Keto Iminol Tautomerism of Protonated Cytidine Monophosphate Characterized by Ultraviolet Resonance Raman Spectroscopy: Implications of C^+ Iminol Tautomer for Base Mispairing

Roberto Purrello,[†] Marina Molina,[‡] Yang Wang,[§] Giulietta Smulevich,[⊥] John Fossella,[∥] Jacques R. Fresco,*[,∥] and Thomas G. Spiro*[,⊗]

Contribution from the Departments of Chemistry and Molecular Biology, Princeton University, Princeton, New Jersey 08544. *Received March 9, 1992*

Abstract: Ultraviolet resonance Raman (UVRR) spectra are reported for cytidine monophosphate in neutral and protonated forms. Aside from shifts in cytosine ring modes, protonation induces the appearance of a new RR band at 1730 cm^{-1} whose excitation profile and temperature-dependent intensity distinguish it from other cytosine bands. The 1730-cm^{-1} band shifts down by 15 cm^{-1} in D_2O, and it is assigned to the C2=N3 stretching mode of an iminol tautomer. Its high frequency is due to the electronic effect of the OH substituent as well as vibrational coupling with the OH bending mode, as shown by comparison with imidic acid model compounds. The iminol tautomer is protonated at N4, not N3, as judged by comparison with UVRR spectra of N^3-methylcytidine, N^4,N^4-dimethyl-2'-deoxycytidine, and cyclocytidine. The UVRR bands are assigned in a consistent fashion for major and minor tautomers of all these species. UVRR spectra of polycytidylic acid (poly(C)) show duplex formation upon hemiprotonation, with concomitant *inhibition* of formation of the iminol tautomer, as expected from the C^+·C H-bonding pattern. Additional protonation dissociates the poly(C) strands of the duplex, showing that C^+·C^+ pairing does not allow an alternative duplex structure. The iminol tautomer of C^+ may play a role in the stabilization of A–C oppositions via wobble pairing.

Introduction

Tautomerization of the nucleic acid bases is a subject of classical importance to molecular biology, since base pairing in the double helix depends on H-bonding between complementary tautomeric structures. The dominant structures for all four bases are the keto and amino tautomers, which enable complementarity between guanine (G) and cytosine (C) and between adenine (A) and thymine (T) or uracil (U). Alternative hydroxy and/or imino tautomers are available at considerable cost in free energy,[1] and the role that these rare tautomers may play in genetic errors via base mispairing has been analyzed extensively.[2] The number of tautomers to be considered is augmented if the possibility of base protonation is taken into account.[2f,3] The pH excursion required for protonation is simply an added term in the free energy cost, which may be compensated by stabilization from base pairing and stacking.

Despite the attention that nucleic acid tautomerization has received, experimental evidence for unfavored tautomers is scanty. It is largely confined to the gas-phase or low-temperature matrices of the bases proper,[4] usually with a proton at the N1 or N9 position in place of the glycosidic bond, which complicates further the tautomerization possibilities. Indirect evidence for the occurrence of unfavored tautomers of the bases has come from comparisons of the pK_as of alkyl-substituted bases[5] analogous to the favored and unfavored tautomeric forms, and from analysis of the electronic spectra of polynucleotide double helices containing particular base mispairs.[2b]

In the present study, we present direct spectroscopic evidence for rare tautomer formation of a nucleotide in aqueous solution, obtained by using ultraviolet resonance Raman (UVRR) spectroscopy. When cytidine monophosphate (CMP) is protonated, the UVRR spectrum contains a band at 1730 cm^{-1} which shifts down 15 cm^{-1} in D_2O. This band has a different excitation profile and a different temperature dependence than the other bands in the spectrum; it is therefore identified with a minor chemical species. Comparison with model systems shows the high frequency and D_2O sensitivity of this band to be diagnostic of a C=N bond with a hydroxy substituent, i.e., an iminol tautomer (see Figure 5).

This assignment is confirmed by examining UVRR spectra of N^3-methyl- and N^4,N^4-dimethylcytidine, both of which show the high-frequency band in their protonated forms, and of cyclocytidine, which is locked in the analogous imino iminoxy structure and which shows a high-frequency, D_2O-insensitive band. Tautomerization of protonated C^+ residues to this form is shown to be inhibited by formation of the duplex structure of hemiprotonated poly(C), an effect attributed to the H-bonding pattern of the base pairs. It is proposed, conversely, that the iminol tautomer of C^+ may help to stabilize A·C mispairs via wobble-type H-bonding.

Experimental Section

Materials. Chemicals were purchased from Aldrich (cyclocytidine hydrochloride, acetamide, *N*-methylacetamide (NMA), and *N*,*N*-dimethylacetamide) and Sigma (cytidine 5'-monophosphate (CMP) sodium salt, N^3-methylcytidine methosulfate, N^4,N^4-dimethyl-2'-deoxycytidine, poly(C), and ethyl acetimidate) and used without further purification. Most samples were 5 mM in the molecule under study and contained 0.3 M Na_2SO_4 as an internal standard. H_3PO_4/NaH_2PO_4 buffer was used for the spectra recorded at pH 2.5 and 3 and NaH_2PO_4/Na_2HPO_4 for those recorded at pH 7; acetate buffer was used for the spectra obtained at pH 4.3, 4.8, and 5. Buffer concentrations were 0.2 M. For D_2O

* Authors to whom correspondence should be addressed.
† Present address: Dipartimento di Scienze Chimiche, Università di Catania, Viale A. Doria 6, 9521 Catania, Italy.
‡ Present address: Instituto de Optica, Serrano 121, Madrid 28006, Spain.
§ Present address: Laser Laboratories, Chemistry Department, Florida State University, Tallahassee, FL 32306.
⊥ Present address: Dipartimento di Chimica, Università di Firenze, Via G. Capponi 9, 50121 Firenze, Italy.
∥ Department of Molecular Biology, Princeton University.
⊗ Department of Chemistry, Princeton University.

(1) (a) Kwiatkowski, J. S.; Pullman, B. *Adv. Heterocycl. Chem.* **1975**, *18*, 199. (b) Dreyfus, M.; Bensaude, O.; Dodin, G.; Dubois, J. E. *J. Am. Chem. Soc.* **1976**, *98*, 6338.

(2) (a) Topal, M. D.; Fresco, J. R. *Nature* **1976**, *263*, 285, 289. (b) Fresco, J. R.; Broitman, S.; Lane, A.-E. In *Mechanistic Studies of DNA Replication and Genetic Recombination*, ICN-UCLA Symposia on Molecular and Cellular Biology; Alberts, B., Ed.; Academic Press: New York, 1980; Vol. XIX, pp 753–768. (c) Strazewski, P. *Nucleic Acids Res.* **1988**, *16*, 9377. (d) Gould, I. R.; Hillier, I. H. *Chem. Phys. Lett.* **1989**, *161*, 185. (e) Cieplak, P.; Bash, P.; Singh, U. C.; Kollman, P. A. *J. Am. Chem. Soc.* **1987**, *109*, 6283. (f) Les, A.; Adamowicz, L.; Bartlett, R. J. *J. Phys. Chem.* **1989**, *93*, 4001.

(3) Hass, E. C.; Mezey, P. G.; Ladik, J. J. *Theor. Chim. Acta* **1981**, *60*, 283.

(4) (a) Szczesniak, M.; Szczepaniak, K.; Kwiatkowski, J. S.; KuBulat, K.; Person, W. B. *J. Am. Chem. Soc.* **1988**, *110*, 8319. (b) Nowak, M. J.; Lapinski, L.; Fulara, J. *Spectrochim. Acta* **1989**, *45A*, 229.

(5) (a) Brown, D. J.; Lyall, J. M. *Aust. J. Chem.* **1962**, *15*, 851. (b) Katritsky, A. R.; Waring, A. J. *J. Chem. Soc.* **1962**, 1540.

0002-7863/93/1515-760$04.00/0

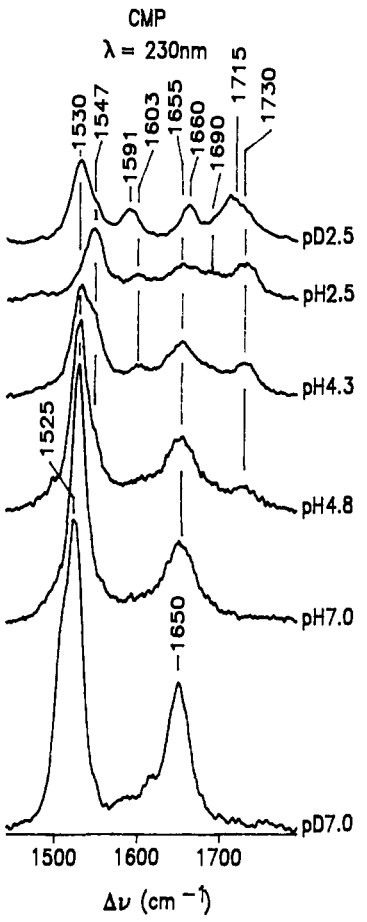

Figure 1. Excited (230-nm) UVRR spectra of 5'-CMP (5 mM) in water at various pH values between 7.0 and 2.5. The limiting spectra for the compound in D_2O are also shown. Samples contained 0.3 M Na_2SO_4, and the ν_1 band of SO_4^{2-} (980 cm^{-1}, not shown) was used as an intensity standard for scaling the spectra.

samples, the readings were corrected for the deuterium isotope effect.[6a] The 5'-CMP was checked for UV absorbing impurities by HPLC and found to contain none. The susceptibility of cytidine to deamination was checked by HPLC for solutions at pH 1, 3, 5, 7, and 14, which were incubated for 6 h at 25 °C and then for 12 h at 37 °C. Only the pH 14 sample showed any trace of uridine formation. Cytidine deamination is known to be accelerated by both acid and base, but catalysis by OH^- is the dominant effect.[6b,c] Under our conditions, deamination is not detected.

Methods. The UVRR spectrometer has been described elsewhere.[7] Spectra for NMA and cyclocytidine were obtained by using a 10-Hz H_2-Raman-shifted Nd:YAG laser. A 300-Hz excimer/dye laser (Lambda Physik LPX130/FL 3002) was used for all other spectra. The scattered photons were dispersed with a SPEX 1269 single monochromator, equipped with UV-enhanced optics and a 2400-grooves/mm holographic grating. The spectra were recorded in spectrograph mode with an intensified diode array detector (Princeton Instruments) or in scanning mode with a solar blind photomultiplier (Hamamatsu) and an electronic integrator (Stanford Research). The diode array response function was recorded with a diffuse white light source and used to correct the spectra. A polarization scrambler was used to correct the intensity variation due to polarization. The spectral data were stored on a computer disk and processed with Lab Calc software. The excitation profiles were determined as peak height ratios relative to sulfate internal standard. All spectra were recorded at room temperature unless otherwise indicated.

Results and Discussion

CMP Protonation Yields Spectra of Two Species. Protonation of 5'-CMP produces major changes in the UVRR spectrum, as shown in Figure 1. The bands in the 1500–1800-cm^{-1} region are

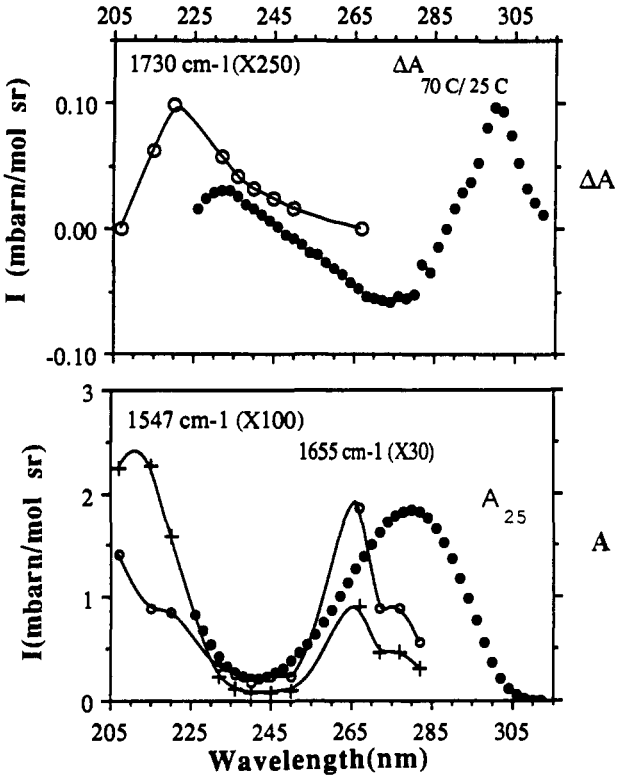

Figure 2. Excitation profiles for three bands in the UVRR spectra of 5'-CMP in water at pH 2.5: the 1547- (+) and 1655- (○) cm^{-1} EPs are shown in the lower panel, and the 1730-cm^{-1} (○) EP is shown in the upper panel. For comparison, the absorption spectrum (●) of CMP at 25 °C is plotted in the lower panel, while the 70–25 °C temperature difference spectrum (●) is plotted in the upper panel. The Raman intensities were measured as peak heights relative to the 980-cm^{-1} peak of SO_4^{2-}, present as an internal standard and expressed in units[17] of mbarn/steradian per mole of CMP. Scaling factors are shown in parentheses.

associated with stretching vibrations of cytosine double bonds and with N–H bending vibrations.[8] The pK_a for ring protonation of CMP is 4.3,[5] and it can be seen that one set of bands is replaced by another (see, e.g., the contour of the 1530/1547-cm^{-1} bands) as the pH is lowered through this value. When the spectra are taken in D_2O, appreciable band shifts are seen, due to replacement of labile N-bound protons by deuterons.

The most interesting feature of the spectra is the 1730-cm^{-1} band that develops as the pH is lowered and that shifts down 15 cm^{-1} in D_2O. The high frequency and D_2O sensitivity identifies this band as the C=N stretching vibration of an iminol unit, as shown in the next section. The 1730-cm^{-1} band responds differently to the excitation wavelength than do the other bands in the low-pH spectrum, as shown in Figure 2. The other bands display excitation profile (EP) maxima at ca. 265 nm and at ca. 210 nm within the two bands in the absorption spectrum. In contrast, the 1730-cm^{-1} band is undetectable at 266 or 204 nm; it reaches its maximum intensity at ca. 220 nm in the low-energy tail of the 200-nm absorption band. Moreover, the 1730-cm^{-1} band has a different temperature dependence than the other bands. Figure 3 shows that the 1730-cm^{-1} band gains in relative intensity as the temperature of the solution is raised. This behavior identifies the 1730-cm^{-1} band with a chemical species distinct from but in equilibrium with the major species responsible for the other UVRR bands. We assign this species as the iminol tautomer of protonated CMP, whose structure is diagrammed in Figure 5 (labeled imH^+/imol). The equilibrium is shifted toward the iminol tautomer with increasing temperature, indicating a positive enthalpy change. This shift in the equilibrium is also reflected in alterations

(6) (a) Gasoe, P. K.; Long, F. A. *J. Phys. Chem.* **1960**, *64*, 188. (b) Garrett, E. R.; Tsau, J. *J. Pharm. Sci.* **1972**, *61*, 1052. (c) Sowers, L. C.; Sedwick, W. D.; Shaw, B. B. *Mutat. Res.* **1989**, *215*, 131.

(7) Su, C.; Wang, Y.; Spiro, T. G. *J. Raman Spectrosc.* **1990**, *21*, 435.

(8) Tsuboi, M.; Takahashi, S.; Harada, I. In *Physico-Chemical Properties of Nucleic Acids*; Duchesne, J., Ed., Academic Press: New York, **1973**; Vol. 2, p 91.

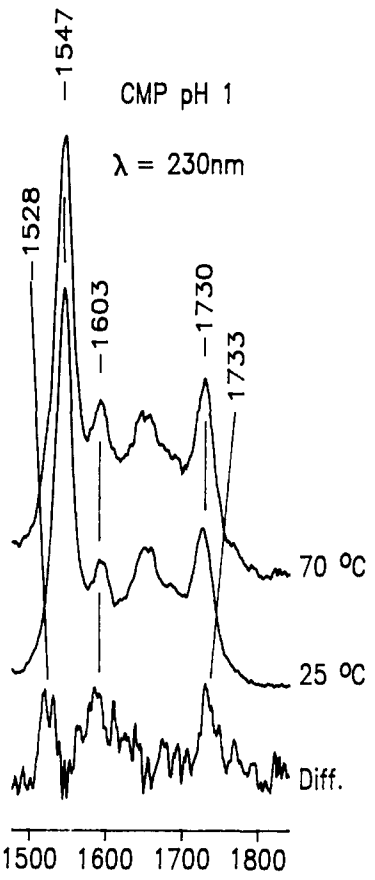

Figure 3. Excited (230-nm) UVRR spectra of protonated CMP at 70 and 25 °C, and the difference spectrum.

Table I. C=N Frequencies (cm^{-1}) for the Indicated Molecules in H_2O and in D_2O

	H_2O	D_2O	Δ^c
N-methylacetimidic acid (Me(HO)C=N$^+$(H)Me)[a]	1710	1686	-24
N,*N*-dimethylacetimidic acid (Me(HO)C=N$^+$(Me)$_2$)[a]	1697	1686	-11
acetimidic acid (Me(HO)C=N$^+$H$_2$)[a]	1712	1662	-50
protonated ethylacetimidate (Me(EtO)C=N$^+$H$_2$)[a]	1682	1646	-36
protonated Schiff base ((Phe)HC=N$^+$(H)(*n*-Bu))[b]	1680	1660	-20
Schiff base ((Phe)HC=N(*n*-Bu))[b]	1646	1646	

[a] This work; from 204-nm excited UVRR spectra in acidic solutions. [b] Reference 9a. [c] Δ: D_2O - H_2O difference.

of the absorption spectrum; the 70–25 °C difference absorption spectrum (Figure 2) shows a minimum at 275 nm, near the absorption maximum for the major amino/keto tautomer, and a maximum at 232 nm, near the 1730-cm^{-1} EP maximum. Thus, the iminol tautomer has an electronic transition lying between those of the amino/keto tautomer. The additional sharp 302-nm maximum in the difference UV absorption spectrum is probably due to hot band contributions to the long wavelength transition of the amino/keto tautomer, rather than to an iminol tautomer transition; no 1730-cm^{-1} RR band enhancement is seen with excitation near 300 nm (not shown).

The 75–25 °C difference UVRR spectrum at low pH (Figure 3, obtained by subtraction with a scale factor that avoids negative peaks) reveals additional bands at 1528 and 1603 cm^{-1}, which are also attributed to the iminol tautomer. In D_2O, the 1603-cm^{-1} band disappears, and the 1528-cm^{-1} band shifts to 1515 cm^{-1} (not shown). We note that the 25 °C spectrum was fully recovered after the solution was cooled. Thus, the 1730-cm^{-1} band is not an artifact of irreversible modification of CMP.

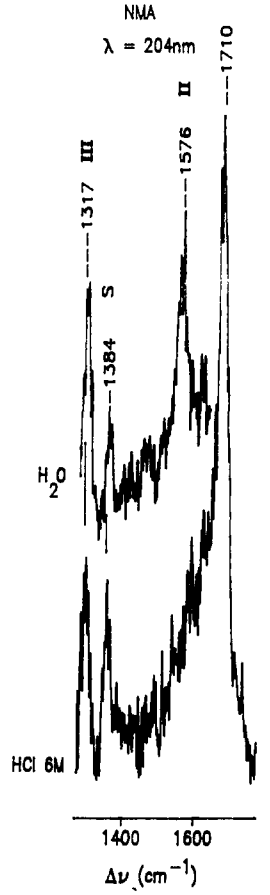

Figure 4. Excited (204-nm) UVRR spectra of NMA in water and in 6 M HCl. The amide II, III, and S assignments are indicated (H_2O). The 1710-cm^{-1} band (6 M HCl) is assigned to the C=N stretching mode of the imidic acid.

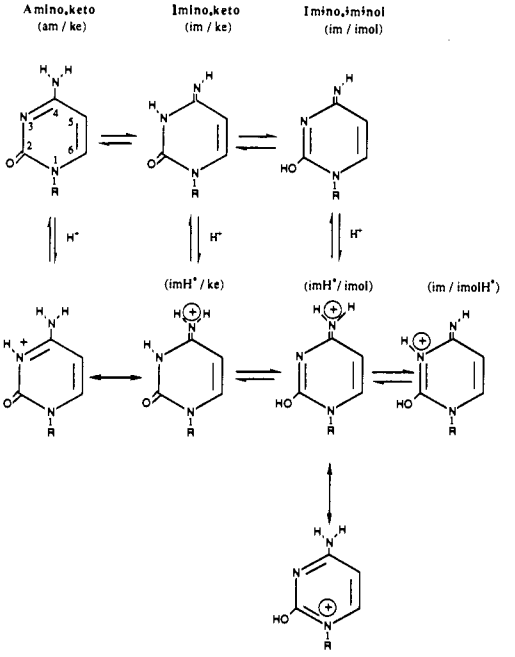

Figure 5. Diagram of tautomeric and resonance Raman forms available to neutral and protonated N1-substituted cytosine.

UVRR Spectroscopy of the Iminol Bond. Assignment of the 1730-cm^{-1} band to the C=N stretch of the iminol tautomer of protonated CMP rests on comparisons with the model compounds listed in Table I. A simple Schiff base imine has a much lower C=N stretching frequency, e.g., 1646 cm^{-1} in the case of *N*-benzylidene-*N*-butylamine.[9a] Upon protonation, this frequency

shifts up to 1680 cm^{-1}. Part of the increase is electronic in origin, due to rehybridization of the C=N bond,[9b] and part is due to interaction with the N–H bending coordinate, as indicated by the 20-cm^{-1} decrease in the N–D isotopomer.[9b]

The frequency increases further when the C=N double bond bears a hydroxyl substituent. Figure 4 shows UVRR spectra of *N*-methylacetamide (NMA) at neutral pH and in 6 M HCl, an acid concentration which is sufficient to protonate the carbonyl group ($pK_a \sim 0.7$)[10] and convert the amide to the imidic acid (structure shown in Table I). At neutral pH, the UVRR spectrum is dominated by the amide II, III, and S modes, which involve coupled C–N stretching, N–H bending, and C–H bending coordinates.[11] The amide I C=O stretching band at 1639 cm^{-1} is very weak, due to minimal resonance enhancement in aqueous solution.[11] The UVRR spectrum of the acid solution shows a single strong band at 1710 cm^{-1}, shifting to 1686 cm^{-1} in D_2O (not shown). This band is assigned to the C=N stretching vibration of the imidic acid, coupled to the motions of the two exchangeable protons on the N and O atoms. Similar spectra were obtained for the imidic acids of acetamide and *N,N*-dimethylacetamide and for protonated ethyl acetimidate. The frequencies of the strongly enhanced C=N stretching UVRR bands are listed in Table I for H_2O and D_2O solutions.

Electronic effects on the C=N bond are best evaluated from the D_2O frequencies, since these are unaffected by kinematic coupling with N–H and/or O–H bending modes. The D_2O frequency is 26 cm^{-1} higher for *N*-methylacetimidic acid than for the protonated Schiff base (Table I). This frequency elevation is attributed to the electronegativity of the OH substituent, which withdraws electron density from the C atom, stabilizing the C=N bond. Alkyl substituents on the N and O atoms also have significant electron effects on the C=N bond strength. Using acetimidic acid as a reference compound, it can be seen that the C=N frequency is raised by alkyl substitution on the N atom but lowered by alkyl substitution on the O atom. These shifts result from the electron-donating effect of alkyl groups, which stabilize the $^-O—C=N^+$ dipole when they are substituted for H at the N atom but destabilize it when substituted at the O atom.

The C=N frequencies in H_2O are elevated from those in D_2O when there are exchangeable protons on the N and/or O atoms. Although the effects are not entirely additive, each of these protons produces a roughly 15 cm^{-1} elevation. In particular, the effect of an OH proton is 14 cm^{-1}, as judged by the difference of the $D_2O - H_2O$ differences of acetimidic acid ($\Delta = -50$ cm^{-1}) and protonated ethyl acetimidate ($\Delta = -36$ cm^{-1}), which differ only by the substitution of H with an ethyl substituent on the O atom. This estimate is in quantitative agreement with the 15-cm^{-1} D_2O shift of the 1730-cm^{-1} RR band of protonated CMP and supports its assignment to the iminol tautomer, which has a proton on the O atom but not on the N atom of the C=N bond (Figure 5). The D_2O frequency itself is higher than that of *N*-methyl acetimidic acid, an effect that may be due to the C=N bond being part of the cytosine ring.

If the molar scattering factors were known for the UVRR bands, then their relative intensities could be used to calculate the tautomer populations. These factors are not known, but we note that the intensity of the protonated CMP band at 1730 cm^{-1} is 25 mbarn/steridian per mole of CMP at its EP maximum (Figure 2), while the intensities of the 1547- and 1655-cm^{-1} bands at their EP maxima range from 50 to 200. These numbers suggest that the fraction of protonated CMP present as the iminol tautomer is in the range of 10–30% if the scattering factors are comparable at the EP maxima, i.e., if the electronic transition moments and the Franck–Condon factors[12] are comparable for the two tautomers. This assumption is far from reliable, however, and the population estimate is merely an educated guess. The increase in the iminol tautomer fraction between 25 and 70 °C is about 30%, as judged from the increase in the 1730-cm^{-1} UVRR band intensity (Figure 3).

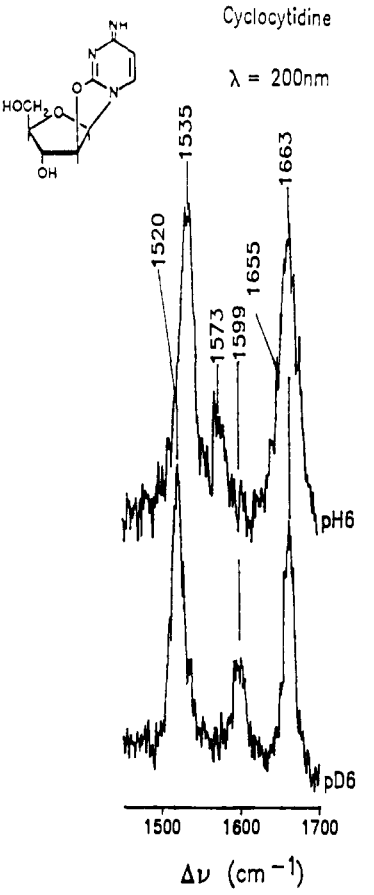

Figure 6. Excited (200-nm) UVRR spectra of cyclocytidine (see inset structure) in water (pH 6) and D_2O (pD 6).

Cytidine Alkylation Illuminates the Tautomer Spectral Pattern. In Figure 5, we enumerate the tautomeric structures accessible to neutral and protonated cytidine residues (C). The most stable tautomer of neutral C is the amino keto (labeled am/ke) tautomer, containing exocyclic amino and keto groups. The next higher energy structure is the imino keto tautomer (im/ke), in which a proton is transferred from the exocyclic amino group to the ring N3 atom, with the attendant rearrangement from a ring to an exocyclic C=N double bond. The energy is higher because the exocyclic N4 is more basic than the ring N3[12b] and because the carbonyl group is left unconjugated. The highest-energy structure is the imino iminol tautomer (im/imol), in which the amino proton is transferred to the carbonyl group, with the attendant shift from an exocyclic C=O to a ring C=N double bond.

Upon protonation of C, the am/ke tautomer remains the most stable one; the im/ke structure is now a (minor) resonance form since it is distinguished from the protonated am/ke structure only by the rearrangement of electrons. On the other hand, two tautomers can be derived by protonation of the im/imol structure, since the proton can be placed on N4 or on N3, yielding imH^+/imol and im/$imolH^+$ tautomers, respectively. The former protonation site is more probable, again because N4 is more basic than N3. Finally, the imH^+/imol tautomer has an energetically accessible benzenoid resonance form, in which the positive charge is shifted from N4 to N1. Anticipating the discussion below, we assign the UVRR spectrum of the major species in protonated CMP to the am/ke tautomer and the spectrum of the minor species to the imH^+/imol tautomer.

The most direct evidence for the imH^+/imol assignment is the UVRR spectrum of cyclocytidine, shown in Figure 6. In this derivative (see structure in Figure 6), the carbonyl oxygen is

(9) (a) Lopez-Gariega, J. J.; Babcock, G. T.; Harrison, J. F. *J. Am. Chem. Soc.* **1986**, *108*, 7241. (b) Lopez-Gariega, J. J.; Hanton, S.; Babcock, G. T.; Harrison, J. F. *J. Am. Chem. Soc.* **1986**, *108*, 7251.
(10) Martin, B. R.; Hulton, W. C. *J. Am. Chem. Soc.* **1973**, *95*, 4752.
(11) Wang, Y.; Purrello, R.; Jordan, T.; Spiro, T. G. *J. Am. Chem. Soc.* **1991**, *113*, 6359.
(12) (a) Spiro, T. G.; Stein, P. *Ann. Rev. Phys. Chem.* **1977**, *28*, 501. (b) Saenger, W. *Principles of Nucleic Acid Structure*, Springer Verlag: New York, 1984, pp 105–115.

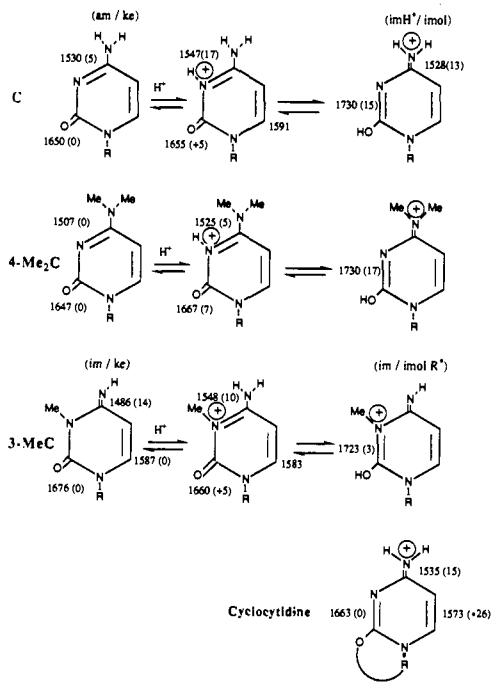

Figure 7. Assignments of UVRR frequencies (cm^{-1}) to modes involving the indicated bonds of neutral and protonated CMP (C), N^4,N^4-dimethyl-2′-deoxycytidine, N^3-methylcytidine, and cyclocytidine (the curved line represents a sugar ring). Downshifts or (+) upshifts in D_2O are given in parentheses.

converted to an ether functionality linked to the sugar ring, thereby forcing a structure analogous to imH^+/imol but with an alkyl substituent replacing the iminol proton. Because of the basicity of N4, the molecule is protonated at neutral pH, the pK_a for deprotonation to the im/imol analog being ca. 12.[5a] The highest-frequency band in the UVRR spectrum is at 1663 cm^{-1}, and it does not shift in D_2O. This band is assigned to the C=N stretch of the RO—C=N unit; the frequency is slightly higher than that of protonated ethyl acetimidate in D_2O, 1646 cm^{-1} (Table I) but substantially lower than the 1715-cm^{-1} frequency assigned to the imH^+/imol tautomer of protonated CMP in D_2O. This frequency depression is consistent with the expected electronic effect of an alkyl substituent on the O atom. The lack of D_2O shift of the 1663-cm^{-1} band rules out the alternative tautomer analogous to im/$imolH^+$, with a proton on N3, since the C=N stretch would then be coupled to N–H bending. The benzenoid resonance form (Figure 5) is judged not to be an important contributor to the structure; if it were, then the 1663-cm^{-1} band would have to be assigned as a benzene ring mode, but substituted benzenes do not have modes at this high a frequency.[13]

When cytidine is methylated at N3, the neutral base is forced into the im/ke structure (Figure 7). The UVRR spectrum of 3-methylcytidine (Figure 8, pH 11) shows the C=O stretch at 1676 cm^{-1}, upshifted from CMP (1655 cm^{-1}) because of the localization of the carbonyl group in the im/ke structure. Protonation at N4 occurs at relatively high pH (pK_a = 8.7),[14] producing the am/ke structure; consequently, the C=O frequency in the pH 5 UVRR spectrum (Figure 8) shifts back down to 1660 cm^{-1} as a result of conjugation with the ring C=N bond (Figure 7). An additional band is seen at 1723 cm^{-1}, however, which is attributed to the iminol stretch of the im/$imolR^+$ structure (Figure 7), the only rare tautomer accessible to 3-methylcytidine.

This tautomer is prevented when the N4 position is dimethylated. For protonated N,N-dimethyl-2′-deoxycytidine, the only rare tautomer is the structure analogous to imH^+/imol, and the UVRR spectrum, shown in Figure 9, contains a 1730-cm^{-1} band at low pH, which shifts to 1713 cm^{-1} in D_2O. This is the

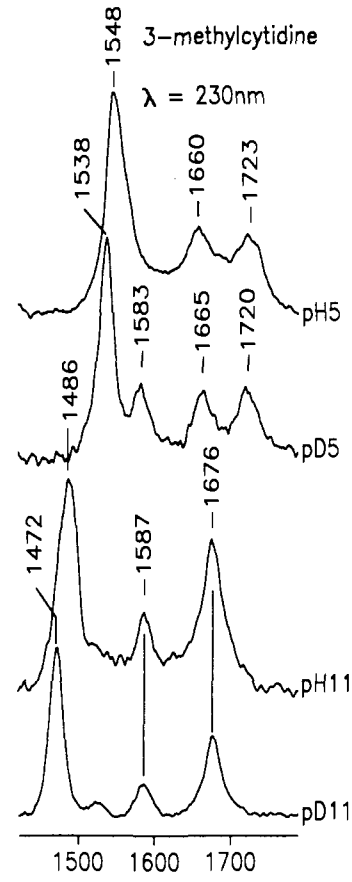

Figure 8. Excited (230-nm) UVRR spectra of N^3-methylcytidine at pH or pD 5 and at pH or pD 11.

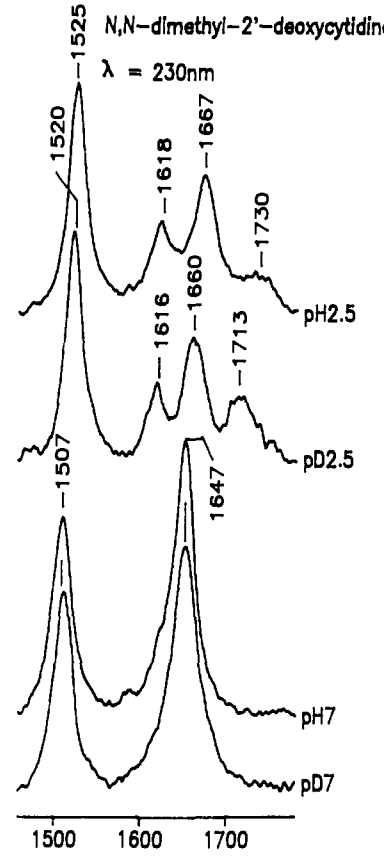

Figure 9. Excited (230-nm) UVRR spectra of N^4,N^4-dimethyl-2′-deoxycytidine in water and D_2O in neutral (pH, pD 7) and acid (pH, pD 2.5) aqueous solutions.

(13) Colthup, N. D.; Daly, L. H.; Wiberley, S. E. *Introduction to Infrared and Raman Spectroscopy*, 3rd ed.; Academic Press: New York, 1990; p 263.
(14) Ueda, T.; Fox, J. J. *J. Am. Chem. Soc.* **1963**, *85*, 4024.

same frequency and the same D_2O sensitivity found for the minor species in protonated CMP, confirming that it is the imH^+/imol tautomer. The relative intensity of the 1730-cm^{-1} band was found to be temperature-sensitive for protonated *N,N*-dimethyl-2′-deoxycytidine (not shown), as it is for protonated CMP.

Assignments. Most of the UVRR bands above 1500 cm^{-1} can be assigned in a consistent way by comparing analogous tautomer structures, as shown in Figure 7. For example, all of the am/ke tautomers have C=O stretching bands at ca. 1660 cm^{-1}, which are essentially insensitive to D_2O. In addition, they each show a strong band between 1500 and 1550 cm^{-1}, which is assigned to a ring stretching mode involving the N3=C4 bond, consistent with normal mode calculations on 1-methylcytosine.[8] This assignment is confirmed by the D_2O shifts, which are large (17 cm^{-1}) when there are protons on both N3 and N4 (protonated C), small (5 cm^{-1}) when only N3 or N4 have protons (neutral C or protonated 4-Me_2C), and zero when neither N has a proton (neutral 4-Me_2C). This mode is much lower in frequency than standard C=N stretches because it is coupled to other bond stretches in the ring, especially the C5=C6 double bond. A mode primarily involving C5=C6 stretching is calculated[8] at 1588 cm^{-1}. Candidate bands are seen at 1591 cm^{-1} for protonated C and a 1583 cm^{-1} for protonated 3-MeC, both in D_2O (Figures 1 and 8). Curiously, these bands are missing in H_2O and are not seen at all for the other am/ke tautomers. Apparently, the mode is too weakly enhanced at 230 nm to be observed in most cases. The only other modes expected above 1530 cm^{-1} for the am/ke tautomers are the δNH_2 scissors mode, seen at 1603 cm^{-1} for protonated C (Figure 1), and the N3H in-plane bend, which is not observed.

Turning to the imino tautomers, we note that the UVRR spectrum is particularly simple for unprotonated 3-methylcytidine (Figure 8; pH 11), the only example of an im/ke tautomer. There are three prominent bands, at 1676, 1587, and 1486 cm^{-1}, and only the last of these is D_2O-sensitive, shifting to 1472 cm^{-1}. The 1676-cm^{-1} band is readily assigned to the C=O stretch, as discussed above, and the 1587-cm^{-1} band is at an appropriate frequency for the C5=C6 stretch. This leaves the 1486-cm^{-1} band to be assigned to the exocyclic C=N4 stretch, consistent with the large shift upon N4H/D exchange. Again, the frequency is much lower than it is for isolated C=N bonds (Table I), an effect which is attributed to coupling with ring modes, especially the stretch of the C5=C6 bond, with which the C=N4 bond is in line and conjugated.

As discussed above, all iminol tautomers have ring (HO)C=N stretches at 1720–1730 cm^{-1}, except for cyclocytidine, whose frequency is lowered to 1663 cm^{-1} by the *O*-alkyl substitution. All of the frequencies are about 20 cm^{-1} higher than for the acyclic analogs (Table I) *N*-methyl- or *N,N*-dimethylacetimidic acid or ethyl acetimidate, an elevation which reflects coupling effects in the ring. For cyclocytidine, the prominent 1535-cm^{-1} band, which shifts down 15 cm^{-1} in D_2O, is assigned to the C4=N4 stretch, the frequency being shifted up from that observed for unprotonated 3-methylcytidine by the positive charge on the protonated N4. The 1528-cm^{-1} band (1515 cm^{-1} in D_2O) in the temperature difference spectrum of protonated CMP (Figure 3) is similarly assigned. Cyclocytidine in D_2O shows a band at 1598 cm^{-1} which is tentatively assigned to the C5=C6 stretch, its frequency elevated slightly from that of unprotonated 3-methylcytidine via coupling to the elevated C4=N4 stretch. In H_2O, however, this band is missing, but a 1573-cm^{-1} band is seen; a shoulder is also seen on the low-energy side of the 1663-cm^{-1} band in H_2O. It is conceivable that this shoulder (ca. 1655 cm^{-1}) is the δNH_2 scissors mode, its frequency shifted up from the CMP value (1603 cm^{-1}) by the positive charge on N4. This mode might couple with the C5=C6 stretch, pushing the frequency of the latter down to 1573 cm^{-1}; the upshift in D_2O would then result from relief of this coupling. Such an effect is actually seen in the 1-methylcytosine normal mode calculation.[5a] The $\delta N4H_2$ and C5=C6 modes are calculated at 1630 and 1588 cm^{-1}, while the latter mode is calculated at 1601 cm^{-1} when the N4H atoms are substituted by D.

Hemiprotonated Poly(C^+·C) Structure Inhibits Formation of the Iminol Tautomer. When the pH of a solution of poly(C) is lowered from neutrality, a hemiprotonated duplex with C^+·C base pairs is formed.[15] As can be seen in Figure 10a, a C residue protonated at N3 can form three H-bonds with an unprotonated C residue in the opposite strand of a parallel-stranded double helix. The stabilization afforded by the stacking of these hemiprotonated base pairs in a helix facilitates protonation of half the C residues between pH 5 and 7, depending upon the ionic strength,[15] well above the intrinsic pK_a of 4.3 for CMP.[5a] This hemiprotonated duplex has a limited pH range of acid stability because, as shown below, protonation of the remaining C residues disrupts the H-bonding scheme for the hemiprotonated base pairs.

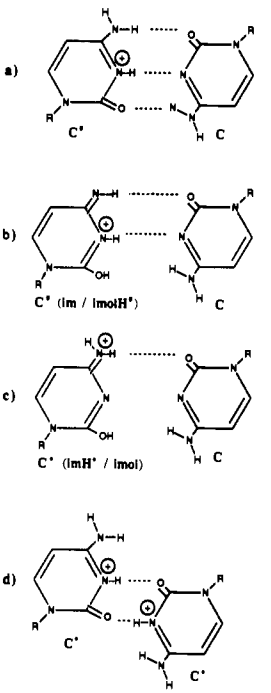

Figure 10. Diagram showing (a) H-bonding scheme for pairing of C^+ and C in a parallel-stranded duplex; (b) the loss of one H-bond if the im/$imolH^+$ tautomer of C^+ is populated; (c) the loss of two H-bonds if the imH^+/imol tautomer of C^+ is populated; and (d) potential H-bonding scheme for pairing of C^+ with C^+ in a parallel-stranded duplex.

As is evident from Figure 10, the H-bonding pattern for C^+·C base pairs requires the amino keto tautomeric form of the bases. It follows that duplex formation should alter the apparent equilibrium constant for tautomerization, for if the protonated residue were to adopt the im/$imolH^+$ or the imH^+/imol tautomer structure, then one or two of the three H-bonds of the base pair would be disrupted (Figures 10b and c). Hence, duplex formation ought to inhibit any tautomerization associated with the protonation of C residues.

Figure 11, which shows UVRR spectra for poly(C) as a function of pH, provides evidence in support of this view. Thus, at pH 7, the UVRR spectrum of poly(C) is nearly the same as that of CMP (Figure 1), with major bands at 1530 and 1655 cm^{-1}. As the pH is lowered, protonation of the C residues is readily monitored by the growth of the 1547-cm^{-1} band, at the expense of the much stronger 1530-cm^{-1} band. However, as the pH is lowered more, further growth of the 1547-cm^{-1} band is retarded relative to that seen for CMP (Figure 1), indicating that further proton uptake does not proceed for poly(C), due to hemiprotonated helix stability. Thus, the 1547-cm^{-1} band is about equal to the 1530-cm^{-1} band at pH 4.3 for CMP but only at pH 3.5 for poly(C). At the intermediate pH values, the total intensity of the poly(C) 1530/1545-cm^{-1} band envelope relative to the 1655-cm^{-1} C=O band is notably weaker than at neutral or low pH. This is in contrast to the behavior of CMP, for which the intensity weakens continuously as the pH is lowered (Figure 1). This behavior is again consistent with duplex formation in hemiprotonated poly(C),

(15) Guschlbauer, W. *Proc. Natl. Acad. Sci., U.S.A.* **1967**, *57*, 1441.

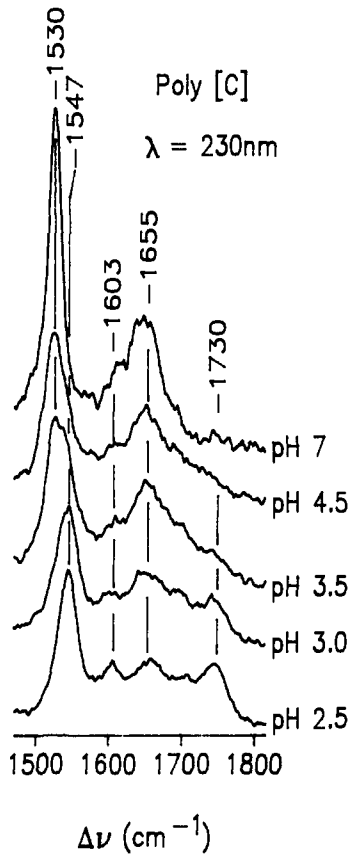

Figure 11. Excited (230-nm) UVRR spectra of poly(C) at different pH values.

since base stacking in the duplex produces additional hypochromism of the strong ring modes.[16] Even more striking, the diagnostic 1730-cm^{-1} band present by pH 4.8 in CMP is not present at intermediate pH values for poly(C). Even at pH 3.5, this band is barely detectable in the poly(C) spectrum. This behavior is consistent with suppression of iminol tautomer formation by the C$^+$·C base pairs of the poly(C) duplex.

At pH 2.5, where the residues of poly(C) are fully protonated, the spectra for polymer and monomer are nearly the same, consistent with the absence of any poly(C) helix. While protonation of the second residue of the base pairs of hemiprotonated poly(C) could conceivably lead to an alternative parallel-stranded duplex (Figure 10d), such an H-bonding scheme is incompatible with the reappearance of the 1730-cm^{-1} band diagnostic of the iminol tautomer when poly(C) is titrated below pH 3. Hence, the presence of the tautomer below the pH range of stability of the hemiprotonated duplex indicates that a double helix with doubly protonated C$^+$·C$^+$ base pairs does not form.

Can the C$^+$ Iminol Tautomer Stabilize A·C Mispairs? It is of interest to consider the implications of iminol tautomer formation for A·C base pairs in Watson–Crick duplexes. The structure of such "mispairs" has been studied by introducing A–C oppositions in oligonucleotides which otherwise form Watson–Crick duplexes.[18,19] Kennard and co-workers[18] found crystallographically that the A and C residues form a "wobble"-type pair,[20] in which the purine and pyrimidine bases are displaced from normal Watson–Crick geometry so that the contacts are between C6–NH$_2$ of A and N3 of C and between N1 of A and C2=O of C (Figure 12). To provide for an H-bond at this latter contact, Kennard et al. proposed that N1 of A becomes protonated (Figure 12a). Gao and Patel[19] found a similar geometry for the A·C mispair in oligonucleotides via 2-D proton NMR but were unable to determine the protonation status due to the possibility of rapid exchange. They did, however, observe a substantial narrowing of the resonances when the pH was lowered from 7.0 to 5.7, suggesting that the A·C pair might be stabilized by protonation.

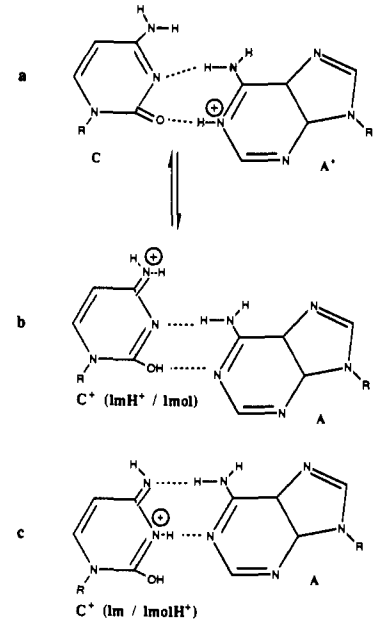

Figure 12. (a) H-bonding scheme for the proposed wobble pair between C and protonated A, and (b) the alternative pair between A and the imH$^+$/imol tautomer of protonated C. (c) H-bonding scheme for a Watson–Crick pair between A and the im/imolH$^+$ tautomer of protonated C.

Adenine can, indeed, be protonated at N1, and protonation would be favored by base pairing, as in the case of the poly(A$^+$·A$^+$) double helix.[21] If, however, the proton is shifted in the A·C structure from N1 of A to C2=O of C, then one has a sterically equivalent base pair between unprotonated A and the iminol tautomer of protonated C (Figure 12b).[22] This alternative base pair structure may actually be the more stable one in view of the greater intrinsic basicity of C. The pK_a of C monomer is nearly 1 unit higher than that of A monomer, so C is nearly 10-fold more basic. If the iminol tautomer population in protonated C is 10% or higher, as our data suggest (see above), then the proton in the base pair would be expected to reside on C2=O of C rather than on N1 of A. This possibility could be checked by monitoring the iminol 1730-cm^{-1} band of A·C pairs in an appropriate oligonucleotide duplex.

Assuming that wobble A·C$^+$ imH$^+$/imol (Figure 12b) is the equilibrium form of an A·C mispair resulting from a transition error during DNA replication, this is the form of the mispair that must be encountered biologically by DNA repair systems. However, this should not be taken to mean that this is also the form of the mispairing event that occurs during DNA replication, as Kennard et al.[18] have suggested. The steric constraints likely to be imposed by the replicative apparatus and the nonequilibrium nature of the DNA replication process probably do not allow base

(16) Fodor, S. P. A.; Spiro, T. G. *J. Am. Chem. Soc.* **1986**, *108*, 3198.
(17) Dudik, J. M.; Johnson, C. R.; Asher, S. A. *J. Chem. Phys.* **1985**, *82*, 1732.
(18) (a) Hunter, W. N.; Brown, T.; Anand, N. N.; Kennard, D. *Nature* **1986**, *320*, 552. (b) Kennard, D. In *Structure and Expression, Vol. 2: DNA and its Drug Complexes*; Sarma, R. H., Sarma, M. H., Eds.; Adenine Press: Guilderland, NY, 1988; pp 1–25.
(19) Gao, X.; Patel, D. J. *Proc. Natl. Acad. of Sci., U.S.A.* **1987**, *35*, 16973.
(20) Crick, F. H. C. *J. Mol. Biol.* **1966**, *19*, 548.
(21) (a) Fresco, J. R.; Klemperer, E. *Ann. N. Y. Acad. Sci.* **1959**, *81*, 730. (b) Fresco, J. R. *J. Mol. Biol.* **1959**, *1*, 106.
(22) Hunter et al.[18a] considered the possibility of wobble pairing between unprotonated A and the im/imol tautomer of unprotonated C (a base pair similar to that in Figure 12b but with a proton removed from the NH$_2$ group of C). Since there is no spectroscopic indication for this tautomer of CMP, it must be less stable in relation to *unprotonated* C than is the imH$^+$/imol tautomer in relation to protonated C. Hence, protonation is expected to stabilize the iminol tautomer A·C$^+$ base pair (Figure 12b) as well as the A$^+$·C structure favored by Hunter et al. (Figure 12a).

mispairs with significant glycosyl bond dihedral angle deviations from those of the standard Watson–Crick pairs.[2a] Therefore, base mispairs are much more likely to contain unfavored tautomers that allow glycosyl bond geometry very close to that of the standard Watson–Crick pairs, i.e., A amino·C imino/keto,[2a] A imino·C amino/keto,[2a] or A amino·C^+ im/imolH^+ (Figure 12c). Hence, what is observed at equilibrium need not be what gives rise to the initial mispairing event.

Conclusions

A minor tautomer of protonated cytidine has a population in aqueous solution possibly in the range of 10–30%, which is detectable by UVRR spectroscopy via its distinctive 1730-cm^{-1} band. On the basis of comparisons with model compounds, the frequency and D_2O sensitivity of this band identify the tautomer as the structure with an iminol bond and a protonated exocyclic imine. Tautomerization is suppressed in hemiprotonated poly(C^+·C) by the H-bonding constraints of the C^+·C base pair in the duplex structure. The reappearance of the tautomer band at low pH indicates that fully protonated poly(C^+·C^+) does not form a base-paired duplex. The iminol tautomer of C^+ may be an important contributor to the observed wobble pairing of A–C oppositions in duplex structures.

Acknowledgment. This work was supported by National Institutes of Health Grants GM 25158 (to T.G.S.) and GM 42936 (to J.R.F.) and by National Science Foundation Grant DMB 8419060 (to J.R.F.).

X-ray Absorption Spectroscopic Studies of the Blue Copper Site: Metal and Ligand K-Edge Studies To Probe the Origin of the EPR Hyperfine Splitting in Plastocyanin

Susan E. Shadle,[†] James E. Penner-Hahn,[§] Harvey J. Schugar,[⊥] Britt Hedman,[*,‡] Keith O. Hodgson,[*,†,‡] and Edward I. Solomon[*,†]

Contribution from the Department of Chemistry and Stanford Synchrotron Radiation Laboratory, Stanford University, Stanford, California 94305. *Received June 15, 1992*

Abstract: X-ray absorption spectra for the oxidized blue copper protein plastocyanin and several Cu(II) model complexes have been measured at both the Cu K-edge and the ligand K-edges (Cl and S) in order to elucidate the source of the small parallel hyperfine splitting in the EPR spectra of blue copper centers. Assignment and analysis of a feature in the Cu K-edge X-ray absorption spectrum at ~8987 eV as the Cu 1s → 4p + ligand-to-metal charge-transfer shakedown transition has allowed for quantitation of 4p mixing into the ground-state wave function as reflected in the 1s → 3d (+4p) intensity at ~8979 eV. The results show that distorted tetrahedral (D_{2d}) $CuCl_4^{2-}$ is characterized by <4% Cu $4p_z$ mixing, while plastocyanin has only Cu $4p_{xy}$ mixing. Thus, the small parallel hyperfine splitting in the EPR spectra of D_{2d} $CuCl_4^{2-}$ and of oxidized plastocyanin cannot be explained by 12% $4p_z$ mixing into the $3d_{x^2-y^2}$ orbital as had been previously postulated. Data collected at the Cl K-edge for $CuCl_4^{2-}$ show that the intensity of the ligand pre-edge feature at ~2820 eV reflects the degree of covalency between the metal half-occupied orbital and the ligands. The data show that D_{2d} $CuCl_4^{2-}$ is not unusually covalent. The source of the small parallel splitting in the EPR of D_{2d} $CuCl_4^{2-}$ is discussed. Experiments at the S K-edge (~2470 eV) show that plastocyanin is characterized by a highly covalent Cu–S(cysteine) bond relative to the cupric–thiolate model complex [Cu(tet *b*)(*o*-$SC_6H_4CO_2$)]·H_2O. Self-consistent-field-Xα-scattered-wave calculations have been used to understand copper–thiolate bonding in this model complex and to quantify the covalency reflected in the S K-edge intensity data. The XAS results demonstrate that the small parallel hyperfine splitting in the EPR spectra of blue copper sites reflects the high degree of covalency of the copper–thiolate bond.

Introduction

The oxidized blue copper protein active site is characterized by unique spectral features compared to those of normal, tetragonal Cu(II) complexes.[1] The development of a detailed understanding of these features has been the goal of many experimental and theoretical studies. The intense blue color of the oxidized blue copper proteins results from an optical absorption at ~600 nm. The extinction coefficient of this feature is 2 orders of magnitude greater than that for absorption bands in the same region in normal tetragonal copper(II) complexes. Polarized absorption and low-temperature magnetic circular dichroism (MCD) studies in combination with self-consistent-field-Xα-scattered-wave (Xα) calculations have definitively assigned this feature as a S(Cys) $3p\pi \rightarrow$ Cu $3d_{x^2-y^2}$ charge-transfer (CT) transition.[2]

The EPR spectra of oxidized blue copper centers exhibit $g_\parallel > g_\perp > 2.00$, indicating that the half-occupied ground-state orbital is $3d_{x^2-y^2}$. The same ground state is found in normal tetragonal copper(II) complexes. However, the EPR spectra of blue copper centers exhibit unusually small parallel hyperfine splitting, the magnitude of which (60×10^{-4} cm^{-1}) is about one-third that of normal copper hyperfine splitting. The half-occupied 3d orbital associated with this EPR signal is involved in the electron-transfer reactivity of the blue copper center. Thus, to gain insight into the function of these proteins, it is essential to have a clear understanding of the electronic structural origin of the small $A_\parallel$ splitting associated with this ground-state wave function.

* Persons to whom correspondence should be addressed.
† Stanford University, Department of Chemistry.
‡ Stanford Synchrotron Radiation Laboratory.
§ Present address: Department of Chemistry, University of Michigan, Ann Arbor, MI 48109-1055.
⊥ Present address: Department of Chemistry, Rutgers University, The State University of New Jersey, New Brunswick, NJ 08903.

(1) (a) Solomon, E. I.; Penfield, K. W.; Wilcox, D. E. In *Structure and Bonding*; Hemmerich, P., et al., Eds.; Springer-Verlag: New York, 1983; pp 1-57. (b) Gray, H. B.; Solomon, E. I. In *Copper Proteins*; Spiro, T. G., Ed.; Wiley: New York, 1980; pp 1-39. (c) Adman, E. T. In *Advances in Protein Chemistry, Vol. 42, Metalloproteins: Structural Aspects*; Anfinsen, C. B., Edsall, J. T., Richards, F. M., Eisenberg, D. S., Eds.; Academic Press, Inc.: New York, 1991; pp 145-192.
(2) (a) Penfield, K. W.; Gewirth, A. A.; Solomon, E. I. *J. Am. Chem. Soc.* **1985**, *107*, 4519-4529. (b) Gewirth, A. A.; Solomon, E. I. *J. Am. Chem. Soc.* **1988**, *110*, 3811-3819.

0002-7863/93/1515-767$04.00/0