) due to limitations in the resolving power of the MALDI-TOF instrument. Direct tandem mass spectrometric methods would suffer from the same limitation unless complex and expensive Fourier-Transform mass spectrometers were employed. In order to generate digest fragments that would be characteristic for either U or C residues, we first explored the possibility of using commercially available base-specific endonucleases such as RNase CL3 and PhyM. Whilst our investigations were underway, similar studies on endonuclease digests were recently published (8). In common with many of their findings, we found that the specificity of the enzymes was not as expected, or depended on the precise reaction conditions (9). An alternative approach was to use direct chemical methods (2,10). Aqueous hydrazine reacts with pyrimidine bases of RNA in the order U>>C (2). The reaction occurs via nucleophilic addition to the 5,6 double bond, and the sites of hydrazinolysis along the RNA strand are then vulnerable to aniline attack (10). Treatment of RNA with aqueous hydrazine, followed by aniline, induces scission at uridines. Here we report our experiences using U/C specific endonuclease digestion and chemical derivatisation methods followed by MALDI-TOF in order to distinguish between the pyrimidine bases of RNA. These results confirm and expand on the findings of others based solely on enzymatic approaches (8), and demonstrate that chemical methods provide a rapid and simple method of unambiguously assigning U and C residues in RNA.

MATERIALS AND METHODS

Materials

RNA was prepared by automated chemical synthesis on an ABI 394 DNA synthesiser using standard protocols (Applied Biosystems Division, Perkin Elmer) with 5′-*O*-dimethoxytrityl-2′-*O*-t-butyldimethylsilyl protected *N*,*N*-diisopropylamino-(2-cyanoethoxy)phosphoramidites protected at the exocyclic amine with a t-butylphenoxyacetyl moiety and 500 A CPG 1 µmol supports derivatised with protected nucleosides (PerSeptive Biosystems, Expedite). Tetrazole activator and other reagents were purchased from Applied Biosystems. Assembled sequences were deprotected with concentrated ammonium hydroxide:ethanol (3:1; 2 ml) at 55°C for 15 min and, after evaporation, the 2′-*O*-t-BDMS was removed with triethylamine trihydrofluoride: dimethylformamide

*To whom correspondence should be addressed. Tel: +44 1279 627888; Fax: +44 1279 627896; Email: David_A_Tolson.@sbphrd.com

(3:1; 2 ml) at 55°C for 1.5 h. The reaction mixture was treated with 0.05 M triethylammonium acetate (0.15 ml), made up to 20 ml with butan-1-ol and precipitated at –20°C overnight. The RNA was collected by centrifugation and purified by ion exchange chromatography (NucleoPac PA-100, Dionex) eluting with a gradient of 25–55% sodium perchlorate in aqueous solution with 25% formamide and 0.02 M Tris at pH 6.8. Pure RNA was dialised (Spectra/Por 6, Medicell International) against water and lyophylised.

Ribonuclease CL3 was purchased from Boeringher Mannheim, Ribonuclease Phy M and pdA12-18 were purchased from Pharmacia Biotech, 3-Hydroxypicolinic acid (3-HPA) was purchased from Aldrich. Snake venom phosphodiesterase (SVP), bovine spleen phosphodiesterase (BSP), hydrazine (55% aqueous soultion), aniline and glycine were purchased from Sigma. Dowex 50 W X 8 cation-exchange resin beads (100–200 mesh) were purchased from Fluka and converted to the ammonium form as described in Ref. 11.

Methods

Ethanol precipitation of RNA. 2.5 vol of 95% ethanol and 0.1 vol of 10 M ammonium acetate was added to 1 vol of RNA solution. The mixture was then chilled at –20°C for 1 h. The mixture was then centrifuged at 12 000 *g* for 30 min. The ethanol supernatant was removed and the precipitate washed with a further 2.5 vol of cold 95% ethanol and recentrifuged at 12 000 *g* for 1 min. The ethanol supernatant was then removed and the sample dried in a stream of argon.

Exonuclease digestion of DNA/RNA. The procedures used for snake venom phosphodiesterase (SVP) and bovine spleen phosphodiesterase (BSP) digests were as described in reference 7. Aliquots of 1 µl were taken at various time intervals for MALDI analysis.

Ribonuclease CL3 digest of RNA. A 3 µl sample of Ribonuclease CL3 (0.02 U/µl) was added to 5 µl of RNA solution (50 pmol/µl) and incubated at 37°C. Aliquots of 1 µl were taken at various time intervals for MALDI analysis.

Ribonuclease Phy M digests. A 2 µl sample of Ribonuclease Phy M (1 U/µl) was added to 2 µl of pH 5.0 ammonium citrate buffer (5 g/100 ml) and 5 µl of RNA solution (50 pmol/µl) and incubated at 55°C. Aliquots (1 µl) were taken at various time intervals for MALDI analysis.

Hydrazine/aniline treatment of RNA. RNA solution (10 µl of 50 pmol/ul) was added to 20 µl of 55% hydrazine (aq) and the mixture was incubated on ice for 30 min. The solution was then ethanol precipitated. The precipitate was taken up in 10 µl of aniline mixture (20 µl aniline plus 12 µl of glacial acetic acid plus 186 µl of water) and incubated for 30 min at 60°C in the dark. The sample was then dried in a vacuum centrifuge and redissolved in 5 µl of water. Aliquots (1 µl) were then taken for MALDI-TOF analysis.

Sample preparation for MALDI-TOF. Sample solution (1 µl) was mixed with 1 µl of 3-HPA solution (75 mg/ml in 1:1 acetonitrile:water) and 1 µl of an aqueous suspension of cation exchange beads (ammonium form) directly on the MALDI target. The mixture was then allowed to dry in a stream of warm air.

Instrumentation. All MALDI-TOF mass spectra were acquired in negative ion linear mode on a TofSpec SE instrument (Micromass, UK) equiped with time-lag focussing using 20 kV accelerating voltage and a sampling rate of 500 MHz. The data was acquired and processed using Masslynx V2.22 software. The instrument was calibrated using pdA12-18.

RESULTS

Exonuclease digestion

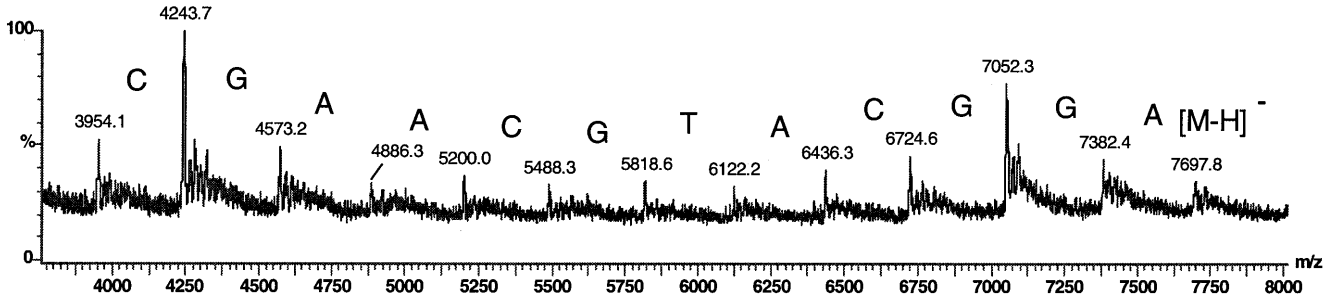
Figure 1 shows the MALDI-TOF mass spectrum of a 25mer DNA subjected to BSP digestion for 1 min. A ladder sequence with fragments containing the original 3′-terminus is generated. The sequence from the 5′-end can be determined by the mass difference between successive peaks as shown. This is possible due to the minimum 9 Da mass difference between the nucleotides dA and dT, which is readily resolved by the MALDI-TOF instrument. By continuing the digest for longer time periods, further sequence information can be obtained. In principle, full sequence information could be obtained from the BSP digest alone, however, smaller fragments (<1000 Da) are not readily observed, either due to incomplete digestion or interference from the matrix. SVP produces similar ladder sequences (see Fig. 2) from the 3′-end. By combining the data from the two digests full sequence information can be obtained. Figure 2 shows the MALDI-TOF mass spectra of a 31mer RNA subjected to SVP digestion for 2 min. A ladder sequence with fragments containing the original 5′-terminus is generated. Sequence information can be derived in a similar manner to that described for DNA above. However, because the U and C residues differ in mass by only 1 Da, the mass of fragments containing these residues cannot be determined with the necessary accuracy, and the resulting assignments therefore contain partial ambiguities where the U and C residues are described as ‘U/C’ as shown.

Endonuclease digestion

Ribonuclease CL3 digest (C specific). Figure 3 shows the MALDI-TOF mass spectrum of the same 16mer RNA subjected to ribonuclease CL3 digestion for 1 min. The major fragment results from cleavage between U(12) and A(13), followed by cleavage at C(3) from this fragment. The masses of the fragments correspond to those containing a 2′–3′ cyclic phosphate at the 3′-end as others have observed (8). This preferential cleavage at U is in contrast to the expected C specificity, and was also observed in the digests of other RNA’s containing the sequence UAG (data not shown).

Ribonuclease Phy M digest (U, A and G specific). Figure 4 shows the MALDI-TOF mass spectrum of a 23mer RNA subjected to ribonuclease Phy M digestion for 2 min. The major fragments result from cleavage at U(13), G(14) and A(17). Fragments containing the original 5′-end (with a 2′–3′ cyclic phosphate) and fragments containing the original 3′-end are observed. The cleavage specificity is in agreement with that published for this enzyme in the absence of urea (14). Only a small proportion of the possible cleavage sites were cleaved, limiting the usefulness of this enzyme for sequence determination. Partial digestion was also observed with other RNA’s (data not shown).

a



b

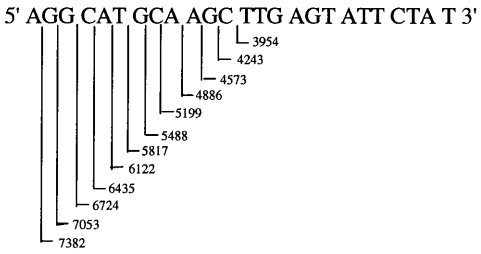
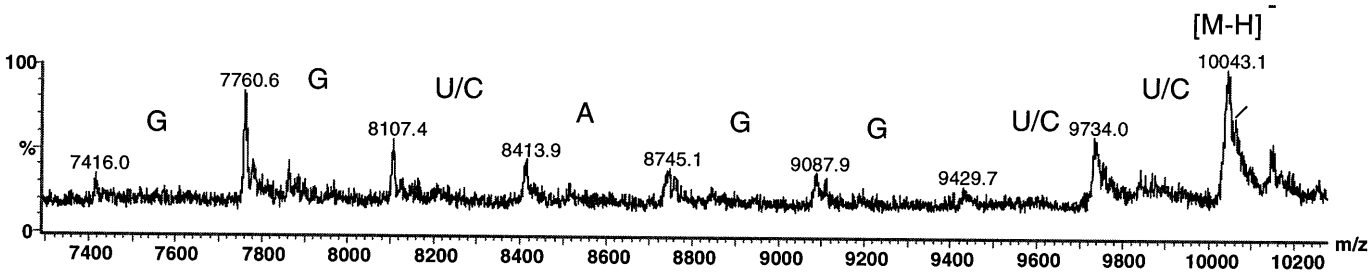


Figure 1. (**a**) Negative ion UV-MALDI mass spectrum of a synthetic 25mer DNA (average molecular weight 7696 Da) subjected to BSP digestion for 1 min. (**b**) Schematic representation of observed cleavages.

a



b

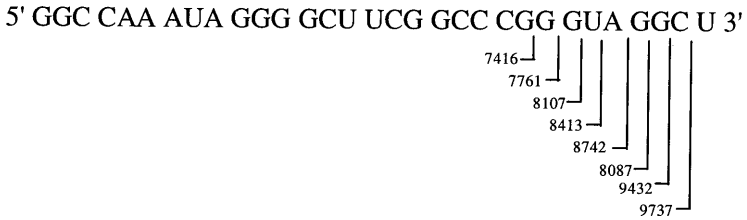


Figure 2. (**a**) Negative ion UV-MALDI mass spectrum of a synthetic 31mer RNA (average molecular weight 10044 Da) subjected to SVP digestion for 2 min. (**b**) Schematic representation of observed cleavages.

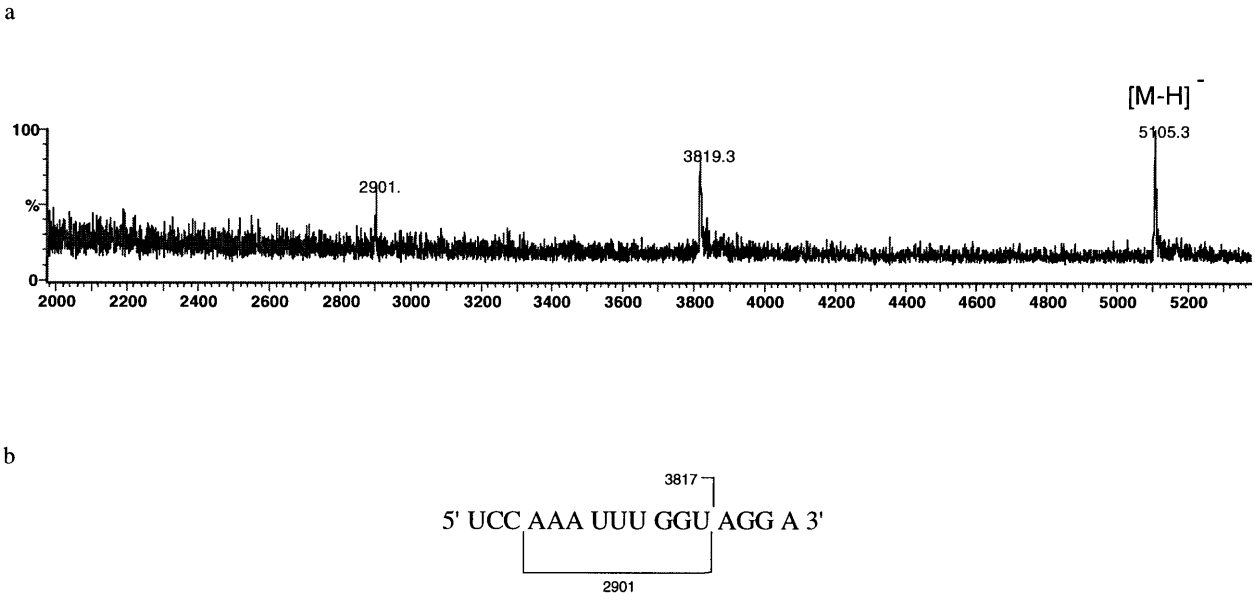


Figure 3. (**a**) Negative ion UV MALDI spectra of a synthetic 16mer RNA (average molecular weight 5106 Da) subjected ribonuclease CL3 digestion for 1 min. (**b**) Schematic representation of observed fragments.

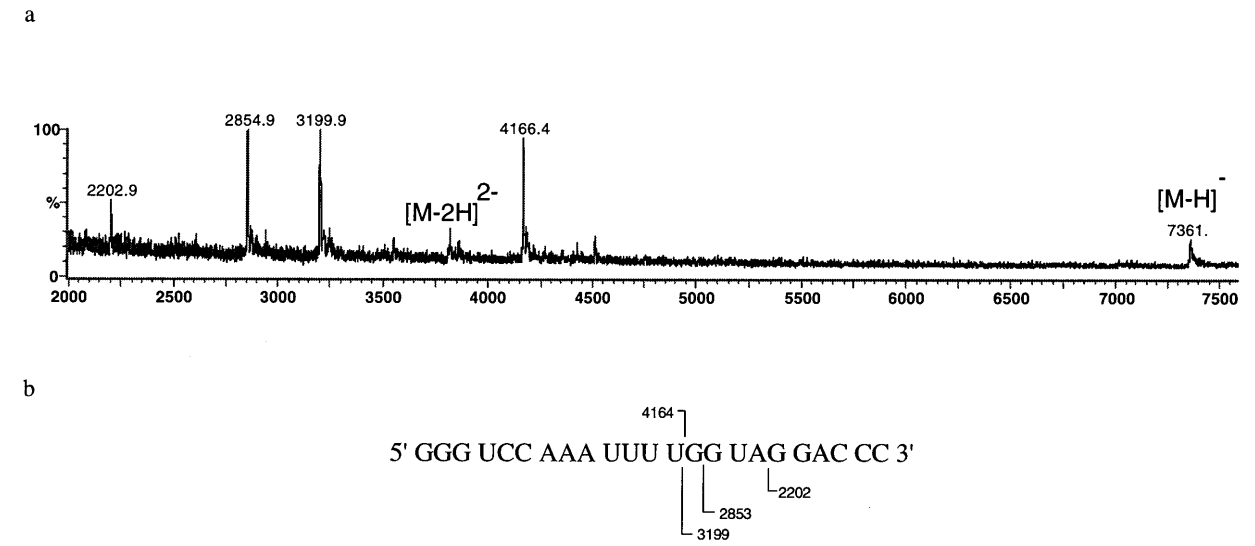
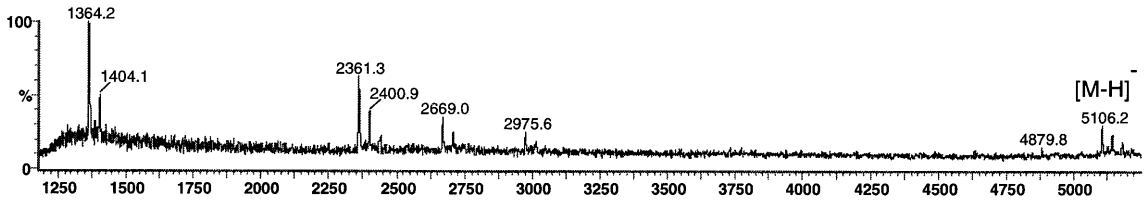


Figure 4. (**a**) Negative ion UV MALDI spectra of a synthetic 23mer RNA (average molecular weight 7363 Da) subjected ribonuclease Phy M digestion for 2 min. (**b**) Schematic representation of the major fragments observed.

Chemical digestion

Hydrazine/aniline treatment. Figure 5 shows the MALDI-TOF mass spectrum of the 16mer RNA subjected to hydrazine/aniline treatment. Four major ions are observed, corresponding to fragments formed by scission at U's, each containing the original 3′-terminus. In agreement with the reaction scheme in reference (10), the mass of the fragments correspond to those possessing a 5′ phosphate. Fragments containing the original 5′-terminus were not observed, in contrast to those obtained from endonuclease digests (8). Figures 6 and 7 show the MALDI-TOF mass spectra of an 18mer RNA and a 27mer RNA subjected to hydrazine/aniline treatment. Again, the major fragments observed were those containing the original 3′-terminus resulting from cleavage at U's. The minor peaks in Figure 7 do not correspond to the missing 5′ fragments. Their origin is unclear and is still under investigation. As can be seen, adjacent uridines are cleaved effectively. Cleavage at cytidines has not been observed.

a



b

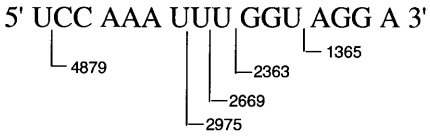
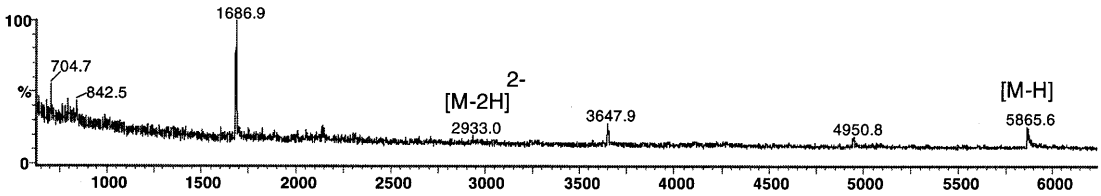


Figure 5. (**a**) Negative ion UV-MALDI mass spectrum of a synthetic 16mer RNA (average molecular weight 5103 Da) subjected to hydazine/aniline treatment. (**b**) Schematic representation of the major fragments observed.

a



b

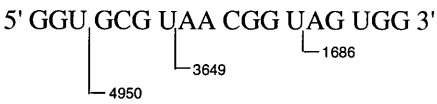
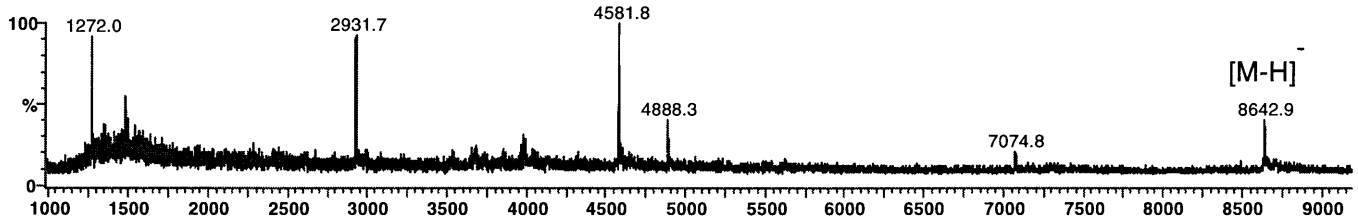


Figure 6. (**a**) Negative ion UV-MALDI mass spectrum of a synthetic 18mer (average molecular weight 5867 Da) subjected to hydrazine/analine treatment. (**b**) Schematic representation of the major fragments observed.

DISCUSSION

The small difference in mass (1 Da) between the pyrimidine bases (U and C) of RNA has been a major limitation to the development of simple mass spectrometric sequencing strategies. Exonuclease digestion results in ambiguous sequence assignments where the pyrimidine bases cannot be distinguished form each other. The purine bases (A and G) can easily be distinguished from each other and from the pyrimidine bases because of larger mass difference (>15 Da), which is within the resolving power of MALDI-TOF instruments such as described here. In order to generate digest fragments that would be characteristic for either U or C residues, we first explored the possibility of using commercially available base-specific enzymes. We found that the specificity of the enzymes was not as expected, possibly due to compromise reaction conditions that had to be used in order to be compatible with MALDI (8). An alternative approach was to use direct chemical methods, which have the advantage of being less susceptible to secondary structure effects (2,10). Hydrazine/aniline treatment of RNA resulted in characteristic fragments formed by scission at U's. In no case have we observed scission at C's. The prevalence of fragments containing the original 3′-terminus is an interesting observation, which contrasts with the observed prevalence for fragments containing the original 5′-terminus in endonuclease digests (8). According to the reaction scheme in reference (10), fragments containing the original 3′-terminus should possess a phosphate at their 5′-end and fragments containing the original 5′-end should consist of a ring opened ribose with an aniline adduct at their 3′-end. The 5′ phosphate on the former

a



b

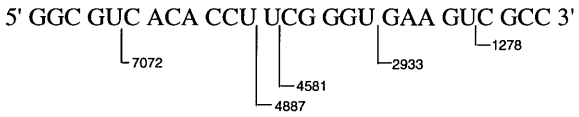


Figure 7. (**a**) Negative ion UV-MALDI mass spectrum of a synthetic 27mer RNA (average molecular weight 8639 Da) subjected to hydrazine/aniline treatment.

series is indeed observed. Since both series of fragments should be formed in equal amounts, the unusual moiety on the 3′-end of the latter series either results in partial exclusion of these fragments from the MALDI matrix, or more probably, in charge suppression during desorption. Whatever the explanation, this behaviour greatly simplifies interpretation of the data without the need for elaborate methods for isolation of a particular series of fragments as others have used (8). The data presented here demonstrates that aqueous hydrazine/aniline treatment of RNA, followed by MALDI-TOF, provides a rapid and simple method of unambiguously assigning U and C residues in RNA sequences obtained from either alkaline hydrolysis or exonuclease digestion. The current limit of molecular size amenable to this technique is estimated to be in the region of 100 nt (7,8). Future studies could involve the use of this and other chemical methods such as dimethyl sulphate alkylation and diethyl pyrocarbonate treatment (2,10) followed by MALDI-TOF to probe the secondary structure of RNA.

REFERENCES

1 Hahn,C.S., Strauss, E.G. and Strauss, J.H. (1989) *Methods Enzymol.*, **180**, 121–130.
2 Peattie, D.A. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 1760–1764.
3 Donis-Keller, H., Maxam, A.M. and Gilbert, W. (1977) *Nucleic Acids Res.*, **4**, 2527–2538.
4 Kuchino, Y. and Nishimura, S. (1989) *Methods Enzymol.*, **180**, 154–163.
5 Nordhoff, E., Kirpekar, F. and Roepstorff, P. (1997) *Mass Spectrom. Rev.*, **15**, 67–138.
6 Pieles, U., Zürcher, W., Schär, M. and Moser, H.E. (1993) *Nucleic Acids Res.*, **21**, 3191–3196.
7 Smirnov, I.P., Roskey, M.T., Juhasz, P., Takach, E.J., Martin, S.A. and Haff, L.A. (1996) *Anal. Biochem.*, **238**, 19–25.
8 Hahner, S., Lüdemann, H., Kirpekar, F., Nordhoff, E., Roepstorff, P., Galla, H. and Hillenkamp, F. (1997) *Nucleic Acids Res.*, **25**, 1957–1964.
9 Donis-Keller, H. (1980) *Nucleic Acids Res.*, **8**, 3133–3142.
10 Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J. and Ehresmann, B. (1987) *Nucleic Acids Res.*, **15**, 9109–9128.
11 Wang, B.H. and Biemann, K. (1994) *Anal. Chem.*, **66**, 1918–1924.