Modified Nucleosides in Transfer RNA

James A. McCloskey*

Departments of Biopharmaceutical Sciences and Biochemistry, University of Utah, Salt Lake City, Utah 84112

Susumu Nishimura*

Biology Division, National Cancer Center Research Institute, Chuo-ku, Tokyo, Japan

Received March 4, 1977

One of the characteristics of transfer RNA (tRNA) is the presence of a variety of atypical or "modified" nucleosides, besides the four that are normally found in RNA. Modified nucleosides also occur in ribosomal RNAs (including 5S RNA), DNA, and eukaryotic messenger RNA, and their functions in these molecules are known to be important.

In the case of tRNA, the proportion of modified nucleosides is much higher than in most nucleic acids. Furthermore, the nucleosides found in tRNA exhibit a wide range of structural variations. To date, more than 50 modified nucleosides have been isolated and characterized.[1,2]

Many of these result from methylation in the base or at O-2′ of ribose, but there are also a number of "hypermodified" nucleosides which reflect more complex modification. Most of these nucleosides are thought to result from enzymatic modification of a parent nucleotide either in the tRNA molecule itself or in its precursor. The structures of the modified nucleosides thus far found in tRNAs, except the 2′-O-methylated nucleosides, are shown in Figure 1.

In the cloverleaf model of tRNA, most modified nucleosides are located in specific positions depending on their structure, rather than at random in the polynucleotide chain.[3] Some modified nucleosides such as ribothymidine (T), pseudouridine (ψ) and dihydrouridine (D) are present in the tRNAs of most organisms. Others, such as 4-thiouridine (s^4U), 1-methyladenosine (m^1A), N^2-methylguanosine (m^2G), N^2,N^2-dimethylguanosine ($m^2{}_2G$) and 5-methylcytidine (m^5C), are only present in prokaryotes, or only in eukaryotes, or in certain specific organisms. Hypermodified nucleosides are located either in the first position of the anticodon or next to the 3′ end of the anticodon. As discussed later, they presumably play an important role in recognition of codon sequences.

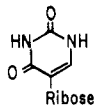
The symbols for modified nucleosides shown in Figure 1 are mostly in accord with the principles set up by the IUPAC–IUB Commission on Biochemical Nomenclature and are generally indicative of chemical structure. However, for some complex hypermodified nucleosides systematic abbreviation is difficult. Trivial names are often given, as for example in the case of "Y" or "Q" (Figure 1).

Since modified nucleosides differ from normal constituents in known specific ways, they can be employed as target sites for selective chemical modification to study the biological function of tRNA and tRNA conformation[4] or to isolate a specific tRNA that contains a reactive modified nucleoside.[5-7] Modified nucleosides are also excellent reporter molecules for analysis of tRNA conformation by physicochemical techniques such as ESR,[8-11] CD,[12,13] NMR,[14-19] fluorescence energy transfer,[20-23] thermal melting,[24] temperature jump,[25] or laser Raman spectroscopy.[26]

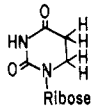
James A. McCloskey is Professor jointly in the Departments of Biopharmaceutical Sciences and Biochemistry at the University of Utah. He was born in San Antonio, Texas, in 1936, and received the B.S. degree from Trinity University and the Ph.D. from Massachusetts Institute of Technology, under Klaus Biemann. Before joining the faculty at Utah in 1974, he was Professor of Chemistry at Baylor College of Medicine. His research interests are in gaseous ion chemistry and applications of organic mass spectrometry, particularly to problems involving constituents of nucleic acids.

Susumu Nishimura was born in Tokyo in 1931 and obtained his B.S., M.S., and Ph.D. degrees from Tokyo University. During 1961–1965 he worked at Oak Ridge National Laboratory and in the Institute for Enzyme Research at the University of Wisconsin. He joined the National Cancer Center Research Institute in Tokyo in 1965, where he is now Chief of the Biology Division. Dr. Nishimura's research concerns the structure and function of transfer RNA, with particular regard to modified nucleosides.

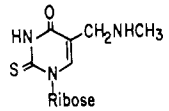
(1) S. Nishimura, *Prog. Nucleic Acid Res. Mol. Biol.*, **12**, 49 (1972).
(2) D. B. Dunn and R. H. Hall in "Handbook of Biochemistry and Molecular Biology", 3rd ed, Nucleic Acids, Vol. 1, G. Fasman, Ed., CRC Press, Cleveland, Ohio, 1975, pp 65, 217.
(3) M. A. Sodd in "Handbook of Biochemistry and Molecular Biology", 3rd ed, Nucleic Acids, Vol. 2, G. Fasman, Ed., CRC Press, Cleveland, Ohio, 1975, p 423.
(4) F. von der Haar, E. Schlimme and D. H. Gauss, *Proced. Nucleic Acid Res.*, **2**, 643 (1972).
(5) R. Solomon, S. Fuchs, A. Aharohov, D. Giveon, and U. Z. Littauer, *Biochemistry*, **14**, 4046 (1975).
(6) T. F. McCutchen, P. T. Gilham, and D. Söll, *Nucleic Acids Res.*, **2**, 853 (1975).
(7) N. Okada, N. Shindo-Okada, and S. Nishimura, *Nucleic Acids Res.*, **4**, 415 (1977).
(8) H. Hara, T. Horiuchi, M. Saneyoshi, and S. Nishimura, *Biochem. Biophys. Res. Commun.*, **38**, 305 (1970).
(9) M. Caron and H. Dugas, *Nucleic Acids Res.*, **3**, 35 (1976).
(10) N. Shalitin and J. Feitelson, *Biochemistry*, **15**, 2092 (1976).
(11) W. E. Daniel, Jr., and M. Cohn, *Biochemistry*, **15**, 3917 (1976).
(12) M. Saneyoshi, T. Ahami, S. Nishimura, and T. Samejima, *Arch. Biochem. Biophys.*, **152**, 677 (1972).
(13) K. Watanabe, T. Oshima, and S. Nishimura, *Nucleic Acids Res.*, **3**, 1703 (1976).
(14) K. L. Wong and D. R. Kearns, *Nature (London)*, **252**, 738 (1974).
(15) L. S. Kan, P. O. P. Ts'o, F. von der Haar, M. Sprinzl, and F. Cramer, *Biochem. Biophys. Res. Commun.*, **59**, 22 (1974).
(16) L. S. Kan, P. O. P. Ts'o, F. von der Haar, and M. Sprinzl, *Biochemistry*, **14**, 3278 (1975).
(17) R. V. Kastrup and P. G. Schmidt, *Biochemistry*, **14**, 3612 (1975).
(18) W. E. Daniel, Jr., and M. Cohn, *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 2582 (1975).
(19) R. J. Hoover, K. F. S. Luk, and A. H. Maki, *J. Mol. Biol.*, **89**, 363 (1974).
(20) C. R. Cantor and T. Tao, *Proced. Nucleic Acid Res.*, **2**, 31 (1972).
(21) R. Langlois, S.-H. Kim, and C. R. Cantor, *Biochemistry*, **14**, 2554 (1975).
(22) C.-H. Yang and D. Söll, *Proc. Natl. Acad. Sci. U.S.A.*, **71**, 2838 (1974).
(23) N. Okabe, K. Suyari, and K.-I. Tomita, *J. Biochem.*, **77**, 1335 (1975).
(24) T. Seno, K. Kobayashi, and S. Nishimura, *Biochim. Biophys. Acta*, **174**, 71 (1969).
(25) S. K. Yang and D. M. Crothers, *Biochemistry*, **11**, 4375 (1972).



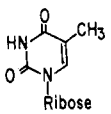
1. Pseudouridine (ψ)



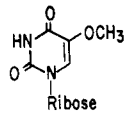
2. Dihydrouridine (D or hU)



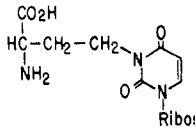
10. 5-Methylaminomethyl-2-thiouridine (mnm^5s^2U)



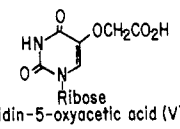
3. Ribothymidine (T)



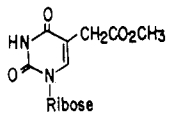
4. 5-Methoxyuridine (mo^5U)



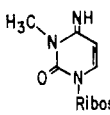
11. 3-(3-Amino-3-carboxypropyl) uridine (acp^3U)



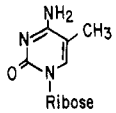
5. Uridin-5-oxyacetic acid (V)



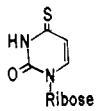
6. 5-(Methoxycarbonylmethyl) uridine (mcm^5U)



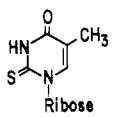
12. 3-Methylcytidine (m^3C)



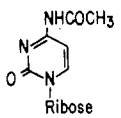
13. 5-Methylcytidine (m^5C)



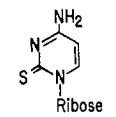
7. 4-Thiouridine (s^4U)



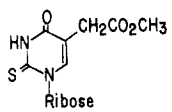
8. 5-Methyl-2-thiouridine (m^5s^2U)



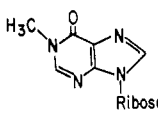
14. N^4-Acetylcytidine (ac^4C)



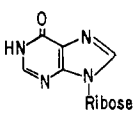
15. 2-Thiocytidine (s^2C)



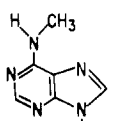
9. 5-(Methoxycarbonylmethyl)-2-thiouridine (mcm^5s^2U)



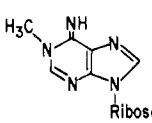
16. 1-Methylinosine (m^1I)



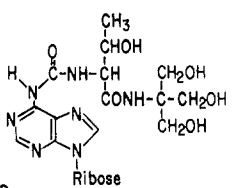
17. Inosine (I)



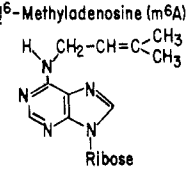
18. N^6-Methyladenosine (m^6A)



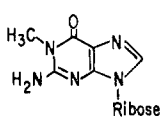
19. 1-Methyladenosine (m^1A)



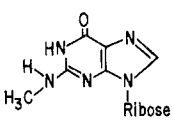
24. N-[N-[(9-β-D-Ribofuranosylpurin-6-yl) carbamoyl] threonyl] 2-amido-2-hydroxymethylpropane-1,3-diol[101]



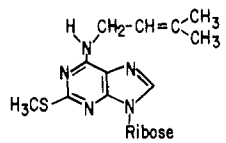
20. N^6-Isopentenyladenosine (i^6A)



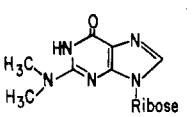
25. 1-Methylguanosine (m^1G)



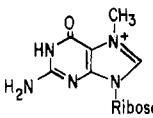
26. N^2-Methylguanosine (m^2G)



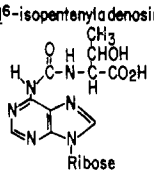
21. 2-Methylthio-N^6-isopentenyladenosine (ms^2i^6A)



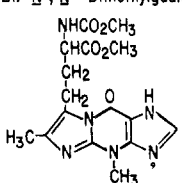
27. N^2, N^2-Dimethylguanosine (m_2^2G)



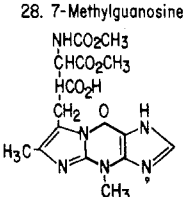
28. 7-Methylguanosine (m^7G)



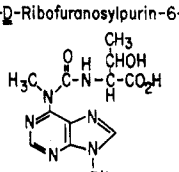
22. N-[N-(9-β-D-Ribofuranosylpurin-6-yl) carbamoyl] threonine (t^6A)



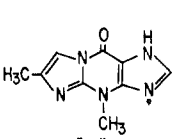
29. Base "Y" (yW)



30. Base "peroxy Y" (oyW)



23. N-[N-(9-β-D-Ribofuranosylpurin-6-yl) N-methylcarbamoyl] threonine (mt^6A)



31. Base "Yt" (W)

32. R=H Q (queuosine or Quo)[46]
R=β-D-mannosyl manQ[35]
R=β-D-galactosyl galQ[35]

Figure 1. Modified nucleosides from transfer RNA. In the case of bases "Y", "Yt", and "peroxy Y" the ribosyl group is presumably attached to N-9. References to characterization of the modified nucleosides are also cited in ref 1 and 2.

However the present Account deals only with the structure elucidation and function of modified nucleosides in tRNA.

Strategy for Structure Elucidation of Hypermodified Nucleosides

Inosine, pseudouridine, 4-thiouridine, and a number of methylated nucleosides were isolated and characterized more than a decade ago.[1,2] These modified nucleosides are present in most tRNA molecules, and so their detection and isolation in relatively large quantities are straightforward, unfractionated tRNA being used as a source. By contrast, hypermodified nucleosides are usually located in the anticodon region, in only one or several species of tRNA from a given source. As a result their detection in unfractionated tRNA is difficult, which is probably one reason that hypermodified nucleosides were not isolated during earlier work. On the other hand, their detection in pure species of tRNA is easier since they constitute at least 1% of the total nucleoside content.

After the presence of a new modified nucleoside in a specific tRNA is established, its isolation from unfractionated tRNA is facilitated by some knowledge of its chemical properties. In some instances, such as that of dihydrouridine from $tRNA^{Ala}$ of yeast,[27] relatively abundant modified nucleosides have been characterized during the course of primary sequence determination. It is often preferable to increase the modified nucleoside content by first separating oligonucleotides which contain the component of interest (an example is given in a later section). Knowledge of the exact location of the modified nucleoside in a specific tRNA molecule is possibly the only way to establish with certainty that it is a legitimate component of tRNA and not an impurity carried through the isolation procedure. When unfractionated tRNA is used as a source, this latter problem can be dealt with by first isolating the corresponding mono- or oligonucleotide before conversion to the nucleoside on which most of the structural work is carried out.

For detection of modified nucleosides in a particular tRNA, the tRNA is first hydrolyzed by RNase T_2 and the resulting nucleotide mixture is subjected to two-dimensional thin-layer or two-dimensional paper chromatography.[1] With this procedure, about 0.1 mg of purified tRNA is sufficient for detection of a modified component present at the level of one residue per tRNA molecule. Enzymatic hydrolysis of tRNA by RNase T_2,[28] snake venom phosphodiesterase, or nuclease P_1[29] is preferable to alkaline hydrolysis because some modified nucleosides are alkali labile. In general, 0.2 A_{260} unit (approximately 10 μg) of a modified nucleoside is sufficient for examination of its UV and mass spectra and measurement of its chromatographic and electrophoretic mobilities.

A known modified nucleoside can be characterized by comparison of data from the above measurements with those from authentic samples. As a minimum criterion for identification, the UV spectra should be compared, preferably at different pH values. Chromatographic mobility is also preferably measured in several solvent systems. Primary reliance should not be placed on elution position in high-pressure liquid chromatography or on electrophoretic mobility of the ^{32}P-labeled nucleotide. For example, 5-methoxyuridine recently isolated from *Bacillus subtilis* was found to exhibit chromatographic behavior identical with that of uridine.[30]

Gas chromatography has an inherent limitation in that the nucleoside must first be converted to a sufficiently volatile derivative,[31] which excludes about 25% of the presently known modified nucleosides from tRNA. However, when gas chromatography is applicable, relative elution positions of nucleosides can be measured with a precision of about one part in 10^3,[32-34] thus providing a means of comparison with an authentic compound that it is substantially more objective than in the case of other chromatographic methods. If the structure has not previously been established, several column chromatographic procedures can be used for isolation. Among the most useful are those involving DEAE-Sephadex A-25 and Dowex-1, as well as paper chromatography.

Limitations in quantity of material pose a great problem. The isolation of 1 to 10 mg of a new nucleoside is usually a laborious process. Physicochemical techniques such as FT NMR and mass spectrometry have proven to be especially valuable when sample quantities are limited. In FT NMR, spectra can be acquired with as little as 50 to 100 μg.[35] Mass spectrometry generally requires 2 to 10 μg for isolated material, or an order of magnitude less in favorable cases.

The sensitivity of mass spectrometry makes the technique well-suited for preliminary survey to differentiate new nucleosides from those whose structures are known, as well as for detailed structural work. The greatest limitation to the use of mass spectrometry results from the high polarity of many modified nucleosides, and thus the requirement that they first be chemically derivatized.[36] The constraints of volatility can often be overcome by field desorption[37] or the innovative technique which has been named plasma desorption mass spectrometry, which was recently demonstrated in the case of the Q* nucleoside[35] (see below). Both techniques are useful for the determination of molecular weight, although in the case of field desorption sensitivity may be adversely affected by the presence of traces of alkali salts.

Satisfactory chemical derivatization requires (a) good yields in the sample range 1–10 μg, (b) products that

(26) Y. Nishimura, A. Y. Hirakawa, M. Tsuboi, and S. Nishimura, *Nature (London)*, **260**, 173 (1976).
(27) J. T. Madison and R. W. Holley, *Biochem. Biophys. Res. Commun.*, **18**, 153 (1965).
(28) F. Egami, K. Takahashi, and T. Uchida, *Prog. Nucleic Acid Res.*, **3**, 59 (1964).
(29) M. Fujimoto, A. Kuninaka, and H. Yoshino, *Agric. Biol. Chem.*, **38**, 777 (1974).
(30) K. Murao, T. Hasegawa, and H. Ishikura, *Nucleic Acids Res.*, **3**, 2851 (1976).
(31) K. H. Schram and J. A. McCloskey, in "GLC and HPLC Determination of Therapeutic Agents", Vol. III, K. Tsuji and W. Morowitz, Ed., Marcel Dekker, New York, N.Y., in press.
(32) S. E. Hattox and J. A. McCloskey, *Anal. Chem.*, **46**, 1378 (1974).
(33) W. C. Butts, *Anal. Biochem.*, **46**, 187 (1972).
(34) A. P. DeLeenheer and C. F. Gelijkens, *Anal. Chem.*, **48**, 2203 (1976).
(35) H. Kasai, K. Nakanishi, R. D. Macfarlane, D. F. Torgerson, Z. Ohashi, J. A. McCloskey, H. J. Gross, and S. Nishimura, *J. Am. Chem. Soc.*, **98**, 5044 (1976).
(36) J. A. McCloskey in "Basic Principles in Nucleic Acid Chemistry", Vol. I, P. O. P. Ts'o, Ed., Academic Press, New York, N.Y., 1974, Chapter 3.
(37) H.-R. Schulten and H. D. Beckey, *Org. Mass Spectrom.*, **7**, 861 (1973).

are relatively volatile and thermally stable, and (c) mass spectra that are structurally informative. To aid in interpretation of the resulting mass spectra, it is helpful also if deuterium-labeled reagents can be employed. Of the various derivatization procedures which have been used for nucleosides,[36] these characteristics are best met by trimethylsilylation[38] and permethylation.[34,39]

When sample quantities permit, the preparation of two different derivatives minimizes risk in case the derivatization yield by one procedure is unexpectedly low, and provides a degree of structural information. For example, exocyclic amino groups (as in adenosine or guanosine) can be recognized by the fact that only one trimethylsilyl (net mass addition 72 amu) but two methyl (net addition 14 amu each) groups are incorporated under normal reaction conditions. Some reactions produce characteristic products for specific structural moieties, for example in the cases of 7-methylpurine nucleosides[40] and 5,6-dihydropyrimidines.[41]

The experimental conditions used for derivatization of an unknown nucleoside may vary depending on what structural features may be known or suspected and on the availability of material. However, a typical sequence of steps is as follows. (1) Two to four samples of 5–10 μg each are extensively dried and then derivatized, the preferred order being trimethylsilylation and then permethylation. (2) A low-resolution mass spectrum is obtained, using direct probe sample introduction as the minimum risk procedure. Gas chromatography–mass spectrometry may be tried in subsequent experiments if sufficient material is available. (3) Based on results at low resolution, the complete high-resolution mass spectrum of one of the two derivatives is obtained. Photographic recording is preferred[42] because a broad range of mass values is usually involved (e.g., 400–900 amu) and because sample vaporization times may be as short as 15 s. (4) The low-resolution mass spectrum is redetermined using deuterium-labeled blocking groups. This permits determination of the number of blocking groups introduced (which corroborates results obtained from use of two derivatives) and leads to the molecular weight of the native molecule. In the case of trimethylsilylation, the deuterium mass shifts permit determination of the number of Si atoms in any given ion,[43] which is valuable in reducing the number of computer-derived elemental composition possibilities derived from exact mass data. Normally data from the above sources enable identification of the basic nucleoside skeleton (among the four normal possibilities), the site and extent of modification, the number of active hydrogen atoms in the base and sugar, and the molecular mass and exact elemental compositions of the base moiety and total nucleoside.[36]

Examples of the procedures used are now outlined for nucleosides Q and Q*. Although in these examples the extent of modification was greater than in most cases, the approaches used are representative.

Structure Elucidation of Nucleosides Q and Q*

A modified nucleoside, designated Q, is found in the first position of the anticodons of *E. coli* $tRNA^{Tyr}$, $tRNA^{His}$, $tRNA^{Asn}$, and $tRNA^{Asp}$, that recognize the codons XA^{U}_{C}.[44] Recently the nucleoside has been named queuosine, and the base queuine. A derivative of Q (named Q*) is located in tRNA from animal tissues.[45]

The structures of nucleosides Q and Q* were determined to be as shown in Figure 1.[35,46] The unique structural features of Q are the presence of the unusual cyclopentenediol side chain and the 7-deazapurine[47] nucleus. This is the first known case of modification of the purine skeleton in a nucleoside from ribonucleic acid. The related Q* nucleoside was found to contain mannose or galactose residues linked to C-4 of the cyclopentene moiety, in an approximate ratio of 3:1. Q* is also the first modified nucleoside containing a sugar different from ribose and is the most structurally complex nucleoside thus far known from any natural source.

Large-scale isolation of Q for structural studies from pure tRNAs was not feasible due to the difficulty in obtaining them in sufficiently large quantity. It was recognized that the four tRNAs which contain Q always possess the common anticodon structure U-Q-U.[44] Therefore, isolation of Q was conveniently carried out via the dinucleotide Q-Up obtained by RNase A hydrolysis of unfractionated tRNA, followed by fractionation by DEAE-Sephadex A-25 column chromatography in the presence of 7 M urea.[46] This procedure separates oligonucleotides in order of their chain length or total negative charge.[48,49] The dinucleotide Q-Up elutes from the column earlier than most mononucleotides, thus providing a means for purification of Q. It should be noted that Q possesses one positive charge at neutral pH due to presence of a secondary amino group.

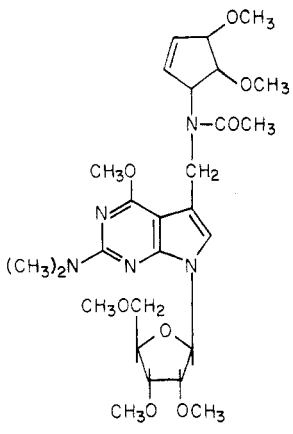
Fractionation of tRNA digests by DEAE Sephadex A-25 or Dowex-1 column chromatography is an excellent procedure for purification of other modified nucleosides. For example, the 3′-phosphate of uridin-5-oxyacetic acid (Vp) elutes immediately after the dinucleotide fraction of DEAE-Sephadex A-25 because it contains one additional acidic charge.[50] The threonine-containing adenosine derivative t^6Ap is eluted from Dowex-1 following the four principal mononucleotides. Hydrolysis of Q-Up by RNase T_2 produces Qp, which is readily isolated by Dowex-1 column chromatography and then enzymatically dephosphorylated to give the nucleoside Q. Approximately 70 A_{260} units (2 mg) was initially obtained in this fashion from 100 000 A_{260} units of unfractionated *E. coli* tRNA.[46] For

(38) J. A. McCloskey, A. M. Lawson, K. Tsuboyama, P. M. Krueger, and R. N. Stillwell, *J. Am. Chem. Soc.*, **90**, 4182 (1968).
(39) D. L. von Minden and J. A. McCloskey, *J. Am. Chem. Soc.*, **95**, 7480 (1973).
(40) D. L. von Minden, R. N. Stillwell, W. A. Koenig, K. J. Lyman, and J. A. McCloskey, *Anal. Biochem.*, **50**, 110 (1972).
(41) R. P. Panzica, L. B. Townsend, D. L. von Minden, M. S. Wilson, and J. A. McCloskey, *Biochim. Biophys. Acta*, **331**, 147 (1973).
(42) K. Biemann and P. V. Fennessey, *Chimia*, **21**, 226 (1967).
(43) J. A. McCloskey, R. N. Stillwell, and A. M. Lawson, *Anal. Chem.*, **40**, 233 (1968).
(44) F. Harada and S. Nishimura, *Biochemistry*, **11**, 301 (1972).
(45) H. Kasai, Y. Kuchino, K. Nihei, and S. Nishimura, *Nucleic Acids Res.*, **2**, 1931 (1975).
(46) H. Kasai, Z. Ohashi, F. Harada, S. Nishimura, N. J. Oppenheimer, P. F. Crain, J. G. Liehr, D. L. von Minden, and J. A. McCloskey, *Biochemistry*, **14**, 4198 (1975).
(47) Although the Q base is a pyrrolo[2,3-*d*]pyrimidine derivative, the conventional numbering system for purines is used for convenience.
(48) R. V. Tomlinson and G. M. Tener, *J. Am. Chem. Soc.*, **84**, 2644 (1962).
(49) G. W. Rushizky, E. M. Bartos, and H. A. Sober, *Biochemistry*, **3**, 626 (1964).
(50) F. Harada, F. Kimura, and S. Nishimura, *Biochemistry*, **10**, 3269 (1971).

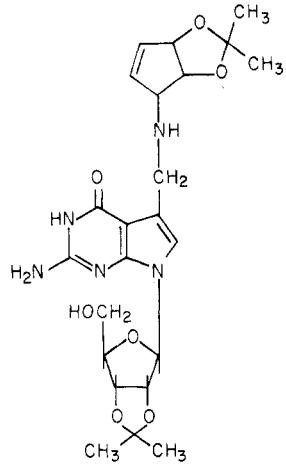
isolation of nucleosides Q and Q* from rabbit liver, the tRNA was hydrolyzed by alkali instead of RNase A because both compounds were known to be stable under alkaline conditions. The 3′(2′)-phosphates of Q and Q* were then isolated by Dowex-1 column chromatography[45] and were separated from each other by paper chromatography. Approximately 10 A_{260} units each of Q and Q* were obtained from 4 g of rabbit liver tRNA.

The structure of Q was deduced from its chemical properties, UV and proton NMR spectra, and the high-resolution mass spectra of several of its chemically prepared derivatives. The most useful of its chemical properties were its stability to acid hydrolysis (1 N HCl, 100 °C for 1 h), which is contrary to the normal behavior of purines, and resistance to exchange of deuterium into position C-8, which also reflects the absence of a normal purine skeleton. The electrophoretic mobility of Q suggested the presence of one positive charge at pH 7.5, while its UV spectrum was similar to that of guanosine, with λ_{max} shifted to longer wavelength (262 nm) and with a second peak appearing at 220 nm. Similarity of the UV spectrum to that of 7-deazaguanine[51] and 7-deazaguanosine[52] supported a 7-deaza structure, which had been indirectly suggested by mass spectrometry (see below).

Several early attempts to prepare volatile derivatives failed, but acetylation followed by permethylation was successful, leading to the derivative shown as **1**. The



1



2

most influential evidence was provided by the high-resolution mass spectrum of **1**. Interpreted in conjunction with the corresponding low-resolution mass spectrum, the elemental compositions generated by the high-resolution spectrum led to the following conclusions: (a) the elemental composition of native Q is $C_{17}H_{32}N_5O_7$, M_r 409; (b) one nitrogen atom is not bound directly to, or part of, the heterocyclic base moiety, which contains four nitrogen atoms; (c) a side chain containing five carbons, two oxygens, and two sites for methylation is present; and (d) the sugar moiety is unmodified.

Further evidence that the side chain contains a *cis*-diol grouping was gained from preparation of the bis-*O*-isopropylidene derivative **2**, whose molecular weight and structural features were determined by mass spectrometry. Additional support for the structure for Q shown in Figure 1 was gained from detailed analysis of the 220-MHz proton NMR spectrum. NMR data revealed 14 nonexchangeable protons, and most important, strongly support the existence of a 2-cyclopentene-4,5-diol moiety and a strongly deshielded diastereomeric methylene group. The observation of a downfield singlet at 7.170, which is shifted upfield about 1 ppm from where the C-8 protons of adenosine or guanosine absorb, is consistent with the behavior of 7-deazaadenosine (tubercidin) or 7-deazainosine.[53] Further, the observation of a single unique set of resonances for the cyclopentenediol ring indicates that the latter is a pure stereoisomer and supports its biological origin.

The structure of nucleoside Q* is the same as that of Q from the sugar through the methylene group at C-7, as shown by the prominent fragment ion m/e 656 in the mass spectrum of the trimethylsilyl derivative.[45] The molecular weight of Q* was established as 571 both by the new technique of plasma desorption mass spectrometry[54] of the underivatized molecule and by conventional mass spectrometry of the *N*-acetyl, permethyl derivative.[35] This value was recognized to correspond in mass to the Q structure plus substitution by an additional hexose unit, the elemental composition of which was confirmed by measurement of exact mass. Definitive information on the new and more complex portion of the molecule—that containing the hexose moiety—was deduced from the 260-MHz proton NMR spectra, analysis of which as a function of pH led to the identification of β-mannosyl and β-galactosyl residues in a ratio of 3:1 and to the site of attachment in the cyclopentenediol moiety, as shown in Figure 1. The identities and approximate ratios of the hexose units were confirmed by analysis of a 1 N H_2SO_4 (100 °C, 4 h) hydrolysate of Q* by thin-layer chromatography and gas chromatography.[35]

Extensive efforts to isolate new nucleosides have primarily involved tRNA from yeast and *E. coli*. However, examination of various other sources should also be fruitful, in particular Gram-positive bacteria, mammalian tissue, and plant cells. It is known that many tRNAs for which the primary structures have been determined contain unknown nucleosides. These unknown compounds are generally located in the anticodon region; determination of their structures will aid understanding of the mechanism of codon–anticodon interactions. The principal experimental difficulty in this regard is the isolation of sufficient amounts of material. The nature of this problem is illustrated by the fact that RNA^{Ile}_{minor} from *E. coli*, which only recognizes the codon AUA, contains a modified nucleoside in the first position of the anticodon.[55] However, the proportion of $tRNA^{Ile}_{minor}$ is less than 5% of major $tRNA^{Ile}$.

The Functions of Modified Nucleosides in tRNA

The structural complexity of many modified nucleosides, their location at specific sites in the tRNA

(51) J. Davoll, *J. Chem. Soc.*, 131 (1960).
(52) R. L. Tolman, Ph.D. Dissertation, University of Utah, Salt Lake City, Utah, 1969, p 135; L. B. Townsend, R. L. Tolman, R. K. Robins, and G. H. Milne, *J. Heterocycl. Chem.*, **13**, 1363 (1976).
(53) J. E. Pike, L. Slechta, and P. F. Wiley, *J. Heterocycl. Chem.*, **1**, 159 (1964).
(54) R. D. Macfarlane and D. F. Torgerson, *Science*, **191**, 920 (1976).
(55) F. Harada and S. Nishimura, *Biochemistry*, **13**, 300 (1974).

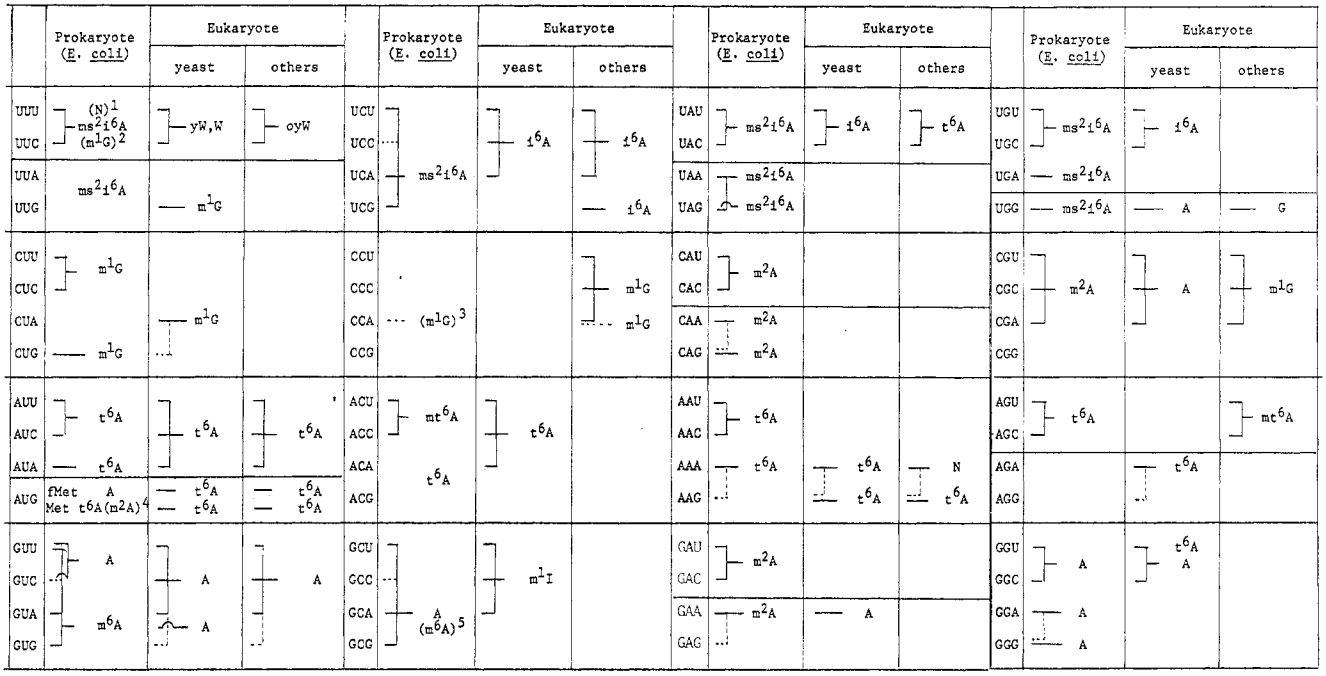


Figure 2. Relationship between the nucleoside located adjacent to the anticodon and codon recognition of tRNA. Data were taken from the references in the text and reports presented at a EMBO/DRC tRNA Workshop at Sandbjerg, Denmark (1976). N, unidentified modified nucleoside; 1, *Bacillus stearothermophilus*; 2, *Mycoplasma*, sp. kid; 3, T_4 phage; 4, *Bacillus subtilis*; 5, *Bacillus subtilis*. Dotted lines indicate uncertain codon recognitions.

molecule, and their ubiquitous presence in a wide range of organisms encourage us to believe that they play an important role in tRNA function, although direct proof is presently lacking. Here we present strong evidence that suggests certain functions for modified nucleosides, together with contradictory data.

Modified Nucleosides Located Adjacent to the Anticodon. Striking regularities are observed between the presence of particular modified nucleosides adjacent to the tRNA anticodon and recognition of the corresponding mRNA codon, as shown in Figure 2. Transfer RNAs which recognize codons starting with U almost always contain hydrophobic modified nucleosides, such as N^6-isopentenyladenosine (i^6A). On the other hand, most tRNAs which recognize codons starting with A contain hydrophilic modified nucleosides, such as *N*-[9-(β-D-ribofuranosyl)purin-6-ylcarbamoyl]threonine (t^6A). Thus it can be said that, so far as tRNA is concerned, the genetic code is read as four, rather than three, letters.

No similar consistency appears in tRNAs which recognize a codon starting with C or G. In these tRNAs, the rather simple methylated purine nucleosides or unmodified adenosine are present. The A·U pair is energically weaker than the G·C pair, since the former contains only two hydrogen bonds and the latter contains three. Hypermodification may be necessary for stabilization of A·U and enhancement of the fidelity of the base pair. In this connection it should be noted that *E. coli* $tRNA^{fMet}$, which contains adenosine instead of t^6A, recognizes GUG and UUG as initiator codons besides the usual AUG.[56] Exceptions to this regularity are known, however, as shown in Figure 2. For example, rat liver $tRNA^{Tyr}$ contains t^6A instead of i^6A,[57] and *B. subtilis* $tRNA^{Met}$ contains m^2A instead of t^6A.[58]

Several workers have shown that the presence of modified nucleosides next to the anticodon is essential for the amino acid transfer function of tRNA. *E. coli* suppressor $tRNA^{Tyr}$, which contains unmodified adenosine adjacent to the anticodon and can be aminoacylated, is nonetheless inactive with respect to amino acid transfer and ribosome binding.[59] Similar behavior has been observed with *E. coli* $tRNA^{Ile}$ which lacks t^6A.[60] Also, the binding of complementary oligonucleotides to the anticodon is enhanced by the presence of modified nucleosides next to the anticodon.[61] Chemically modified tRNAs, in which only modified nucleosides adjacent to the anticodons are transformed, are also inactive both in template-dependent binding of tRNA to ribosomes and in amino acid transfer.[62,63] However, it is known that tRNAs from *Lactobacillus acidophilus* and *Mycoplasma* sp. (kid), both of which lack i^6A, function normally in amino acid transfer reactions.[64-66] It has also been reported that removal of "Y" base from yeast $tRNA^{Phe}$ has no effect on the formation of a ternary complex with GTP and elongation factor.[67]

Modified Nucleosides Located in the First Position of the Anticodon. A number of hypermodified nucleosides are present in the first position of the anticodon (Figure 3) and are involved in codon–anticodon interactions. When located in this position, they

(56) J. G. Files, K. Weber, and J. H. Miller, *Proc. Natl. Acad. Sci. U.S.A.*, 71, 667 (1974).
(57) R. Brambilla, H. Robb, and M. Staehelin, *Nature (London)*, **263**, 167 (1976).
(58) H. Ishikura, personal communication.
(59) M. L. Gefter and R. L. Russell, *J. Mol. Biol.*, **39**, 145 (1969).
(60) J. P. Miller, Z. Hussoin, and M. P. Schweizer, *Nucleic Acids Res.*, **3**, 1185 (1976).
(61) O. Pongs and E. Reinwald, *Biochem. Biophys. Res. Commun.*, **50**, 357 (1973).
(62) F. Fittler and R. H. Hall, *Biochem. Biophys. Res. Commun.*, **25**, 441 (1966).
(63) R. Thiebe and H. G. Zachau, *Eur. J. Biochem.*, **5**, 546 (1968).
(64) H. Hayashi, H. Fisher, and D. Söll, *Biochemistry*, **8**, 3680 (1969).
(65) M. D. Litwack and A. Peterkofsky, *Biochemistry*, **10**, 994 (1971).
(66) M. E. Kimball and D. Söll, *Nucleic Acids Res.*, **1**, 1713 (1974).
(67) K. Ghosh and H. P. Ghosh, *J. Biol. Chem.*, **247**, 3369 (1972).

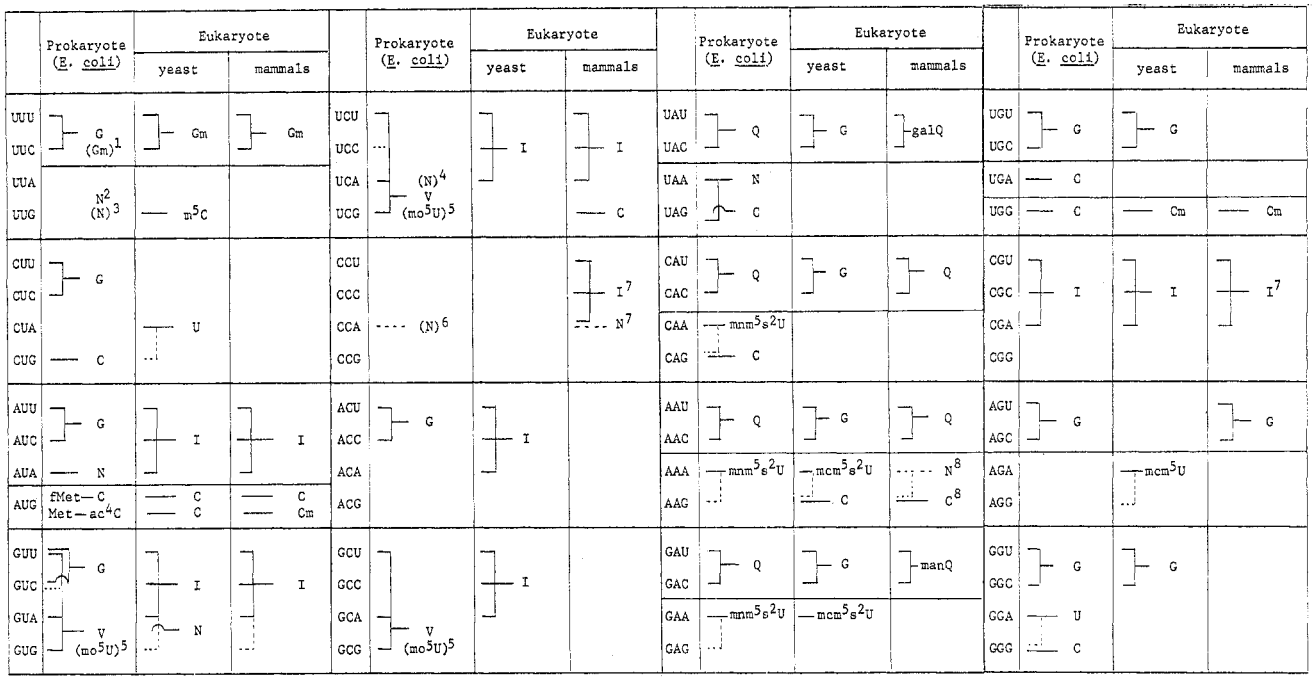


Figure 3. Relationship between the nucleoside located in the first position of the anticodon and codon recognition of tRNA. Data were taken from the references in the text and reports presented at a EMBO/DRC tRNA Workshop at Sandbjerg, Denmark (1976). N, unidentified modified nucleoside; 1, *Bacillus stearothermophilus*; 2, Z. Ohashi and S. Nishimura, unpublished; 3, T_4 phage; 4, T_4 phage; 5, *Bacillus subtilis*; 6, T_4 phage; 7, F. Harada, personal communication; 8, H. J. Gross, personal communication. Dotted lines indicate uncertain codon recognitions.

have been described by Crick as "wobbling bases".[68] Inosine (I) is the first such nucleoside, pairing with U, C, and A in the third position of the codon sequence. Uridin-5-oxyacetic acid (V), found in *E. coli*, recognizes A, G, and U in the third position of the codon, when assayed by binding of tRNA to ribosomes.[69] Recently 5-methoxyuridine (mo^5U) was found instead of V in *B. subtilis* tRNA.[30,70] Organisms may be classified into three groups depending upon whether V, mo^5U, or I acts as a "wobbling base" in the anticodon of tRNA for valine, serine, or alanine.

Contrary to the "wobble base" hypothesis, 2-thiouridine achieves strict base-pairing with A, but not with G, in the third letter of the codon sequence, as clearly shown in the case of yeast $tRNA^{Glu}_3$ which contains 5-(methoxycarbonylmethyl)-2-thiouridine (mcm^5s^2u).[71] Preferential recognition of A in the first position of the anticodon has been found in other 2-thiouridine-containing tRNAs, such as *E. coli* $tRNA^{Glu}$ [72] and $tRNA^{Gln}$,[73] and rabbit (reticulocyte and liver) $tRNA^{Lys}$, $tRNA^{Glu}$, and $tRNA^{Gln}$.[74,75] The *E. coli* $tRNA^{Ile}_{minor}$, which contains an unidentified modified nucleoside, recognizes only AUA;[55] this minor species presumably regulates the rate of synthesis of specific proteins which use only AUA codon as a code word.[76]

Thus it initially appeared that the function of modified nucleosides in the first position of the anticodon was clearly understood. However, it has been shown recently that rabbit liver $tRNA^{Val}$, which contains inosine, recognizes all four codons for valine, with preference being given to G-U-G.[77] Also, the $tRNA^{Val}$s from yeast and *E. coli* recognize the four valine codons in MS2 coat protein messenger, regardless of whether $tRNA^{Val}$ contains I, V, or G.[78]

With very few exceptions, almost all tRNAs so far sequenced contain modified U and A in the first position of the anticodon. Apparently nature does not allow the occurrence of normal U or A, which is presumably lethal to cells if present. In the cases of $tRNA^{Leu}$ (for the UU series), and $tRNA^{Gln}$, $tRNA^{Lys}$, $tRNA^{Glu}$, and $tRNA^{Arg}$ (AG series), the pairing of U in the first position of the anticodon with U or C in messenger RNA may cause miscoding which would be lethal to cells. Modification of U to 2-thiouridine derivatives may prevent such mispairing. It should be noted that 2-thiouridine derivatives have been found only in some of the tRNAs mentioned above.

The modified nucleoside Q has a wobbling property similar to that of G, although it has a greater affinity for U than C.[44] Thus the function of Q in protein biosynthesis is presently unclear.

Modified Nucleosides Present in Other Positions of the tRNA Molecule. There are data which suggest functions for modified nucleosides located in regions other than the anticodon. They are: (a) involvement of ribothymidine and pseudouridine in binding of tRNA in ribosomes,[79] (b) stabilization of the conformation of

(68) F. H. C. Crick, *J. Mol. Biol.*, **19**, 548 (1966).
(69) T. Takemoto, K. Takeishi, S. Nishimura, and T. Ukita, *Eur. J. Biochem.*, **38**, 489 (1973).
(70) M. Albani, W. Schmidt, H. Kersten, K. Geibel, and I. Luderwald, *FEBS Lett.*, **70**, 37 (1976).
(71) T. Sekiya, K. Takeishi, and T. Ukita, *Biochim. Biophys. Acta*, **182**, 411 (1969).
(72) Z. Ohashi, M. Saneyoshi, F. Harada, H. Hara, and S. Nishimura, *Biochem. Biophys. Res. Commun.*, **40**, 866 (1970).
(73) M. Yaniv and W. R. Folk, *J. Biol. Chem.*, **250**, 3243 (1975).
(74) K. Hilse and E. Rudloff, *FEBS Lett.*, **60**, 380 (1975).
(75) E. Rudloff and K. Hilse, *Hoppe-Seyler's Z. Physiol. Chem.*, **356**, 1359 (1975).
(76) W. M. Jou, M. V. Montagu, and W. Fiers, *Biochem. Biophys. Res Commun.*, **73**, 1083 (1976).
(77) P. Jank, N. Shindo-Okada, S. Nishimura, and H. J. Gross, *Nucleic Acids Res.*, in press.
(78) S. K. Mitra, F. Lustig, B. Åkesson, U. Largerkvist, and L. Strid, *J. Biol. Chem.*, **252**, 471 (1977).
(79) M. Sprinzl, T. Wagner, D. Lorenz, and V. A. Erdmann, *Biochemistry*, **15**, 3031 (1976).

tRNA,[13,80-82] (c) enhancement of resistance of tRNA to attack by RNase and nuclease, and (d) enhancement of specificity of the recognition of aminoacyl-tRNA synthetase.[83]

Other data, however, contradict the above indications. Mutants of *E. coli* have been isolated with tRNAs completely lacking 7-methylguanosine, 4-thiouridine, or ribothymidine,[84-86] yet they grow normally. Purified methyl-deficient $tRNA^{fMet}$ of *E. coli* functions normally in aminoacylation, formylation, recognition of code, and interaction with initiation factor.[87] *E. coli* $tRNA^{Val}_1$, in which T and ψ in the GTψC loop are replaced by 5-fluorouridine, can function in aminoacylation and in vitro protein synthesis.[88,89] Moreover, several amino acid specific tRNAs from mammalian or plant tissues do not contain ribothymidine.[3] Finally, some species of microorganisms almost completely lack ribothymidine in their tRNA.[90-93]

A mutant of *E. coli* has been found which lacks ribothymidine but appears to be normal with respect to growth rate and other criteria. However, when the mutant was grown together with the original wild-type strain of *E. coli*, the mutants were overcome by the wild type after several generations.[93] It is possible that the presence of ribothymidine slightly enhances tRNA function in a way difficult to detect by an in vitro biochemical assay. Nevertheless just such a slight advantage gained by having modified nucleosides may be decisive for maintaining organisms in their natural environment.

(80) T. Igo-Kelmers and H. G. Zachau, *Eur. J. Biochem.*, 18, 292 (1971).
(81) W. Wintermeyer and H. G. Zachau, *FEBS Lett.*, 58, 306 (1975).
(82) K. Watanabe, T. Oshima, M. Saneyoshi, and S. Nishimura, *FEBS Lett.*, 43, 59 (1974).
(83) B. Roe, M. Michael, and B. Dudock, *Nature (London), New Biol.*, 246, 135 (1973).
(84) I. Svensson, K. Isaksson, and A. Henningsson, *Biochim. Biophys. Acta*, 238, 331 (1971).
(85) M. G. Marinus, N. R. Morris, D. Söll, and T. C. Kwong, *J. Bacteriol.*, 122, 257 (1975).
(86) T. V. Ramabhadran, T. Fossum, and J. Jagger, *J. Bacteriol.*, 128, 671 (1976).
(87) J. B. Marmor, H. W. Dickerman, and A. Peterkofsky, *J. Biol. Chem.*, 246, 3464 (1971).
(88) J. Horowitz, C.-N. Ou, M. Ishaq, J. Ofengand, and J. Bierbaum, *J. Mol. Biol.*, 88, 301 (1974).
(89) J. Ofengand, J. Bierbaum, J. Horowitz, C.-N. Ou, and M. Ishaq, *J. Mol. Biol.*, 88, 313 (1974).
(90) L. Johnson, K. Hayashi, and D. Söll, *Biochemistry*, 9, 2823 (1970).
(91) A. N. Delk, J. M. Romeo, D. P. Nagle, Jr., and J. C. Rabinowitz, *J. Biol. Chem.*, 251, 7649 (1976).
(92) J. Ramakrishnan and S. Nishimura, unpublished results.
(93) G. R. Björk and P. C. Neidhardt, *J. Bacteriol.*, 124, 99 (1975).

The Possible Role of Modified Nucleosides in the Regulatory Function of tRNA

Pseudouridine in $tRNA^{His}$ from *Salmonella* has been found to be important in the function of tRNA in regulatory expression of the histidine operon.[94] Possibly other modified nucleosides are similarly involved. Modified nucleosides must constitute suitable sites for specific interaction of tRNA with protein or other components related to regulation, as reflected by their differences in structure, charge, and presence of certain reactive groups, when compared with normal nucleosides. It has been suggested that the Q nucleoside in *Drosophila* tRNA is related to cell differentiation.[95] Perhaps inhibition of the enzymatic activity of tryptophan pyrrolase by *Drosophila* $tRNA^{Tyr}$ from a vermilion mutant depends on whether the $tRNA^{Tyr}$ contains Q or guanosine in the first position of the anticodon,[96] despite contradictory evidence.[97]

In tRNA from tumor cells, the levels of Q-containing tRNAs are greatly changed.[98-101]

We have recently shown that rabbit liver $tRNA^{Asp}$ contains a Q* species that contains mannose, while $tRNA^{Tyr}$ contains a Q* species which bears galactose. The sugars in Q* therefore appear not to be incorporated at random, but rather in specific tRNAs. It has also been demonstrated that plant agglutinin, such as concanavalin A or *Ricinus communis* lectin, interacts with tRNA containing Q*.[7] It is noteworthy that mannose and galactose are components that specify the properties of glycoprotein receptors in membranes. The presence of these sugars in specific tRNAs may therefore indicate a role for these tRNAs in membrane function.

We gratefully acknowledge the support of this work by the National Science Foundation (OIP 74-22813 and BMS 74-20123).

(94) C. E. Singer, G. R. Smith, R. Cortese, and B. N. Ames, *Nature (London), New Biol.*, 238, 72 (1972).
(95) B. N. White, G. M. Tener, J. Holden, and D. T. Suzuki, *J. Mol. Biol.*, 74, 635 (1973).
(96) K. B. Jacobson, *Nature (London) New Biol.*, 231, 17 (1971).
(97) D. Mischke, P. Kloetzel, and M. Schwochau, *Nature (London)*, 254, 79 (1975).
(98) B. S. Baliga, E. Borek, I. B. Weinstein, and P. R. Srinivasan, *Proc. Natl. Acad. Sci. U.S.A.*, 62, 899 (1969).
(99) T. Sekiya and K. Oda, *Virology*, 47, 168 (1972).
(100) J. R. Katze, *Biochim. Biophys. Acta*, 383, 131 (1975).
(101) H. Kasai, K. Murao, S. Nishimura, J. G. Liehr, P. F. Crain, and J. A. McCloskey, *Eur. J. Biochem.*, 69, 435 (1976).