[CONTRIBUTION FROM THE DEPARTMENT OF CHEMISTRY, EMORY UNIVERSITY, ATLANTA 22, GEORGIA]

A Nuclear Magnetic Resonance Investigation of Tautomerism and Substituent Effects in Some Pyrimidines and Related Nucleosides

BY J. P. KOKKO, J. H. GOLDSTEIN AND LEON MANDELL

RECEIVED NOVEMBER 28, 1960

The n.m.r. spectra of uracil, thymine, cytosine, cytosine hydrochloride, 5-bromouracil and certain related nucleosides have been studied using dimethyl sulfoxide as the solvent. The spectrum of cytosine suggests pronounced tautomerism leading to a zwitterionic form, but the other bases appear to occur as the diketo form. The spectra of the nucleosides indicated that in all the cases studied, tautomeric structures were either absent or undetectable. Substitution of Br or CH_3 at the C_5 position in uracil affects the resonances of the N_1 and N_3 protons to varying degrees. These effects may be interpreted in terms of the lability of these protons and, by inference, of the relative susceptibility of the bases to tautomeric shifts. Some possible biological implications of these results are discussed.

Introduction

The tautomeric behavior of the various purines, pyrimidines and related nucleosides and nucleotides is of interest in connection with such questions as the structure of the nucleic acids,[1-3] as well as their replication, and the mechanism of mutagenesis.[4] This problem has been previously investigated by infrared spectroscopy for both the purines and pyrimidines[5] and for the nucleosides.[6-8] A nuclear magnetic resonance (n.m.r.) study[9] of these compounds has been carried out recently, using D_2O as the solvent at various pD values, but this did not include a consideration of tautomerism.

The utilization of n.m.r. techniques in the study of tautomeric equilibria presents certain advantages. An example of such an application which is relevant to the present investigation is the recently reported n.m.r. study of the system 4-hydroxypyridine $\rightleftarrows$ 4-pyridone.[10] In the case of the purines, pyrimidines and their nucleosides, however, the use of water or heavy water as a solvent produces obvious complications which are best avoided by going over to a suitable aprotic solvent. After surveying a wide range of solvents, completely deuterated dimethyl sulfoxide was finally selected for this purpose because it provided adequate solubility without proton transfer and with very little spectral interference.

Experimental

All spectra were obtained at 40 megacycles using a Varian Model 4300 B high-resolution spectrometer equipped with a field homogeneity control unit. Because of the marked hygroscopic character of dimethyl sulfoxide, solution of the samples was carried out under dry nitrogen gas. The solutions were transferred, again under dry nitrogen, to 5 mm. sample tubes containing tetramethylsilane (TMS) reference capillaries and sealed at the temperature of Dry Ice. Calibrations were performed by the usual side-band modulation technique.[11] The audio oscillator, a Krohn-Hite Model 440-B, was checked at various times against a Hewlett-Packard Model 522B electronic counter.

To assist in the assignments, parallel spectra were also obtained for the partially deuterated bases and nucleosides. These were prepared simply by dissolving the samples in heavy water, allowing the solutions to stand for about 24 hr. and finally evaporating the solutions to dryness under vacuum. Only the labile protons were exchanged by deuterium in this process.

The compounds used in this study were the commercially available materials. It was found that the n.m.r. spectra provided a convenient and reliable test for impurities. No change was observed in the spectra of these samples after storing for two months at 20°.

Spectra of the Pyrimidines.—The spectra of the four pyrimidines studied are shown in Fig. 1, together with that of the base analog, 5-bromouracil. The spectra of the nucleosides are shown in Fig. 2. The abscissa gives the chemical shifts in cycles per second (c.p.s.) relative to an external TMS capillary at 20° using d_6-DMSO as the solvent.

A distinctive feature of most of these spectra is the occurrence of a broad peak or peaks in the vicinity of −450 c.p.s. These peaks obviously arise from protons attached to the pyrimidine ring nitrogen atoms. This assignment was verified by comparing the area of these peaks with that of the methyl peak or of the vinylic proton peaks and also by comparing these areas with those in the related nucleosides. The effect on the spectrum of methyl substitution at C_5, as in thymine, is to separate the N_1–H and N_3–H peaks, the former being shifted upfield by about 14 c.p.s. relative to uracil, while the latter peak is essentially unchanged. In 5-bromouracil, on the other hand, both N–H peaks are shifted downfield relative to uracil, but the two protons are separated by roughly the same amount as in thymine. In both molecules the higher field N–H peak is a doublet which is most reasonably attributed to coupling with the C_6 proton. Hence these doublets are to be assigned to the protons at the N_1 positions.

The peaks occurring in the vicinity of −300 to −330 c.p.s. must be due to C_6–H since it is the doublet at higher field which disappears upon substitution of Br or CH_3 at the C_5 position. In uracil the C_6–H resonance is a quartet showing intervals of 5.7 and 7.6 c.p.s., the latter obviously being due to coupling with C_5–H since this is the magnitude of the major splitting in the higher-field ethylenic group. The splitting of 5.7 c.p.s., which must then arise from coupling with the N_1 proton, also occurs in the C_6–H pattern of thymine and 5-bromouracil in which the C_5 proton is absent. In thymine and bromouracil the broad doublet in the N–H region is also split by very nearly 5.7 c.p.s., supporting its previous assignment to the N_1 proton. The thymine spectrum has a relatively large, sharp peak at −81.7 c.p.s., from TMS which is due to the methyl substituent at the C_5 position. In all of the spectra, measurements of relative areas of the peaks are in excellent agreement with these assignments.

In the case of formamide, it has been shown that the coupling constant is 2.1 c.p.s. when the protons are *cis* and 13 c.p.s. when they are *trans* to each other.[12] It will be observed that for our compounds, which are cyclic

(1) L. Pauling and R. B. Corey, *Arch. Biochem. Biophys.*, 164 (1956).

(2) R. Langridge, H. R. Wilson, C. W. Hooper, M. H. F. Wilkins and L. D. Hamilton, *J. Mol. Biol.*, **2**, 19 (1960).

(3) R. Langridge, D. A. Maroin, W. E. Seeds, H. R. Wilson, M. H. F. Wilkins and L. D. Hamilton, *ibid.*, **2**, 38 (1960).

(4) E. Freese, *ibid.*, **1**, 87 (1959).

(5) J. R. Lacher, J. L. Bitner, D. H. Emery, M. E. Seffl and J. D. Park, *J. Phys. Chem.*, **59**, 615 (1955).

(6) H. T. Miles, *Biochim. Biophys. Acta*, **22**, 247 (1956).

(7) H. T. Miles, *ibid.*, **27**, 46 (1958).

(8) H. T. Miles, *ibid.*, **30**, 324 (1958).

(9) C. D. Jardetzky and O. Jardetzky, *J. Am. Chem. Soc.*, **82**, 222 (1960).

(10) R. A. Y. Jones, A. R. Katritzky and J. M. Lagowski, *Chem. and Ind. (London)*, No. **27**, 870 (1960).

(11) J. T. Arnold and M. G. Packard, *J. Chem. Phys.*, **19**, 1608 (1951).

(12) L. H. Piette, J. D. Ray and R. A. Ogg, *J. Mol. Spectroscopy*, **2** (1958).

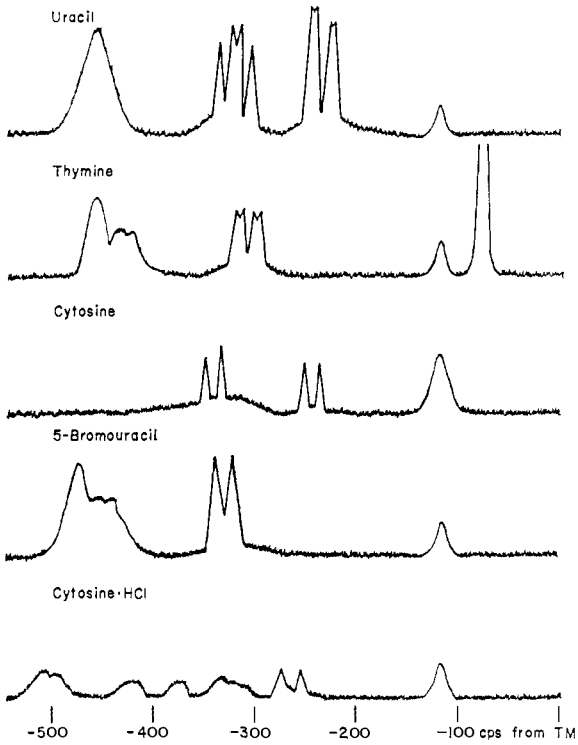

Fig. 1.—40 Mc. n.m.r. spectra of pyrimidines taken at 20° in d_6-dimethyl sulfoxide, relative to an external TMS capillary reference.

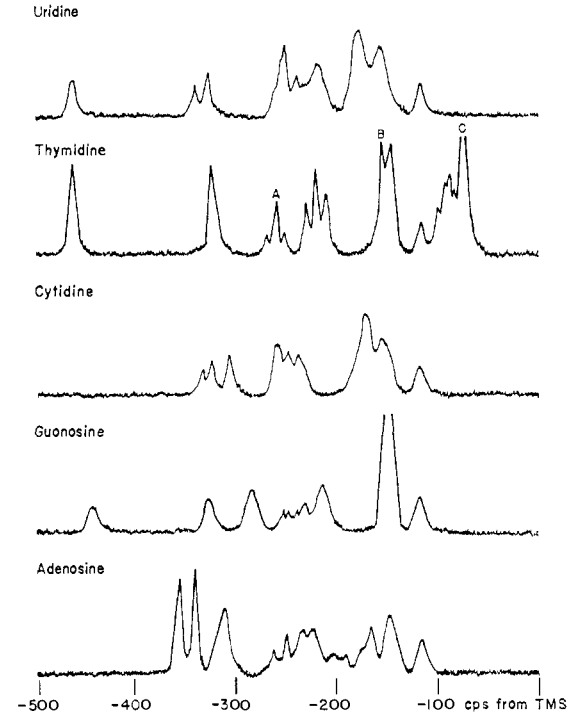

Fig. 2.—40 Mc. n.m.r. spectra of nucleosides, taken at 20° in d_6-dimethyl sulfoxide, relative to an external TMS capillary reference.

amides, an intermediate value, 5.7 c.p.s., is achieved. Since the area of the high-field N–H region corresponds quite closely to one proton, it seems unlikely that keto–enol tautomerism is involved. Therefore, and if the formamide results are applicable to these complex amides, some other explanation must be sought for the low value of the H_1–H_6 coupling constant, possibly a non-coplanar arrangement of the N_1–H and C_6–H bonds. If, on the other hand, the bases are thought of as essentially aromatic in character, the observed value of the coupling constant is reasonable, since it would then be analogous to the *ortho* couplings in benzene compounds.

The spectrum of cytosine presents some unique features for this series of bases. It will be observed that the typical broad N–H peak below −400 c.p.s. is entirely absent and that, instead, the doublet near −300 c.p.s. is superposed on a diffuse and somewhat unsymmetrical peak. Moreover this doublet, due to C_6–H, is coupled only to C_5–H as is indicated by the splitting of 7.1 c.p.s. occurring in both groups. The only plausible explanation for these observations is that cytosine in dimethyl sulfoxide solution does not exist in the keto structure, as do uracil and thymine. Two possible alternate forms are the enol and zwitterion structures shown in Fig. 3.

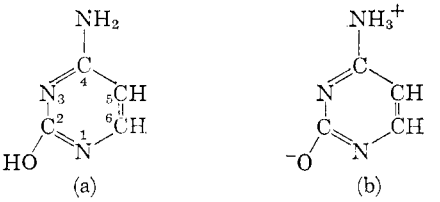

Fig. 3.—Alternate structures for cytosine in DMSO solution: (a) enol and (b) zwitterion.

Presumably, the pseudo-aromatic character of both these structures would tend to stabilize them relative to the keto form. In the zwitterion structure the three amine protons are expected to appear as a broad triplet as a result of coupling with the nitrogen nucleus. To test this expectation, a small amount of HCl gas was bubbled through a solution of cytosine to form the hydrochloride (see Fig. 4).

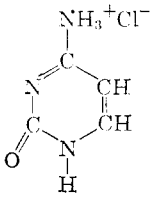

Fig. 4.—Cytosine hydrochloride.

The spectrum of cytosine hydrochloride in DMSO does indeed show a symmetrical triplet between about −400 and −300 c.p.s., each of whose peaks corresponds in area to a single proton. In addition, the N_1 proton has now returned as a broad doublet, whose area corresponds to one proton, with a separation of about 5.7 c.p.s.

Area measurements of the diffuse peak in cytosine show that it corresponds to very nearly three protons, allowance being made for the superposed peaks of the single C_6 proton. Moreover, there is no indication of either the hydroxyl peak or the NH_2 peaks that would be expected for the enolic form shown in Fig. 3.

These observations suggest very strongly that cytosine (in DMSO) exists predominantly in the zwitterion form. The diffuse character of the peaks in the amine region of the cytosine spectrum can be attributed to an exchange of the protons involved. The n.m.r. spectrum of aqueous methylammonium chloride shows an analogous change as the *p*H is varied, and this has been attributed to exchange processes of which several are conceivable in the aqueous system.[13] In the present case the spectra suggest that the N_1 (or enolic) proton may be involved through an intermolecular or intramolecular exchange process. Unfortunately, the solubility of cytosine in DMSO is too low to permit reliable studies of the effect of concentration upon its n.m.r. spectrum, so that this approach fails to provide information concerning intermolecular association.

(13) E. Grunwald, A. Loewenstein and S. Meiboom, *J. Chem. Phys.*, **27**, 630 (1957).

However, no significant changes were observed in the n.m.r. spectra of uracil and thymine as their concentrations were varied.

Spectra of the Nucleosides.—The n.m.r. spectra of the nucleosides studied in d_6-DMSO are shown in Fig. 2. A recently reported n.m.r. investigation by Jardetzky and Jardetzky[9] of a series of purines, pyrimidines, and their nucleotides and nucleosides in aqueous solution provided assignments for the non-exchangeable protons with which the present results are in agreement. In addition we have studied thymidine, which was not included in the above mentioned work. This is the only nucleoside in the group which contains the 2′-deoxyribose ring, and its spectrum contains a characteristic triplet (group A in the thymidine spectrum of Fig. 2) which is assigned to the C_1' proton of the deoxyribose ring. The triplet structure results from spin–spin coupling with the two protons at the C_2 position on the ring. In all the other nucleosides the C_1 proton was split into a doublet by the single C_2 proton, but this is obscured by the ring hydroxyl peaks. Upon deuteration the latter disappears revealing the underlying doublet. (The spectra of the deuterated compounds are not shown here.) The C_2 proton peaks are at B in the thymidine trace of Fig. 2. Peak C is the methyl peak shifted a few cycles lower than the methyl peak of thymine, presumably as a result of the inductive effect of the ring at N_1.

Our principal interest in these nucleosides was the possible occurrence of tautomeric structures. Uridine and thymidine are considered to be in the diketo form by arguments which parallel those given for the related pyrimidines. Guanosine appears to exist as the keto structure since its spectrum shows a peak at −447 c.p.s. which is ascribed to the N_1 proton. In the case of guanosine, adenosine and cytidine, an alternate imino-form was considered, as shown in Fig. 5.

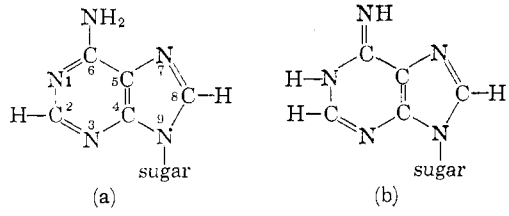

Fig. 5.—Possible tautomeric forms in adenosine: (a) amino, (b) imino.

However, the unmistakable NH_2 peak, near −300 c.p.s., rules out the imino-structure.

Discussion

The previously described upfield shift of the N_1 proton in thymine signifies that this proton is in a region of greater electron density, *i.e.*, that it is less acidic than is its counterpart in uracil. In 5-bromouracil both the N_1 and N_3 protons are, by similar reasoning, in regions of lower electron density as compared with uracil, so that they are correspondingly more acidic. Nevertheless, in bromouracil, as in thymine, the N_3 proton is apparently more acidic than the N_1 proton. These observations would appear to be relevant to the fact that 5-bromouracil is capable of replacing thymine in the DNA structure as well as to the ability of 5-bromouracil to induce mutations. The admittedly simplified spectral interpretation invoked here suggests a comparable difference in the lability of the N protons for both molecules. Moreover, one might expect tautomeric forms to be somewhat more attainable, although still in low abundance, for the bromouracil molecule, since both its N protons are in regions of relatively smaller electron density. The role of tautomeric shifts in inducing mutations through mistaken base pairings has been discussed by Freese,[4] who also credits Meselson with the suggestion that the higher electronegativity of Br, compared to CH_3 or H, should increase the frequency of tautomeric shifts in bromouracil relative to thymine and uracil. The n.m.r. interpretations offered here are not only consistent with the ideas of Freese and Meselson, but in addition they offer at least the possibility of quantitatively assessing proton lability in the DNA bases.

It is interesting that, of the bases studied here, and under the conditions employed, only in the case of cytosine have we found evidence for extensive tautomerism. It is conceivable, of course, that very minor predispositions to tautomerism possible in the other bases would not be manifest in their n.m.r. spectra but would still be able to influence their behavior in biological processes.

One further observation is suggested by the appearance of the spectra of cytosine and its hydrochloride. The fact that the $-NH_3^+$ peaks are more diffuse in the base than in the hydrochloride can be ascribed to the difference in acid strength of the acid in cytosine (essentially phenolic in character) and that in the salt (hydrogen chloride). In the former, the weaker acidity permits a more general exchange of protons which gives rise to the broad proton resonance, whereas in the latter the protons are more firmly fixed, and as a consequence the triplet structure is relatively more pronounced.

Jardetzky[14] has studied the ribose structure in various nucleosides by n.m.r. methods and has concluded that the conformation of the ribose is altered by attachment to a purine or pyrimidine. Our principal interest in the nucleosides during this study was the possible occurrence of tautomeric forms. From our spectra it is established readily that uridine and thymidine exist as the diketo structures. The appearance of the C_6 proton peak shows that the N_1 proton is missing; this is, of course, the point of attachment for the ribose ring. In the cytidine spectrum we see that the NH_2 has reappeared (near −300 c.p.s.), so that, unlike the case of cytosine, the zwitterionic form does not occur. In both cytidine and adenosine the appearance of an NH_2 peak at a normal position for this group also rules out the possibility of significant amounts of imino structures, which have been considered as likely to be involved in some cases of mistaken base pairings.[4]

It should be borne in mind that the above conclusions and inferences refer to the compounds as studied in dimethyl sulfoxide solution. Nevertheless, the n.m.r. spectra indicate an absence of serious perturbations arising from solute–solvent interaction and that, as a consequence, the spectrum of each substance is approximately typical of the isolated solute molecules. Translation of the results to the more common aqueous environment, especially on a comparative basis, thus appears to be a fairly safe procedure.

The authors wish to express their appreciation to the Public Health Service for their support of this work (Research Grant A—2397(C2) and Experimental Training Grant 2-R-7). They are also indebted to Dr. John M. Reiner for his assistance in various phases of this investigation.

(14) C. D. Jardetzky, *J. Am. Chem. Soc.*, **82**, 229 (1960).