Nucleic Acids Research, 1996, Vol. 24, No. 8 **1489–1496**

2′-*O*-methyl-5-formylcytidine (f^{5}Cm), a new modified nucleotide at the 'wobble' position of two cytoplasmic tRNAs^{Leu}(NAA) from bovine liver

Jean-Paul Païs de Barros, Gérard Keith[1], Chakib El Adlouni[1,+], Anne-Lise Glasser[§], Gérard Mack, Guy Dirheimer[1] and Jean Desgrès*

Laboratoire de Biochimie Médicale, Equipe de Recherche Biochimie–Biologie Cellulaire, Faculté de Médecine et Centre Hospitalier Universitaire, Université de Bourgogne, 21034 Dijon, France and [1]Structure des Macromolécules Biologiques et Mécanismes de Reconnaissance, Unité Propre de Recherche no. 9002, Institut de Biologie Moléculaire et Cellulaire du CNRS et Université Louis Pasteur, 67084 Strasbourg, France

Received December 28, 1995; Revised and Accepted March 4, 1996

ABSTRACT

The nucleotide analysis of a cytoplasmic tRNA^{Leu} isolated from bovine liver revealed the presence of an unknown modified nucleotide N. The corresponding N nucleoside was isolated by different enzymatic and chromatographic protocols from a partially purified preparation of this tRNA^{Leu}. Its chemical characterization was determined from its chromatographic properties, UV-absorption spectroscopy and mass spectrometric measurements, as well as from those of the borohydride reduced N nucleoside and its etheno-trimethylsilyl derivative. The structure of N was established as 2′-*O*-methyl-5-formylcytidine (f^{5}Cm), and its reduced derivative as 2′-*O*-methyl-5-hydroxymethylcytidine (om^{5}Cm). By sequencing the bovine liver tRNA^{Leu}, the structure of the anticodon was determined as f^{5}CmAA. In addition, the nucleotide sequence showed two primary structures differing only by the nucleotide 47c which is either uridine or adenosine. The two slightly differing bovine liver tRNAs-^{Leu}(f^{5}CmAA) are the only tRNAs so far sequenced which contain f^{5}Cm. The role of such a modified cytidine at the first position of the anticodon is discussed in terms of decoding properties for the UUG and UUA leucine codons. Recently, precise evidence was obtained for the presence of f^{5}Cm at the same position in tRNAs^{Leu}(NAA) isolated from rabbit and lamb liver. Therefore, the 2′-*O*-methyl-5-formyl modification of cytidine at position 34 could be a general feature of cytoplasmic tRNAs^{Leu}(NAA) in mammals.

INTRODUCTION

The primary structures of several cytoplasmic mammalian isoaccepting tRNAs^{Leu} are already known as carrying the following anticodons: IAG from bovine liver (1) and cow mammary gland (2), CAG from cow mammary gland (3), and MmAA from rat Morris hepatoma (4) and human HeLa cells (5) where the unknown nucleotide called Mm was described as a 2′-*O*-methylated modified pyrimidine.

In order to determine whether or not this modified nucleotide Mm was identical to the already published 2′-*O*-methyl-5-carbamoylmethyluridine (ncm^{5}Um) found in yeast tRNA^{Leu}(ncm^{5}UmAA) (6,7), we purified a bovine liver tRNA^{Leu} corresponding to the above tRNAs^{Leu}(MmAA), and analyzed its modified nucleotides. Some preliminary studies showed the presence of an unknown nucleotide that could be located at the 'wobble' position of the anticodon. This unknown nucleotide, that we called N, behaved on thin-layer or high performance liquid chromatography like a 2′-*O*-methylated modified pyrimidine, as previously described for Mm in tRNAs^{Leu}(MmAA) from rat Morris hepatoma (4) and human HeLa cells (5).

We report here on the isolation of the unknown N nucleoside, and its structure determination by the combined means of thin-layer chromatography (TLC), high performance liquid chromatography (HPLC) with real-time UV-absorption spectroscopy, and gas chromatography-mass spectrometry (GC-MS) analysis (13,14). By improving some of the previously used sequencing methods (8–12), the complete nucleotide sequence of the tRNA^{Leu} isolated from bovine liver was further determined. The latter is also reported in this paper.

Preliminary studies recently conducted in our laboratories have shown that N nucleotide was also present in pure tRNAs^{Leu}(NAA) isolated from rabbit and lamb liver. The coding properties of this modified nucleotide located at the 'wobble' position of these mammalian tRNAs^{Leu} are discussed.

MATERIALS AND METHODS

Purification of bovine tRNA^{Leu}(NAA)

Total bovine tRNA was prepared according to Fournier *et al.* (9). It was further chromatographed on BD-cellulose column (10). The fraction containing the tRNA^{Leu}(NAA) was detected by dot-blot hybridization (12) with a probe complementary to the 3′-end (nucleotides 59–73) of already sequenced tRNAs-^{Leu}(MmAA) from rat Morris hepatoma (4) and human HeLa cells

* To whom correspondence should be addressed

Present addresses: +Université Chouaib Doukhali, Faculté des Sciences, El Jadida, Maroc and §Université d'Auvergne Clermont 1, I.U.T., Biologie Appliquée, BP 86, 63172 Aubière Cédex, France

(5). This partially purified preparation of tRNA was used for the structure determination of the unknown N nucleoside, as described below. Further purification steps of $tRNA^{Leu}$(NAA) were carried out by chromatography on Sepharose 4B column (10), followed by two dimensional polyacrylamide gel electrophoresis in non-denaturating conditions (8,11). To obtain a tRNA suitable for sequencing, the $tRNA^{Leu}$(NAA) was further purified on a denaturing polyacrylamide gel at 55–60°C, in order to remove partially cut molecules (10,11).

Isolation of the unknown N nucleoside

The N nucleoside was isolated from the partially purified preparation of bovine liver $tRNA^{Leu}$(NAA), using the analytical procedure previously described for the isolation of other unknown nucleosides (13,14). This procedure can be summarized as follows: (i) nuclease P1 digestion followed by bacterial alkaline phosphatase (BAP) treatment of the $tRNA^{Leu}$(NAA) sample (~50 A_{260} units in 200 µl water); (ii) isolation and desalting by HPLC of a dinucleotide resistant to nuclease P1 action; (iii) combined hydrolysis by snake venom phosphodiesterase (SV-PDE) and BAP of the dinucleotide sample, leading to the liberation of two nucleosides; (iv) isolation and desalting of N nucleoside by HPLC; and (v) drying of N nucleoside under vacuum at room temperature.

Analysis, isolation and desalting of the ribonucleosides by HPLC were performed using the chromatographic experimental conditions previously described (13,15–17).

Mass spectrometry

The conversion of N nucleoside (~3 µg) to trimethylsilyl (TMS) derivative, and the electron impact-mass spectrometry (EI-MS) of silylated N by gas chromatography-mass spectrometry (GC-MS) analysis, using a Hewlett Packard GC-mass spectrometer, model 5971A, were carried out as previously reported (7,13,14). Identical GC-MS conditions were also applied on the TMS and etheno-trimethylsilyl (etheno-TMS) derivatives of the reduced product resulting from borohydride treatment of N nucleoside.

Borohydride reduction of N nucleoside

Borohydride reduction of N nucleoside was performed according to Matsuda *et al.* (18). To ~1 µg of the dry N nucleoside were added 100 µl 5 mg/ml aqueous natrium borohydride solution ($NaBH_4$) and 100 µl 20 mM orthophosphoric acid. The mixture was incubated for 1 h at 65°C and further analyzed by HPLC. The resulting product was isolated, desalted and dried as described above for the isolation of N nucleoside. The dry compound was finally converted either to TMS derivative, or to etheno-TMS derivative for structure characterization by GC-MS analysis.

Etheno-trimethylsilyl (etheno-TMS) procedure

The etheno-TMS derivatives of nucleosides were prepared using an adaptation of the procedures previously described by others (19–22). A 50% aqueous chloracetaldehyde solution was adjusted to pH 4.0 using 2 M sodium hydroxide solution. To 50 µl of aqueous solution containing ~1 µg of N nucleoside were added 10 µl of the above chloracetaldehyde solution, and the mixture was stirred overnight at room temperature. In order to isolate the resulting product (i.e. the etheno derivative), and to remove the remaining chloracetaldehyde, the above mixture was chromatographed on HPLC, using the same chromatographic conditions as for the HPLC analysis of tRNA nucleosides (13,14). After drying under vacuum at room temperature, the pure etheno derivative was submitted to TMS derivatization, and it was further characterized by GC-MS analysis.

Sequencing procedures

To sequence the $tRNA^{Leu}$(NAA), and to characterized the modified nucleotides therein, we used the random single hit sequencing ^{32}P post-labelling procedure that has already been described (11). Prior to sequencing, the highly purified tRNA (~5 µg in 150 µl) was renatured by heating for 5 min at 65°C in presence of 20 mM $MgCl_2$, and cooled down slowly step by step: 10 min at 37°C followed by 10 min at room temperature. To remove the excess of $MgCl_2$, the sample was desalted by centrifugation on Sephadex G25 columns poured into 1 ml syringes (11).

After desalting, the sample was dried and hydrolyzed in 10 µl deionised formamide at 95°C for 4 min (10,11). The digestion products were labelled by T4 polynucleotide kinase and [γ-^{32}P]ATP, deproteinized by phenol, and desalted on Sephadex G25 as described above. The labelled digest was denatured (5 min at 95°C in loading buffer), and directly loaded onto pre-heated (by pre-electrophoresis at 1500 V for 2 h) denaturing 15% polyacrylamide gels containing 10% deionised formamide and 7 M urea. To keep the fragments in denatured form during loading, the buffer in the wells was heated by injecting, prior to loading, hot (95°C) running buffer. Separation of the fragments was performed at 1500 V.

RESULTS

Isolation of the unknown N nucleoside

HPLC nucleoside analysis of the partially purified bovine liver $tRNA^{Leu}$(NAA) after exhaustive nuclease P1 digestion followed by BAP hydrolysis (Fig. 1a) gave the following modified nucleosides (in moles per mole of tRNA): D, 1.6; ψ, 4.3; m^1A (+m^6A), 1.0; m^5C, 1.2; T, 1.0; Um, 1.0; m^1G, 0.9; ac^4C, 0.7; m^2G, 1.8; $m^2{}_2G$, 1.1; Cm, 0.1; i^6A, 0.1; m^7G, 0.3; Gm, 0.3; and a peak eluted at 53 min retention time, and which should correspond to the peak N mentioned at 31 min retention time on the HPLC chromatogram of nucleosides from bovine $tRNA^{Leu}$ previously published elsewhere (23). These results suggested that the $tRNA^{Leu}$(NAA) preparation was contaminated with several cytoplasmic tRNAs ($tRNA^{Ser}$, $tRNA^{Phe}$, ...).

However, the peak that eluted at 53 min retention time was isolated, and designated as compound N_{53}. This compound exhibited a UV-absorption spectrum (Fig. 1b) not yet described in the tRNAs so far sequenced. It was nuclease P1-resistant, but it yielded two nucleosides upon snake venom phosphodiesterase (SV-PDE) plus BAP treatments (Fig. 2a). One of those was easily identified as an unmodified adenosine (A) by comparison of its HPLC retention time (~38 min) and UV-absorption spectrum with those of authentic adenosine.

The second nucleoside that eluted at ~42 min HPLC retention time was designated as N_{42}. It was collected, desalted and concentrated. The UV-absorption spectrum of this compound (Fig. 2b) was very similar to that of the 5-formylcytidine (f^5C) identified in the methionine tRNA isolated from bovine liver mitochondria (mt $tRNA^{Met}$) (24,25). From this observation, it was supposed that the unknown N nucleoside from bovine liver $tRNA^{Leu}$(NAA) could have a structure closely related to f^5C. Further investigations by mass spectrometry on N_{42} and its derivative upon reductive treatment are described below, and confirmed this hypothesis.

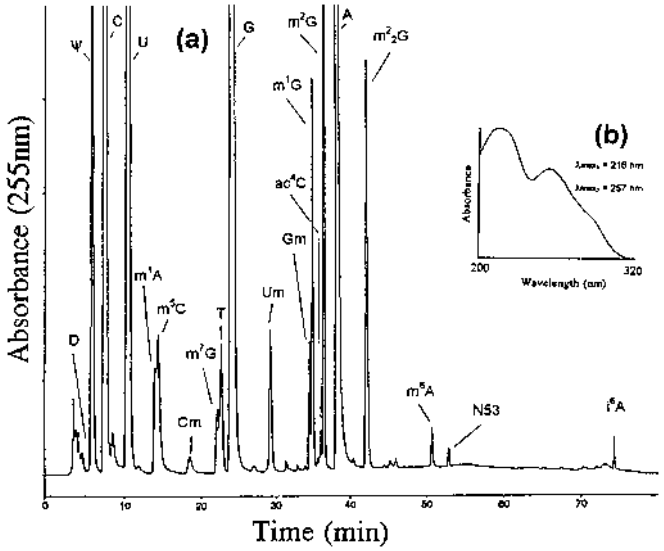

Figure 1. (**a**) HPLC chromatogram of nucleosides resulting from exhaustive nuclease P1 plus bacterial alkaline phosphatase digestion of the partially purified cytoplasmic $tRNAs^{Leu}$ (NAA) isolated from bovine liver. The peak N_{53} corresponds to an unknown nuclease P1-resistant dinucleotide NpA eluted at 53 min retention time. (**b**) Real-time UV absorption spectrum of the unknown N_{53} dinucleotide.

Characterization of N_{42} as a modified 2′-*O*-methylcytidine by mass spectrometry

The EI-mass spectrum of N_{42} as TMS derivative is presented in Figure 3a, and several ion series are summarized in Table 3. The ion peaks at m/z 501, 486, 469 and 398 were assigned to the molecular ion M^+ and to the fragments $M^+ –CH_3$, $M^+ –CH_3OH$ and $M^+ –CH_2OSi(CH_3)_3$, respectively. Resulting from N–C glycosylic bond cleavage and fragmentations in the sugar moiety (26), the ion peaks at m/z 341, 312, 284, 268, 251, 240, 223 and 212 were assigned to contain the base moiety (B) of 210 Da plus portions of the sugar moiety. Finally, numerous ions such as m/z 290, 258, 187 and 159 were characteristic of ribonucleosides having a 2′-*O*-methylated ribose.

According to the fragmentation processes described by McCloskey's group (26–28) for EI-MS of silylated ribonucleosides, all these ion series from N_{42} as TMS derivative were consistent with an EI fragmentation pathway of a silylated 2′-*O*-methyl-ribonucleoside with a molecular mass value of 501, i.e. 28 Da greater than that (m/z 473, results not shown) of 2′-*O*-methylcytidine (Cm) as $(TMS)_3$ derivative. The chemical structure of N_{42} was thus postulated to be a Cm bearing a 28 Da modification located on the cytosine moiety.

Such a mass increment corresponds in principle to net addition of either CO, or two methyl groups. The 28 Da modification was characterized as a CO group because it was a reducible 28 Da group as described below.

Characterization of unknown N as formyl-Cm by borohydride reduction

When submitted to borohydride reduction, the pure N_{42} nucleoside led to a single product eluting at ~22 min HPLC retention time. This product was designated as N_{22} upon HPLC purification. Pure N_{22} exhibited a UV-absorption spectrum nearly identical to that of 5-hydroxymethylcytidine (om^5C) previously found in canine serum (14): same profile for both UV spectra, and λ_{max} for N_{22} 4 nm longer than λ_{max} for cytidine (not shown). The EI-mass spectrum of N_{22} as TMS derivative is shown in Figure 3b. From the corresponding ion series summarized in Table 4, the chemical structure of the silylated N_{22} was established as the $(TMS)_4$ derivative of 2′-*O*-methyl-hydroxymethylcytidine (omCm). This result showed clearly that the 28 Da modification on N_{42} was not constituted by two methyl groups, but by a formyl group reducible into a hydroxymethyl group upon borohydride treatment.

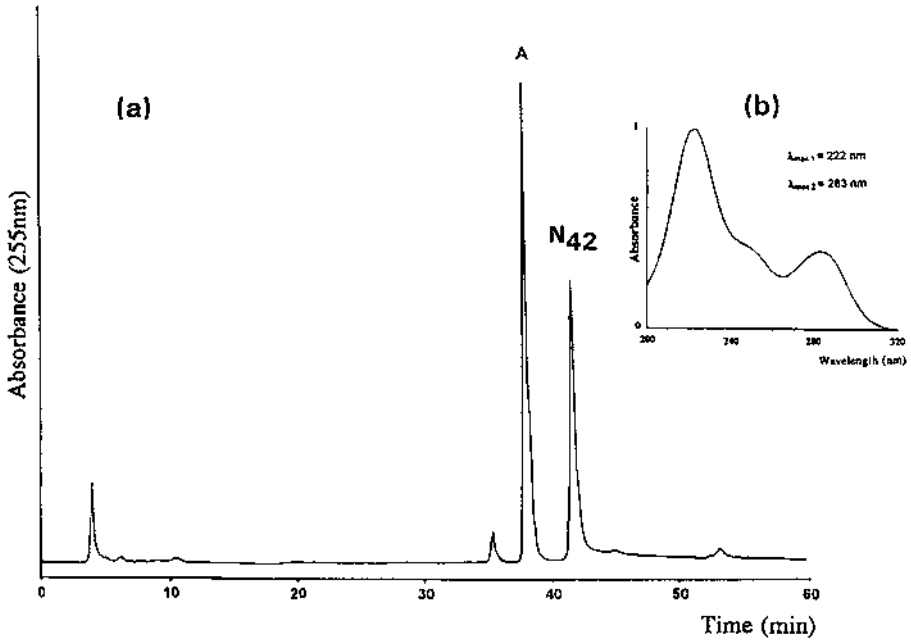

Figure 2. (**a**) HPLC chromatogram of nucleosides resulting from complete snake venom phosphodiesterase (SV-PDE) plus bacterial alkaline phosphatase hydrolysis of unknown N_{53} dinucleotide, i.e. NpA, isolated by HPLC from bovine liver $tRNAs^{Leu}$ (NAA). (**b**) Real-time UV-absorption spectrum of the unknown N_{42} nucleoside.

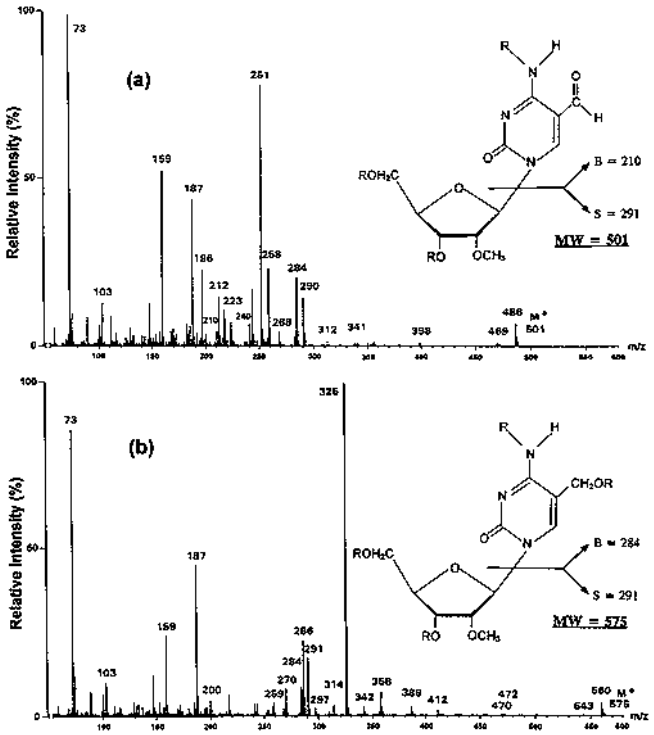

Figure 3. Electron impact-mass spectrum (EI-MS) of trimethylsilyl (TMS) derivative of N_{42} nucleoside, i.e. the unknown N nucleoside isolated from bovine liver $tRNAs^{Leu}$ (NAA) [upper panel (**a**)], as compared to that of (TMS)-derivative of the compound N_{22} resulting from reducing borohydride treatment of N_{42} [lower panel (**b**)].

The structure of N_{42} was therefore deduced as a 2′-*O*-methyl-formylcytidine, i.e. fCm.

However, the location of the formyl group linkage on fCm stayed to be defined between the alternative C-5, N^4 or N-3 positions of the base moiety. Because the too small amount of the pure fCm nucleoside did not allow an accurate analysis by nuclear magnetic resonance (NMR) spectroscopy, the formyl group location was determined by an ethenylation procedure, using the chloracetaldehyde ($ClCH_2CHO$) reaction on cytosine- or adenine-containing nucleosides.

Determination of the formyl group location at C-5 on fCm by ethenylation procedure

General features of nucleoside derivatives obtained from ethenylation reaction. As previously described by others (19–22), the ethenylation reaction on unmodified-, ribose modified-, and C-5 substituted-cytidines, or on unmodified-, and ribose modified-adenosines, leads to the formation of either 3,N^4-etheno-cytidine or 1,N^6-etheno-adenosine derivatives, respectively. However, it is totally inefficient on the modified cytidines substituted at N-3 (e.g. 3-methylcytidine) or on the modified adenosines substituted at N-1 (e.g. 1-methyladenosine) (20–22). When applied on N^6-substituted adenosines, like N^6-isopentenyladenosine (20–22) or N^6-methyladenosine (personal results), the ethenylation procedure leads to adenosine derivatives bearing a 1,N^6-hydroxyethyl-ring and a positive charge on the N-1. It can be supposed that structurally similar derivatives, i.e. cytidine derivatives bearing a 3,N^4-hydroxyethyl ring and a positive charge on their N-3, would be obtained from ethenylation procedure on N^4 substituted cytidines.

Table 3. Fragment-ion series from electron impact-mass spectrum of N nucleoside (f^5Cm) isolated from the bovine liver leucine tRNAs(NAA), as $(TMS)_3$ derivative

Mass	From molecular ion $(TMS)_3$	
(m/z)	Ion^+	Composition
501	M	Molecular ion $(TMS)_3$
486	M –15	M $-CH_3$
469	M –32	M $-CH_3OH$
398	M –103	M $-CH_2OTMS$
341	B +131	B $+C_2H_2O_2TMS$
312	B +102	B +CHO +TMS
284	B +74	B +H +TMS
268	B +58	B $+Si(CH_3)_2$
251	B +41	B $+C_2HO$
240	B +30	B $+CH_2O$
223	B +13	B +CH
212	B +2	B $+H_2$
210	B	Base moiety
196	B –14	B $-CH_2$
290	S –1	S –H
258	S –33	S –H $-CH_3OH$
187	S –104	S –H $-CH_2OTMS$
159	S –132	$C_4H_6O_2TMS$
103	–	CH_2OTMS

B, base moiety of $(TMS)_3$ derivative.
S, sugar (2′-*O*-methyl-ribose) moiety.
TMS, trimethylsilyl group = $Si(CH_3)_3$.

HPLC and GC-MS analysis of fCm and omCm as etheno-TMS derivatives. When submitted to ethenylation procedure, the fCm nucleoside yielded an unknown compound which eluted in HPLC at ~36 min, i.e. 6 min earlier than the fCm precursor. This compound was totally retained, at once on the HPLC column when chromatographed in desalting conditions, and on the GC column when analyzed by GC-MS upon TMS derivatization. Because of such chromatographic behaviours, any structural information on fCm could be obtained from this product which was presumably a very polar compound formed by chemical reaction between the chloracetaldehyde and the reactive CO group of fCm.

A more informative result on the formyl group location in fCm was finally given by etheno-TMS derivatization and GC-MS analysis of the borohydride-reduced fCm nucleoside, i.e. the hydroxymethyl-Cm (omCm) described above.

The reaction of omCm with chloracetaldehyde led to a single product which eluted in HPLC at ~52 min, i.e. 30 min later than omCm. The EI-mass spectrum of this product upon TMS derivatization is presented in Figure 4. It shows an EI-fragmentation pathway with several essential properties previously described as characteristic of the 3,N^4-etheno-TMS derivatives of cytosine-containing nucleosides (22), and which can be summarized as follows: (i) the values of the molecular ion M^+ (m/z 527) and of the fragment M $-CH_3$ (m/z 512) are consistent with a $(TMS)_3$ derivative of omCm bearing unsubstituted 3,N^4-etheno function on the base moiety, (ii) the molecular ion abundance of this etheno-$(TMS)_3$ derivative is markedly greater than that of the corresponding $(TMS)_4$ derivative shown in Figure 3b, and (iii) the

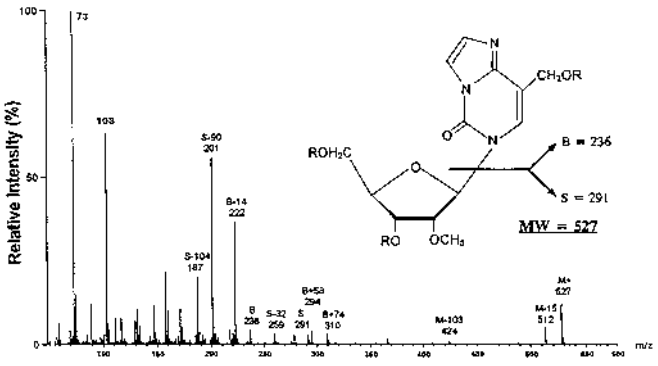

Figure 4. Electron impact-mass spectrum (EI-MS) of the reduced compound N_{22}, i.e. omCm, as etheno-trimethylsilyl (etheno-TMS) derivative. R = H: not silylated molecule; R = TMS: silylated molecule for GC-MS analysis, with M = molecular ion, B = base moiety, S = sugar (2′-*O*-methyl-ribose) moiety.

fragment-ion at base +41 a.m.u. (m/z 277), characteristic of cytidine analogs as TMS derivatives, is absent from the mass spectrum because the free imino group at C-4 required for its formation (26,28) is blocked by the etheno function.

Definitive identification of N nucleoside in bovine tRNA^{Leu}(NAA) as f^{5}Cm. Since the ethenylation reaction was efficient on omCm, it was assumed that the hydroxymethyl group was not linked at the N-3 position. According to the general features of the ethenylation derivatization mentioned above for N^6-substituted adenosines or N^4-substituted cytidines, a linkage of the hydroxymethyl group at the N^4 position of omCm should lead by chloracetaldehyde reaction to the 3,N^4-hydroxyethyl-N^4-hydroxymethyl-Cm structure bearing a positive charge on the N-3. In that case, the corresponding persilylated compound would be a $(TMS)_4$-derivative with a molecular mass value of 617, i.e. 90 Da greater than the one observed in the EI-mass spectrum of omCm as etheno-TMS derivative (Fig. 4). The hydroxymethyl group of omCm, and consequently the corresponding formyl group of fCm, can therefore only be linked at the C-5 position to give the etheno-$(TMS)_3$ derivative obtained from omCm.

These results definitively established the 2′-*O*-methyl-5-formyl-cytidine (f^{5}Cm) structure for the unknown N nucleoside present in bovine liver tRNA^{Leu}(NAA) (Fig. 3a), and the 2′-*O*-methyl-5-hydroxymethyl-cytidine (om^{5}Cm) structure for the reduced nucleoside obtained by borohydride treatment of N (Fig. 3b).

Primary structure of the tRNA^{Leu}(NAA) from bovine liver

New improvements in sequencing procedures. When the classical single hit sequencing procedure hydrolysis (95°C for 4 min in formamide) is applied on tRNAs highly purified by successive ion-exchange chromatographies and gel electrophoresis (10,11), the cuts are seldom randomly distributed. These tRNAs are often preferentially cut in the variable loop and in the anticodon loop, and only very little random cuts occur elsewhere in the molecules of tRNA. In addition, the regions rich in G·C pairs are often not cut at all. They behave like tRNAs from which most Mg^{2+} and other divalent cations or heavy metals have been removed by EDTA treatment (unpublished results). This means that the cleavages during the random hydrolysis of the tRNA could be due to Mg^{2+} or other cations which could have been removed from the tRNA during the numerous purification steps involving ion-exchange or electrical fields.

Table 4. Fragment-ion series from electron impact-mass spectrum of the (TMS)-compound resulting from reducing borohydride treatment of N nucleoside isolated from bovine liver tRNAs^{Leu}(NAA), i.e. om^{5}Cm as $(TMS)_4$ derivative

Mass (m/z)	From molecular ion $(TMS)_4$	
	Ion$^+$	Composition
575	M	Molecular ion $(TMS)_4$
560	M –15	M –CH_3
543	M –32	M –CH_3OH
472	M –103	M –CH2OTMS
470	M –105	M –CH_3 –TMSOH
412	B +128	B +C_3H_3OTMS
386	B +102	B +CHO +TMS
358	B +74	B +H +TMS
342	B +58	B +$Si(CH_3)_2$
325	B +41	B +C_2HO
314	B +30	B +CH_2O
297	B +13	B +CH
286	B +2	B +H_2
285	B +1	B +H
284	B	Base moiety
270	B –14	B –CH_2
291	S	Sugar moiety
259	S –32	S –CH_3OH
200	S –91	S –H –TMSOH
187	S –104	S –H –CH_2OTMS
159	S –132	$C_4H_6O_2$TMS
103	–	CH_2OTMS

B, base moiety of $(TMS)_4$ derivative.
S, sugar (2′-*O*-methyl-ribose) moiety.
TMS, trimethylsilyl group = $Si(CH_3)_3$.

In order to reintroduce Mg^{2+} onto the tRNA, we renatured in presence of $MgCl_2$ the highly purified tRNA^{Leu}(NAA) from bovine liver after elution from denaturating gels, and removed the excess of salt on small Sephadex G25 columns as described in Materials and Methods (10). The sample was hydrolyzed by random single hit cutting and post-labelled as already described above. Finally, the labelled digest was either separated on two-dimensional polyacrylamide gel electrophoresis, or on one-dimensional denaturating gel electrophoresis (11). Autoradiography was used to detect the complete set of randomly distributed 5′-^{32}P-labelled fragments going from each potential cleavage point inside the tRNA molecule up to the 3′-CCA end which is common to all of these fragments. The labelled fragments were then eluted, and the end nucleotides were characterized as previously described (10,11).

Nucleotide sequence of tRNA^{Leu}(NAA) from bovine liver. The primary structure of the bovine tRNA^{Leu}(NAA) could be deduced completely by analysing the above 5′-^{32}P-labelled fragments as follows. (i) By determining the 5′-^{32}P-labelled end nucleotides by one-dimensional thin-layer chromatography (not shown), and two-dimensional thin-layer chromatography in the case of the separation of the modified nucleotides. The composite drawing of the separation of the modified mononucleotides is presented in Figure 5. This figure shows in particular the positions of pf^{5}Cm,

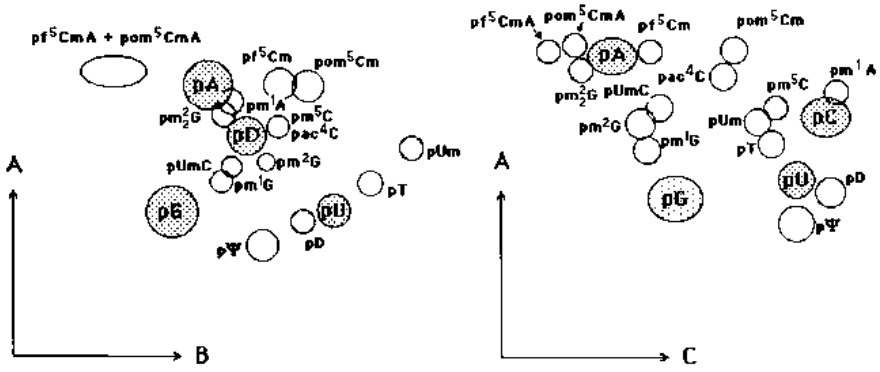

Figure 5. Drawings of two-dimensional thin layer separations on cellulose plates (F1440 or G1440; Schleicher and Schüll, Dassel, Germany) of $tRNA^{Leu}$(NAA) mononucleotides and nuclease P1-resistant dinucleotides. The spots corresponding to the major ribonucleotides are shaded. Solvents were: A, isobutyric acid–25% NH_4OH–H_2O (50/1.1/28.9, by vol); B, 0.1 M sodium phosphate pH 6.8–ammonium sulphate–*n*-propanol (100/60/2, v/w/v); C, HCl–isopropanol–H_2O (15/70/15, by vol).

pom^{5}Cm, and the corresponding nuclease P1- (and RNase-, and alkali-) resistant dinucleotides pf^{5}CmpA and pom^{5}CmpA whose positions were assigned while preparing the unlabelled nucleotides for their above chemical structure determinations. (ii) By studying the end oligonucleotidic structure by electrophoresis-homochromatography of some fragments just upstream of two compressed zones which were not directly solved in the sequencing gel, or of positions in which two different nucleotides were found in two places of the primary structure of the tRNA. These analyses concerned the 5′-end of the whole tRNA (Fig. 6A), the region between the D-arm and the anticodon-arm (Fig. 6B and C), and the region between anticodon-arm and extra-arm (Fig. 6F and G), in which the sequencing gel gave compressed fragments (i.e. several fragments in the same band). They were also performed at two heterogeneously occupied positions: nucleotide 47c (A or U) (Fig. 6F and G), and nucleotide 34 (f^{5}Cm or om^{5}Cm) (Fig. 6D and E) that have already been seen in the end nucleotide analysis of the fragments reported above.

As far as position 47c is concerned, the experiments on the first heterogeneously occupied position confirmed the analysis of the 5′-end nucleotides of the fragments by showing the separation of the nucleotide track into two different tracks at that position (Fig. 6F and G), because of the difference in net charges between U and A located there. This result allowed to assign the existence of two $tRNAs^{Leu}$(NAA) differing by the presence of U or A in position 47c of the extra loop.

The second heterogeneously occupied position concerned the ‘wobble’ nucleotide of the anticodon where the sequencing of the ends of the fragments, using exhaustive nuclease P1 and venom phosphodiesterase digestions, showed the presence of either f^{5}Cm or om^{5}Cm. However, on electrophoresis-homochromatography, because of the presence of the *O*-methylation on the ribose moieties, the single hit sequencing hydrolysis led to two end-dinucleotide wandering spot jumps (Fig. 6D and E): pf^{5}CmpAp which behaved like pUmpAp and pUpAp (probably because of the formyl group in f^{5}Cm), and pom^{5}CmpAp which behaved like pCmpAp and pCpAp (not shown). This result was further confirmed by electrophoresis on Whatman DE81 paper in presence of 7% formic acid, where the nuclease U2 digest led to two dinucleotides: pf^{5}CmpAp and pom^{5}CpAp which behaved like pUpAp and pCpAp, respectively (not shown).

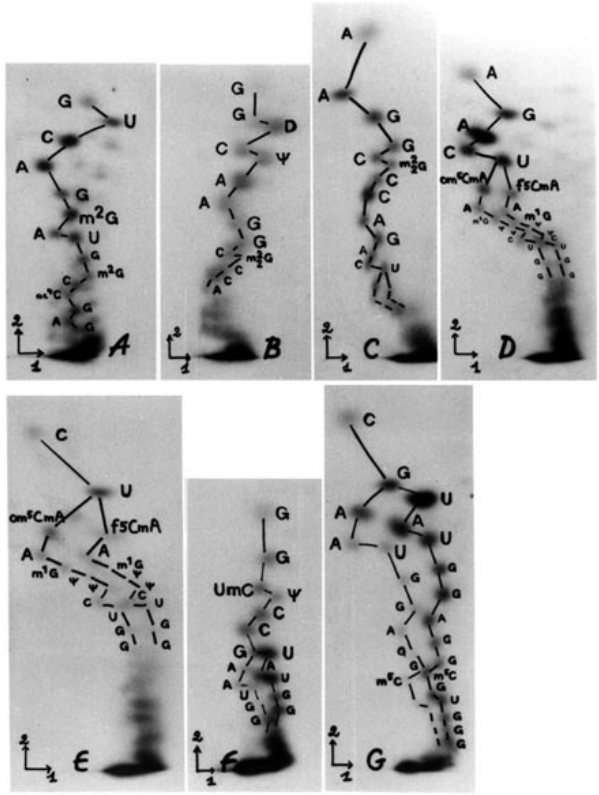

Figure 6. Electrophoresis-homochromatography performed on fragments eluted from two dimensional sequencing gel electrophoresis (not shown) for resolving the sequences near (i) the 5′-end of the tRNA, (ii) compression zones, (iii) heterogeneously occupied positions, or (iv) modified nucleotides: (**A**), nucleotide N1 to nucleotide N15; (**B**), N18–N29; (**C**), N21–N33; (**D**), N29–N43; (**E**), N32–N43; (**F**), N42–N47g; (**G**), N47a–N53.

However, the two nucleotides f^{5}Cm and om^{5}Cm were found in very variable ratios depending of the studied samples: number of purification steps for preparing the bovine $tRNA^{Leu}$, times of storage of the tRNA or oligonucleotide preparations, types of the analytical methods used for sequencing these preparations. Since only f^{5}Cm

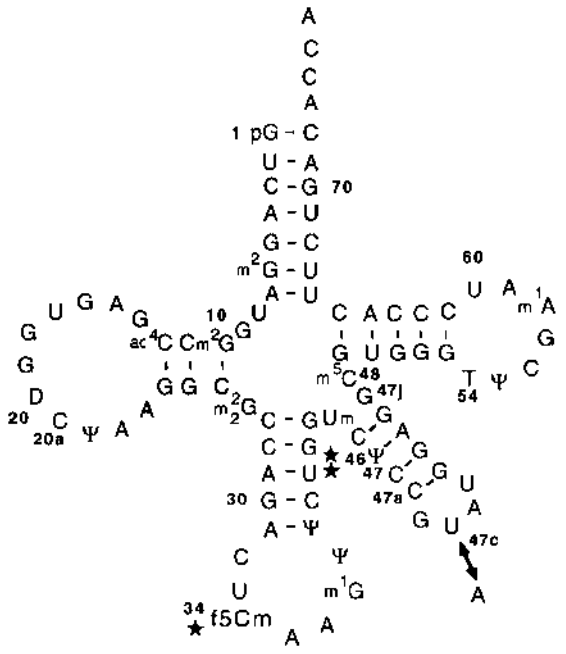

Figure 7. Cloverleaf drawing of the two bovine isoacceptor tRNAs-^{Leu}(f^{5}CmAA) differing by one nucleotide at position 47c (shown by an arrow). *, f^{5}Cm is the natural occurring nucleoside in the 'wobble' position 34. However, variable amounts of om^{5}Cm were also found in that position during sequencing (see text). **, Position 46 was undermodified: it contained both Y and unmodified U.

could be detected by HPLC nucleoside analysis in partially purified preparations of bovine tRNA^{Leu}(NAA) (Fig. 1a), as well as in crude preparations of unfractionated bovine tRNAs (results not shown), we assume that f^{5}Cm is the only nucleotide present at the 'wobble' position in the two native bovine tRNAs^{Leu}(NAA). Thus, the fluctuating amounts of om^{5}Cm found by sequencing can be only provided by a f^{5}Cm reduction, similar to that chemically performed in our above structural studies of f^{5}Cm, but induced here in variable levels by the numerous treatments or storages prior to sequencing.

According to these results, the final cloverleaf drawing of the two slightly differing tRNAs^{Leu}(NAA) isolated from bovine liver is presented in Figure 7.

DISCUSSION

In terms of primary structures, there are two cytoplasmic tRNAs-^{Leu}(f^{5}CmAA) in bovine liver that differ only by the nucleotide 47c which is either uridine or adenosine, whereas they carry the same f^{5}CmAA anticodon. Among all the tRNAs so far sequenced, these bovine tRNAs^{Leu} are the first tRNAs which contain 2′-*O*-methyl-5-formylcytidine (f^{5}Cm). However, a closely-related nucleoside, the 5-formylcytidine (f^{5}C), has been already described in the tRNA^{Met} isolated from mitochondria of the same mammalian tissue (bovine liver) (24,25), and of the nematode *Ascaris suum* (29).

When comparing the primary structures of the two bovine tRNAs^{Leu}(f^{5}CmAA) with those of the already published mammalian tRNAs^{Leu}(NAA) from rat Morris hepatoma (4) and human HeLa cells (5), several differences are displayed for some nucleotides of the extra loop: (i) in rat Morris hepatoma, N44 is mentioned to be an unknown modified U, but not Um, (ii) N47c is either A or U in both bovine liver and rat Morris hepatoma, whereas this heterogeneity in position 47c is replaced by a unique G47c in human HeLa cells, and (iii) G47j is absent from both Morris hepatoma and HeLa cells, while C45 is absent from only Morris hepatoma. These extra loop differences between the previously published primary structures of cytoplasmic tRNAs^{Leu}(MmAA) from rat Morris hepatoma (4) and human HeLa cells (5), and those of tRNAs^{Leu}(f^{5}CmAA) from bovine liver determined in the present paper, are underlined in the following schematic in-line models:

	43	44	46	47	47c					47j	48
Morris hepatoma	G	N	(Ψ/U)	C	C	G	(U/A)	A	U G G A G		m^{5}C
HeLa cells	G	Um C	Ψ	C	C	G	G	A	U G G A G		m^{5}C
Bovine liver	G	Um C	(Ψ/U)	C	C	G	(U/A)	A	U G G A G	G	m^{5}C

However, the most noteworthy feature of the mammalian tRNAs^{Leu}(NAA) is the presence of a new modified nucleotide found at their 'wobble' positions. This new nucleotide was partially characterized in rat Morris hepatoma as a probably C-5 substituted Cm, called Mm (4), and as the same Mm modified nucleotide in human HeLa cells (5), while it has been fully identified as f^{5}Cm in the present paper. These results suggest that the N34 nucleotides carried by the cytoplasmic tRNAs^{Leu}(NAA) of the three above mammalian species could have a unique chemical structure: f^{5}Cm. To argue for this hypothesis, studies were recently conducted in our laboratories on other purified tRNAs^{Leu}(NAA) isolated from rabbit and lamb liver. Using our analytical procedure, we obtained precise evidence for the presence of f^{5}Cm in the 'wobble' position of these tRNAs. Therefore, the 2′-*O*-methyl-5-formyl hypermodification of cytidine at position 34 could be a specific feature of the cytoplasmic mammalian tRNAs^{Leu}(NAA).

With regard to codon–anticodon interactions, the role and the significance of f^{5}Cm in the two cytoplasmic bovine tRNAs-^{Leu}(f^{5}CmAA) remain to be determined. However, they can be discussed in terms of decoding properties for the leucine codons, as compared with those previously described in different eucaryotic species, like yeasts and plants.

According to the genetic code, the leucine codons are located in two codon boxes: the CUN box which contains only leucine codons, and the UUN box which is shared between two UUPy phenylalanine codons and two UUPu leucine codons. In yeast, the latter two corresponding anticodons are carried by two tRNAs^{Leu}(XAA) which differ by 31 nucleotides in their primary structures, and by the length of their variable loop: 15 versus 13 nucleotides (6,30). In addition, the X modified nucleotide located at the 'wobble' position of yeast tRNAs^{Leu}(XAA) is either m^{5}C (30), or ncm^{5}Um (7), which restrict the recognition of the leucine codons to UUG or UUA, respectively (7). An identical situation, i.e. two different tRNAs reading each one of the UUPu leucine codons, could exist in the two cytoplasmic tRNAs^{Leu}(ZAA) isolated from plants (31,32). However, the Z modified nucleotides in position 34 of these tRNAs have not yet been identified.

Regarding the bovine liver cytoplasm, we show in the present paper that two slightly differing isoacceptor tRNAs^{Leu} carry the same anticodon, f^{5}CmAA. The question then arises as to whether the modified nucleotide f^{5}Cm could give to these two tRNAs^{Leu} the ability to translate both the UUG and UUA leucine codons. As compared with the conformational properties described for the structurally-related f^{5}C nucleotide of bovine mitochondria tRNA^{Met} (25), f^{5}Cm should take a rigid C3′-endo form similar to that of f^{5}C,

and that rigidity could even be re-enforced by the 2′-*O*-methylation on the ribose moiety (33,34). In addition, it should also present an intramolecular hydrogen bond between the carbonyl of the 5-formyl group and the 4-amino function (25). These conformational properties would avoid f^5Cm in the first position of the anticodon to form base pairs with U and C (33,35,36). Therefore, one function of f^5Cm in bovine $tRNAs^{Leu}(f^5CmAA)$ would be to prevent an incorrect recognition of the UUU and UUC phenylalanine codons. Another function of f^5Cm could be to restrict the decoding property of the two cytoplasmic bovine $tRNAs^{Leu}(f^5CmAA)$ to the only UUG leucine codon, similarly to what has been shown for f^5C by Takemoto *et al.* (37) in *in vitro* translation studies on $tRNA^{Met}(f^5CAU)$ from bovine mitochondria.

Therefore, and similarly to the situation mentioned above for the yeast and plant species, a second cytoplasmic bovine liver $tRNA^{Leu}(NAA)$, with N different from f^5Cm, would be involved in the decoding of the UUA leucine codon. Investigations to detect such a $tRNA^{Leu}$ in bovine and other mammalian species are underway in our laboratories. Other studies are devoted to confirm or not the f^5Cm structure for the unknown Mm nucleotide previously mentioned at the 'wobble' position in the $tRNAs^{Leu}(MmAA)$ from rat Morris hepatoma and human HeLa cells (4,5). The aim of these studies would be to accurately define the coding function(s) of this new modified cytidine, and to clarify the translation process of both the UUG and UUA leucine codons in mammals.

ACKNOWLEDGEMENTS

The expert technical assistance of C. Fix in preparing and sequencing $tRNA^{Leu}$ samples is warmly acknowledged. We are grateful to A. Athias who conducted the MS studies in the Laboratoire de Biochimie des Lipoprotéines (Université de Bourgogne, Dijon, France) managed by Prof. P. Gambert. We are also indebted to Dr B. Rayner, Université de Montpellier II, for technical advice in the borohydride reduction procedure on nucleosides, and Prof. J. A. McCloskey, University of Utah, for fruitful discussions in mass spectrometry. This research had financial support from Association pour la Recherche sur le Cancer, Ministère de la Recherche et de la Technologie, Ligue Bourguignonne contre le Cancer, and Centre National de la Recherche Scientifique.

REFERENCES

1 Pirtle,R., Kashdan,M., Pirtle,I. and Dudock,B. (1980) *Nucleic Acids Res.* **8**, 805–815.
2 Vasil'eva,I.G., Tukalo,M.A., Krikli'yi,I.A. and Matsuka,G.K. (1985) *Mol. Biol.* (USSR) **18**, 1073–1076.
3 Tukalo,M.A., Vasil'eva,I.G., Matsuka,G.K. and Vlassov,V.V. (1984) *Bioorganickaya Chimia* (USSR) **10**, 58–67.
4 Randerath,E., Gupta,R.C., Morris,H.P. and Randerath,K. (1980) *Biochemistry* **19**, 3476–3483.
5 Harada,F., Matsubara,M. and Kato,N. (1984) *Nucleic Acids Res.* **12**, 9263–9269.
6 El Adlouni,C., Desgrès,J., Dirheimer,G. and Keith,G. (1991) *Biochimie* **73**, 1355–1360.
7 Glasser,A.-L., El Adlouni,C., Keith,G., Sochacka,E., Malkiewicz,A., Santos,M., Tuite,M.F. and Desgrès,J. (1992) *FEBS Lett.* **314**, 381–385.
8 Fradin,A., Grühl,H. and Feldmann,H. (1975) *FEBS Lett.* **50**, 185–189.
9 Fournier,M., Dorizzi,M., Sarger,C. and Labouesse,J. (1976) *Biochimie* **58**, 1159–1167.
10 Keith,G., Pixa,G., Fix,C. and Dirheimer,G. (1983) *Biochimie* **65**, 661–672.
11 Keith,G. (1990) In Gehrke,C.W. and Kuo,K.C. (eds) *Chromatography and Modifications of Nucleosides.* Elsevier Chromatography Library Series, The Netherlands, Vol. 45A, pp. A103–A141.
12 Heitzler,J., Marechal-Drouard,L., Dirheimer,G. and Keith,G. (1992) *Biochim. Biophys. Acta* **1129**, 273–277.
13 Desgrès,J., Keith,G., Kuo,K.C. and Gehrke,C.W. (1989) *Nucleic Acids Res.* **17**, 865–882.
14 Gehrke,C.W., Desgrès,J., Keith,G., Gerhardt,K.O., Agris,P.F., Gracz,H., Tempesta,M.S. and Kuo,K.C. (1990) In Gehrke,C.W. and Kuo,K.C. (eds) *Chromatography and Modifications of Nucleosides.* Elsevier Chromatography Library Series, The Netherlands, Vol. 45A, pp. A159–A223.
15 Gehrke,C.W., Kuo,K.C., Davis,G.E., Suits,R.D., Waalkes,T.P. and Borek,E. (1978) *J. Chromatogr.* **150**, 455–476.
16 Gehrke,C.W., Kuo,K.C., McCune,R.A., Gerhardt,K.O. and Agris,P.F. (1982) *J. Chromatogr., Biomed. Applic.* **230**, 297–308.
17 Gehrke,C.W. and Kuo,K.C. (1989) *J. Chromatogr.* **471**, 3–36.
18 Matsuda,A., Inada,M., Nara,H., Ohtsuka,E. and Ono,A. (1993) *Bioorg. Med. Chem. Lett.* **3**, 2751–2754.
19 Kochetkov,N.K., Shibaev,V.N., Kost,A.A. and Zelinsky,N.D. (1971) *Tetrahedron Lett.* **22**, 1993–1971.
20 Barrio,J.R., Secrist III,J.A. and Leonard,N.J. (1972) *Biochem. Biophys. Res. Commun.* **46**, 597–604.
21 Secrist III,J.A., Barrio,J.R., Leonard,N.J. and Weber,G. (1972) *Biochemistry* **11**, 3499–3506.
22 Schram,K.H., Taniguchi,Y. and McCloskey,J.A. (1978) *J. Chromatogr.* **155**, 355–361.
23 Gehrke,C.W. and Kuo,K.C. (1990) In Gehrke,C.W. and Kuo,K.C. (eds) *Chromatography and Modifications of Nucleosides.* Elsevier Chromatography Library Series, The Netherlands, Vol. 45A, pp. A3–A71.
24 Moriya,J., Yokogawa,T., Wakita,K., Ueda,T., Nishikawa,K., Crain,P.F., Hashizume,T., Pomerantz,S.C., McCloskey,J.A., Kawai,G., Hayashi,N., Yokoyama,S. and Watanabe,K. (1994) *Biochemistry* **33**, 2234–2239.
25 Kawai,G., Yokogawa,T., Nishikawa,K., Ueda,T., Hashizume,T., McCloskey,J.A., Yokoyama,S. and Watanabe,K. (1994) *Nucleosides Nucleotides* **13**, 1189–1199.
26 Pang,H., Schram,K.H., Smith,D.L., Gupta,S.P., Townsend,L.B. and McCloskey,J.A. (1982) *J. Org. Chem.* **47**, 3923–3932.
27 McCloskey,J.A. (1985) *Anal. Chem. Symp. Ser.* **24**, 521–546.
28 McCloskey,J.A. (1990) *Methods Enzymol.* **193**, 825–842.
29 Watanabe,Y., Tsurui,H., Ueda,T., Furushima,R., Takamiya,S., Kita,K., Nishikawa,K. and Watanabe,K. (1994) *J. Biol. Chem.* **269**, 22902–22906.
30 Chang,S.H., Kuo,S., Hawkins,E. and Miller,N.R. (1973) *Biochem. Biophys. Res. Commun.* **51**, 951–955.
31 Green,G.A., Marechal-Drouard,L., Weil,J.-H. and Guillemaut,P. (1987) *Plant Mol. Biol.* **10**, 13–19.
32 Marechal-Edouard,L., Weil,J.-H. and Guillemaut,P. (1988) *Nucleic Acids Res.* **16**, 4777–4788.
33 Kawai,G., Hashizume,T., Yasuda,M., Miyazawa,T., McCloskey,J.A. and Yokoyama,S. (1992) *Nucleosides Nucleotides* **11**, 759–771.
34 Kawai,G., Yamamoto,Y., Kamimura,T., Tasegui,T., Serine,M., Matta,T., Limori,T., Watanabe,T., Miyazawa,T. and Yokoyama,S. (1992) *Biochemistry* **31**, 1040–1046.
35 Yokoyama,S., Watanabe,T., Murao,K., Ishikura,H., Yamaizumi,Z., Nishimura,S. and Miyazawa,T. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4905–4909.
36 Yokoyama,S. and Miyazawa,T. (1990) In Gehrke,C.W. and Kuo,K.C. (eds), *Chromatography and Modifications of Nucleosides.* Elsevier Chromatography Library Series, The Netherlands, Vol. 45B, pp. B159–B223.
37 Takemoto,C., Koike,T., Yokogawa,T., Benkowski,L., Spremulli,L.L., Ueda,T.A., Nishikawa,K. and Watanabe,K. (1995) *Biochimie,* **77**, 104–108.