Tautomerism of Nucleic Acid Bases. I. Cytosine[1]

Gregory C. Y. Lee, J. H. Prestegard, and Sunney I. Chan*

Contribution No. 4241 from the Arthur Amos Noyes Laboratory of Chemical Physics, California Institute of Technology, Pasadena, California 91109.
Received April 22, 1971

Abstract: It has long been noted that the H_5 proton of cytosine and its related derivatives exhibits unusually broad resonances in the pmr spectrum under certain conditions of temperature and solution pD. We have examined this phenomenon as a function of temperature, concentration, and solution pD, as well as the external magnetic field, and have shown that the line broadening arises from chemical exchange between the amino and imino tautomers of the cytosine base. The observation of sharp H_5 resonances in cytosine derivatives existing in only the amino or imino tautomeric structure supports this interpretation. This tautomeric exchange was found to be catalyzed by OD^- as well as the solvent molecules (D_2O). Analysis of the temperature and pD dependences of the H_5 line width for 5′-CMP led to the following rate law at 30°, $-(d[A]/dt)_t = 1.2 \times 10^{10}[OD^-][A] + 1.2[D_2O][A]$ (M sec^{-1}), where [A] represents the major tautomeric species of 5′-CMP. Activation energies of 3.5 and 7.8 kcal/mol were obtained for the OD^- and D_2O catalyzed steps, respectively. The imino tautomer was estimated to be present to the extent of 15 ± 3% at room temperature in neutral aqueous solution.

The transfer of genetic information during the replication of DNA and during RNA transcription for protein synthesis relies on the unique pairing of the complementary nucleic acid bases.[2,3] It has, however, been suggested that these bases can exist in other tautomeric forms in addition to the principal tautomeric structure, and the presence of these minor base tautomers has been postulated as an important source of the imperfect pairing which can lead to genetic mutation.[4] The cytosine base, for example, may exist in either its normal amino or abnormal imino form.[5,6] It is well known that the normal amino form pairs with guanine to form a stable Watson–Crick G–C base pair; the imino tautomer, by contrast, has the appropriate electronic structure for pairing with the adenine base to form the abnormal A–C base pair (Figure 1). In view of the important biological implications of the above tautomerism, it is desirable to obtain a detailed understanding of the thermodynamic as well as kinetic properties of the amino–imino system. The possibility of this tautomerism has been examined by uv as well as ir spectrophotometric studies of cytosine and its derivatives,[7–9] and, more recently, there have been some theoretical attempts to predict the existence of this tautomerism and to ascertain the energetics involved.[10–14] In spite of these efforts, our understanding of the problem is still far from satisfactory, and, in particular, experimental data are lacking on the kinetics of the tautomerism.

It is shown in this paper that proton magnetic resonance spectroscopy can be an effective method for the examination and the elucidation of the amino–imino tautomerism in cytosine. It has frequently been observed that the H_5 resonance in the pmr spectra of cytosine and related compounds in aqueous solution exhibits unusual broadening under certain conditions of temperature and pH.[15] This line broadening can be shown to arise from tautomeric exchange between the amino and the imino species, and from a detailed study of the temperature and pH dependences of the H_5 line broadening, we have been able to determine the equilibrium as well as the dynamical properties of this tautomerism.

Experimental Section

Materials. A number of cytosine nucleosides and nucleotides were selected for this work. These compounds were obtained from various sources: cytosine and cytidine were purchased from Calbiochem, Los Angeles, Calif.; cytidine 2′-monophosphate (2′-CMP) was obtained from P-L Biochemicals, Inc., Milwaukee, Wis.; and 3′-CMP was obtained from Sigma Chemical Co., St. Louis, Mo.; 3-methylcytidine (methosulfate salt) was obtained from Mann Research Laboratory, New York, N. Y.; 4-*N*,*N*-dimethylcytosine was synthesized from 4-ethoxy-2-hydroxylpyrimidinone provided by Cyclo Chemical Corp., Los Angeles, Calif., according to the procedure of Wempen, *et al.*;[16] cytidine 5′-triphosphate (5′-CTP) was also purchased from this company. 5′-CMP, which was used as a model compound in our kinetic studies, was purchased from the following three sources: (1) Sigma Chemical Co., (2) Calbiochem, and (3) Cyclo Chemical Corp.

The cytosine H_5 resonance of various derivatives was found to vary with the source of the sample as well as the method of purification indicating the presence of contaminants. However, after the original sample was recrystallized in a water–ethanol mixture[16] and then passed through a Dowex 50W-X8 cation exchanger which was washed thoroughly with EDTA before use, the line width was reproducible from sample to sample. The absence of paramagnetic impurities [Mn^{2+}, Ni^{2+}, Co^{2+}, Cu^{2+}] at significant levels after purification was confirmed by atomic absorption analysis of the sample by Schwarzkopf Microanalytical Laboratory, Woodside, N. Y. The experimental results reported in this work have been obtained on samples purified in the above manner.

(1) This work was supported in part by Grant GM 14523-04 and -05 from the National Institute of General Medical Sciences, U. S. Public Health Service, and Grant GP-8540 from National Science Foundation.
(2) J. D. Watson and F. H. Crick, *Nature*, 171, 737 (1953).
(3) J. D. Watson and F. H. Crick, *ibid.*, 171, 964 (1953).
(4) J. D. Watson and F. H. Crick, *Cold Spring Harbor Symp. Quant. Biol.*, **18**, 123 (1953).
(5) G. W. Kenner, C. B. Reese, and A. R. Todd, *J. Chem. Soc.*, 855 (1955).
(6) A. R. Katritzky and A. J. Waring, *ibid.*, 1540 (1962); 3046 (1963).
(7) D. M. Brown and M. J. E. Hewlins, *ibid.*, 2050 (1968).
(8) R. V. Wolfenden, *J. Mol. Biol.*, **40**, 307 (1969).
(9) H. T. Miles, *Proc. Nat. Acad. Sci.*, **47**, 791 (1961).
(10) A. Pullman and B. Pullman, *Biochim. Biophys. Acta*, **64**, 403 (1962).
(11) A. Pullman and B. Pullman, *ibid.*, **75**, 269 (1963).
(12) P. O. Löwdin, "Advances in Quantum Chemistry," Vol. II, Academic Press, New York, N. Y., 1965, pp 213–360.
(13) T. A. Hoffman and J. Ladik, "Advances in Chemical Physics," Vol. 7, Interscience, New York, N. Y., 1964, Chapter 2.
(14) N. Bodor, M. J. S. Dewar, and A. J. Harget, *J. Amer. Chem. Soc.*, **92**, 2929 (1970).
(15) B. W. Bangerter and S. I. Chan, *ibid.*, **91**, 3910 (1969).
(16) I. Wempen, R. Duschinsky, L. Kaplan, and J. J. Fox, *ibid.*, 83, 4755 (1961).

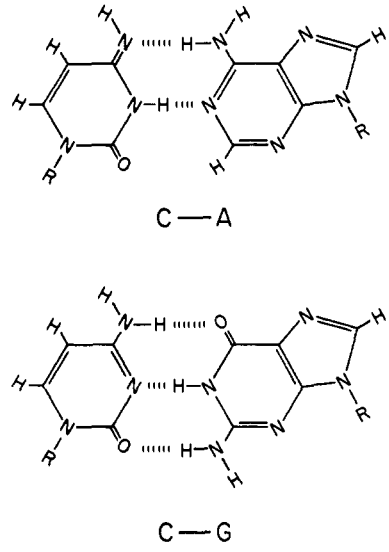


Figure 1. Abnormal A–C base pair and normal G–C base pair.

Method. Sample solutions were prepared at concentrations of 0.01–0.05 *M* in 99.7% D_2O supplied by Columbia Organic Chemicals, Columbia, S. C. Some of the samples were lyophilized from D_2O prior to sample preparation, in order to reduce the magnitude of the HDO peak in the spectra. The pD of each solution was measured with a Leeds and Northrup 7401 pH meter, equipped with miniature electrodes, and was calculated from the observed pH meter reading plus 0.4 (the standard correction).[17] pD adjustments were made with either 1 *M* DCl or 1 *M* NaOD.

The pmr spectra of the various nucleoside and nucleotide solutions were recorded on Varian HA-100 and HR-220 nmr spectrometers. A C-1024 time-average computer was used to enhance the signal-to-noise ratio. Chemical shifts were measured relative to an external TMS capillary, which also provided the field/frequency lock signal for HA-100 operation. The error in the measurement of the chemical shifts and line widths is ±0.5 Hz. The ambient probe temperature was 30 ± 1° in the case of the HA-100 spectrometer and 17 ± 1° for the HR-220. In the variable-temperature studies, the sample temperature was controlled to ±1° by means of a variable temperature controller and was measured using the methanol and ethylene glycol standards together with the calibration curves supplied by Varian.

Results

The pmr spectra of cytosine and many of its nucleoside and nucleotide derivatives have previously been reported and discussed.[15,18,19] In our present work, we shall primarily be concerned with the spectral behavior of the H_5 and H_6 resonances of the cytosine base. These resonances appear as doublets due to the spin–spin coupling between these protons and are readily identified by the spin–spin splitting of 8 Hz. The H_5 doublet is found 6 ppm downfield from TMS, and the H_6 doublet 8 ppm downfield from TMS. The H_6 resonances are always well defined and relatively sharp; by contrast, the H_5 resonances are frequently broadened, often so much as to escape observation. This H_5 line broadening is both pD and temperature dependent. The elucidation of this previously unexplained line broadening is the principal objective of this work.

pD Studies. As illustrated in Figures 2 and 3, the spectral positions of the H_5 and H_6 resonances of cytosine and its nucleoside and nucleotide derivatives are pD dependent. The chemical shifts of these protons, for example, change abruptly in the pD region 4–5, where ring protonation is known to occur. Above pD

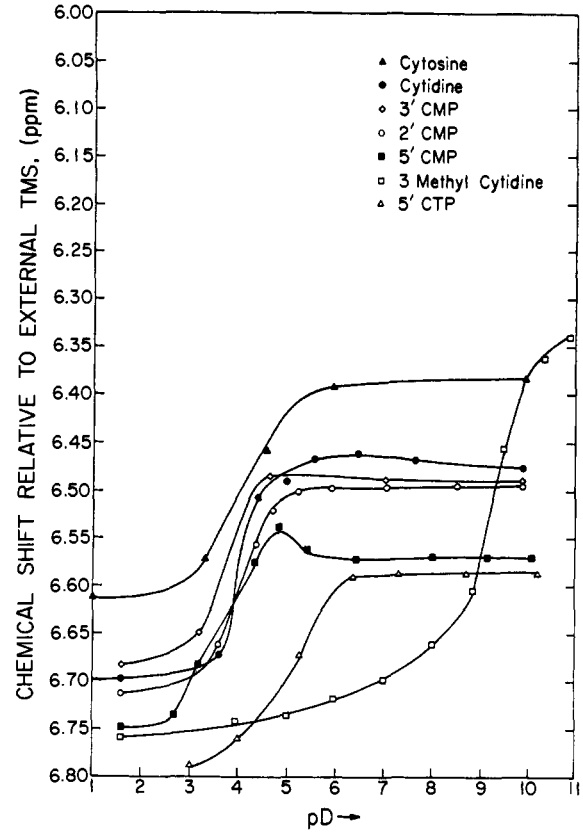


Figure 2. pD dependence of the chemical shifts of the H_5 resonances of cytosine and related compounds at 30°.

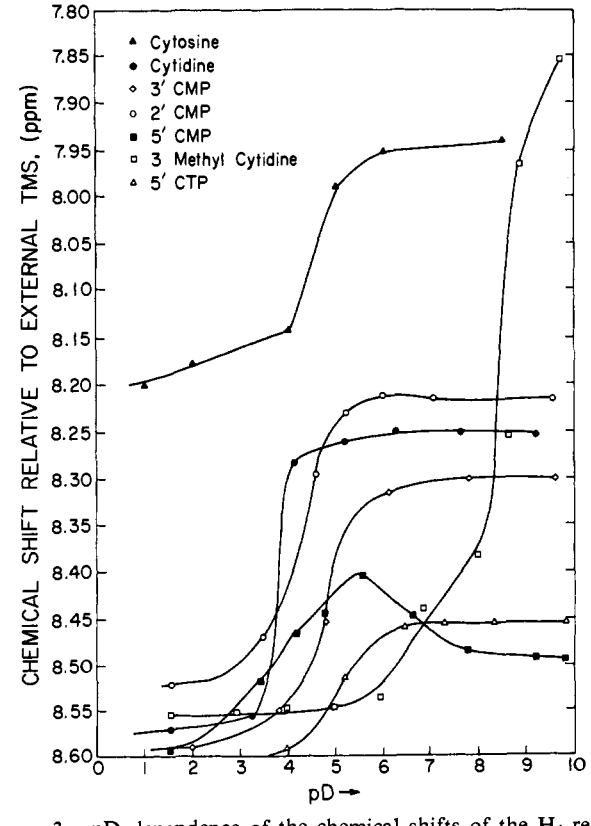


Figure 3. pD dependence of the chemical shifts of the H_6 resonances of cytosine and related compounds at 30°.

(17) R. Lumry, E. L. Smith, and R. R. Glantz, *J. Amer. Chem. Soc.*, 73, 4335 (1951).
(18) M. P. Schweizer, A. D. Broom, P. O. P. Ts'o, and D. P. Hollis, *ibid.*, **90**, 1042 (1968).
(19) S. S. Danyluk and F. E. Hruska, *Biochemistry*, 7, 1038 (1968).

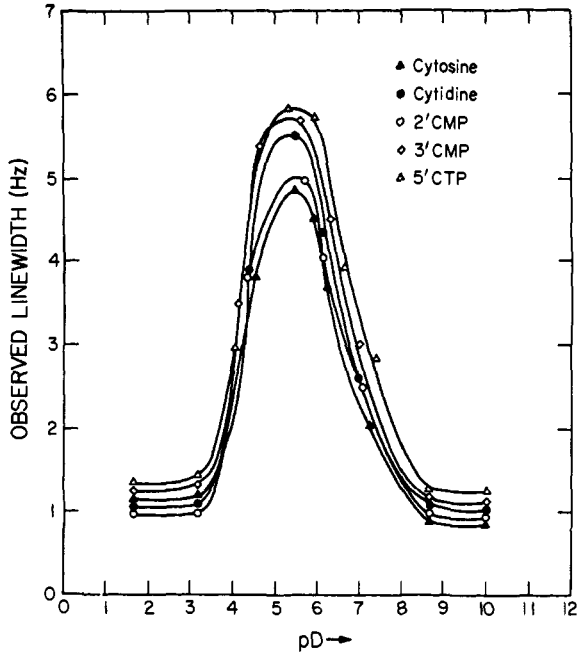


Figure 4. pD dependence of the line widths of the H_5 resonances of cytosine and related compounds at 30° and 100 MHz.

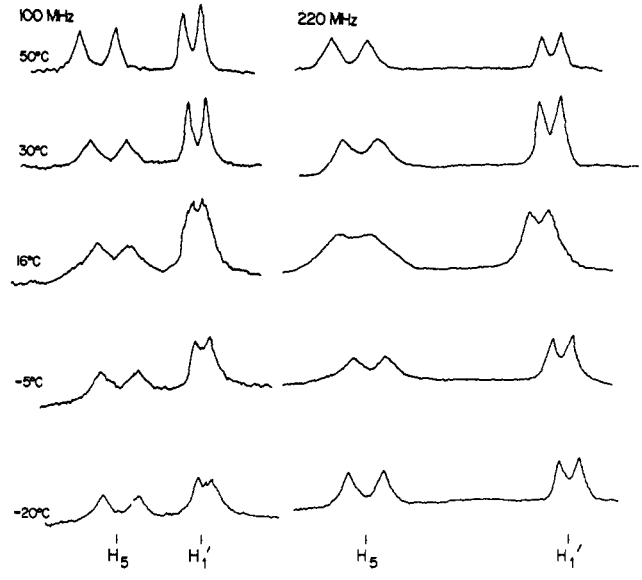


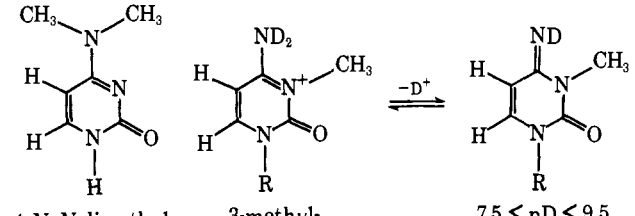
Figure 5. The 100- and 220-MHz pmr spectra of the H_5 resonance of 5'-CMP at various temperatures.

5, however, the chemical shifts are relatively independent of pD, except in the case of 5'-CMP, where effects due to ionization of the phosphate group at pD 6 are apparent. The phosphate group ionization is important, since in pyrimidine nucleotides, the conformation of the base about the glycosidic bond is expected to be predominantly anti,[18-20] and in this base conformation, the phosphate group is in juxtaposition with that part of the cytosine base bearing the H_5 and H_6 protons. As expected, the deshielding effect of the negatively charged phosphate group is more pronounced for the H_6 than for the H_5 proton because of the closer proximity of the phosphate group to the H_6 proton.

In contrast to the essentially constant chemical shifts, the line width of the cytosine H_5 resonance is strongly pD dependent in the pD range 5–7. In the case of cytosine, for example, the H_5 line width increases abruptly in the pD range 3–5, reaching a maximum of 4.8 Hz at pD 5.5, and then decreasing to a width of 1.5 Hz at pD's above 8.0. The width of the H_6 resonance is also pD dependent but to a much lesser extent. Its pD behavior is qualitatively similar to that observed for H_5 except that the H_6 line width reaches a maximum of only 1.4 Hz at pD 5.0.

As shown in Figures 4 and 11, the extent of the broadening of the H_5 resonance, as well as the pD range over which the broadening is important, depends only slightly on the compounds examined. The addition of a ribose moiety to the base (cytidine) as well as the addition of a phosphate group to the nucleoside (2'-CMP, 3'-CMP, and 5'-CMP) has little effect on the qualitative behavior of the H_5 line width. Similar line broadening of the H_5 resonance of the cytosine base was also observed in 2'-deoxycytidine, 5'-dCMP, and CpC. The absence of a 2'-OH group on the sugar moiety in the 2'-deoxynucleoside and 2'-deoxynucleotide as well as base stacking between the two cytosine bases in the dinucleotide CpC has no significant effect on the H_5 line broadening.

For two cytosine derivatives, 4-*N*,*N*-dimethylcytosine and 3-methylcytidine methosulfate, both the H_5 and H_6 resonances were observed to be sharp throughout the pD range of interest. Here only one tautomeric structure can exist, the amino form in the case of 4-*N*,*N*-dimethylcytosine and the imino form in the case of 3-methylcytidine. Although the line width was studied from pD 2 to 12 for comparative purposes, only the range in which the neutral tautomeric form exists is of interest here. In the case of 4-*N*,*N*-dimethylcytosine, this range is essentially that of the normal cytosine base



(pD 4–10). However, in the case of 3-methylcytidine, the neutral species exists only in a narrow pD range beginning around pD 8.0, at which the amino proton is lost to form an imino tautomer, and ending near pD 10.0, where the second amino proton is lost. The existence of two p*K*'s involved in this latter titration process is apparent from the width of the pD titration curve. The p*K* values of 8.2 and 10.5 deduced from the pH titration curve were confirmed by analysis of uv spectral data.

Effects of Temperature and Magnetic Field. We have investigated the temperature and frequency dependence of the H_5 line width in several of the cytosine derivatives. These studies were made in aqueous D_2O solutions containing high concentrations of salt (NaCl, LiCl, CsCl) at pD 5.8 ± 0.2 over the temperature range −25–90° at both 100 and 220 MHz. The use of electrolytes permitted the temperature range to be extended

(20) J. H. Prestegard and S. I. Chan, *J. Amer. Chem. Soc.*, 91, 2843 (1969).

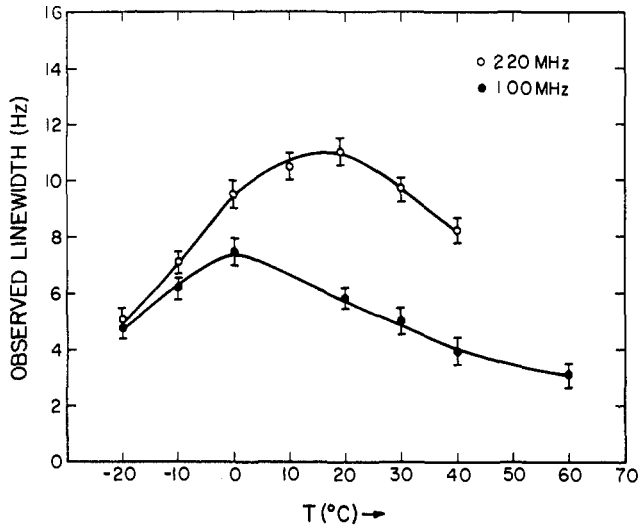


Figure 6. Frequency and temperature dependence of the line width of the H_5 resonance of a 0.02 *M* 5′-CMP solution containing 5 *M* NaCl in D_2O at pD 6.0.

below ~0°. The observations were affected somewhat by the added salt, but for a given salt concentration and at a given temperature, there did not appear to be a noticeable line width dependence on the nature of the electrolyte.

The results for 5′-CMP in 5 *N* NaCl are summarized in Figures 5 and 6. For sake of discussion, we have divided the temperature range into three regions. In the high temperature region (>20°), the H_5 resonance of 5′-CMP was observed to be strongly temperature and field dependent. The H_5 line width is much broader at 220 MHz than at 100 MHz and decreases gradually with increasing temperature. In the low temperature region (<0°), the H_5 line width becomes essentially independent of field and decreases with decreasing temperature. The greatest H_5 line broadening was observed in the intermediate temperature region, where the broadening was so severe that the H_5 doublet was frequently unresolved. At 220 MHz, the maximum H_5 line width is 11.5 Hz at 17° and at 100 MHz, the H_5 resonance reaches a maximum width of 7.5 Hz at 5°.

In conjunction with the temperature studies, we have also noted a progressive downfield shift of the H_5 resonance relative to that of $H_{1'}$ as the temperature was increased. For 5′-CMP in 5 *N* NaCl, this differential shift, $\Delta(\delta_{H_5}-\delta_{H_{1'}})$, amounted to -2.1 ± 0.5 Hz at 100 MHz and -4.0 ± 0.5 Hz at 220 MHz over the temperature range −20–50°. These observations may be contrasted with the almost zero differential shifts observed for the H_5 and $H_{1'}$ resonances of uridine 5′-monophosphate.[21]

Intensity Measurements. Experimentally we have observed that the intensity ratio of the H_5 and $H_{1'}$ doublets changes with temperature. The results of these intensity measurements for 5′-CMP are summarized in Table I. Above 20°, the ratio of the H_5 and $H_{1'}$ intensities was found to be unity within experimental error. At low temperatures (−10 to −25°), this intensity ratio is reduced to 0.90 ± 0.03 at 100 MHz and 0.93 ± 0.02 at 220 MHz. Two methods of obtaining these intensity ratios were used: spectral area measurement using a planimeter and comparison of the experimental spectra with those simulated on a digital computer.

(21) G. Lee, unpublished data.

Table I. Intensity Measurements of H_5 and $H_{1'}$ Resonances[a] at Different Temperatures

Field	Temp, °C	Intensity, arbitrary units H_5	$H_{1'}$	Intensity ratio, %	Av, %
HA-100	30	148	141	105	
	30	55	57	96.5	
	30	99	102	97.0	
	50	78	81	96.3	
	50	68	69	98.6	98.6 ± 3.1
HA-100	−10	126	139	90.64	
	−10	102	118	86.44	
	−20	96	110	92.30	
	−15	62	69	89.85	89.79 ± 3.1
HR-220	−20	48.3	50.70	95.26	
	−20	94.08	100.27	93.8	
	−20	64.10	69.30	92.49	
	−20	22.78	25.35	89.95	
	−20	13.22	14.06	94.0	93.06 ± 2.09

[a] Compound: 5′-CMP, pD 5.8.

Discussion

Mechanism of Line Broadening. A number of mechanisms could cause broadening of the H_5 resonance of cytosine and its derivatives in aqueous solution. These include contamination by paramagnetic ions, spin coupling of the H_5 proton with neighboring quadrupolar nitrogen-14 nuclei, extensive cytosine aggregation, and chemical exchange processes with rates observable by nmr.[22]

In our present work, we can rule out paramagnetic ion complexation as a source of line broadening in view of our extensive efforts to remove such possible contaminants in the preparation of our samples. The absence of a concentration dependence of the H_5 line width over the concentration range 0.002–0.1 *M* excludes broadening by molecular aggregation. We can also eliminate effects arising from indirect spin–spin coupling of the H_5 and H_6 protons to the various quadrupolar ^{14}N nuclei, since the H_5 line width is independent of the viscosity of the solution and there is no broadening of the H_5 resonance in 4-*N*,*N*-dimethylcytosine and 3-methylcytidine, two cytosine derivatives with seemingly rather similar electronic structure to the parent molecule.

These considerations[23] leave chemical exchange processes as the only possible source of the H_5 line broadening. When a molecule can exist in two distinct molecular environments, the chemical shift(s) and the line width(s) of the resulting nmr signal(s) will be determined by the ratio of the rate of chemical exchange to the chemical shift difference between the two environments (the nmr time scale). When the exchange is fast compared to the nmr time scale, only the weighted average signal is observed and its line width is given by[24]

$$\Delta\nu_{obsd} = \Delta\nu_0 + 4\pi p_A^2 p_B^2 \Delta^2 (\tau_A + \tau_B) \qquad (1)$$

(22) A. Abragam, "The Principles of Nuclear Magnetism," Oxford University Press, London, 1962.
(23) G. C. Y. Lee, J. H. Prestegard, and S. I. Chan, *Biochem. Biophys. Res. Commun.*, 43, 435 (1971).
(24) J. A. Pople, W. G. Schneider, and H. J. Bernstein, "High Resolution Nuclear Magnetic Resonance," McGraw-Hill, New York, N. Y., 1959, pp 218–225.

where $\Delta\nu_{obsd}$ is the observed line width; $\Delta\nu_0$ is the weighted average line width in the absence of chemical exchange; p_A and p_B are the relative populations of the two exchanging species; τ_A and τ_B denote their pre-exchange lifetimes; and Δ is the chemical shift difference of the proton between the two environments. It is clear that in this limit, increasing the temperature will decrease the line width, since both τ_A and τ_B decrease with increasing temperature. In the slow exchange limit, *i.e.*, when the exchange is slow on the nmr time scale, we expect to see two separate signals, provided that the populations of both species are large enough and their resonances narrow enough to enable them to be observed. The line width of the resonances due to species A is, for example, given by[24]

$$(\Delta\nu_{obsd})_A = (\Delta\nu_0)_A + 1/\pi\tau_A \qquad (2)$$

Here $(\Delta\nu_0)_A$ is the resonance line width of species A in the absence of chemical exchange. Note that because of the reciprocal dependence on τ_A, the line width increases with temperature in this limit. As expected, in the intermediate exchange region, where the exchange is taking place at a rate comparable to the nmr time scale, maximum broadening occurs. Thus, as the exchange rate increases from the slow to the fast exchange limit, the resonance broadens to a maximum width and then decreases in width.

The observed variation of the H_5 line width of 5′-CMP with temperature (Figure 6) is characteristic of the spectral behavior depicted above and strongly suggests that the H_5 broadening is due to chemical exchange of the cytosine base between two molecular environments. Presumably the three temperature regions cited earlier in the Results section correspond to the regions of fast, intermediate, and slow chemical exchange. This conclusion is clearly borne out by the effect of frequency and/or magnetic field on the broadening, where the nmr time scale has been altered by varying the magnetic field. Since Δ, the chemical shift difference of the H_5 resonance between the two magnetic environments, is directly proportional to the magnetic field strength in the limit of fast exchange, we expect the exchange line width to be proportional to the square of the magnetic field (*cf.* eq 1). In the limit of slow exchange, however, the line width should be independent of the magnetic field (*cf.* eq 2). Our experimental data show that in the high temperature region, the line width of the H_5 resonance of 5′-CMP increases by a factor of two to four (depending on temperature) as the magnetic field is increased from 23.5 (100) to 51.7 kG (220 MHz). This agreement with prediction, together with the absence of field dependence in the low temperature region, provides strong evidence in support of broadening by chemical exchange. The results of our intensity measurements are also consistent with this interpretation. At low temperatures, we observed a significant reduction in the $H_5/H_{1'}$ intensity ratio, and from this result we concluded that the major species is being monitored and that the minor component is present to the extent of the order of 10%. (Note that in this analysis we are making the assumption that the exchange rate is fast compared with Δ for the $H_{1'}$ proton in the two environments over the temperature range under consideration.) The H_5 resonance of the minor species is presumably not observed, either because of its relatively low population and/or because the resonance is broad as might be expected. At high temperatures, the observed H_5 to $H_{1'}$ intensity is unity, as expected for rapid exchange.

A change in the magnetic environment of the H_5 proton can occur in a number of ways, but the fact that the line broadening occurs in cytosine and nucleosides as well as nucleotides indicates that neither the phosphate group nor the sugar moiety is responsible for the chemical shift difference of this proton between the two species. Thus the magnetic nonequivalence must arise from some chemical modification localized in the cytosine base. Since the observed broadening is most pronounced for H_5, indicating that the chemical shift difference between the two molecular environments is greatest for this proton, we surmise that the structural difference must be confined to that part of the cytosine base which is in close proximity to this proton. The absence of an observed concentration dependence of the H_5 line width suggests that the chemical exchange process is unimolecular in cytosine, that the difference in the molecular environment does not arise from the molecular association, and hence the species involved must be tautomeric structures of the neutral cytosine base. An equilibrium between the amino and imino tautomers of cytosine immediately comes to mind. *The absence of line broadening of the H_5 resonance in two methylated derivatives, 4-N,N-dimethylcytosine, where the cytosine base can exist only in the amino form, and 3-methylcytidine, where the base is frozen in the imino structure*, provides plausibility for this hypothesis. In any case, this latter observation strongly suggests that structural differences at the amino group and the N-3 position of the cytosine base are responsible for the chemical shift difference of the H_5 proton in the two species involved in the chemical exchange. We shall later show that an equilibrium between the amino and imino tautomers of cytosine can indeed account for all our nmr observations, and we shall present a kinetic analysis which bears on the plausibility of this hypothesis.

If a chemical shift nonequivalence exists for the H_5 proton between the two species involved in the chemical exchange, we might expect a chemical shift difference between the H_6 protons as well, albeit a smaller one. In this connection, we note that a slight broadening is indeed observed for the H_6 resonance. If the two species involved in the chemical exchange correspond to the amino and imino tautomers of cytosine, the chemical shifts of the protons in the species, δ_A and δ_B, can be estimated from their chemical shifts in model compounds which exist predominantly in one tautomeric structure. Cytidine is believed to be predominantly in the amino form, so that the observed chemical shifts should approximate δ_A. At pD 8.5, the nonexchangeable methyl group of 3-methylcytidine freezes the cytosine base in the imino form, and the chemical shifts here may be used to approximate δ_B. At 100 MHz, the H_5 and H_6 resonances of 3-methylcytidine are 20 and 4 Hz, respectively, downfield from those of cytidine. Since Δ is 20 Hz for H_5 and only 4 Hz for H_6, the effect of exchange broadening on the line width of the H_5 resonance should be much more pronounced than that of H_6, as observed. Thus if our hypothesis regarding the amino–imino equilibrium is correct, we would also

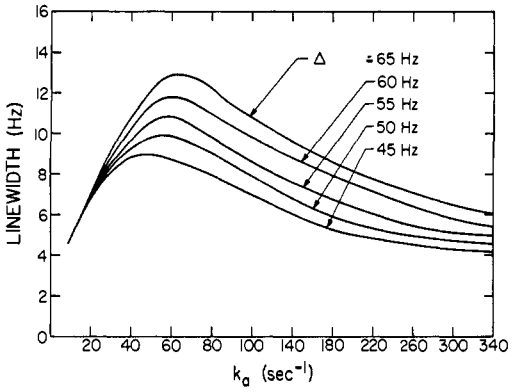


Figure 7. Calculated dependence of the H_5 line width at 220 MHz on the exchange rate constant $k_a(T)$ for a range of chemical shift difference (Δ) of the H_5 proton between the two tautomers assuming 14% minor tautomer.

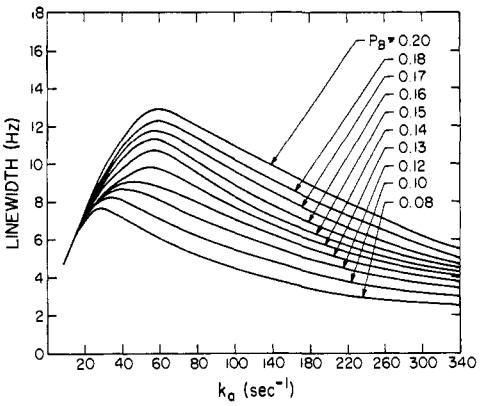
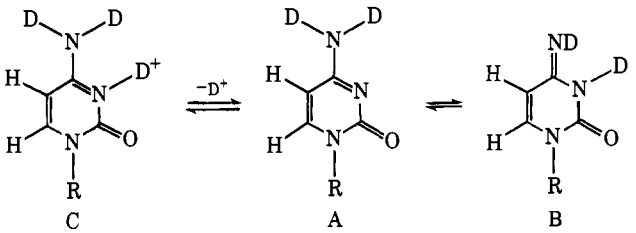


Figure 8. Calculated dependence of the computed H_5 line width at 220 MHz on the exchange rate constant $k_a(T)$ for a range of populations of the minor tautomer assuming a Δ of 55 Hz.

have a satisfactory rationale for the greater exchange width observed for H_5 compared to H_6 in normal cytosine.

Before concluding this section, it would seem appropriate to comment on the progressive downfield shifts of the H_5 resonance relative to that of $H_{1'}$ observed with increasing temperature. It is likely that these shifts are also associated with the dynamic equilibrium between the amino and imino tautomers. If at low temperatures the tautomeric exchange is slow, then the spectral position of the H_5 resonance should correspond to that of the amino tautomer. As the kinetics become more rapid with increasing temperature, we expect gradual collapse of the signals due to the two species (even though in this case the signal due to the minor tautomer is not apparent) and a shift of the resonances toward a spectral position determined by the weighted average chemical shift, $\delta_{mean} = \delta_A p_A + \delta_B p_B$. It is possible that the $\delta_{H_5}-\delta_{H_{1'}}$ differential shifts observed with increasing temperature merely reflect the influence of these time-dependent factors on the nmr spectrum. At 100 MHz, the H_5 resonance of the imino tautomer is expected to appear ~20 Hz downfield from that of the amino tautomer. Our data indicate that the percentage of minor tautomer is ~15% at 50° (*vide infra*). Thus, if $\delta_A-\delta_B$ remains unchanged over the temperature range under discussion, we expect a downfield shift of ~3 Hz between −20 and 50°, if the two temperatures adequately represent the two exchange limits. This calculated shift may be compared with the observed $\delta_{H_5}-\delta_{H_{1'}}$ differential shift of -2.1 ± 0.5 Hz (at 100 MHz) over the same temperature range. The agreement is satisfactory, considering that we have ignored other contributions to $\delta_{H_5}-\delta_{H_{1'}}$, and that the exchange is probably not totally quenched at −20°. In this connection, we note that almost no differential $\delta_{H_5}-\delta_{H_{1'}}$ shifts were observed with temperature in 5′-UMP, where tautomerism of the pyrimidine base is known to be absent.[21]

Kinetic Analysis. To make more plausible our contention that the observed line broadening of the H_5 resonance arises from chemical exchange between the amino and imino tautomers, we have undertaken a kinetic analysis of this problem. In this analysis, we have considered the following equilibrium



where A, B, and C represent the amino and imino tautomers and the protonated cytosine species, respectively. It is necessary to include in this discussion the protonated species in addition to the two neutral tautomers, since our data extend over a wide range of pH. However, the chemical exchange between the protonated species and the amino tautomer is expected to be fast, so that the problem can still be treated as a two-site exchange problem.

At pD ~6.0, the extent of protonation of the cytosine base is negligible. If we can then assume that the line broadening arises from chemical exchange between the two neutral tautomers (A and B), then the line width of the H_5 resonance can be calculated for a given k_a (preexchangerate constant for A), if the relative populations between the two tautomers as well as the chemical shift difference, Δ, is known. We have carried out such calculations for the H_5 resonance using the Dynamic nmr program developed by Binsch and Kleier.[25] Typical results of this calculation are summarized in Figures 7 and 8. In Figure 7, the calculated line width is given as a function of k_a for a range of the chemical shift difference, Δ, assuming 14% minor tautomer. Similarly, in Figure 8 the calculated line width is given as a function of k_a for several populations of the minor tautomer in the range 0.08–0.20 assuming a Δ of 55 Hz at 220 MHz.

From plots of this type, it is possible to extract the rate constant k_a from the experimental line width for a given assumed Δ and $\Delta G°$, where $\Delta G°$ is the standard free energy difference between the two tautomers. At a given temperature, the observed line widths at 100 and 220 MHz must, however, yield the same k_a and this correspondence must apply over the whole temperature range, if Δ and $\Delta G°$ are to be unique. (We are as-

(25) G. Binsch and D. A. Kleier, "The Computation of Complex Exchange-Broadened Nmr Spectra Computer Program DNMR," Quantum Chemistry Program Exchange, 1969.

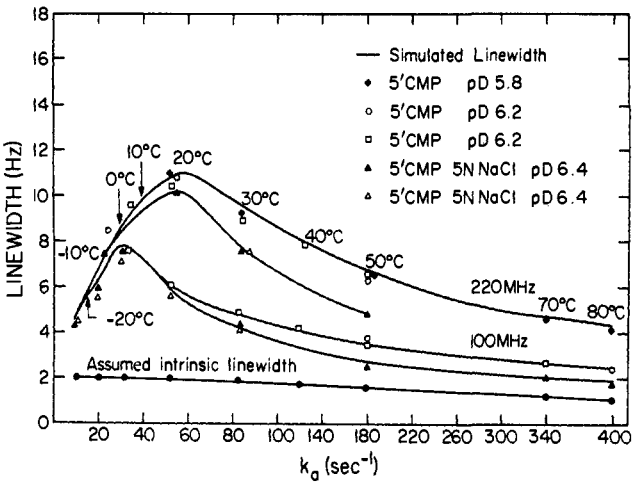


Figure 9. A comparison of the calculated and observed H_5 line widths at 100 and 220 MHz for 5'-CMP in aqueous solution near pD ~6.0 and for 5'-CMP in 5 *N* NaCl at pD ~6.4.

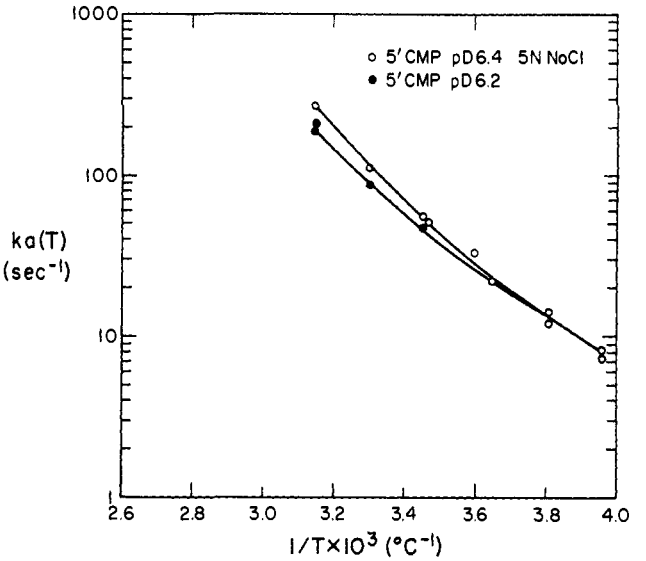


Figure 10. Log k_a *vs.* $1/T$.

suming here that Δ and $\Delta G°$ are invariant over this temperature range.) As a result of a number of iterations, we have found that Δ ~55 Hz at 220 MHz and $\Delta G°$ ~1.1 kcal/mol gave the best fit to our data. A comparison of the observed line widths at the two frequencies with the theoretically expected line width is depicted in Figure 9 for 5'-CMP near pD ~6.0 and for 5'-CMP in 5 *N* NaCl at pD ~6.4. We recall that our intensity measurements of the H_5 resonance had previously suggested that the percentage of the minor tautomer is 10 ± 2% at −20°, which corresponds to a $\Delta G°$ ~ 1.1 kcal/mol, and studies of model compounds had suggested that the chemical shift difference between the two tautomers, Δ, is ~0.20 ppm. These values agree well with our kinetic analysis and indicate that to a good approximation $\Delta S° \approx 0$ and $\Delta H° \approx \Delta G° \approx$ 1.1 kcal/mol.

In Figure 10, we have plotted the log of the rate constant, log k_a, so extracted *vs.* the reciprocal absolute temperature for 5'-CMP at pD ~6 both in the presence and absence of NaCl. If the chemical exchange between the two tautomers involves only a single mechanism, log k_a *vs.* $1/T$ should yield a straight line with slope given by $\Delta E^{\pm}/R$, where $\Delta E^{\pm}$ is the activation energy for the exchange process. Experimentally, we have found that this plot is not linear throughout the whole temperature range; instead, k_a appears to be the sum of two exponentials. This suggests the possibility that two activation energies are involved and the exchange between the two tautomers might involve two parallel paths. A careful study of the relative contributions of the two terms indicates that the higher activation energy component is pH independent, but solvent dependent, *e.g.*, the salt concentration, whereas the lower activation energy component is pD dependent only. In view of these observations, and the pronounced pD dependence of the H_5 line width, particularly at higher pD's, it is reasonable to propose that the chemical exchange in one of these paths is base (OD^-) catalyzed and that of the other kinetic pathway involves the solvent. If this kinetic model is correct, then

$$-(d[A]/dt)_f = k_B[A][OD^-] + k_S'[A][D_2O] = k_B[A][OD^-] + k_S[A] \quad (3)$$

and

$$k_a(T) = 1/\tau_{ab} = -\left(\frac{1}{[A]}\frac{d[A]}{dt}\right)_f = k_B[OD^-] + k_S = A_B[OD^-]e^{-\Delta E_B^{\pm}/RT} + A_S e^{-\Delta E_S^{\pm}/RT} \quad (4)$$

where $-(d[A]/dt)_f$ is the preexchange rate of the major tautomer, k_B and k_S denote the rate constants for the base- and solvent-catalyzed steps, respectively, $k_S = k_S'[D_2O]$, and $\Delta E_B^{\pm}$ and $\Delta E_S^{\pm}$ denote the activation energies for the base- and solvent-catalyzed steps. Analysis of the data in Figure 10 yielded the following activation energies: $\Delta E_B^{\pm} = 3.5 \pm 0.5$ kcal/mol, $\Delta E_S^{\pm} = 7.8 \pm 0.6$ kcal/mol (in the absence of salt), $\Delta E_S^{\pm} = 10.8 \pm 0.5$ kcal/mol (in the presence of 5 *N* NaCl).

We have examined the base-catalyzed step of this tautomeric exchange in detail. The pD dependence of the line width comes about in two ways. First, the exchange rate $k_a(T)$ involves the base (OD^-) concentration directly. However at lower pD's, where the cytosine base can become protonated, the population of the neutral tautomers p_A and p_B varies with the solution pD. p_A and p_B can be expressed in terms of the solution pD and the protonation constant K_a. Since

$$p_A + p_B + p_C = 1 \quad (5)$$

$$K_a = p_A[D^+]/p_C \quad (6)$$

and

$$p_A = p_B e^{\Delta H°/RT} \quad (7)$$

$$p_B = \{1 + (1 + [D^+]/K_a)e^{\Delta H°/RT}\}^{-1} \quad (8)$$

The effective time scale of the nmr experiment, *i.e.*, Δ_{eff}, is also pH dependent as a result of the rapid chemical exchange between the major tautomer A and the protonated cytosine species C and the effect of this exchange on the spectral position of the weighted averaged signal due to these species. Δ_{eff} is readily shown to be

$$\Delta_{eff} = \delta_B - \delta_{AC} = \delta_B - (p_A\delta_A + p_C\delta_C)/(p_A + p_C) \quad (9)$$

A knowledge of $p_A + p_C$, p_B, Δ_{eff}, and the rate expres-

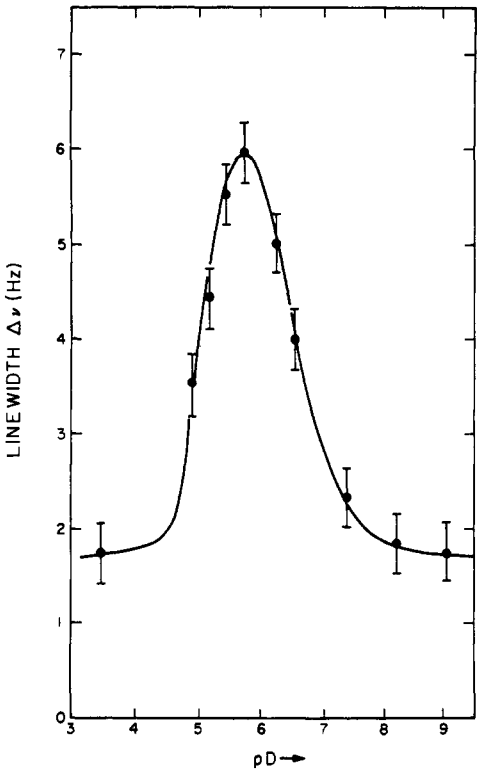


Figure 11. A comparison of the calculated and observed pD dependence of the H_5 line width of 5′-CMP in D_2O (— theoretical line width, ‡ observed line width).

sion $k_a(T)$ enables us to simulate the pD dependence at a given temperature, when the relative contribution of the two mechanisms is known, *i.e.*, if we know the ratio of the preexponential factors, A_B/A_S, or the ratio of the rate constants k_B/k_S at a given temperature.

We have determined the ratio of the preexponential factors for the two exchange mechanisms, $\eta = A_B/A_S$, from the pD value corresponding to the observed maximum line width at 30°. To do this, we have used the following approximation for the line width

$$\Delta\nu_{obsd} = \Delta\nu_0 + \frac{4\pi p_B^2}{p_A}(1 - p_B)^2\Delta_{eff}^2\tau_{ab} \quad (10)$$

which we have derived from the Bloch equation for three site exchange, *i.e.*, C ⇌ A ⇌ B, in the limit of very fast exchange between C and A and moderately fast exchange between A and B, a condition which is satisfied at 30°, pD 6.0, and at 100 MHz. We have shown that under these conditions this equation predicts a line width which is in good agreement (within 10%) with that obtained by the density matrix method employed in the Dynamic nmr program.[25] Substituting eq 6–9 into 10, we obtain the following pD dependence for the H_5 resonance at a given temperature T under conditions where eq 10 is applicable and $p_B \ll p_A$

$$\Delta\nu_{obsd} = \Delta\nu_0 + (4\pi/A_S)\Delta_{eff}^2 e^{(\Delta E_S^{\pm} - 2\Delta H^{\circ})/RT} \times [D^+]/\{1 + [D^+]/K_a\}\{[D^+] + \eta K_w e^{(\Delta E_S^{\pm} - \Delta E_B^{\pm})/RT}\} \quad (11)$$

K_w is the ionization constant of heavy water ($K_w = 1.95 \times 10^{-15}$). The pD corresponding to the maximum line width can be obtained by differentiating the above expression with respect to $[D^+]$ and setting the resultant derivative to zero. Experimentally, the maximum H_5 line width was observed at pD ~6.0 at 30°, from which we ascertain that

$$\eta = 2 \times 10^8 e^{(\Delta E_B^{\pm} - \Delta E_S^{\pm})/0.6} \quad (12)$$

The calculated pD dependence of the H_5 line width at 30° in D_2O solution is depicted in Figure 11. The overall agreement between the theoretical curve and the experimentally observed line widths can be seen to be excellent.

The results of this kinetic analysis yielded the following rate constants for the tautomeric exchange process at 30°. For the base-catalyzed step, $k_B \approx 1 \times 10^{10}$ l. mol^{-1} sec^{-1}; for the step involving the solvent, $k'_S \approx 1.2$ l. mol^{-1} sec^{-1}. Both k_B and k'_S are of the proper order of magnitude[26] for proton abstraction from an amino group by OD^- and D_2O.

EDTA as a Catalyst for the Tautomeric Exchange. In our previous paper,[23] we noted that the addition of 10^{-6} *M* EDTA is sufficient to sharpen the broad H_5 resonance. This observation might imply that the broadening is due to paramagnetic ion contaminants present at the concentration level of 10^{-6} *M*, and that the added EDTA is merely serving as a chelating agent. However, in our earlier work,[23] we argued that it is unlikely that paramagnetic ion contaminants at this concentration level should lead to the line broadening observed. Moreover, if this were the case, other efficient chelating agents should be effective in sharpening the broadened H_5 resonance as well. The fact that ethylenediamine does not sharpen the H_5 resonance even at concentrations as high as 10^{-3} *M* would seem to rule out this possibility. This leaves us with the other only alternative, which is that EDTA serves as a catalyst for the tautomeric exchange between the amino and imino tautomers of the cytosine base. The manner in which EDTA can promote the tautomeric exchange is not exactly clear, but near neutral pD, EDTA is both a proton acceptor as well as a proton donor, and hence we suspect it acts through a concerted mechanism involving simultaneous proton abstraction from the amino group by one of the four carboxylate anions and proton transfer from one of two protonated EDTA nitrogen sites[27] to the N-3 of the cytosine base. If this mechanism is correct, the efficiency with which EDTA catalyzes the tautomeric exchange indicates that the kinetics is almost diffusion controlled.

Conclusions

We believe we have presented experimental evidence to show that the unusual H_5 broadening observed in the pmr spectra of cytosine and its related derivatives is due to chemical exchange between the amino and imino tautomers of the cytosine base. We have studied this line broadening phenomenon as a function of temperature, concentration, solution pD, and magnetic field as well as for a number of cytosine derivatives, including those capable of existing in only the amino or imino tautomeric structure, and insofar as we can make out, the experimental observations are consistent with this hypothesis. A quantitative treatment of the line width data has been presented in the case of 5′-CMP, and we have shown that this broadening of the H_5 reso-

(26) E. Grunwald, A. Loewenstein, and S. Meiboom, *J. Chem. Phys.*, 27, 630, 641 (1957).
(27) R. J. Kula, D. T. Sawyer, S. I. Chan, and C. M. Finley, *J. Amer. Chem. Soc.*, 85, 2930 (1963).

nance is indeed very sensitive to the amount of the imino tautomer present. On the basis of this analysis, we have ascertained the kinetics of this tautomerism, have demonstrated that the tautomeric exchange is catalyzed by base (OD^-) as well as by the solvent molecules (D_2O), and have concluded that the imino tautomer is present to the extent of $15 \pm 3\%$ at room temperature in neutral aqueous solution.

The results of this pmr study indicate that the cytosine base exists in significant percentages in the minor imino tautomeric form. This result is contrary to what is generally accepted, as all previous investigations of this tautomerism had provided estimates which are significantly lower than our present indications. From uv and pK_a studies of cytosine and its derivatives, Kenner, *et al.*,[5] estimated in 1955 that K_T = [amino]/[imino] = $10^{4.7}$, indicating that the population of the imino tautomer is negligible under ordinary conditions. A similar conclusion was reached by Miles in 1961, who used ir spectroscopy to elucidate the detailed molecular structure of nucleic acid bases in D_2O solution.[9] More recently, however, Brown and Hewlins[7] concluded on the basis of uv and ir measurements that $K_T \approx 25$ (in water) for the amino–imino tautomerism in 5,6-dihydrocytosine, a result which, perhaps, is more in line with our present estimates for cytosine. In any case, it should be noted that accurate determination of the amino–imino equilibrium in these earlier studies is limited by the sensitivity and the resolution of the methods used. Although uv spectroscopy is a highly sensitive method for the characterization of electronic structures of molecules, the spectral similarity of nucleic acid bases and their alkylated anlogs and the serious overlap of component bands make its application to the elucidation of the tautomeric structures of nucleic acid bases inconclusive and unreliable. Similar difficulties are inherent in the ir spectroscopic studies. The determination of tautomeric equilibria by the comparison of pK_a's of nucleic acid bases with alkylated analogs is also complicated by the multiplicity of ionizable groups as well as protonation sites.

Our results are not exactly incompatible with the theoretical predictions of Pullman and Pullman, Löwdin, and Bodor, *et al.*, all of whom have predicted that the amino form of the cytosine base is more stable than the imino structure. The Hückel MO calculations of Pullman and Pullman[10] and Löwdin[12] have shown that the resonance energy of the imino form is 2.1 kcal/mol higher than that of the amino tautomer. More recently, Bodor, *et al.*,[14] predicted on the basis of semiempirical SCF–MO calculations that the difference in the heats of atomization between the two tautomers of cytosine is 2.2 kcal/mol higher in the case of the amino structure. In this work, we have estimated that the enthalpy difference between the two tautomers is $\sim$1.1 kcal/mol. Considering that the theoretical calculations are made for the molecules in their vacuum states, and hence have neglected solvation effects, these theoretical results may be considered to be in reasonable agreement with our experimental results.

The X-ray work of Barker and Marsh[28] has shown that cytosine exists predominantly as the amino form in the crystalline state, as expected. However, it is not unreasonable that the cytosine amino/imino equilibrium should exhibit somewhat different thermodynamic properties in solution from that in the crystalline state. In fact, we expect this tautomerism to be sensitive to the solvent, as the recent work of Brown and Hewlins[7] on 5,6-dihydrocytosine has shown. These workers have reported that $K_T \approx 2.5$ in H_2O, but the imino structure is more stable in $CHCl_3$ with $K_T \approx 0.1$.

(28) D. L. Barker and R. E. Marsh, *Acta Crystallogr.*, 7, 1581 (1964).