Photochemistry and Photobiology, 2011, 87: 818–832

# Opened-Ring Adducts of 5-Methylcytosine and 1,5-Dimethylcytosine with Amines and Water and Evidence for an Opened-Ring Hydrate of 2′-Deoxycytidine

**Martin D. Shetlar* and Janet Chung§**

Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, CA

**Received 21 December 2010, accepted 15 April 2011, DOI: 10.1111/j.1751-1097.2011.00936.x**

## ABSTRACT

**A variety of nucleic acid components and related compounds undergo photoreaction with water to form so-called "photohydrates" (*e.g.* uracil forms 6-hydroxy-5,6-dihydrouracil). However, the corresponding hydrates of 5-methylcytosine (a minor nucleobase in eukaryotic DNA) and related compounds have not been characterized. We report the preparation of opened-ring forms of such products for 5-methylcytosine (m5C) and 1,5-dimethylcytosine (DMC). This was accomplished *via* thermal reaction of ring-opened amine adducts (*e.g.* *N*-carbamoyl-3-amino-2-methylacrylamidine (IVa) or *N*-(*N′*-methylcarbamoyl)-3-amino-2-methylacrylamidine (IVb)) produced by photo-induced reactions of m5C with ammonia or methylamine. When these adducts were treated with dilute trifluoroacetic acid, the amino group at the 3-position was replaced with a hydroxyl group; with IVa, *N*-carbamoyl-3-hydroxy-2-methylacrylamidine (Va) was formed, while reaction of IVb led to *N*-(*N′*-methylcarbamoyl)-3-hydroxy-2-methylacrylamidine (Vb). These compounds are ring-opened isomers of 5,6-dihydro-6-hydroxy-5-methylcytosine (Ia and IIa) and 5,6-dihydro-6-hydroxy-1,5-dimethylcytosine (Ib and IIb). Compounds Va and Vb each undergo thermal ring closure reactions to form two unstable compounds with chemical and UV spectral properties expected for Ia and IIa (or Ib and IIb). The latter compounds have been identified as minor products in UV-irradiated aqueous solutions of m5C and DMC. Evidence is also presented that the 2′-deoxycytidine photohydrates coexist with an opened-ring form, possibly similar in nature to Vb.**

## INTRODUCTION

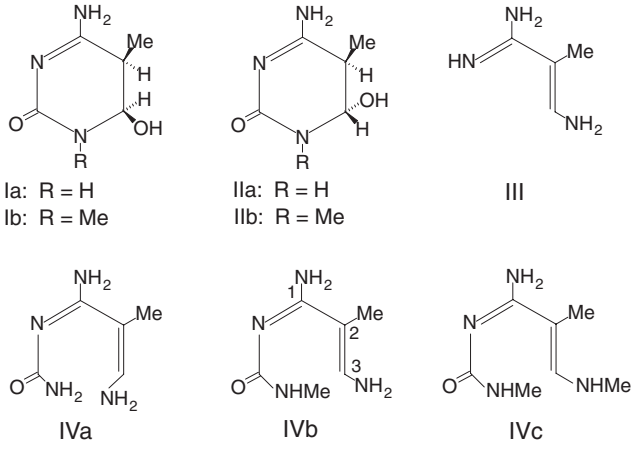
UV-irradiation of living cells leads to DNA photoproducts that can act as precursors to mutagenesis, carcinogenesis and cell death (1). A number of these products involve photoreaction of pyrimidine nucleobases within DNA to form compounds of various types. For example, the pyrimidine nucleobases can react with one another to form cyclobutane dimers and so-called (6-4) adducts. In solution, reactions of nucleobases within DNA can occur with surrounding solvent molecules (*e.g.* with water to form so called "photohydrates") or with reactive groups in neighboring complexed proteins (*e.g.* with the ε-amino moieties of lysine contained in histones). This latter type of reaction leads to DNA-protein cross-linking. Progress in understanding the photochemistry occurring within the cell has drawn heavily from the results of a large number of studies on model systems, including those on photochemistry of the component pyrimidine nucleobases and nucleosides. These studies have been extensively reviewed; for older reviews on cyclobutane dimers, (6-4) adducts and hydrates see, for example, Fisher and Johns (2,3), Wang (4), Cadet and Vigny (5), Davies (6) and Ruzsicska and Lemaire (7). More recent reviews, including within their purview the photochemistry of DNA at the cellular level, are given by Ravanat *et al.* (8) and Cadet *et al.* (9,10). Reviews of photo-induced nucleic acid–protein cross-linking and the chemical nature of the adducts responsible are given by Shetlar (11), Saito and Sugiyama (12) and Meisenheimer and Koch (13).

Among the earliest studies of the photochemistry of pyrimidine nucleobases and related compounds, in which the product was isolated and characterized, was the reaction of 1,3-dimethyluracil with water to form the "photohydrate" 5,6-dihydro-1,3-dimethyl-6-hydroxyuracil (14). Since that time, there have been a large number of studies of the photohydration reactions of uracil, cytosine, thymine and related compounds, including nucleosides (for comprehensive reviews covering much of this literature, see Fisher and Johns [2] and Cadet and Vigny [5]).

In addition to thymine and cytosine, 5-methylcytosine (m5C) is also found in eukaryotic DNA as a minor, but important component base; it is involved in gene regulation and other chromosomal transactions. (For an overview of the roles of m5C in eukaryotic DNA function and of the chromosomal machinery associated with its presence, see Robertson [15].) Interestingly, the isolation and characterization of hydrates similar in structure to **Ia** and **Ib** (Scheme 1) have not been reported in the chemical literature for m5C or for related compounds (*e.g.* nucleosides such as 5-methyl-2′-deoxycytidine [m5dC]) or derivatives of m5C (*e.g.* 1,5-dimethylcytosine [DMC]). Based on indirect evidence, it has been proposed that such species exist transiently as intermediates in observed chemical processes. For example, it has been suggested (16) that unstable m5C hydrate (*e.g.* **Ia**) may be an intermediate in a deamination reaction leading to thymine that occurs when m5C is irradiated at a concentration of 0.1 mM in

*Corresponding author email: shetlar@cgl.ucsf.edu (Martin D. Shetlar)
§Current address: Janet Chung, Molecular and Cell Biology, Beckman Research Institute, City of Hope, 1500 East Duarte Road, Duarte, CA, USA.
© 2011 The Authors
Photochemistry and Photobiology © 2011 The American Society of Photobiology 0031-8655/11

unbuffered aqueous solution. Vairapandi and Duker (17), in a comment in passing, also indicated that irradiation of m5C as a free base, using unspecified reaction conditions, resulted in formation of thymine hydrates; they likewise suggested that such hydrates resulted from deamination of an unstable intermediate m5C hydrate.



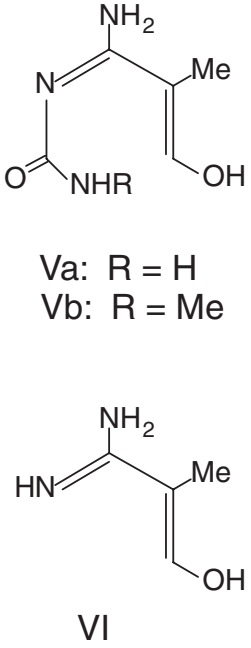

**Scheme 1.**

It has been suggested that m5C hydrates are formed in a polynucletide context. Vairapandi and Duker (17) indicated that irradiation of poly (dG-m5dC): poly (dG-m5dC), (a double stranded polydeoxyribonucleotide containing alternating deoxyguanine and m5dC residues) produced m5C hydrates. The evidence for this conclusion was that treatment of the irradiated polynucleotide with *Escherichia coli* endonuclease III released a mixture of the *cis* and *trans* isomers of the thymine hydrate. It was suggested that these two hydrates resulted from deamination of the *cis* and *trans* isomers of the corresponding hydrates of m5C (**Ia** and **IIa**) within the milieu of the polynucleotide and that these precursors were produced by photochemical reaction of m5C residues, contained in the polynucleotide, with water. However, Zuo *et al*. (18) examined the photochemical reactions in the same polydeoxyribonucleotide and found no evidence for release of thymine hydrates upon treatment with *E. coli* endonuclease III. The reasons for these differing results are unclear.

Detailed studies of the photochemical reactions of m5C in water have revealed that the predominant product in this system is very different in nature than the type of hydrates depicted as **Ia** and **IIa**. An early spectroscopic study (19) showed that irradiation of m5C in aqueous solution leads to rapid production of a compound with an absorption maximum at $\lambda = 285$ nm. In later work (20), the compound responsible for the observed spectroscopic behavior was isolated and characterized; it was identified as 3-amino-2-methylacrylamidine (**III**). Similar types of product were shown to form in the corresponding photoreactions of m5dC with water. Compounds with properties corresponding to those expected of **Ia** and/or **IIa** or their nucleoside analogs were not identified in these studies.

When m5C and related compounds are irradiated in the presence of amines, ring-opened adducts are formed. Indeed, such opened-ring compounds can be detected spectroscopically upon irradiation of m5C (and Cyt) in the presence of the lysine analog $N^{\alpha}$-acetyl-L-lysine ([21], fig. 1); ring closure of such adducts produces the type of photoexchange adducts first observed by Saito and co-workers (22) in the analogous thymine-lysine system.

Reaction of m5C with methylamine (MA) leads to *N*-(*N*′-methylcarbamoyl)-3-amino-2-methylacrylamidine (**IVb**) (23). Similarly, reaction with methanol leads to *N*-carbomethoxy-3-amino-2-methylacrylamidine (24). In the case of the latter compound, the hydroxy analog, namely *N*-carbomethoxy-3-hydroxy-2-methylacrylamidine has also been isolated and characterized (25). (For the numbering system in this type of opened-ring compound, see **IVb** in Scheme 1.) The existence of such compounds suggests that compounds with structures similar to **IVa** and **IVb**, but with a hydroxyl group replacing the amine functionality (as shown in **Va** and **Vb** in Scheme 2), might be stable enough for characterization if they could be prepared.




**Scheme 2.**

In the following, we describe the results of experiments that prove this supposition correct. Using authentic samples of such opened-ring water adducts, it was then determined that, in fact, detectable amounts of such hydrates were not produced when m5C or DMC are irradiated in water. However, in other experiments, we found that standing of solutions of **Va** and **Vb** at −20°C slowly produced unstable compounds with chemical and UV spectroscopic properties expected of closed-ring water adducts. Below, we describe experiments indicating these latter compounds are formed as minor products when m5C and DMC are irradiated in water. We also describe results of work indicating that an opened-ring isomer of the photohydrates of 2′-deoxycytidine (dCyd) coexists with parent hydrates; it may have a structure similar to those depicted in **Va** and **Vb**. Furthermore, we present results of some experiments suggest-

ing that hydrates of cytosine (Cyt) and 1-methylcytosine (1MC) can undergo ring-opening reactions.

## MATERIALS AND METHODS

*General aspects*. 5-Methylcytosine hydrochloride, 1,5-dimethylcytosine, cytosine, 1-methylcytosine and 2′-deoxycytidine were purchased from Sigma (St. Louis, MO), while 5-methyl-2′-deoxycytidine was from R. I. Chemical (Orange, CA). Methylamine was from Aldrich (Milwaukee, WI). Ammonium hydroxide and HPLC solvents were from Fisher (Fair Lawn, NJ); NMR solvents were from Aldrich (Milwaukee, WI). Preparative separations were done on a Shiseido Capcell UG120 10 × 250 mm reverse phase column (5 $\mu$m particle size [Yokohama, Japan]); in the following, this column is termed Column A. Analytical HPLC used a Capcell UG120 4.6 × 150 mm reverse phase column (5 $\mu$m particle size); hereafter we call this Column B. The HPLC system used was a Rainin binary gradient pumping system (Emeryville, CA) coupled to a Hewlett-Packard 1040A diode array HPLC detector (Palo Alto, CA). Prior to injection of HPLC samples, they were subjected to spin filtration, using Costar Spin-X micro-centrifuge filter tubes containing a 0.2 $\mu$m nylon filter (Corning Incorporated, Corning, NY). UV spectra were run on a Hewlett-Packard 8452A diode array spectrometer or using the "on the fly" spectral capture capability of the Hewlett-Packard diode array HPLC detector.

NMR spectra were run at 600 MHz on a Varian INOVA NMR spectrometer (Palo Alto, CA). Electrospray ionization (ESI) mass spectra were run on either a Waters Micromass ZQ4000 instrument (Beverly, MA) or a Sciex API300 triple quadrupole electrospray instrument (Toronto, Canada). High resolution mass spectra were run on a Thermo Electron Corp. LTQ Orbitrap XL (Waltham, MA).

*Irradiation methods*. Preparative irradiations were done in the cold with light centered at 254 nm, which was provided by unfiltered Spectronics BLE-1T155 15 watt lamps (Spectronics, Westbury, NY) housed in Spectroline XX-15A lamp holders. The solutions to be irradiated were placed in 187 mL cylindrical quartz vessels (47.6 cm length × 2.6 cm OD) from Southern New England Ultraviolet Company (Branford, CT). These vessels fit snugly between the two lamps in a Spectroline XX-15A lamp holder placed in a "lamps up" orientation and photoreactions were carried out on 160 mL of solution with the irradiation vessels placed in this position; one end of the lamp housing was elevated at about a 10° angle to the horizontal to avoid leakage of the solution out the top of the stoppered vessels. During irradiation, the reaction vessels were stoppered with appropriate glass taper seal stoppers. All irradiations were done under air in a cold room at about 5°C.

Some irradiations were carried out at 254 nm in a Southern New England Ultraviolet Company RP-100 reactor. In this case, the solution, contained in a 187 mL quartz tube, was placed at the center of the reactor; during irradiation the reaction vessel was inserted into a cylindrical Vycor shield.

*Preparation of m5C-ammonia, m5C-methylamine and DMC-methylamine solutions for irradiation*. Commercially available m5C is usually in the form of the hydrochloride salt. However, we conducted our irradiations at pH 7 or slightly above. We prepared 20 mM stock solutions of m5C·HCl for use in making up solutions for irradiation. This solution was kept refrigerated when not in use. (The concentration was verified by running the UV spectrum of a diluted sample in phosphate buffer [10 mM, pH 7.5] [$\varepsilon_{273}$ of m5C is 6230 (26)].) Using this stock solution, 500 mL portions of the solution to be irradiated were made up at a time. Twenty five milliliter of stock m5C·HCl was diluted with 50 mL of 1 M $NH_3$ or 50 mL of 1 M MA, each of which had been adjusted to have a pH of 7.0. The pH of the resulting solution (with an initial pH between 3.5 and 4) was then adjusted with 1 M NaOH until it was near pH 7. The volume of this solution was brought to 150 mL in a graduate cylinder and additional pH adjustment was done by adding small amounts of 1 M NaOH until the pH was above 7. The solution was then diluted with 350 mL of HPLC grade acetonitrile; the measured pH of this final solution, containing both water and acetonitrile, was above 7 in all cases. (Use of this mixed solvent medium significantly increases the yield of the desired amine adducts relative to the yield of 3-amino-2-acrylamidine (**III**), which is the dominant product when m5C is irradiated in pure water [20].)

Stock solutions of aqueous DMC (20 mM) were similarly used for preparation of solutions containing DMC (1 mM) and either ammonia (100 mM) or methylamine (100 mM). As above, the reactants were dissolved in a mixture of acetonitrile/water (70%/30% vol/vol) and the pH values of the solutions to be irradiated were near 7.

*Product preparation, isolation and purification in the m5C-ammonia system*. In a typical run, 160 mL of solution 1 mM in m5C and 100 mM in ammonia (pH 7.2) (contained in 70%/30% vol/vol $CH_3CN/H_2O$) was irradiated for 4 h as described above. The UV spectrum of a sample of the photoreaction mixture, after five-fold dilution in water, showed greatly enhanced absorbance centered at 305 nm. One milliliter samples of unirradiated solution and the irradiated solution were reduced to dryness by rotatory evaporation and redissolved in distilled water. The initial analysis of the m5C-$NH_3$ system was carried out on Column B using the following gradient: 0 min: 0% MeOH; 1 min: 0% MeOH; 4 min: 25% MeOH; 7 min: 25% MeOH; 7 min 15 s: 0% MeOH; 12 min: 0% MeOH. The aqueous eluent was 100 mM NaCl; peak detection was at 274 nm. Under these conditions, the parent m5C peak elutes at 3.5 min, while **III** in the irradiated sample elutes at 2.3 min and the desired putative adduct, termed $P_A$, elutes as the major peak at 4.5 min; no other product peaks, aside from very minor amounts of the ureidoacrylonitriles of m5C (27), were noted in the HPLC elution profile. (The identification of the peak corresponding to m5C-ammonia adduct was based on its absorption maximum at $\lambda$ = 305 nm, while identification of **III** was based on the chromatographic behavior and UV spectroscopic properties of an authentic sample prepared as described in Celewicz and Shetlar [20].) It was also determined that about 65% of parent m5C had been converted to product, based on areas of the parent peaks after 0 and 4 h irradiation.

After reducing the remaining irradiated solution to dryness *via* rotatory evaporation, the residue was redissolved in 1.5 mL of distilled water to give a cloudy solution. This solution was filtered and a trial run was made on Column A using a 20 $\mu$L injection with 98% (10 mM sodium phosphate buffer, pH 7.5)/2% MeOH flowing at 4 mL $min^{-1}$ as eluent. Under these conditions, **III** eluted at 5.9 min, m5C eluted at 8.9 min and the desired $P_A$ eluted as a broad peak at 9.6 min. The concentrated reaction mixture was then partitioned in 0.5 mL batches on Column A using the same elution conditions. With the greater injection volume, the m5C peak eluted as a sharp peak centered at 8.8 min, within the envelope of the broad peak corresponding to $P_A$, which eluted between 7.2 and 10.5 min. The two materials eluting in this latter time interval were collected separately and the fraction containing $P_A$ was purified multiple times to eliminate the sharp peak corresponding to m5C. (A variety of chromatographic separation conditions were explored in attempts to find conditions to cleanly separate m5C from $P_A$ when 0.5 mL volumes of concentrated reaction mixture were injected. However, while several worked well with small injection volumes, none gave clean separations with the larger volume.)

Prior to further studies, salt was removed from $P_A$. The collected purified $P_A$ was reduced to dryness *via* rotatory evaporation. The residue was then shaken with 10 mL of methanol and the resulting suspension was placed in a 15 mL plastic centrifuge tube; this was repeated with a 5 mL portion of MeOH. After spinning insoluble salt down to a pellet, the methanolic supernatant containing $P_A$ was removed and stored in solution in a freezer at −20°C. Samples of $P_A$ were removed as needed for spectroscopic and chemical studies.

*Product preparation, isolation and purification in the m5C-methylamine system and the DMC-methylamine system*. Since the protocols for the preparation, isolation and purification of the m5C-MA adduct (previously shown to be **IVb** [23]) and the DMC-MA adduct are similar to those given for the m5C-$NH_3$ system, they are provided in Appendix S1, rather than in the body of this paper.

*Studies on the conversion of $P_A$ to $P_H$, a putative opened-ring isomer of 5,6-dihydro-6-hydroxy-5-methylcytosine*. In spectrophotometric studies of the behavior of $P_A$ when dissolved in 0.1% trifluoroacetic acid (TFA), we found that the characteristic spectrum of this compound ($\lambda_{max}$ = 305 nm) disappeared over a period of 16 min and was replaced by a low absorbance band with a maximum at about 276 nm. During this time, the absorbance at 306 nm decreased from 2.11 to 0.05; the absorbance at 275 nm after 16 min was about 0.2. When $P_A$ was treated with 0.01% TFA, the rate of disappearance of parent compound was much smaller; after 48 min of incubation, the absorbance due to $P_A$ was 0.46. When $P_A$ was treated with methanolic

0.1% TFA (prepared by diluting concentrated TFA in methanol), the rate of disappearance was very slow; over a period of 8 min, about 1.6% of the absorbance at 306 nm had disappeared. Disappearance of $P_A$ upon incubation in either 1 M ammonium formate (pH = 7) or 100 mM phosphate buffer (pH 7.3) was likewise very slow. These results suggest that (i) the rate of this reaction is enhanced by increasing the concentration of $H_3O^+$ and (ii) water is required in order for the reaction to proceed at appreciable rates.

We examined the HPLC of $P_A$ that had been incubated in 0.1% aqueous TFA for various times at room temperature and then diluted three-fold with 100 mM sodium phosphate buffer (pH 7.3). For this analysis, we used Column B with the following gradient: 0 min, 0% MeOH; 0.75 min, 0% MeOH; 4 min, 15% MeOH; 5 min, 15% MeOH; 5 min 15 s, 0% MeOH; 9 min, 0% MeOH. The aqueous eluent was 10 mM $Na_2SO_4$ and the rate of flow was 2 mL $min^{-1}$. Under these conditions, the parent $P_A$ elutes at 3.1 min. The HPLC of samples taken after various times of incubation (1.5, 13.5, 24 min) indicated that the area of the peak associated with $P_A$ decreased with increasing incubation time and that the area of a new peak, corresponding to a compound of unknown structure eluting at 6.3 min with $\lambda_{max}$ = 313 nm (termed $P_H$ in the following discussion), correspondingly increased. After 24 min, the peak associated with $P_A$ had almost completely disappeared from the HPLC chromatogram and $P_H$ dominated the HPLC trace. (Note that this reaction cannot be followed cleanly by UV spectroscopy, as both $P_A$ and $P_H$ absorb strongly in the same region of the spectrum.)

*Preparation, isolation and purification of $P_H$ for spectroscopic characterization*. For purposes of preparation of sufficient amounts of the dominant product $P_H$ for mass spectral and NMR studies, we used the following protocol. About 5 $\mu$mol of $P_A$ was taken to dryness by rotatory evaporation at 40°C. To the residual film in the evaporation flask was added 500 $\mu$L of distilled water and then 1 mL of 0.1%TFA; the mixture was incubated at ambient room temperature. The reaction was followed on Column B by HPLC using 90% (100 mM NaCl)/10% MeOH flowing at 2 mL $min^{-1}$. Samples were diluted with equivalent amounts of 100 mM phosphate buffer, pH 7.3, prior to injection. Under these conditions, $P_A$ eluted at 1.3 min and $P_H$ at 3.7 min. While most of the parent $P_A$ had been consumed after 19 min, the reaction mixture was allowed to stand for 75 min, after which very little parent reactant remained. The reaction mixture was then taken to dryness by rotatory evaporation; the residue was then dissolved in 0.5 mL of distilled water and filtered. The resulting material was chromatographed on column B, using 76% (100 mM NaCl)/24% MeOH flowing at 2 mL $min^{-1}$ as eluent. Prior to injection, 100 $\mu$L portions of sample were diluted with 100 $\mu$L of 100 mM sodium phosphate buffer (pH 7.3). Peak $P_H$, eluting between 1.7 and 2.2 min, was collected. Using rotatory evaporation at 40°C, the collected material was evaporated to dryness and then extracted with 5 mL of MeOH. After centrifugation and decanting of the methanolic extract, analytical HPLC of the extract on Column B, after dilution with 100 mM sodium phosphate (pH 7.3) and using 90% (100 mM NaCl)/10% MeOH flowing at 2 mL $min^{-1}$ for elution, indicated that the product was quite pure. The amount of $P_H$ isolated was estimated to be about 1.6 $\mu$mol *via* UV spectroscopy of an aqueous dilute solution; this corresponds to an isolated yield of 32%. In this calculation, it was assumed that the $\varepsilon_{max}$ for **IVb** (39050) is a good approximation for the corresponding $\varepsilon_{max}$ of the product (see section A3 in Appendix S1). No attempts were made to optimize the reaction and workup conditions with a view of achieving higher yields of isolated $P_H$.

For the purposes of NMR study, an appropriate volume of methanolic abstract was rotatory evaporated to dryness at ambient temperature using a water aspirator as a pump; the resulting film was used for preparation of NMR samples. NMR samples were prepared within a couple of hours of having their spectra run; samples in $D_2O$ were kept on ice until shortly before they were placed in the spectrometer probe.

*Conversion of **IVb** to $P_J$, a putative opened-ring isomer of 5,6-dihydro-6-hydroxy-1,5-dimethylcytosine and its isolation and purification.* Hydrolysis studies analogous to those done for $P_A$ were done for **IVb** as well. The results of this study, as well as the protocols for isolation, purification and sample preparation for $P_J$, the main product of this reaction, are similar to those discussed above for the acid hydrolysis product of $P_A$; these matters are discussed in Appendix S1.

## RESULTS AND DISCUSSION

### Introduction

Study of the implications of an observation in our laboratory, concerning the spectroscopic behavior of the m5C-methylamine adduct **IVb** when it is treated gently with dilute acid, has provided a route to compounds that can be regarded as opened-ring isomers of 5,6-dihydro-6-hydroxy-5-methylcytosine (having *cis* and *trans* forms, depicted in Scheme 1 as **Ia** and **IIa**) and 5,6-dihydro-1,5-dimethyl-6-hydroxycytosine (**Ib** and **IIb**). Our original observation was spectroscopic in nature and indicated that a new product was gradually formed upon incubation of **IVb** with 0.1% TFA. When **IVb** when dissolved in 0.1% TFA, the characteristic spectrum of this compound ($\lambda_{max}$ = 305 nm) disappeared over a period of a few min and was replaced by a low absorbance band with a weak maximum at about 276 nm. Neutralization led to recovery of absorbance at long wavelengths, however, the $\lambda_{max}$ was shifted to the red, as compared with that of **IVb**, suggesting that a new compound was present. (More detail about these observations will be given later.) In the following, we describe the studies that led to preparation and characterization of ring-opened isomeric forms of the m5C and DMC hydrates.

### Characterization of $P_A$ isolated from the m5C-ammonia system

The UV spectrum of $P_A$ in distilled water showed that $\lambda_{max}$ = 305 nm and $\lambda_{min}$ = 254 nm; the ratio $A_{305}/A_{254}$ was measured to be about 60. The ESI mass spectrum of $P_A$ displayed peaks at 143.1 [M + $H^+$] and 165.1 [M + $Na^+$], indicating that this product contained both m5C and $NH_3$. The high resolution ESI mass spectrum was consistent with protonated $P_A$ having the molecular formula $C_5H_{11}N_4O$ (calculated: 143.0927; observed: 143.0923). These mass spectral observations are consistent with $P_A$ being identified as **IVa** (Scheme 1).

The NMR spectral data, contained in Table 1, are also consistent with identification of $P_A$ as **IVa** (*N*-carbamoyl-3-amino-2-methylacrylamidine). Note that the structure **IVa** is displayed in the Z conformation. The actual conformation has not been definitively established. Results of quantum mechanical calculations (Table 2) indicate that this conformation and the alternative E conformation (displayed for esthetic reasons in structures **IVb** and **IVc** in Scheme 1) have comparable energies and, thus, do not supply definitive guidance on this question.

When **IVa** is heated in 0.1 M HCl for 15 min at 100°C, it is completely converted into m5C (major product) and Thy (minor product). The HPLC analysis demonstrating this (carried out after dilution of the heated sample with two volumes of 100 mM phosphate buffer, pH 7.3) was done on Column B using 100 mM $Na_2SO_4$ as eluent flowing at 2 mL $min^{-1}$. Under these conditions, m5C eluted at 2.5 min and Thy eluted at 4.4 min. This is consistent with analogous observations, reported in (23), that found that **IVb**, the corresponding methylamine adduct, is converted into DMC and 1-methylthymine after similar treatment.

### Product characterization in the m5C-methylamine system

The opened-ring conjugate, formed in the reaction between m5C and methylamine, has been previously isolated and

**Table 1.** Proton NMR chemical shift data for the photoproducts produced in the reactions of m5C with ammonia (**IVa**) and methylamine (**IVb**) and of DMC with methylamine (**IVc**) and the opened-ring hydrates **Va** and **Vb**.

| | IVa* | IVb† | IVc | Va | Vb |
|---|---|---|---|---|---|
| C2C**H**$_3$ | 1.71(3) | 1.71(3) | 1.71 (3) | 1.74 (3) | 1.73 (3) |
| C3**H** | 8.17(t,1, $J$ = 11.4) | 8.13 (t,1, $J$ = 11.4) | 8.37(d,1, $J$ = 14) | 8.35 (1) | 8.32 (1) |
| C3N**H**$_2$ | 8.05(b, 0.5) | 8.04 (b,0.4) | | | |
| C3N**H**CH$_3$ | | | 8.00 (b,m,1) | | |
| CON**H**$_2$ | 8.81(b,0.8) | | | | |
| CON**H**CH$_3$ | | 8.82 (b,0.8) | 8.69 (b,0.6) | | |
| C3NHC**H**$_3$ | | | 3.01(d,3, $J$ = 4.4) | | 2.76 (3) |
| CONHC**H**$_3$ | | 2.68 (d,3, $J$ = 3.6) | 2.68 (d,3, $J$ = 4.4) | | |

*Roman numerals correspond to the structures given in Schemes 1 and 2, while m denotes a multiplet of complex structure and b indicates a broadened peak.

†Spectra of **IVa**, **IVb** and **IVc** were run in $d_6$-DMSO, while the spectra of **Va** and **Vb** were run in $D_2O$. The chemical shift values ($\delta$) are relative to tetramethylsilane (TMS) in $d_6$-DMSO or TSP (3-(trimethylsilyl)propionic-2,2,3,3-$d_4$ acid, sodium salt) in $D_2O$ and are given in ppm; $J$ values are in Hz. Integrated values are rounded to nearest integer, except when such values deviate significantly from that expected (for some of the broad NH proton resonances); then the actual measured integrated intensity is given. In all cases run in $d_6$-DMSO, addition of a small amount of $D_2O$ to the NMR sample resulted in disappearance or, in same cases, considerable loss in size of peaks assigned as hydrogens attached to nitrogen; under these conditions, spin-spin splittings due to coupling to such hydrogens also disappeared.

**Table 2.** Calculated aqueous energies of various compounds referenced in the text.

| Compound | Energy† (hydrated), kJ mol$^{-1}$ |
|---|---|
| **Ia (trans)** | −1340854.58 |
| **Ib (cis)** | −1444042.93 |
| **IIa (trans)** | −1340850.76 |
| **IIb (trans)** | −1444052.71 |
| **IVa (Z)** | −1288676.17 |
| **IVa (E)** | −1288672.79 |
| **IVb (Z)** | −1391880.28 |
| **IVb (E)** | −1391875.75 |
| **Va (Z)** | −1340840.70 |
| **Va (E)** | −1340811.85 |
| **Vb (Z)** | −1443999.47 |
| **Vb (E)** | −1443995.64 |
| **VIIa** | −1289879.59 |
| **VIIb** | −1289874.07 |
| **VIIc** | −1289873.97 |
| **VIId** | −1289814.88 |
| **VIIIa** | −1342028.69 |
| **VIIIb** | −1341975.42 |
| **IX** | −1340826.81 |

†These values given are those evaluated at 0 K using Spartan 2010 for the Macintosh. The methods used for these calculations were as follows. A set of conformers was generated *via* molecular mechanics using the molecular mechanics force field implementation of Spartan. This was followed by optimization of the equilibrium geometry of each of the conformers generated, using semi-empirical calculations employing a PM3 model. After eliminating duplicate conformations using the alignment tool built into Spartan, an optimized energy calculation in water was made using a 3-21G basis set. After elimination of high-energy conformations, the remaining conformations were subjected to an energy optimization in water using Hartree-Fock calculations with the 6-31G* basis set. In a final stage, the set of values obtained *via* this calculation were subjected to a final optimization in water utilizing B3LYP density functional calculations, again using the 6-31G* basis set. The energies of the lowest energy conformation of each molecular species are those tabulated above. Spartan 2010 calculates the energy of solvated species by applying the SM8 model of A. V. Marenich, R. M. Olsen, C. P. Kelly, C. J. Cramer and D. G. Truhlar (2007) Self-consistent reaction field model for aqueous and nonaqueous solutions based on accurate polarized partial charges. *J. Chem. Theory Comput.*, **3**, 2011–2033, which adjusts wave functions to account for solvent effects.

characterized as having the structure **IVb** (23). The protocols for its preparation, isolation and purification are given in Appendix S1. (It should be noted that the overall protocol given in Appendix S1 is a modified and improved version of that given in [23].) Consistent with, but differing very slightly from previously reported data (23), the UV spectrum of $P_B$ in distilled water showed $\lambda_{max}$ = 305 and 228 nm and $\lambda_{min}$ = 256 nm. The ESI mass spectrum of $P_B$ displayed peaks at 157.1 [M + H$^+$] and 179.1 [M + Na$^+$], indicating that this product contained both m5C and methylamine. The high resolution ESI mass spectra confirmed that protonated $P_B$ has a molecular formula of $C_6H_{13}N_4O$ (calculated: 157.1084; observed: 157.1080). The NMR spectra run in DMSO and contained in Table 1, along with the UV and mass spectral data given above, are all consistent with the previous identification of $P_B$ as *N*-(*N*′-methylcarbamoyl-3-amino-2-methylacrylamidine (**IVb** in Scheme 1). Note that the peak at $\delta$ = 8.13, corresponding to C3**H** moiety is a triplet. This verifies that a $NH_2$ group is also attached to C3. The presence of a doublet, in the signal corresponding to the CONHC**H**$_3$ protons, provides confirmation that these protons are incorporated into the amide group. The chemical shift values of the nonexchangeable protons given in Table 1 for **IVb** are near the values of those found previously (23) where the NMR of this compound was run in $D_2O$.

As was the case for **IVa**, based on the data at our disposal, we are unable to definitively assign this compound as either the E form (shown in Scheme 1) or Z form. However, as discussed below, results of quantum chemical calculations (Table 2) suggest that the Z form is somewhat more stable than the E form.

### Product characterization in the DMC-methylamine system

As for the m5C-MA system, the detailed protocols for preparation, isolation and purification of $P_C$, the m5C-methylamine adduct, are given in Appendix S1. The UV spectrum of $P_C$ in distilled water showed that $\lambda_{max}$ = 319 nm; there was a very broad $\lambda_{min}$ centered at about 262 nm. The ESI mass spectrum of $P_C$ displayed peaks at 171.2 [M + H$^+$] and

193.1 [M + $Na^+$], indicating this product contained the elements of both DMC and methylamine. The high resolution ESI mass spectra of this compound indicated that protonated $P_C$ has a molecular formula of $C_7H_{15}N_4O$ (calculated: 171.1240; observed: 171.1232); this is consistent with identification of $P_C$ as **IVc**. The NMR spectral data, contained in Table 1, along with the UV and mass spectral data, are also consistent with identification of $P_C$ as *N*-(*N′*-methylcarbamoyl)-3-methylamino-2-methylacrylamidine (**IVc**). Note that the peak at $\delta = 8.37$, corresponding to C3**H** moiety is a doublet. This verifies that a hydrogen is contained in the amine group that is also attached to C3. The presence of a doublet, in the signal corresponding to the CONHC**H**$_3$ protons, likewise provides confirmation that these protons are incorporated into the amide group.

### Yields of IVa and IVb

The % yields of **IVa** and **IVb**, based on the amount of parent consumed, were estimated to be about 58% (**IVa**) and 59% (**IVb**). A detailed description of the protocols used for these measurements will be found in Appendix S1.

### The product $P_H$, formed via acid treatment of IVa, is *N*-carbamoyl-3-hydroxy-2-methylacrylamidine (Va)

The UV spectrum of purified $P_H$ in water displays, in addition to the maximum at 313 nm, an absorption minimum at $\lambda = 257$ nm; the ratio $A_{313}/A_{257}$ is about 9.9. ESI mass spectrometry shows [M + $H^+$] at 144.1, corresponding to the molecular mass expected for a compound with the structure shown as **Va** in Scheme 2. Comparing this structure to that given in Scheme 1 for **IVa**, the parent compound in the acid induced reaction, it can be seen that the $NH_2$ at C3 in **IVa** has been replaced by an OH in **Va**. A smaller peak occurs at 127.1, corresponding to the [M + $H^+$] expected for thymine, while another yet smaller peak is seen at 126.1, which corresponds to the [M + $H^+$] value predicted for m5C. The high resolution ESI mass spectra indicated that protonated $P_H$ has the expected molecular formula of $C_5H_{10}N_3O_2$ (calculated: 144.0764; observed: 144.0768). The proton NMR spectrum (Table 1), run in $D_2O$, is also consistent with $P_H$ being assigned the structure given by **Va**; it displays the expected singlet methyl resonance at 1.74 ppm and a singlet C3H resonance at 8.35 ppm. Both of these resonances are quite close to those seen in the NMR spectrum of **IVa**, the amino analog of **Va**. In addition to the resonances corresponding to **Va**, there are weak resonances at $\delta = 1.95$ and 7.35; these resonances have the same $\delta$ values as those observed by Sulkowska (28) for the C5C**H3** and C6**H** protons of m5C in $D_2O$. A chemical argument, similar to that outlined below in the section containing discussion of the DMC hydrate, can be used to rule out an alternative structure to **Va**, in which an amino group is attached to C3 and a hydroxyl group is attached to the terminal nitrogen.

### The product $P_J$, isolated by acid hydrolysis of IVb, is *N*-(*N′*-methylcarbamoyl)-3-hydroxy-2-methylacrylamidine (Vb)

The compound $P_J$, which is putatively an opened-ring form of 5,6-dihydro-1,5-dimethyl-6-hydroxycytosine, was prepared by using a protocol similar to that used for the preparation of **Va** (Materials and Methods). Details of this preparation are given in Appendix S1.

The UV absorption spectrum displayed a maximum at 313 nm and a minimum at 257 nm; the ratio of $A_{313}/A_{257}$ is about 7.8. Molecular mass peaks, determined by ESI mass spectrometry, were observed at 158.1 [M + $H^+$] and 180.1 [M + $Na^+$]). The high resolution ESI mass spectra indicated that protonated $P_J$ has a molecular formula of $C_6H_{12}N_3O_2$ (calculated: 158.0924; observed: 158.0921); this value is that expected for a species with structure **Vb** (Scheme 2). Examination of the proton NMR spectrum (see Table 1) also provides data consistent with the structure of $P_J$ being **Vb**. The NMR spectrum contains a proton resonance at $\delta = 8.32$. This value is quite close to the values for the C3H resonances in **IVa**, **IVb** and **IVc**, in which an $NH_2$ or $NHCH_3$ group is likewise attached to C3. This indicates that this resonance should also be assigned to C3H and supports the supposition that $P_J$ contains a vinyl group as exhibited by **Vb**. There are two methyl group resonances, one at $\delta = 1.73$ and the other at $\delta = 2.76$. The chemical shift of the first is near the values for the C2C**H$_3$** group observed for **IVb**, its amino analog ($\delta = 1.71$), indicating that this methyl group is attached to the vinyl linkage; the lack of spin-spin splitting in this peak indicates that there is not a nonexchangeable proton attached to the carbon adjacent to the methyl group. The chemical shift of the second methyl resonance is consistent with this group being attached to nitrogen; however, this alone does not rule out the possibility that the attachment site of the $NHCH_3$ group could be at C3. A chemical argument discounts this. Consider the stability of the alternative structure to **Vb**, with the $NHCH_3$ moiety at C3 and an OH group attached to the carbonyl group (*i.e.* an *N*-substituted carbamic acid). As discussed in (20), this type of compound would be expected to be labile in aqueous solution, decomposing to form 3-methylamino-2-methylacrylamidine and carbon dioxide. Indeed, the analogous compounds, in which $NHCH_3$ at C3 is replaced with either an amino group or with an NH-2′-deoxyribosyl group, cannot be isolated. They undergo rapid reaction to form 3-amino-2-methylacrylamidine and 3-(2-*erythro*-D-pentopyranos-1-yl)amino-2-methylacrylamidine respectively (20). An additional argument, based on results obtained by treatment of a crude reaction mixture from photoreaction of DMC and ammonia with TFA (see below), lends additional evidence supporting the conclusion that the $NHCH_3$ moiety is positioned as shown in **Vb**.

### NMR spectra of Va and Vb in $d_6$-DMSO

When the NMR properties of either **Va** or **Vb** are studied in $d_6$-DMSO (using TMS as standard), rather than $D_2O$, the complexity of the resulting spectra is greater than in $D_2O$. In the case of **Va**, two peaks appear at $\delta = 1.65$ and 1.74 with an integrated area ratio of 3.14. The latter $\delta$ value is close to that seen for the methyl group when **Va** is run in $D_2O$. In addition there is a pair of peaks with $\delta = 8.54$ and 8.68 with an area ratio of 3.06. One possibility is that these two pairs of resonances correspond to C2 methyl and C3H groups in two different structural isomers of **Va**. One reasonable interpretation is that they correspond to the C2 methyl and C3H

resonances in the E and Z structural isomers of **Va**. Similarly, when **Vb** is run in $d_6$-DMSO, there appears to be two sets of peaks relating to two forms of the same compound. One set displays $\delta$ = 1.65 (s, 3), 2.58 (broad, 2.3) and 8.51 (s, 1) and the second has $\delta$ = 1.74 (s, 0.83), 2.63 (broad doublet, 0.97) and 8.67 (s, 0.24). (Integrated areas for the various peaks for **Vb** are referenced to a value of 3.0 for the $\delta$ = 1.65 peak.) Again, one possible interpretation of this data is that these two sets of spectra correspond to E and Z structural isomers of **Vb**.

The NMR spectra of **Va** and **Vb** in $d_6$-DMSO both display peaks that disappear when a drop of $D_2O$ is added to the sample. These likely correspond to the expected exchangeable protons in these compounds (*i.e.* protons attached to nitrogen or oxygen).

### Incubation of *N*-(*N'*-methylcarbamoyl)-3-methylamino-2-methylacrylamidine (IVc) with 0.05% TFA

We also examined the effects of incubation of **IVc** with TFA. A portion of aqueous purified **IVc** was diluted with an equal amount of 0.1% TFA and allowed to stand at room temperature for 6 min. After addition of an equal volume of sodium phosphate buffer (100 mM, pH 7.3), a sample was injected on Column B using 76% (100 mM NaCl)/24% MeOH flowing at a rate of 2 mL $min^{-1}$ as eluent. The product peak eluted at 3.4 min. Injection of a sample of **IVb** that had been simultaneously treated in the same manner yielded a peak that eluted after 3.2 min. The UV spectra of the dominant product contained in each of the two samples were identical. Since the product from the incubation of **IVb** was identified as **Vb**, it can be concluded that **IVc** hydrolyzes to the same product.

### Preliminary results from study of the TFA-mediated hydrolysis of a photoreacted 1,5-dimethylcytosine-ammonia system

When 1 mM DMC is irradiated in the presence of 100 mM $NH_3$ at pH 7 in 70% acetonitrile/30% water, UV spectroscopic and HPLC studies, conducted as described above, indicated that an analogous opened-ring adduct was formed in good yield. After 10 h irradiation at 254 nm, about 64% conversion of parent to products was achieved, as evidenced by calculations involving peak areas for the parent compound in unirradiated and irradiated solution. (Column B was used for the analysis, utilizing 94% [100 mM NaCl]/6% MeOH flowing at 2 mL $min^{-1}$. Under these conditions, DMC elutes at 2.3 min and the desired adduct elutes at 2.8 min.) However, DMC elutes as a much broader peak than m5C on this column. When preparative runs were attempted on Column A using the same solvent system, it was found that the peak corresponding to parent DMC completely enveloped the adduct peak and separation of these two compounds was not possible. Indeed, we were not able to find suitable preparative separation conditions for this system utilizing the various column and eluent systems available to us. Isolation and characterization of DMC-$NH_3$ adduct, putatively analogous to **IVa**, but with $NHCH_3$ attached to C3, remains to be accomplished.

Although we were unable to isolate a pure DMC-$NH_3$ adduct, we were able to obtain evidence strongly suggesting that hydrolysis of the putative ammonia adduct results in formation of **Va**. A portion of concentrated crude reaction mixture, resulting from photoreaction of DMC with ammonia, was made 0.05% in TFA and incubated at ambient room temperature. After varying times, samples were withdrawn and neutralized by mixing with an equal portion of 100 mM sodium phosphate buffer (pH 7.3). The HPLC of 10 $\mu$L of the resulting mixture was then run on Column B using 90% (100 mM NaCl)/10% MeOH flowing at 2 mL $min^{-1}$ as eluent. Under these conditions, parent DMC eluted at 2.0 min, the putative DMC-$NH_3$ adduct at 2.2 min and **Va** eluted at 3.7 min. After 60 min, about 75% of the parent adduct, as measured by area, had disappeared and a new peak at 3.7 min appeared, containing material with $\lambda_{max}$ at 313 nm. The identity of the HPLC retention behavior and the UV spectroscopic properties of the material eluting at 3.7 min with the corresponding properties of **Va** confirms that the TFA-mediated hydrolysis of the putative DMC-$NH_3$ adduct leads to **Va**.

In a discussion above, chemical reasoning was used to reach the conclusion that TFA-mediated hydrolysis of **IVb** results in loss of the alkylamino group attached to C3. Although the results of these experiments described immediately above are not definitive, as they were not done with purified adduct, they do provide suggestive supporting evidence for this conclusion; hydrolysis at the alternative position, next to the carbonyl group, would leave an $NCH_3$ group at C3 and the resulting product would not be **Va**.

### Effect of treatment of 3-aminoacrylamidine (III) with TFA

The results of an earlier unpublished study in our laboratory on the effects of dilute acid on the spectrophotometric properties of **III** (20) indicated (in retrospect) that this compound behaves similarly to the m5C-amine adducts. In view of this similarity in behavior, we conducted further studies to see if indeed the chemistry occurring is related. The results of these studies suggest that **III**, as well as its nucleoside analog (20), undergo reaction to form the compound **VI** or its Z isomer (or a mixture of these two isomers). However, since the problematic HPLC properties of the putative **VI** produced preclude isolating sufficient material for NMR, this identification must be regarded as tentative. The details of these studies on acid hydrolysis of **III** and its nucleoside analog are contained in Appendix S1.

### The hydrates Va and Vb are not formed when m5C and DMC are irradiated in water

To determine if **Va** is produced in significant amounts when the parent compound m5C is irradiated in distilled water, we irradiated a 1 mM solution of this compound (pH = 7.6) at room temperature for 2 h at 254 nm (68% conversion to products) and immediately analyzed it *via* HPLC. Using Column B with 76% (100 mM NaCl)/24% MeOH flowing at 2 mL $min^{-1}$, authentic **Va** elutes at 1.7 min. We found no evidence for the presence of this compound in the reaction mixture. A similar study with 1 mM DMC in distilled water at pH 6.9 (41% conversion to products), using the same HPLC conditions (**Vb** elutes at 3.0 min) yielded similar results, suggesting that this compound is not a significant photoproduct under these reaction conditions.

**Ib and IIb may be intermediates in the thermal decomposition of Vb to DMC, while Ia and Ib may similarly be intermediates in the decomposition of Va**

During storage of **Va** and **Vb** in methanolic solution at −20°C, decomposition of these compounds occurs slowly to form the parent compounds m5C and DMC respectively. In each case, however, there were also small amounts of other compounds present with UV spectra reminiscent in profile to that of the photohydrates of 2′-deoxycytidine (see below). It can be hypothesized that these compounds are **Ia** and/or **IIa** in the case of the decomposition of **Va** and **Ib** and/or **IIb** in the corresponding decomposition of **Vb**. One test of this hypothesis would be to determine if, in fact, these other compounds have **Va** or **Vb** as a direct precursor and whether these compounds are direct precursors of the m5C or DMC also seen in the solutions stored at −20°C. It might be expected that ring closure of an opened-ring hydrate (*e.g.* **Vb** → **Ib** and/or **IIb**) would precede elimination of water to form the final pyrimidine product. We carried out such experiments by following the thermal decomposition of **Va** and **Vb** as a function of time of incubation at 40°C in aqueous media. In the case of **Vb**, we rotatory evaporated 100 μL of the methanolic solution of **Vb** to dryness and took up the resulting residue in 300 μL of 100 mM NaCl. The resulting solution was incubated in a water bath maintained at 40°C and 30 μL samples were removed after successive time intervals and subjected to HPLC on Column B using 100 mM NaCl and MeOH flowing at 2 mL $min^{-1}$ as gradient components. The gradient (Gradient M) was as follows: 0 min: 0% MeOH; 2 min: 0% MeOH; 5 min: 25% MeOH; 6.75 min: 25% MeOH; 7 min: 0% MeOH; 10 min: 0% MeOH. The peak areas for the peaks of interest in each HPLC chromatogram were measured at 313 nm (**Vb**), 243 nm ($\lambda_{max}$ for putative **Ib** and **IIb**) and 279 nm (DMC). These peak areas were plotted as a function of incubation time. We observed that (1) **Vb** (elution time [$t_e$]: 7.3 min) disappeared *via* 1st order kinetics ($t_{1/2}$ = 10 min) over a 80 min time interval, (2) the sum of the two peak areas corresponding to the putative **Ib** and **IIb** ($t_e$: 1.5 min [$\lambda_{max}$ = 243 nm] and 2.0 min [$\lambda_{max}$ = 245 nm] or vice versa) at first increased and then steadily decreased and (3) the area corresponding to DMC ($t_e$: 5.1 min) correspondingly increased with time at 40°C. These results indicate that the reaction in the incubated system can be represented by the consecutive reaction sequence **Vb** → (putative **Ib**, **IIb**) → DMC. While more detailed structural characterization of the two unstable intermediate products must be done before they can be definitively identified, these results are consistent with the interpretation that they can be identified as **Ib** and **IIb** formed *via* ring closure reactions of **Vb**. (It can be noted that over the final 30 min period of incubation, after most of **Vb** had disappeared, *ca* 24% of the summed peak areas for putative **Ib** and **IIb** were converted to DMC. Thus, preparation and maintenance of pure samples of these compounds for NMR studies could require special care.)

Analogous kinetic behavior was observed when **Va** was similarly incubated at 40°C. Using Gradient M (see above), the dominant putative closed-ring hydrate ($\lambda_{max}$ = 233 nm) eluted at 0.85 min, m5C at 2.3 min and **Va** at 6.0 min; a minor putative closed-ring hydrate eluted at about 1.3 min and had an area about 11-fold less than that of the dominant putative hydrate peak eluting at 0.85 min. The $t_{1/2}$ for disappearance of **Va** was about 5 min at this incubation temperature. The putative closed-ring hydrates also displayed instability at 40°C; about 22% of the summed area for the hydrates disappeared over the 20 min incubation interval after the **Va** initially present had disappeared.

**Putative closed-ring hydrates are present in freshly irradiated aqueous m5C and DMC solutions**

As noted in the Introduction, closed-ring photohydrates of m5C (**Ia** and/or **Ib**) have not been directly observed as products, although indirect evidence suggesting formation of these compounds has been presented. We have used the HPLC properties of the putative closed-ring hydrates, described in the previous section, to answer the following question. Are products with the properties of putative hydrates formed when either m5C or, alternatively, DMC is irradiated in aqueous solution?

We irradiated 10 mL samples of m5C and DMC, each 1 mM in concentration, for 32 min in a cold room (about 4°C) at 254 nm; the samples were placed next to an unfiltered Spectronics BLE-1T155 15 watt lamp. The samples were then analyzed using Gradient M, described above. In the case of the DMC system (5% conversion to product), three product peaks were observed, in addition to small late eluting peaks corresponding to the ureidoacrylonitriles of DMC (27). The main product eluted at about 5.2 min (just prior to the DMC peak eluting at 5.4 min) and had an absorption spectrum similar in profile ($\lambda_{max}$ = 299 nm) to that expected for an aminoacrylamidine (20). Two minor products eluted at 1.5 and 2.0 min, with the area of peak at 2.0 min, measured at 243 nm, being about four times as large as the peak eluting at 1.5 min. The HPLC retention times and UV absorption spectra corresponding to these two peaks were identical to those of the putative **Ib** and **IIb** described above. The summed area of the two peaks, measured at 243 nm, was about 19% of the area of the putative acrylamidine peak, measured at 299 nm. Based on these observations, it can be stated that compounds with the properties expected of **Ib** and **IIb** are indeed formed when DMC is irradiated aqueous solution. An additional experiment showed that, after heating for 10 min at 100°C, the peaks at 1.5 and 2.0 min disappeared from the HPLC chromatogram. This indicates that the corresponding compounds are unstable to elevated temperature (as, in general, are photohydrates of pyrimidine bases [2]) and provides additional support for the idea that the two putative hydrates can, in actuality, be identified as **Ib** and **IIb**.

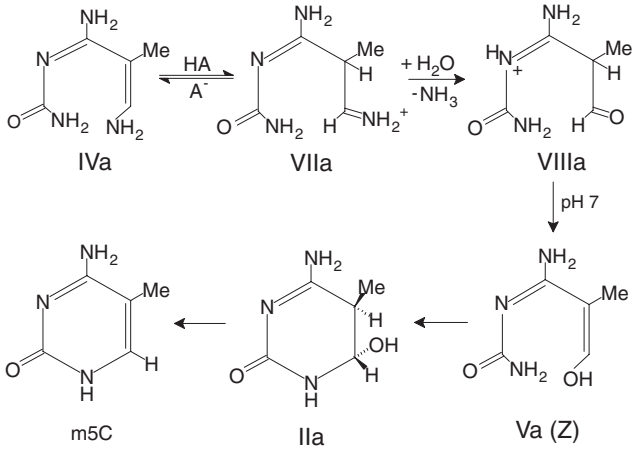
A similar analysis was conducted for the irradiated m5C system. In this case, three product peaks eluting prior to m5C were observed, two of them corresponding to the previously observed isomeric forms of the aminoacrylamidine (**III**) of m5C (20). (As for the DMC system, small peaks corresponding to ureidoacrylonitriles [27] were also seen.) The two aminoacrylamidine peaks eluted at 1.7 (minor) and 1.8 min (major), while parent m5C eluted at 2.2 min. The third product peak eluted at 0.88 min and had a UV spectrum identical to that of the dominant putative hydrate ($t_e$ = 0.85 min) described above for the system in which **Va** was incubated at 40°C. The peak area, measured at 233 nm, was about 0.9% of the sums of the areas of the aminoacrylamidine peaks, measured

at 285 nm. These results are suggestive that either putative **Ia** or, alternatively, **Ib** is indeed produced upon irradiation of m5C in aqueous solution, albeit with a very low yield. This latter result is perhaps not surprising; the quantum yield for the photohydration reaction of Thy (also containing a methyl group at C5) at pH 6 is $3 \times 10^{-6}$ ([2], p.180).

The results of an additional HPLC experiment, in which the irradiated m5C solution was heated at 100°C for 10 min prior to injection, showed that the peak at 0.88 min disappeared from the chromatogram, consistent with this peak being a photohydrate with a structure given by either **Ia** or **IIa**.

### On the mechanism of acid-promoted transformation of amine adducts of m5C to opened-ring hydrates of m5C, then closed-ring hydrates and finally m5C

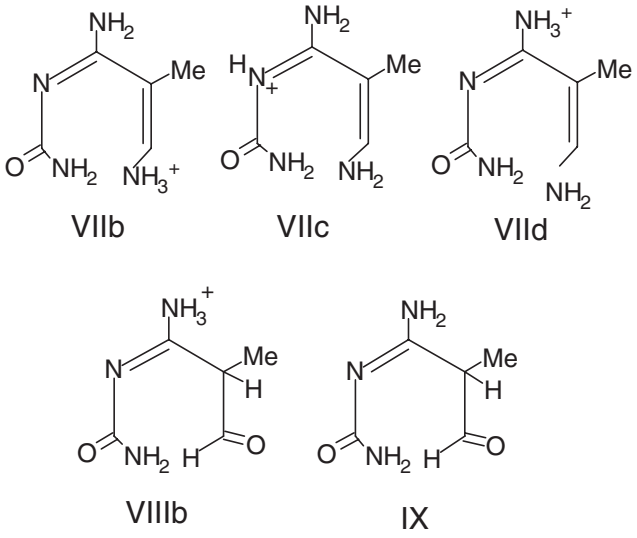
The experimental results discussed above provide several hints about the pathway by which amine adducts of m5C are converted to the corresponding water adducts in the presence of dilute acid. In the following, we propose a partial mechanism (shown in Scheme 3) consistent with what we have observed for this reaction. For the sake of simplifying the discussion, we will focus on the specific case of conversion of **IVa**, the m5C-ammonia adduct to **Va**, the ring-opened water adduct of m5C. However, the discussion can likely be generalized to similar reactions of the various other amine adducts of m5C and DMC.




**Scheme 3.**

In constructing this mechanism, we used the calculated solvated energies of various species as an approximate guide in selecting appropriate structures of the various intermediates shown in Scheme 3. These energies were evaluated by quantum mechanical calculations using Spartan 10 for the Macintosh. Details are given in Table 2. (While values of the Gibbs free energy at 298 K for the various solvated species listed in Table 2 would provide a more rigorous guide for constructing this mechanism, calculated values for this thermodynamic quantity are not readily evaluated in aqueous solution.) In the latter part of the mechanism, we show steps containing the observed products, in which **Va** transforms to a mixture of **Ia** and **IIa** and finally m5C, as is observed experimentally in solutions of **Va** upon standing, either in the cold or upon being heated (see above).

In the first reaction in Scheme 3, **IVa** becomes partially protonated in the presence of dilute acid (*e.g.* 0.1%TFA, pH = 1.9) to form the iminium cation **VIIa**; as will be discussed below, **IVa** and **VIIa** are postulated to coexist in a pre-equilibrium. Iminium cations often have a p$K_a$ in the range between 3 and 4 (29), while evidence has been recently presented that an iminium cation with a p$K_a$ in the neighborhood of 2.5 is involved in the ring closure of thymine-ammonia adducts to form thymine hydrates (30).




**Scheme 4.**

There are other possible structures resulting from *N*-protonation of **IVa**, resulting, for example, in the cations shown in **VIIb**, **VIIc** and **VIId** in Scheme 4; however, the calculated solvated energies of these species (Table 2) suggest that **VIIb** and **VIIc** are somewhat less stable energetically than **VIIa** in aqueous solution, while **VIId** is much less stable; these results predict that **VIIa** will be the dominant protonated form. Other structures analogous to **VIIa**, but containing an unprotonated imine (HC=NH) moiety and a protonation site on another nitrogen, are much less stable than **VIIa**; the solvated energies for two such cases are $E = -1289838.86$ kJ ($NH_2CONH^+CNH_2CHMeCHNH$) and $E = -1289802.46$ kJ ($NH_2CONCNH_3^+CHMeCHNH$).

The next step in Scheme 3 is an irreversible hydrolysis reaction of **VIIa** to form **VIIIa**, resulting in loss of ammonia and gain of water. This step likely involves addition of water to the iminium ion functionality to form a protonated amino-carbinol species, which then undergoes loss of ammonia to form **VIIIa**. (It can be noted in Table 2 that **VIIIa** is calculated to be considerably more stable than the alternative **VIIIb**, in which the exocyclic amino group is protonated.) It is this conversion of iminium moiety to carbonyl that gradually depletes the concentration of **VIIa**, as well as that of **IVa** existing in pre-equilibrium with **VIIa**. Experimentally, this is manifested by the disappearance (over a period of 16 min at room temperature) of the absorption spectrum characteristic

of opened-ring adduct **IVa** that is seen after it is placed in 0.1% TFA. Neutralization of the solution containing **VIIIa** is followed by conversion of this compound to the Z form of **Va**, the product isolated immediately after this treatment; examination of Table 2 predicts that this isomer is considerably more stable than the alternate E form. Upon standing, **Va** is converted to more stable closed-ring hydrates. These two hydrates have comparable calculated solvated energies; thus the final product is predicted to consist of a mixture of closed-ring hydrates **Ia** and **IIa**. The $\Delta E$ for the reaction of **Va** to form **Ia**, the most stable hydrate, can be estimated using data in Table 2, from which a value of $\Delta E = -13.88$ kJ is obtained. Finally, in a final irreversible thermal reaction, the two hydrates gradually decompose to produce m5C.

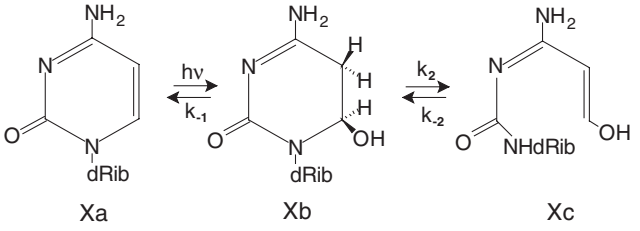
Facile conversion of **Va**(E) to **Va**(Z) could be mediated by the aldehyde species **IX**. Table 2 indicates that the computed solvated energy of this species lies intermediate between the corresponding energies of these two compounds.

As mentioned above, the half-lives of **Va** and **Vb** at 40°C in distilled water were about 5 and 10 min respectively. It was also noted that at the same temperature, over a 20 min period when little **Va** was present, about 22% of combined putative **Ia** and **IIa** disappeared; in a similar situation, about 24% of combined putative **Ib** and **IIb** disappeared over a 30 min period. These kinetic results suggest that free energy barriers of significant size exist at each step along the reaction pathways **Va** → (putative **Ia**, **IIa**) → m5C and **Vb** → (putative **Ib**, **IIb**) → DMC. Study of the rates of individual steps in these reactions as a function of temperature would be useful, so as to quantify the enthalpies and entropies of activation for each of these steps.

It is possible that other plausible mechanisms can be proposed to explain the experimental results described in this paper. However, the mechanism put forth in Scheme 3 appears to be a good starting point, as it is consistent with experimental data. In addition, the various proposed protonated intermediates (*i.e.* **VIIa**, **VIIIa**), appearing in the mechanism are predicted by quantum chemical calculation to be the energetically most stable in aqueous solution, as compared to various alternative isomeric species that could be proposed in their place.

### Evidence for an opened-ring form of the 2′-deoxycytidine hydrate

In the following section, we present preliminary results that suggest that the two 2′-deoxycytidine hydrates each undergo a thermal reaction to produce the same opened-ring hydrate isomer. Such a reaction is illustrated in Scheme 5, where **Xb**, one of the hydrate diastereomers (in this case with OH attached in the R configuration to the pyrimidine ring) reacts to form a putative ring-opened product (perhaps with a structure given by **Xc**).




**Scheme 5.**

An extensive examination of the photohydration products of cytidine (Cyd) was published in 1978 (31); in this study, exchangeable hydrogens were replaced with deuterium prior to irradiation. Irradiation was carried out in $D_2O$ and the reaction followed by proton NMR. In this work, it was shown that deuterium oxide added across the 5,6-double bond to form a hydrate with OD attached to the 6-position of the nucleosidic pyrimidine base; evidence was presented indicating that two diastereomeric products were present in the NMR sample. Experimental evidence in the literature is consistent with the supposition that water also adds across the 5,6-double bond in dCyd to give the corresponding diastereomeric hydrates (for reviews see [2,5]). These two products can be partially resolved *via* HPLC on an analytical reverse phase column (see below). However, isolated purified dCyd hydrates have not been studied in detail using NMR and other tools for structural characterization. Their instability towards reversion to parent compound and deamination to give 2′-deoxyuridine hydrates appears to have precluded structural studies; these matters are discussed by Douki *et al.* (32).

In the course of follow-up experiments, building on studies of the photoreactivity of dCyd with aliphatic amines in aqueous solution (33), it was noted that irradiation of dCyd in the absence of amines led to formation of a compound that was strongly retained on a reverse phase column. This compound was found to have a spectrum very similar in profile to that of adducts of dCyd with ethylamine, but with a different $\lambda_{max}$. In the following, we present the results of some exploratory studies, including UV spectroscopic and HPLC experiments, which suggest that this compound corresponds to an opened-ring form of dCyd hydrate.

When the progress of photoreaction of an aqueous 0.2 mM solution of dCyd (**Xa**) is followed using UV spectroscopy, three phenomena are observed. The first is the loss of absorbance at 270 nm, the $\lambda_{max}$ of dCyd and the second is the increase in an absorbance band with $\lambda_{max} = 241$ nm (corresponding to production of photohydration products of dCyd). However, in addition to these two absorption changes, a weak absorption band appears that tails to the red; this band extends to about 330 nm.

When HPLC of a 250 $\mu$L sample of a 0.2 mM solution that had been irradiated for 90 min at 254 nm was run immediately after photolysis, the chromatogram displayed in Fig. 1 was obtained. This chromatographic separation was made on Column B, using the following gradient with 100 mM NaCl and MeOH as eluents: 0 min: 0% MeOH; 2 min: 0% MeOH; 6 min: 12% MeOH; 7 min: 12% MeOH; 7 min 15 s: 0% MeOH; 10 min: 0% MeOH. The flow rate was 2 mL $min^{-1}$. In Panel (a), the HPLC trace using a detection wavelength of 241 nm (the absorption maximum of the putative dCyd hydrates), dCyd elutes at 4.92 min and two diastereomeric hydrates elute at 1.01 and 1.24 min. (In the following, we will term these two products hydrates, even though definitive characterization by NMR remains to be done. This is based on the fact that each of these compounds, upon heating, is converted to dCyd [along with a small amount of 2′-deoxyuridine].) Two compounds with absorption spectra similar to those of the hydrates elute at 1.75 and 3.40 min; in some runs, the latter peak appears to be considerably broader than shown in Fig. 1. These latter two compounds may be products analogous to the $O^6$,5′-cyclo-5,6-dihydro-2′-deoxyuridine

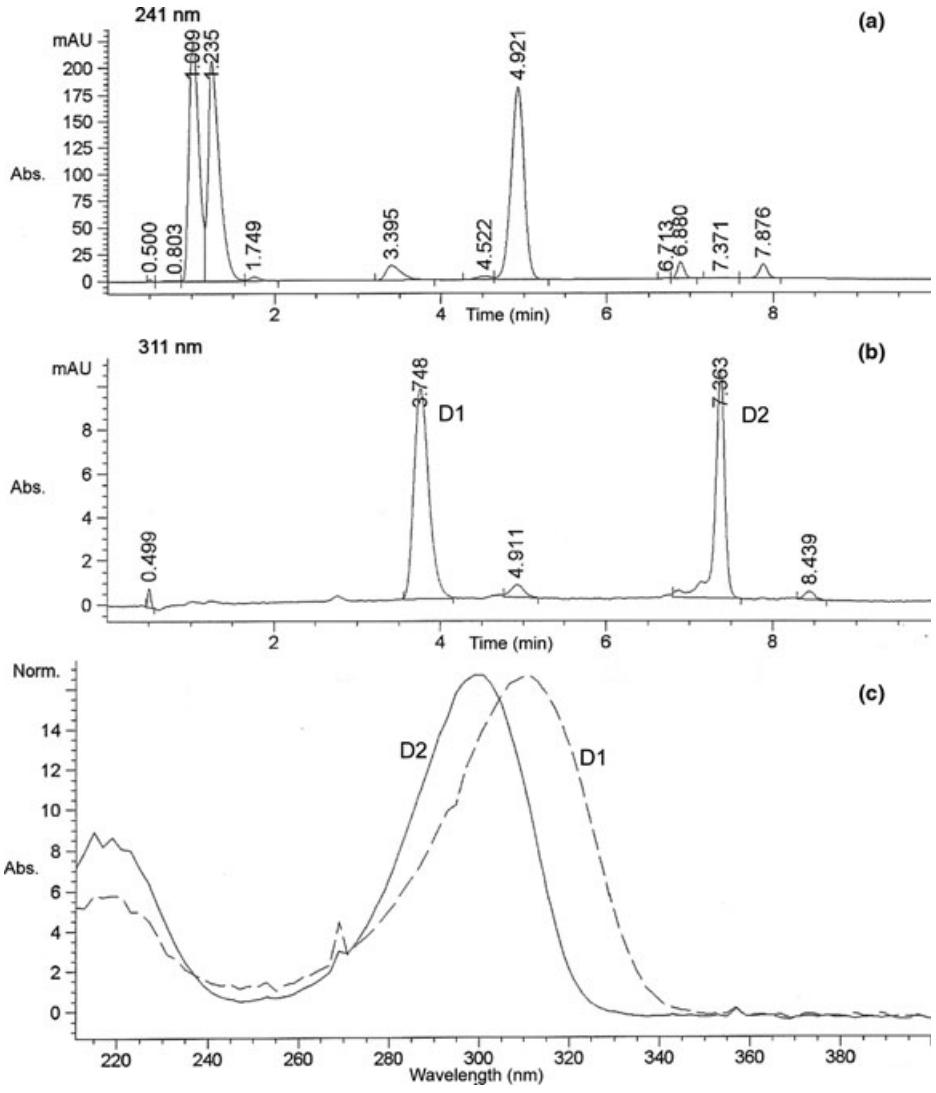



**Figure 1.** The HPLC chromatogram resulting from injection of 250 $\mu$L of 0.2 mM aqueous dCyd that had undergone photoreaction. The injection was made immediately after irradiation of the parent solution at 254 nm for 90 min; a Shiseido Capcell UG120 4.6 × 150 mm column was used for separation. (a) HPLC trace when detection was done at $\lambda$ = 241 nm, while (b) shows the corresponding trace with detection at 311 nm. (The identities of the various peaks, as well as the HPLC gradient and flow rate conditions used, are discussed in the text.) (c) Spectra of the peaks shown as eluting at 3.7 and 7.4 min in (b). The spectra in (c) are normalized, thus appearing on the same scale.

products reported by Cadet *et al*. (34); however, because of the small amounts in which they are produced and their instability towards the conditions used in our work-up procedures, we were unable to isolate them for further characterization. Of particular interest here, however, are the two main peaks in the chromatogram shown in Panel (b), where the detection wavelength is 311 nm. The absorption spectra corresponding to these two peaks (shown in Panel [c]), obtained by utilizing the "on the fly" spectral capabilities of the diode array detector, have maxima at wavelengths where they could contribute to the redshifted weak absorption band that was observed in the UV spectrum of the reaction mixture described above. Indeed, the two spectra display a remarkable resemblance in profile to the corresponding spectra observed for compounds **Va** and **Vb**.

To determine if the dCyd hydrates could be precursors to these two products, hereafter termed D1 ($t_{elution}$ = 3.7 min) and D2 ($t_{elution}$ = 7.4 min), we irradiated 1 L of 0.2 mM dCyd for 90 min at 254 nm in the cold in 165 mL batches as described in Materials and Methods. This solution was taken to dryness in stages using rotatory evaporation at 40°C; the residue was dissolved in 2 mL of distilled water. The resulting solution was partitioned on Column A using the following gradient with water serving as the aqueous eluent: 0% MeOH; 4 min: 0% MeOH; 6 min: 12% MeOH; 18 min: 12% MeOH; 21 min 15 s: 0% MeOH; 25 min: 0% MeOH. The flow rate was 4 mL min$^{-1}$. The fractions (dCH1 and dCH2), corresponding to the two hydrates and eluting in peaks centered at 4.2 and 5.1 min, were collected and the HPLC of each freshly purified fraction was run on Column B, using the following gradient at 2 mL min$^{-1}$: 0% MeOH; 4 min: 0% MeOH; 5 min: 18% MeOH; 9 min: 18% MeOH; 9 min 18 s: 0% MeOH; 14 min: 0% MeOH. In each case, in addition to the parent hydrate peak, a small amount of the second hydrate

and a small peak corresponding to D2, eluting at 7.2 min, came off as well. The UV spectrum of the material eluting in this peak was identical to that of the peak eluting at 7.4 min in Panel (b) in figure ($\lambda_{max}$ = 299 nm). The peak D1, shown in Panel (b) of Fig. 1 as eluting at 3.75 min, did not appear in these chromatograms. We did some exploratory experiments that showed that D1 is the same compound described by Urata and Akagi (35); these workers found that irradiation of dCyd in 0.2 mM solution at pH 4 resulted in production of 2-deoxyribolactone. In addition, UV spectral monitoring of the progress of the reaction indicated that an unidentified product with a $\lambda_{max}$ at 311 nm increased in amount with irradiation time. We repeated this work; however, in addition to following the reaction by UV spectroscopy, we analyzed the reaction mixture using HPLC. Analysis of the reaction mixture resulting from 32 min irradiation at 254 nm (3 mL of 0.2 mM dCyd in a quartz spectrophotometer cuvette, pH 4, placed next to a BLE-IT155 lamp) on Column B was carried out (elution gradient: 0 min: 0% MeOH; 2 min: 0% MeOH; 6 min: 12% MeOH; 7 min: 12% MeOH; 7 min 15 s: 0% MeOH; 10 min: 0% MeOH; flow rate: 2 mL $min^{-1}$). The main peak observed, when detection was done at either 311 nm or 299 nm, was that corresponding to D1 ($t_{elution}$ = 3.5 min), as shown by identity of UV spectra and *via* coinjection with a similarly irradiated solution made up in distilled water. Only a trace of material corresponding to D2 was present in the solution irradiated at pH 4. We did not pursue the identification of D1 further.

To obtain additional support for the hypothesis that the material corresponding to peak D2 was chemically related to the known hydrates, we isolated this substance and followed its thermal reaction *via* spectroscopy and chromatography. A solution of dCyd in distilled water (2 mM, 50 mL) was irradiated at 254 nm in a cold room (about 4°C) for 16 h in a Southern New England Ultraviolet Company photochemical reactor. Forty-five milliliter of the irradiated solution was concentrated to 0.5 mL, using rotatory evaporation at 32°C. Thirty microliters of this material was injected on Column B and a separation was made using distilled water flowing at 2.25 mL $min^{-1}$. The material corresponding to D2, eluting at 10.2 min, was collected in a 0.6 mL quartz spectrophotometer cell and this cell was then immediately situated in a cell holder whose temperature was controlled by a circulating water bath. After temperature equilibration, the HP diode array spectrometer was used in the kinetic mode to collect UV spectra at appropriate intervals. A representative set of raw data obtained at 25.3°C, with a spectrum run every 3 min, is shown in Figure S1. This figure shows that, over a period of 60 min, the absorbance at 299 nm decreases with time, while the absorbance at 242 nm ($\lambda_{max}$ for dCyd hydrate) concurrently increases. Thus, D2 is converted to compounds having UV spectral traces very similar to those of the hydrates H1 and H2.

The UV spectrum of the material corresponding to D2 is very similar in profile to the corresponding spectra seen for the amine adducts (**IVa**, **IVb** and **IVc**) and hydrates (**Va** and **Vb**) of m5C and DMC. This indicates that extensive double bond conjugation must be present in the structure of this product and strongly suggests that the structure has an opened-ring structure. Although we have not established a definitive structure for D2, it seems likely that the structure may be or have similarities to the structure displayed as **Xc** (Scheme 5). A discussion of some of the challenges that will need to be met to obtain a definitive structure of the material corresponding to D2 is given below.

Could dCyd hydrates similarly exist within the context of DNA as ring-closed forms (**Xb**) in pseudoequilibrium with an opened-ring form, as shown in the second reaction in Scheme 5 (drawn with the unproven assumption that the material corresponding to D2 has the structure **Xc**)? If so, then achieving a complete understanding of the photochemistry and photobiology of cytosine residues within DNA (or RNA) could require dealing with an unanticipated degree of complexity in terms of the photoproducts present. For example, repair of hydrates that exist within DNA in both closed-ring and opened-ring structures could occur at different rates or, perhaps, could even require differing modes of enzymatic detection and removal within damaged DNA; similarly, interactions of other DNA-binding proteins (*e.g.* those involved in transcription, replication or establishment of DNA structure within the chromosome) with damaged DNA could be different for the two forms of hydrate. One question that can be asked is whether an opened-ring hydrate could continue to exist for a reasonable length of time after a closed-ring hydrate was removed by repair or decomposition. Answering the question for photodamaged DNA *in vitro* or *in vivo* will require sophisticated experimentation. However, the type of data displayed in Figure S1 does allow an answer to be given to this question at the level of the nucleoside. In particular, we have determined the enthalpy of activation and entropy of activation for reversion of nucleosidic D2 to closed-ring hydrates in the doubly distilled water eluting from Column B (measured pH = 8.8) and in sodium phosphate buffer at near physiological pH. First, we repeated the experiment run at 25.3°C (see above) at three additional temperatures in the distilled water eluting from the HPLC column. Using measured absorbance values at $\lambda_{max}$ = 299 nm as data points, we found that the reversion reaction at each temperature obeyed first order kinetics (*i.e.* graphs of the natural log of the measured absorbance at 299 nm *versus* time were linear). The first order rate constants determined at the four temperatures were as follows: 12.8°C, 0.0151 $min^{-1}$; 18.7°C, 0.0254 $min^{-1}$; 25.3°C, 0.0472 $min^{-1}$; 29.9°C, 0.0637 $min^{-1}$. Another set of experiments was conducted in 4 mM sodium phosphate buffer (pH 7.54). In these experiments, 960 $\mu$L of freshly collected material was diluted with 40 $\mu$L of 100 mM phosphate and put in a 1.2 mL cuvette. After temperature equilibration, four sets of overlaid spectra similar to those shown in Figure S1 were obtained. From these runs, the following set of first order rate constants was calculated: 15.1°C, 0.0287 $min^{-1}$; 18.2°C, 0.0415 $min^{-1}$; 21.2°C, 0.0552 $min^{-1}$; 26.3°C, 0.0858 $min^{-1}$. In each of the runs described above, the correlation coefficient was 0.989 or greater.

It can be noted the rate constant for reversion of Cyd hydrate to Cyd in unbuffered solution at 25°C and pH 8 has been determined by deBoer *et al.* (36) to be 0.0036 $min^{-1}$; at pH 9, the corresponding value was 0.0045 $min^{-1}$. If it is assumed that similar numbers obtain for the dCyd hydrate, then conversion of D2 to hydrate appears to be on the order of 10 times faster than the conversion of hydrate to parent nucleoside in unbuffered water. Likewise, the rate constant for reversion of dCyd hydrate to parent at 26°C has been evaluated to be 0.0066 $min^{-1}$ (37) in 4 mM phosphate buffer

at pH 7. This compares to the value of 0.0858 $min^{-1}$ given above for conversion of D2 to hydrate in 4 mM phosphate buffer at pH 7.5 at 26.3°C. Thus, it can be estimated that the dCyd hydrate decomposes to dCyd in the neighborhood of 13 times slower under these conditions than the reversion of D2 to hydrate. Rate constants for conversion of dCyd hydrate to D2 at corresponding temperatures and pH values remain to be determined. Such rate constants are required for rigorous kinetic analysis of the dCyd hydrate-D2-dCyd system. However, detection of D2 in the dCyd hydrate system, along with the implications of the kinetic parameters available at the nucleoside level, suggest that it should not be ruled out that dCyd hydrates, contained within DNA, could have a sufficiently long lifetime to establish a pseudoequilibrium with D2 before their eventual reversion to dCyd.

The Eyring transition state theory of chemical kinetics indicates that the temperature dependence of the rate constant for a reaction in solution can be mathematically described as shown in the equation given immediately below (see, for example, Chang [38], pp. 476–480). In this equation $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ are, respectively, the enthalpy of activation and entropy of activation, $\kappa$ is the transmission coefficient, $R$ is the gas constant, $T$ is the temperature in K, $N$ is Avogadro's number and $h$ is Planck's constant.

$$k = \kappa(\mathrm{RT}/\mathrm{Nh})\mathrm{e}^{\Delta S^{\ddagger}/R}\mathrm{e}^{-\Delta H^{\ddagger}/\mathrm{RT}}$$

If $\kappa$ is set equal to unity, this equation can be rewritten in the form of $\ln (k/T) = \ln (\mathrm{R}/\mathrm{Nh}) + \Delta S^{\ddagger}/\mathrm{R} - \Delta H^{\ddagger}/\mathrm{RT}$. According to this equation, the value of $\Delta H^{\ddagger}$ can be evaluated from the slope of a graph of $\ln (k/T)$ *versus* $1/T$, while the y-intercept of such a graph yields a quantity from which $\Delta S^{\ddagger}$ can be calculated. We analyzed the kinetic data for reversion of D2 to hydrate using this equation; the resulting plot yielded values of $\Delta H^{\ddagger} = 62.6$ kJ and $\Delta S^{\ddagger} = -94.6$ J $K^{-1}$ for the reaction converting D2 to hydrate in distilled water. The correlation coefficient describing the linear fit of the rate constant data for the Eyring plot was 0.996. A similar analysis of the data obtained in phosphate buffer indicates that $\Delta H^{\ddagger} = 78.6$ kJ and $\Delta S^{\ddagger} = -36.2$ J $K^{-1}$ with correlation coefficient 0.994. Note that these two sets of activation parameters cannot necessarily be directly compared, as they were obtained at different pH values. Furthermore, the data for the phosphate buffer may reflect inclusion of catalytic pathways for conversion of D2 to hydrate that do not occur in the experiments in distilled water. As deBoer *et al*. (36) have shown, the rate of decomposition of dCyd hydrates is increased in the presence of various buffers, as compared to the corresponding rates when buffers were absent. (Phosphate was not among the buffers studied, although it was shown that the rate of loss of the photohydrate of cytidylic acid is greatly enhanced, as compared to the corresponding rate for Cyd hydrates [36].) The effects of the various buffers on the kinetics of these reactions were analyzed in terms of acid-base catalysis. Conceivably, the presence of buffer components could similarly introduce new pathways for conversion of D2 to hydrate.

From the values obtained for $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ the rate constant and half-life ($t_{1/2}$) for reversion of D2 to hydrate can be evaluated at other temperatures of interest (*e.g*. 0°C and 37°C). In distilled water, the desired values are $k = 4.43 \times 10^{-3}$ $min^{-1}$, $t_{1/2} = 156$ min (0°C) and $k = 0.12$ $min^{-1}$, $t_{1/2} = 5.8$ min (37°C). In 4 mM phosphate, the corresponding values are $4.39 \times 10^{-3}$ min, $t_{1/2} = 158$ min (0°C) and $k = 0.282$ $min^{-1}$, $t_{1/2} = 2.5$ min (37°C). If the value of $t_{1/2}$ in phosphate buffer at 37°C can be extrapolated to ring-opened dCyd hydrates contained in DNA in the environment of a cellular nucleus, then this value suggests that the dCyd opened-ring hydrate, once formed, could have a sufficiently long lifetime for it to be regarded as a separate lesion in a biological context.

We now return to the challenges of obtaining a definitive structure for the compound corresponding to D2. Ideally such a structure should be based on NMR spectroscopic and mass spectrometric characterization of D2. The rate constant data discussed above suggests that elegant experimentation will be required to obtain such information. For example, the path to obtaining a pure sample of D2 suitable for proton or C-13 NMR has difficulties. Based on the kinetic data discussed above, if purified D2 is to be maintained in a pure state, then the mode of sample preparation should avoid elevated temperatures; otherwise conversion to closed-ring hydrate will occur. However, even such a gentle technique as lyophilization of aqueous solvent from the purified frozen sample requires at least several hours of time. Assuming a 3-h lyophilization period in which an aqueous sample is maintained in the frozen state at 0°C, a little bit more than 50% of the original compound will have reverted to closed-ring hydrate at the end of the lyophilization process. Procedures for putting the freeze-dried sample into a suitable NMR solvent (*e.g*. $D_2O$) are likely to lead to even more loss, while extended NMR examination, particularly those involving two-dimensional techniques, will likewise lead to an even greater degree of decomposition. Use of a technique such as stopped flow HPLC-NMR could be useful in obtaining a proton NMR spectrum of putative **Xc**. Beyond preparation of a sample of D2 free of parent hydrate for NMR purposes, an additional problem exists in obtaining a definitive molecular mass for D2, particularly if it has the structure **Xc**. The closed-ring hydrate and D2 would then have the same molecular mass; thus, parent D2 cannot be distinguished from its closed-ring hydrate decomposition product(s) *via* simple mass spectrometric measurements. Utilization of HPLC/mass spectrometry or other advanced methods may allow the problem of molecular mass determination to be solved. The observation that D2 can be eluted from a Capcell reversed phase column with distilled water could be helpful in such studies. Summarizing, definitive structural characterization of D2 *via* physical techniques could provide a worthy challenge for future investigators.

### Preliminary results suggest that the cytosine photohydrate and 1-methylcytosine photohydrate also undergo ring-opening to yield analogs of D2

To examine the possibility that the cytosine hydrate can also exist in an alternative ring-opened form, we examined the HPLC of freshly irradiated Cyt. It is interesting to note that, as was the case with the hydrates of dCyd, definitive structural studies of the photohydrates of cytosine and 1-methylcytosine (*e.g*. involving NMR and mass spectrometric studies of purified compound) have evidently not been published, probably because their instability makes such studies difficult. Therefore, based on their thermal reversibility (as described in the relevant literature [reviewed in 2,5]), we have assumed that the dominant

product arising from UV irradiation of each of these compounds at low concentration is indeed a hydrate analogous to that shown in **Xb**. Irradiation of 100 mL of 0.2 mM Cyt at 254 nm was carried out for 60 min as described above for dCyd. After concentration to 1.4 mL *via* rotatory evaporation at 40°C, 500 $\mu$L of the irradiated solution was chromatographed on Column B using the following gradient with 100 mM NaCl and MeOH as eluents: 0 min: 0% MeOH; 1 min: 0% MeOH; 5 min: 15% MeOH; 5 min 30 s: 15% MeOH; 5 min 45 s: 0% MeOH; 9 min: 0% MeOH. The flow rate was 2 mL $min^{-1}$. A peak eluting between 0.6 and 1.1 min corresponded to Cyt hydrate. A variety of additional peaks also eluted, including at least one ureidoacrylonitrile product (27). Of particular interest was a relatively small peak (compared with the size of the hydrate and Cyt peaks) eluting at 3.7 min; this peak contained material with a UV absorption spectrum that displayed a $\lambda_{max}$ = 297 nm; the profile of the absorption band was very similar to that observed for D2 in the case of dCyd. Reinjection of 250 $\mu$L of the collected Cyt hydrate, utilizing the same column and gradient conditions, again led to an HPLC chromatogram that contained a small peak with a retention time of 3.5 min and containing material with $\lambda_{max}$ = 297 nm.

The same experimental protocol was applied to the 1-methylcytosine system. In this case, the hydrate eluted between 0.8 and 1.6 min, while material eluting in a small peak at 6.6 min had the UV spectroscopic character of a compound similar in nature to D2 ($\lambda_{max}$ = 297 nm). Once again, a number of other products also eluted from the column, including previously characterized ureidoacrylonitriles (27). The HPLC chromatogram resulting from reinjection of 250 $\mu$L of collected hydrate, using the same column and elution conditions as previously, showed a small peak eluting at 6.6 min with the same spectral profile and value of $\lambda_{max}$.

The small amount of putative D2 analog present in either the isolated Cyt hydrate or the corresponding 1-methylcytosine hydrate samples, as revealed by HPLC, precluded isolating sufficient amounts of these putative opened-ring hydrates to conduct studies to observe the reversal of these compounds to closed-ring hydrates. Thus, the results of the studies described for these two compounds must be regarded only as suggestive.

## CONCLUDING REMARKS

There are reports in the literature indicating that each of the three major pyrimidine nucleobases (Ura, Thy and Cyt) forms a photohydrate when irradiated in water. However, a hydrate with a similar structure has not been reported for m5C. We have prepared an opened-ring structural isomer of this compound; this synthesis relied on treatment of an m5C-ammonia adduct with dilute acid. It has been shown that this opened-ring structure (*N*-carbamoyl-3-hydroxy-2-methylacrylamidine (**Va**)) decomposes thermally to form products that are putatively *cis* and *trans* 5,6-dihydro-6-hydroxy-5-methylcytosine (**Ia** and **IIa**), one of which can be detected in irradiated aqueous m5C. Analogous results are described for **Vb**, the opened-ring water adduct corresponding to **Ib** and **IIb**. Other work showed that the photohydrates of dCyd in aqueous solution coexist with a compound that exhibits the UV spectroscopic properties expected of the corresponding ring-opened water adduct. While it is well known that amine adducts of m5C, Cyt, dCyd and related compounds can readily be isolated in opened-ring forms, this report gives the first indication that water adducts of such compounds can similarly exist as such isomeric forms in amounts detectable by HPLC.

*Acknowledgements*—We gratefully acknowledge NSF Grant CHE-0131203 and services provided by the Bio-organic, Biomedical Mass Spectrometry Resource (A. L. Burlingame, Director), supported by the NIH Division of Research Resources Grant RR 01614 and especially thank David Maltby who took special pains to obtain needed mass spectra in a timely manner.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Figure S1.** A set of UV spectral traces showing the decay of the intensity of the spectrum corresponding to D2, the putative opened-ring form of the hydrate, and the corresponding increase in the intensity of the spectrum corresponding to the closed-ring dCyd hydrate(s).

**Appendix S1.** Experimental details concerning the isolation of **IVb**, **IVc** and **Vb**, as well as a description of the methods for determining the yields of **IVa** and **IVb** and the preparation of **VI.**

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

## REFERENCES

1. Friedberg, E. C., G. C. Walker, W. Seide and R. D. Wood (2005) *DNA Repair and Mutagenesis*, 2nd edn. ASM Press, Washington, DC.
2. Fisher, G. J. and H. E. Johns (1976) Pyrimidine photohydrates. In *Photochemistry and Photobiology of Nucleic Acids*, Vol. 1 (Edited by S. Y. Wang), pp. 169–224. Academic Press, New York.
3. Fisher, G. J. and H. E. Johns (1976) Pyrimidine photodimers. In *Photochemistry and Photobiology of Nucleic Acids*, Vol. 1 (Edited by S. Y. Wang), pp. 225–294. Academic Press, New York.
4. Wang, S. Y. (1976) Pyrimidine bimolecular photoproducts. In *Photochemistry and Photobiology of Nucleic Acids*, Vol. 1 (Edited by S. Y. Wang) pp. 295–356. Academic Press, New York.
5. Cadet, J. and P. Vigny (1990) The photochemistry of nucleic acids. In *Bioorganic Photochemistry*, Vol. 1 (Edited by H. Morrison), pp. 1–272. John Wiley and Sons, New York.
6. Davies, R. J. H. (1995) Ultraviolet radiation damage in DNA. *Biochem. Soc. Trans.* **23**, 407–418.
7. Ruzsicska, B. P. and D. G. E. Lemaire (1995) DNA photochemistry. In *CRC Handbook of Organic Photochemistry and Photobiology* (Edited by W. M. Horspool and P.-S. Song), pp. 1289–1315. CRC Press, Boca Raton, FL.
8. Ravanat, J.-L., T. Douki and J. Cadet (2001) Direct and indirect effects of UV radiation on DNA and its components. *J. Photochem. Photobiol. B* **63**, 88–102.
9. Cadet, J., S. Courdavault, J.-L. Ravanat and T. Douki (2005) UVB and UVA radiation-mediated damage to isolated and cellular DNA. *Pure Appl. Chem.* **77**, 947–961.
10. Cadet, J., E. Sage and T. Douki (2005) Ultraviolet radiation-mediated damage to cellular DNA. *Mutation Res.* **571**, 3–17.
11. Shetlar, M. D. (1980) Cross-linking of proteins to nucleic acids by ultraviolet light. *Photochem. Photobiol. Rev.* **5**, 105–197.
12. Saito, I. and H. Sugiyama (1990) Photoreactions of nucleic acids and their constituents with amino acids and related compounds. In *Bioorganic Photochemistry*, Vol. 1 (Edited by H. Morrison), pp. 317–340. John Wiley and Sons, New York.

13. Meisenheimer, K. M. and T. H. Koch (1997) Photocross-linking of nucleic acids to associated proteins. *Crit. Rev. Biochem. Mol. Biol.* **32**, 101–140.
14. Moore, A. M. and C. H. Thomson (1957) Ultraviolet irradiation of pyrimidine derivatives: I. 1,3-dimethyluracil. *Can. J. Chem.* **35**, 163–169.
15. Robertson, K. D. (2003) DNA methylation and chromatin-unraveling the tangled web. *Oncogene* **21**, 5361–5369.
16. Privat, E. and L. Sowers (1996) Photochemical deamination and demethylation of 5-methylcytosine. *Chem. Res. Toxicol.* **9**, 745–750.
17. Vairapandi, M. and N. J. Duker (1994) Excision of ultraviolet-induced photoproducts of 5-methylcytosine from DNA. *Mutation Res.* **315**, 85–94.
18. Zuo, S., R. J. Boorstein, R. P. Cunningham and G. W. Teebor (1995) Comparison of the effects of UV irradiation on 5-methyl-substituted and unsubstituted pyrimidines in alternating pyrimidine-purine sequences in DNA. *Biochemistry* **34**, 11582–11590.
19. Wierzchowski, K. L. and D. Shugar (1960) Further studies on the photochemistry of pyrimidines, with special reference to 5- and 6-substituted derivatives in relation to photoreactivation in the T-even bacteriophages. *Acta Biochim. Pol.* **7**, 63–84.
20. Celewicz, L. and M. D. Shetlar (1992) The photochemistry of 5-methylcytosine and 5-methyl-2′-deoxycytidine in aqueous solution. *Photochem. Photobiol.* **55**, 823–830.
21. Dorwin, E. L., A. A. Shaw, K. Hom, P. Bethel and M. D. Shetlar (1988) Photoexchange products of cytosine and 5-methylcytosine with N-α-acetyl-L-lysine and L-lysine. *J. Photochem. Photobiol. B* **2**, 265–278.
22. Saito, I., H. Sugiyama, S. Ito, N. Furukawa and T. Matsuura (1981) A novel photoreaction of thymidine with lysine. Photoinduced migration of thymine from DNA to lysine. *J. Amer. Chem. Soc.* **103**, 1598–1600.
23. Shetlar, M. D., K. Hom, S. Distefano, K. Ekpenyong and J. Yang (1988) Photochemical reactions of cytosine and 5-methylcytosine with methylamine and n-butylamine. *Photochem. Photobiol.* **47**, 779–786.
24. Shaw, A. A. and M. D. Shetlar (1989) Ring-opening photoreactions of cytosine and 5-methylcytosine with aliphatic alcohols. *Photochem. Photobiol.* **49**, 267–271.
25. Shaw, A. A. and M. D. Shetlar (1989) The photochemistry of 2-alkoxycytosines in phosphate buffer and its link to cytosine photochemistry in alcoholic solution. *Photochem. Photobiol.* **49**, 273–277.
26. Shugar, D. and J. J. Fox (1952) Spectrophotometric studies of nucleic acid derivatives and related compounds as a function of pH. *Biochim. Biophys. Acta* **9**, 199–218.
27. Shaw, A. A. and M. D. Shetlar (1990) 3-Ureidoacrylonitriles: Novel products from the photoisomerization of cytosine, 5-methylcytosine and related compounds. *J. Amer. Chem. Soc.* **112**, 7737–7742.
28. Sulkowska, A. (1997) Effect of temperature on the stability of association of pyrimidine bases with serum albumin: Proton NMR study. *Appl. Spectroscopy* **51**, 428–432.
29. Knittel, J. and R. Zavod (2002) Drug design and relationship of functional groups to pharmacologic activity. In *Foye's Principles of Medicinal Chemistry* (Edited by D. A. Williams and T. L. Lemke), pp. 37–67. Lippincott Williams and Wilkins, Philadelphia, PA.
30. Shetlar, M. D., J. Chung and K. Hom (2010) Photoinduced ring opening reactions and subsequent thermal ring closure reactions in the thymine-ammonia system and related systems. *Photochem. Photobiol Sci.* **9**, 1128–1243.
31. Liu, F.-T. and N. C. Yang (1978) Photochemistry of cytosine derivatives. 2. Photohydration of cytosine derivatives. Proton magnetic resonance study on the chemical structure and property of photohydrates. *Biochemistry* **17**, 4877–4885.
32. Douki, T., G. Vadesne-Bauer and J. Cadet (2002) Formation of 2′-deoxyuridine hydrates upon exposure of nucleosides to gamma radiation and UVC-irradiation of isolated and cellular DNA. *Photochem. Photobiol. Sci.* **1**, 565–569.
33. Hom, K. (1991) Photochemical reactions of 2′-deoxycytidine, 2′-deoxyuridine and related pyrimidine nucleosides and nucleobases with ethylamine. Ph.D. dissertation, University of California, San Francisco.
34. Cadet, J., L.-S. Kan and S. Y. Wang (1978) $O^6$,5′-Cyclo-5, 6-dihydro-2′-deoxyuridine. Novel deoxyuridine photoproducts. *J. Amer. Chem. Soc.* **100**, 6715–6720.
35. Urata, H. and M. Akagi (1991) Photo-induced formation of the 2-deoxyribonolactone-containing nucleotide for d(ApCpA); effects of neighboring bases and modification of deoxycytidine. *Nucleic Acids Res.*, **19**, 1773–1778.
36. DeBoer, G., O. Klinghoffer and H. E. Johns (1970) Reversal mechanisms for the photohyrates of cytosine and its derivatives. *Biochim. Biophys. Acta* **213**, 253–268.
37. Schott, H., U. Hommel and G. Gauglitz (1988) Examination of the photokinetics of 1-β-D-arabinofuranosylcytosine and deoxyribocytidine. *Photochem. Photobiol.* **47**, 517–522.
38. Chang, R. (2000) *Physical Chemistry for the Chemical and Biological Sciences*. University Science Books, Sausalito, CA.