Photochemistry and Photobiology, 2012, 88: 336–343

The (α-4) Photoconjugates of 5-Methylcytosine, 1,5-Dimethylcytosine, 1-Methylthymine and Thymidine

Martin D. Shetlar*,§ and Janet Chung

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

Received 17 October 2011, accepted 21 December 2011, DOI: 10.1111/j.1751-1097.2011.01070.x

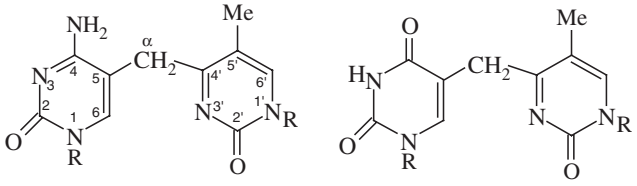
ABSTRACT

The pyrimidine nucleobases contained in DNA undergo a variety of photoinduced reactions in which two moieties become joined to form a product (*e.g.* formation of cyclobutane dimers and [6-4] adducts). Herein, we describe a new type of photoconjugation reaction that has been shown to occur for 5-methylcytosine (5-MeC), 1,5-dimethylcytosine (1,5-diMeC), 1-methylthymine and thymidine; in this reaction the 5-methyl group of one nucleobase (or nucleoside) becomes attached to the 4-position of the second moiety. For example, 5-MeC forms α-4′-(5′-methylpyrimidin-2′-one)-5-methylcytosine. The various (α-4) conjugates are produced upon irradiation of the parent compound in frozen aqueous solution at −78.5°C. The UV spectra of these compounds display a characteristic "double humped" profile, similar to that expected from overlaying the spectrum of parent nucleobase with that of a 2′-pyrimidone moiety. Preliminary results suggest that thymine and 5-methyl-2′-deoxycytidine (5-MedCyd) form analogous photoproducts. A variety of other previously unreported photoproducts are described as well for the 5-MeC, 1,5-diMeC and 5-MedCyd systems.

INTRODUCTION

The photochemistry of DNA pyrimidine nucleobases and their nucleosides has received extensive study over the past 5 decades. A variety of types of products have been isolated and characterized. Among these are cyclobutane pyrimidine dimers (CBDs), the pyrimidine (6-4) pyrimidone photoproducts ([6-4]PPs), (and their Dewar valence isomers), photohydrates and the so-called "spore" photoproducts. (Extensive reviews relevant to these products are available. For older reviews see [1-6], whereas more recent reviews, including information on the photochemistry of DNA at the cellular level, are given in [7–9].) A number of these photoproducts have been implicated as contributing factors to DNA photodamage, ultimately leading to mutagenesis, carcinogenesis and cell death (10). In addition to photoreactions involving only the nucleobase components of DNA, photoexcited nucleobases contained within DNA can react with molecules complexed with DNA in its chromosomal state (*e.g.* structural proteins, such as histones or enzymes involved in DNA replication or transcription). Reviews of photoinduced nucleic acid-protein cross-linking and the chemical nature of the adducts responsible are given in (11–13).

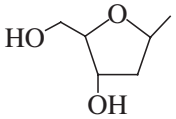
During the course of studies of the photochemical reactions occurring for a number of compounds related to 5-methylcytosine (5-MeC) and thymine (T), we often noted that a product containing a characteristic "double humped" UV spectrum (termed a "dh" product below) was produced. Herein, we report the isolation and characterization of a number of dh nucleobase, nucleoside and related photoproducts. We have found that these compounds fall into a class that can be termed as "(α-4) conjugates," in analogy to the terminology for (6-4) and (5-4) adducts. In particular, we describe (α-4) conjugates resulting from UV iradiation of 5-methylcytosine, 1,5-dimethylcytosine (1,5-diMeC), 1-methylthymine (1-MeT) and thymidine (Thd) in frozen aqueous solution at −78.5°C and present suggestive evidence pointing toward this type of compound being formed in the 5-methyl-2′-deoxycytidine (5-MedCyd) and T systems. In these (α-4) products, one nucleobase (or nucleoside) is joined to a second *via* photoinduced linkage of the methyl group of one nucleobase to the 4-position of the second, as shown in the compounds displayed in Scheme 1. In addition, we describe several other previously uncharacterized photoproducts formed in the 5-MeC, 1,5-diMeC and 5-MedCyd systems.



Ia: R = H
Ib: R = Me
Ic: R = dRib

IIa: R = H
IIb: R = Me
IIc: R = dRib

dRib:



Scheme 1.

§Current address: Department of Molecular and Cell Biology, Beckman Research Institute of the City of Hope, Duarte, CA, USA.
*Corresponding author email: shetlar@cgl.ucsf.edu (Martin D. Shetlar)

MATERIALS AND METHODS

General aspects. 5-Methylcytosine hydrochloride (5-MeC.HCl) and 1,5-diMeC were obtained from Sigma (St. Louis, MO), whereas 5-MedCyd was purchased from R. I. Chemical (Orange, CA); T, 1-MeT and Thd were from Sigma. HPLC solvents were purchased from Fisher (Fair Lawn, NJ), whereas NMR solvents were from Aldrich (Milwaukee, WI). Preparative HPLC separations were carried out on a Shiseido Capcell UG120 10 × 250 mm reverse phase column (5 μm particle size [Yokohama, Japan]); hereafter, this column is termed as Column A. Analytical HPLC was carried out on Column B, a Capcell UG120 4.6 × 150 mm reverse phase column (5 μm particle size). Column C, a Microsorb Phenyl column (4.6 × 250 mm, 5 mm particle size; Varian, Walnut Creek, CA), was used in some purifications. The HPLC system employed was a Rainin binary gradient pumping system (Emeryville, CA) connected to a Hewlett-Packard 1040A diode array HPLC detector (Palo Alto, CA). Before injection, HPLC samples were subjected to spin filtration using Costar Spin-X micro-centrifuge filter tubes containing a 0.2 μm nylon filter (Corning Incorporated, Corning, NY). UV spectra were obtained 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 Sciex API300 triple quadrupole electrospray instrument (Toronto, Canada).

Irradiation methods. Preparative irradiations were carried out 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.

Irradiations in the frozen state at 254 nm were carried out on 250 mL portions of aqueous solution. Each batch was placed in a 13 × 9 inch nonstick baking pan (Bradshaw International, Rancho Cucamonga, CA) and frozen on two 25 × 25-cm dry ice slabs (−78.5°C) placed side by side. During freezing the pan was covered with a 15 × 10 inch Pyrex baking dish; this minimized frosting of the surface of the solution during the freezing process. Four lamps, contained in the two Spectroline X-Series lamp housings, were placed across the top of the pan containing the solutions; irradiations were carried out with the pan resting on top of the pair of dry ice slabs. The surface exposure of the frozen aqueous layer in the pan averaged about 25 J m^{-2} sec^{-1}, as measured using a Spectronics DM-254N Ultraviolet Meter (Westbury, NY).

In general, solutions to be irradiated were prepared by simple dilution of 20 mM stock solutions of the parent compounds. In the case of 5-MeC, the solutions for irradiation were prepared by a somewhat more complex protocol, as described in Section A1 in Appendix S1 in the Supporting Information. Solutions were not deoxygenated prior to irradiation.

Photochemical reactions in the 1,5-dimethylcytosine system: an example protocol for photoproduct preparation, isolation and purification. In the following, we provide a sample protocol used for preparation, isolation and purification of the 1,5-diMeC photoproducts of interest for structural study. As the corresponding procedures focusing on 5-MeC, T, 1-MeT, Thd and 5-MedCyd photoproducts are similar in design (but different in detail), they are not presented in the body of this article, but are given in Appendix S1. The structural characterizations of a number of these latter compounds, as well as the various photoproducts of 1,5-diMeC, are discussed in Results and Discussion. For those cases where less-complete characterization was accomplished, discussions of tentative photoproduct identifications are provided in Appendix S1.

In a representative experiment, we irradiated 1250 mL of aqueous 0.2 mM 1,5-diMeC in the frozen state for 32 min in 250 mL batches at 254 nm as described elsewhere in the text. Analytical HPLC indicated that about 20% of the parent compound was consumed. The resulting solution was concentrated to 1.5 mL by rotatory evaporation at 40°C; samples to be injected were subjected to spin filtration. Injections of 250 μL samples on Column A were made, using the following linear HPLC water/methanol gradient flowing at 4 mL min^{-1}: 0 min, 3% MeOH; 15 min, 20%; 15.3 min, 3%; 20 min, 3%. Ten fractions were collected as follows: G1, 2.7–6.8 min; G2, 6.8–7.1 min; G3. 7.1–8.7 min; G4, 8.7–9.1 min; G5, 9.1–9.9 min; G6, 9.9–10.8 min; G7, 10.8–13.4 min; G8, 13.4–14.0 min; G9, 14.0–18.4 min; G10, 18.4–19 min. After rotatory evaporation of each of these fractions to 1 mL at 40°C, the analytical HPLC of each fraction was run on Column B. For G1 and G2, 100% 10 mM sodium phosphate, pH 7.5 was used, whereas for G3–G6, 96% (10 mM sodium phosphate)/4% MeOH was the eluent. (In the remainder of this article, sodium phosphate will be taken to imply sodium phosphate buffer, pH 7.5.) For G7 and G8, the eluent was 94% (10 mM sodium phosphate)/6% MeOH and for G9 and G10, 91% (10 mM sodium phosphate)/9% MeOH was employed. Four fractions (G2, G4, G8 and G10) contained compounds of particular interest; G6 contained most of the parent 1,5-diMeC, whereas G8 contained the 1,5-diMeC dh of primary interest. The remaining fractions contained a variety of other compounds, none present in dominating amounts. Each of the fractions selected for further study was further purified for purposes of mass spectrometry and NMR spectroscopy. (The HPLC chromatogram resulting from injection of the above concentrate is shown in Fig. 1, along with the correlation between the various peaks and compounds described in the following paragaraphs.)

Fraction G2: The 1 mL of G2 concentrate was injected in 500 μL batches on Column B using 10 mM sodium phosphate, pH 7.5, flowing at 2 mL min^{-1} as eluent. The peak eluted as a doublet; the latter component, centered at 1.85 min, was collected and is hereafter termed as PG_2. Reinjection showed it contained only a trace amount of the early eluting component, which was shown to be 5-MeC *via* coinjection and by identity of its UV spectrum with that of an injected authentic sample of 5-MeC. This 5-MeC was probably due to its presence as a trace impurity in the parent 1,5-diMeC used; HPLC inspection of the unirradiated 1,5-diMeC solution showed a very small peak eluting with the same retention time and UV spectrum as 5-MeC.

Fraction G4: The 1.5 mL of G4 was purified by injection of one 500 μL batch and one 1000 μL batch on the same column, using 96% (10 mM sodium phosphate)/4% MeOH eluting at 2 mL min. The material eluting between 2.5 and 3.0 min (PG_4) was collected; a later eluting small shoulder was not collected. Analytical HPLC indicated that it was pure.

Fraction G8: One mL of concentrated G8 was injected in 500 μL batches on Column B using 94% (10 mM sodium phosphate)/6% MeOH flowing at 2 mL min^{-1}. The desired product, 1,5-diMeC dh (PG_8), eluted as a large peak between 5.3 and 6.5 min; this material was pure, as evidenced by analytical HPLC using the same conditions.

Fraction G10: The 1000 μL of concentrate of G10 was ejected as one batch on Column B, using 91% (10 mM sodium phosphate)/9% MeOH flowing at 2 mL min^{-1} as eluent. The desired material (PG_{10}) eluted between 6.5 and 7.5 min. Analytical HPLC indicated that it was pure.

The purified compounds listed above were stored in the frozen state until samples were taken for mass spectral and NMR studies. Prior to such studies, each of the compounds was repurified and then desalted by running preparative HPLC of the purified compound on Column B, using the solvent compositions described elsewhere in the text (but using distilled water, rather than sodium phosphate). In these desalting runs, the peaks in some cases were somewhat broader and misshapen, as compared with those found when phosphate buffer was a component of the eluent.

Photochemistry in the 5-MeC, 5-MedCyd, T, 1-MeT and Thd systems. The reaction conditions and the isolation and purification protocols used for these systems are similar to those described above for the 1,5-diMeC system. Due to these similarities, we have placed the corresponding detailed protocols in Appendix S1.

Elsewhere in the text, we denote the HPLC fractions and photoproducts in the 1,5-diMeC system using the nomenclature Gx to signify fractions and PG_X to represent products. In the remainder of this article and Supporting Information, we designate HPLC fractions and photoproducts for each system studied using similar nomenclatures. In particular, we will use letter prefixes as follows: 5-MeC: Fx, (PF_x); 5-MedCyd: Mx (PM_x); T: Jx (PJ_x); 1-MeT: Hx (PH_x); Thd: Lx, (PL_x). A complete listing of the compounds isolated in this study, along with the compound designation (*e.g.* PG_4), structural designation (*e.g.* **IIIa**) and compound or compound type (*e.g.* 5-formyl-1-methylcytosine) will be found in Table S1 in the Supporting Information.

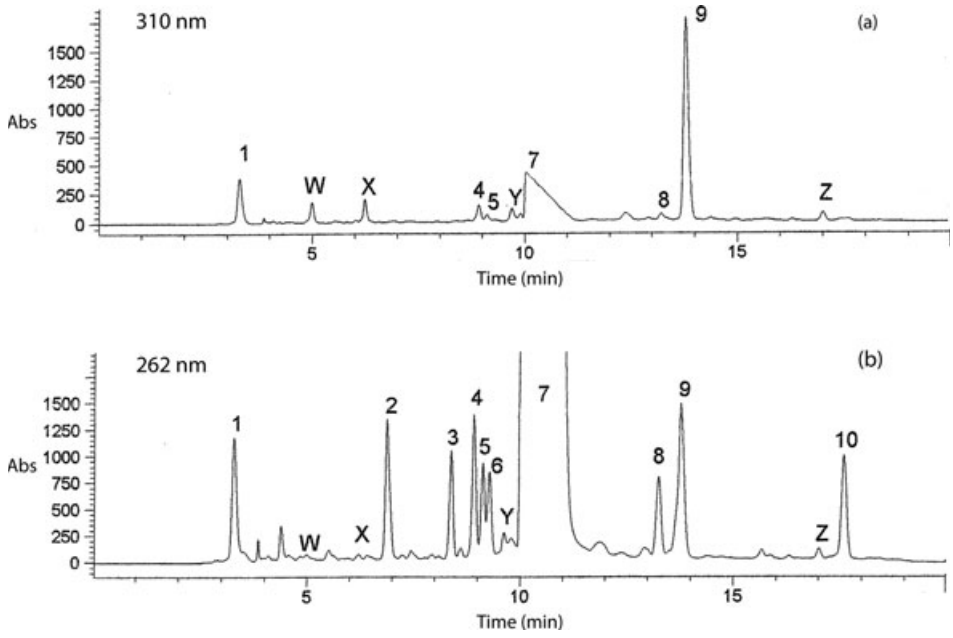


Figure 1. The HPLC chromatogram resulting from injection of 250 μL of a solution (1250 mL) of 0.2 mM 1,5-diMeC that had been irradiated for 32 min on dry ice and concentrated 833-fold by rotatory evaporation. Detailed chromatographic conditions, using Column A, are described in the Materials and Methods section. Panel (a) shows the chromatogram seen with detection at 310 nm, whereas panel (b) shows the same chromatogram with detection at 262 nm. The compounds of interest in this study eluted in the peaks denoted by 2 (PG_2), 4 (PG_4), 9 (PG_8) and 10 (PG_{10}). Probably, peaks 3, 5, 6 and 8 correspond to CBDs, based on the profiles of their absorption spectra, whereas the substances eluting in peaks W and X have spectra characteristic of (6-4) adducts. Peak 7 corresponds to the parent 1,5-diMeC. The spectrum of material eluting in peak Z has a ''triple hump'' profile, similar to that shown in the middle panel of Fig. S7.

RESULTS AND DISCUSSION

In the following, we discuss the characterization of the various ''dh'' products from the relevant HPLC fractions for each irradiated compound, collected and purified as described in the Materials and Methods section or in Appendix S1. In particular, we present lines of evidence (from mass spectral, proton NMR and UV spectroscopic studies) that convincingly identify four of these dh products (1,5-diMeC, 5-MeC, 1-MeT and Thd) as (α-4) products; for two others (T, 5-MedCyd), the corresponding evidence is strongly suggestive that this is the case. In addition, for several of the compounds studied, we isolated and characterized additional previously unreported photoproducts. The characterizations of these compounds are also discussed in the following paragraphs.

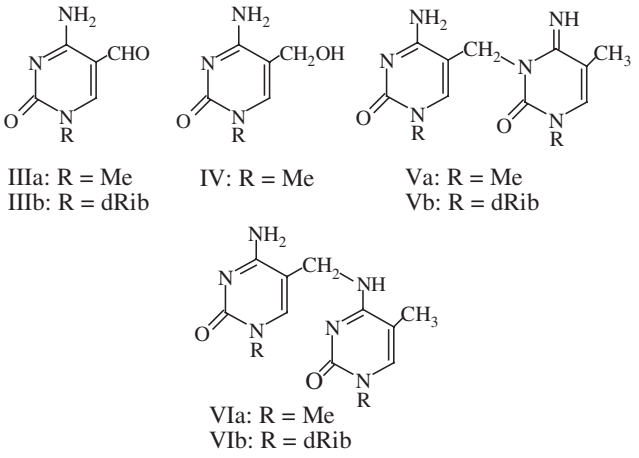
Photoproducts in the 1,5-dimethylcytosine (1,5-diMeC) system

Herein, we will discuss the identification of four photoproducts, termed as PG_2, PG_4, PG_8 and PG_{10}, that were isolated after 1,5-diMeC was irradiated in the aqueous frozen state. The details of the isolation and purification of these compounds was described in the Materials and Methods section.

The (α-4) conjugate of 1,5 dimethylcytosine. We established the structure of the product termed as PG_8 using UV spectroscopy, ESI mass spectrometry and proton NMR. The UV spectrum showed a ''double humped'' appearance (see Fig. S1 in the Supporting Information), although the long wavelength ''λ_{max}'' at 306 nm is more like an extended shoulder than a maximum. A short wavelength λ_{max} occurred at 286 nm and there was a λ_{min} at 253 nm. This spectrum corresponds to that expected if one superimposed the spectrum of 1,5-diMeC upon that of a 1,5-dimethylpyrimidin-2-one ring, the latter being analogous to that seen in (6-4)PPs. This suggests that the spectrum of PG_8 could result from two independent sets of conjugated double bonds, as displayed in the structure **Ib** in Scheme 1. The ESI mass spectrum indicated that the molecular mass of [M + H^+] is 262.1, which is that expected for a protonated conjugate of two 1,5-diMeC moieties in which the equivalent of an ammonia molecule has been lost. The proton NMR, run in D_2O (with TSP as reference), shows proton peaks at δ = 2.13 (C5′C**H**$_3$, 3), δ = 3.38 and 3.56 (N_1 and N_1′ C**H**$_3$, s, 3), 3.78 (C5C**H**$_2$, s, 2) and 7.46 and 7.93 (C6 and C6′ CH, each s, 1). (Here, and in the following, all chemical shift values (δ) are in ppm and all spin–spin coupling constants [J] are in Hz.) The above data are consistent with PG_8 having the structure shown by **Ib** in Scheme 1; this compound can be named as α-4′-(1′,5′-dimethylpyrimidin-2′-one)-1,5-dimethylcytosine.

The yield of purified **Ib** was estimated to be about 1.7% *via* UV spectroscopy, using an estimated value for ε_{286} of 8000 L mol^{-1} cm^{-1}.

Identification of PG_4 as 5-formyl-1-methylcytosine (IIIa) and PG_2 as 5-hydroxymethyl-1-methylcytosine (IV). The product PG_4, purified as described elsewhere, displays a UV spectrum with λ_{max} at 251 nm and 285 nm and λ_{min} at 242 nm and 266 nm. (The UV spectrum is shown in the top panel of Fig. S2.) Although the UV spectrum for PG_4 has multiple humps, the profile is significantly different than that of the 1,5-diMeC (α-4) product. The ESI mass spectrum showed that the mass of [M + H^+] was 154.0, which is consistent with a molecular formula of $C_6H_7N_3O_2$. The proton NMR spectrum,



Scheme 2.

run in D_2O with TSP as reference, displays resonances at δ = 3.53 (N1C**H**$_3$, s, 3), 8.49 (C6**H**, s, 1) and 9.49 (C**H**O, s, 1). These latter two assignments are consistent with those given for the corresponding protons in the NMR for 5-formyl-2′-deoxycytidine (14), also run in D_2O. Thus, PG_4 is assigned to have the structure **IIIa** in Scheme 2.

The purified compound PG_2 displayed a UV spectrum with λ_{max} at 275 nm and λ_{min} at 251 nm. (The UV spectrum is shown in the middle panel of Fig. S2.) The ESI mass spectrum showed that the mass of $[M + H^+]$ was 156.0, which is consistent with a compound with a molecular formula of $C_6H_9N_3O_2$; the presence of an additional oxygen, as compared to the parent formula $C_6H_9N_3O$, suggests that a methyl group has been converted to a CH_2OH moiety. The proton NMR spectrum, run in D_2O, shows resonances at δ = 3.40 (N$_1$C**H**$_3$, s, 3), δ = 4.43 (C5C**H**$_2$, s, 2) and δ = 7.65 (C6**H**, s, 1). The above spectral data are consistent with identification of PG_2 as 5-hydroxymethyl-1-methylcytosine (**IV**), shown in Scheme 2. This compound has been previously synthesized (15) and the NMR chemical shifts for nonexchangeable hydrogens, obtained in d_6-DMSO, are consistent with those obtained here in D_2O (although, as would be expected, various resonances differed somewhat in chemical shift value). Very similar chemical shifts have been observed for C5C**H**$_2$ and C6**H** in 5-hydroxymethyl-2′-deoxycytidine run in D_2O (14).

Using assumed values of ε_{275} (PG_2) and ε_{285} (PG_4) of 8000 L mol^{-1} cm^{-1}, we employed UV spectroscopy to estimate the yield of purified product to be about 0.5% for each of these compounds.

Tentative assignment of PG_{10} to be the α-N_3 conjugate of 1,5-dimethylcytosine. The product corresponding to PG_{10} displayed a UV spectrum with λ_{max} at 283 nm and λ_{min} at 253 nm. (The UV spectrum is shown in the bottom panel of Fig. S2.) The ESI mass spectrum provided a value of 277.0 for the mass of $[M + H^+]$, consistent with the loss of two hydrogen atoms from a dimeric form of 1,5-diMeC. The proton NMR spectrum, run in D_2O, showed proton peaks at δ = 1.91 (C5C**H**$_3$, s, 3), δ = 3.37 and 3.39 (N$_1$ and N$_1$′ C**H**$_3$, each s, 3), 4.41 (C5′C**H**$_2$, s, 2) and 7.40 and 7.65 (C6 and C6′ C**H**, each s, 1). The NMR data indicates that a 5-methyl group of the parent compound has been converted to a methylene linkage. However, as vinyl hydrogens at C6 and C6′ and an intact methyl group at C5 are present in the product, the site of attachment must be elsewhere. As two oxygens are still present in PG_{10} (from the molecular mass measurement) and as the N1 methyl groups are intact (from proton NMR), the only remaining site on the ring for attachment of the methylene group is N3. A similar attachment site between 5′-methylene and N3 has been observed in the compound 3-(α-thymidyl)thymidine, which is formed when thymidine in frozen aqueous solution is irradiated with gamma rays (16); the chemical shift for the corresponding methylene protons in this compound was about 4.3 ppm. On the basis of above reasoning, we tentatively identify PG_{10} to be **Va** in Scheme 2. However, another possible structure cannot be ruled out by the available spectroscopic data, namely that shown in Scheme 2 as **VIa**. In this structure, the methylene bridge is attached to the exocyclic amino group, rather than to a ring nitrogen. Appropriate NMR studies in an aprotic solvent could provide conclusive evidence favoring either **Va** or **VIa** as the structure in best agreement with the resulting experimental data.

It can be noted that **Va** may exist in aqueous solution near neutral pH in a form that has the imino group protonated. The iminium form of the related compound 1,3-dimethylcytosine has a pK_a of 9.3 (17); the λ_{max} for the protonated form of this compound was found to be 281 nm, whereas that of the neutral species was 272 nm. Compound **VIa** would not be predicted to have a functional group with a pK_a having a similar value; indeed the pK_a of 1-methylcytosine (1-MeC) is about 4.6 (17). The λ_{max} value for protonated 1-MeC is 282 nm and that of unprotonated form is 273 nm (17). If similar correlations between λ_{max} and protonation state exist for the corresponding 5-MeC derivatives, then studies of the acid-base properties of PG_{10} could also yield evidence useful for distinguishing between **Va** and **VIa** as potential structures for this compound.

Using an assumed value for ε_{283} of 16000 L mol^{-1} cm^{-1}, UV spectroscopy was used to estimate the yield of puified PG_{10} to be about 0.4%. This roughly estimated ε value takes into account that the proposed structure **Va** (as well as the alternate structure **VIa**) contains two rings that would be expected to have UV absorption properties similar to parent 1,5-diMeC.

One question that can be asked is whether each of the four compounds identified above as photoproducts in the 1,5-diMeC system are primary photoproducts. For example, it could be possible that some of these compounds are formed during work-up from other components of the original reaction mixture. To answer this question, we irradiated parent 1,5-diMeC (0.2 mM) on dry ice under the same conditions as described previously, but with smaller sample volumes and times of irradiation. We injected 200 μL of the various irradiated samples on Column B immediately after thawing, using the following linear water/methanol HPLC gradient flowing at 2 mL min^{-1}: 0 min, 3% MeOH; 7 min, 20%; 7.25 min, 3%; 10 min, 3%. We found that PG_8 could be readily identified in freshly irradiated and thawed solution after 4 min irradiation *via* its retention time and characteristic UV spectral profile. Compounds with the properties of PG_2 and PG_{10} could be easily detected after 8 min irradiation, based on their retention times and on the shape of their UV absorption spectral profiles (as measured "on the fly" with the diode array detector). The amounts of each compound increased with time of irradiation. We confirmed the identifications of PG_2 and PG_{10} by

rotovaporating 2 mL portions of the 4 and 32 min samples to 200 μL and injecting the resulting volumes on the same column; considerably larger HPLC peaks resulted for these compounds after this procedure. Overlay of the UV spectral profiles corresponding to these enhanced peaks with those of authentic samples resulted, in each case, in a close match. It can be concluded that both PG_2 and PG_{10}, as well as PG_8, are primary photoproducts. As a result of overlapping of the retention time of PG_4 with a considerably larger putative 1,5-diMeC CBD peak, we were not able to draw conclusions about whether this compound is a primary photoproduct.

Photochemistry in the 5-MeC system

The formation of CBDs in the 5-MeC system has been described previously (18). Herein, we discuss the characterization of two new primary photoproducts and a secondary product formed *via* thermal decomposition of one of the primary photoproducts.

The product PF_6 of 5-MeC can be identified as an (α-4) product. The preparation, isolation and purification protocols for the three products of interest, termed as PF_6, PF_{21} and PF_{22}, are described in detail in Appendix S1. In this subsection, we will discuss the identification of the product PF_6. The UV spectrum of PF_6 in distilled water showed maxima at 278 and 310 nm and minima at 252 and 295 nm. The ESI mass spectrum of PF_6 showed a strong $[M + H^+]$ peak at 234.1, corresponding to the protonated molecular mass of a compound containing two 5-MeC moieties from which the equivalent of one NH_3 has been lost. Weaker peaks at 253.0 and 268.9 correspond to $[M + Na^+]$ and $[M + K^+]$. The proton NMR, run in d_6-DMSO with TMS as a standard, shows nonexchangeable proton peaks at δ = 1.98 (C′5C**H**$_3$, s, 3), δ = 3.52 (C5C**H**$_2$, s, 2), and overlapping broadened resonances centered at about δ = 7.14 (C**H**, b, 2) corresponding to the vinyl protons at C5 and C5′. In addition, peaks at δ = 7.05, 7.68 and 9.03 were evident that disappeared upon addition of a small amount of D_2O. These must correspond to exchangeable protons attached to nitrogen. The addition of D_2O also led to separation of the composite of overlapping peaks at 7.14 into two peaks, one at δ = 7.13 and the second at δ = 7.15. Taken as a whole, the UV spectral, mass spectral and proton NMR spectral evidence are consistent with PF_6 having the structure **Ia** in Scheme 1.

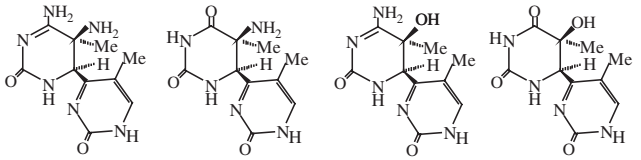
Other photoproducts of 5-MeC. As mentioned elsewhere in the text, two additional 5-MeC products were isolated, namely PF_{21} and PF_{22} (see Section A1 in Appendix S1 for details). Both of these compounds have UV absorption spectra that are characteristic of (6-4) (or [5-4]) adducts. One of these (PF_{21}) is produced *via* thermal decomposition of the other (PF_{22}). We characterized both of these compounds by mass spectrometry. We also obtained a reasonable proton NMR spectrum for PF_{21}, but not for PF_{22}. Due to the thermal instability of the latter compound, we were unable to obtain a sufficient amount of pure sample to yield a high-quality NMR spectrum.

The product PF_{22} displayed λ_{max} at 315 nm and λ_{min} at 272 nm. ESI mass spectrometry yielded a parent peak with a mass value of 251.14 for $[M + H^+]$, which is consistent with the PF_{22} being a dimeric form of 5-MeC; also present was a peak at 234.12, which is the value of the mass expected for $[M + H^+]$ if the parent dimer had lost NH_3. The product PF_{21} had λ_{max} at 316 nm and λ_{min} at 255 nm. Mass spectrometry gave a parent peak with a value of $[M + H^+]$ = 252.1. This is consistent with PF_{21} being a product arising from PF_{22} *via* replacement of an equivalent of ammonia with an equivalent of water. The proton NMR of PF_{21} displayed resonances at the following δ values in d_6-DMSO: 1.39 (C5C**H**$_3$, s, 3), 2.06 (C5′C**H**$_3$, s, 3), 4.18 (C6**H**, s, 1), 7.75 (C6′**H**, s, 1). In addition, there was a broad peak at δ = 7.68 (b, 0.74) and a sharper peak at δ = 9.94 (s, 0.5); both of these peaks disappeared upon addition of a drop of D_2O. The lack of spin–spin splitting in the C6**H** peak suggests that this proton is not located adjacent to OH in the structure of PF_{21}; this points toward this compound and its precursor being identified as (6-4)PPs rather than (5-4) adducts, such as have been recently observed as photoproducts when 1-methylthymine (1-MeT) and thymidine (Thd) are irradiated in the frozen aqueous state (19). Assuming this is true, PF_{22} can likely be identified as **VIIa** and PF_{21} as either **VIIb** or **VIIc**, probably the former (see Scheme 3). This latter statement is justified by the observation that the UV absorption spectrum of PF_{21} greatly resembles the corresponding spectrum of **VIId**, the thymine (6-4) adduct; in particular, it does not display a shoulder of absorption extending up to 270 nm, such is seen for the 5-MeC CBDs (18), which contain a saturated 5,6 bond (and indeed, PF_{22}, tentatively identified as **VIIa**).

Formation of (α-4) conjugates in the 1-MeT and T systems

In the following section, we discuss the detailed characterization of the 1-MeT (α-4) conjugate and the evidence that suggests, but does not prove, that a corresponding (α-4) product is formed when T is irradiated in frozen aqueous solution. The photochemistry of both of these systems has been extensively reported in previous studies and will not be discussed here in detail (reviews of the photochemistry of T, including studies carried out in frozen solution, are provided by [1,2,4], whereas previous work on the photoproducts formed when 1-MeT is irradiated is described in [19,20]).

The (a-4) conjugate of 1-MeT. The production and isolation of a 1-MeT photoproduct with a double-humped spectrum is described in another study dealing with the isolation of (5-4) and (6-4)PPs of this compound (19). This compound was termed as "P9" in that article; for consistency of notation, it is termed as PH_9 here. A brief description of the isolation protocol for PH_9 is given in Appendix S1. The UV spectrum, shown in the lower panel of Fig. S1, displays maxima at λ = 275 and 309 nm and minima at 241 and 298 nm. The ESI mass spectrum displayed peaks at 263.2 $[M + H^+]$, 285.1 $[M + Na^+]$ and 301.1 $[M + K^+]$. This is consistent with PH_9 being a compound with the formula $C_{12}H_{14}N_4O_3$, which is equivalent to a dimeric form



VIIa: R = H **VIIb**: R = H **VIIc**: R = H **VIId**: R = H

Scheme 3.

of 1-MeT that has lost an equivalent of H_2O. The proton NMR spectrum in D_2O showed peaks at δ = 2.15 (5CC**H**$_3$, s, 3), 3.38 and 3.54 (N1 and N1′C**H**$_3$, each s and 3), 3.72 (C5′C**H**$_2$, s, 2), 7.48 and 7.89 (C6**H** and C6′**H**, each s, 1). This spectral information is consistent with PH_9 being α-4′-(1′,5′-dimethylpyrimidin-2′-one)-1-methylthymine (**IIb** in Scheme 1).

UV spectroscopic and mass spectral evidence suggests that α-4′-(5′-methylpyrimidin-2′-one)-thymine (IIa) is formed when T is irradiated in ice. Isolation and purification of a compound with a double-humped spectrum, from frozen aqueous solutions that had been irradiated at 254 nm, was accomplished as described in Section A3 of Appendix S1; this compound is termed PJ_{64} there. This compound displays a UV absorption spectrum with maxima at 264 and 308 nm, whereas there are absorption minima at 238 and 294 nm. The ratio of A_{264}/A_{308} was 1.90. The spectrum of this compound differs from that of the compound obtained by dehydrating the C5C6 bond of the thymine (6-4) adduct, for which the observed λ_{max} occur at 257 and 319 nm (see table 7 in the review by Wang [3]). The mass spectrum, run in the MALDI mode, gave a molecular mass for $[M + H^+]$ of 235.09, whereas the ESI spectrum, run in the negative ion mode, gave an $[M - H]^-$ peak with a molecular mass of 233.07. Both of these values suggest that this compound corresponds to a dimer of thymine that has lost one mole of water. Because only a very small amount of this compound could be isolated from rather large preparative runs (see Appendix S1), we were unable to obtain enough of PJ_{64} to obtain an NMR spectrum. The similarity of the UV absorption spectra to those of the various (α-4) adducts described elsewhere in the text, along with the fact that the measured values of the molecular mass are consistent with expectation, provides strong suggestive evidence that this compound can be identified as α-4′-(5′-methylpyrimidin-2′-one)-thymine (**IIa** in Scheme 1).

The (α-4) conjugate of thymidine

In a recent study of the (5-4) adducts of Thd (19), evidence was presented for production of a compound that showed a double-humped UV spectrum; as discussed in that article, this compound eluted in a fraction termed F11′ under HPLC conditions used in that investigation. Herein, we term this compound PL_{11}. (An abbreviated description of the protocols used for isolation of this product is included in Appendix S1.) Elsewhere in the text, we discuss the structural characterization of this molecule.

Identification of PL_{11} as an (α-4) conjugate of Thd. The UV spectrum of PL_{11} shows λ_{max} at 268 and 310 nm along with λ_{min} at 240 and 296 nm; the ratio A_{268}/A_{310} was 1.53 (see the bottom panel in Fig. S3). This spectrum suggests that PL_{11} contains both a thymine ring and a 2-pyrimidone ring. The ESI mass spectrum of PL_{11} showed peaks at 467.2 $[M + H^+]$, 489.2 $[M + Na^+]$ and 505.2 $[M + K^+]$. The peak at 467.2 corresponds to the protonated form of a compound with the molecular formula $C_{20}H_{26}N_4O_9$ and is thus consistent with PL_{11} having the structure shown by **IIc** in Scheme 1. The δ and J values, obtained from detailed analysis of the proton NMR spectrum (shown in Fig. S4), are given in Table 1. Of particular interest is the observation that PL_{11} contains one methyl and one methylene group, as well as two vinyl protons.

Photochemistry in the 5-methyl-2′-deoxycytidine system

The HPLC of 5-MedCyd, after irradiation in frozen aqueous solution, indicates that the resultant reaction mixture contains a variety of products (see Fig. S5). One of these compounds as been definitively identified as 5-formyl-2′-deoxycytidine (**IIIb** in Scheme 2) *via* comparison of its UV, mass spectral and proton NMR properties with those previously published for this compound (14). This identification is discussed elsewhere in the text.

The identification of the other compounds is more tentative. One of the isolated photoproducts displays a "double humped" spectrum analogous to those seen for 5-MeC and 1,5-diMeC. This material was isolated in a fraction (termed as M7) and purified to yield a product labeled PM_7; this compound has been tentatively identified to be an (α-4) product with the structure given by **Ic** in Scheme 1. Details concerning methods, along with the data supporting this identification, are given in Section A5 of Appendix S1.

Other compounds were also isolated and subjected to study. One of these has been tentatively identified as being a α-N_3 adduct of 5-MedCyd (**Vb**). Two other products may be analogs of the "spore" adducts of Thd (4, p. 96); because of difficulties in obtaining stable pure samples of these photoproducts (possibly due to sugar isomerisation), these identifications are tentative. Details concerning these other 5-MedCyd products are given in Section A6 in Appendix S1 and in Figs. S6 and S7. In addition to these products, HPLC peaks corresponding to CBDs and (6-4)PPs were identified. The structural characterization of some of these products will be given in another place.

Fraction M6 contains 5-formyl-2′-deoxycytidine. The fraction M6 contained several photoproducts with one being of particular interest, as its UV spectrum resembled that given in Fig. S2 for 5-formyl-1-methylcytosine. After purification as described in Appendix S1, this product, termed as PM_6, was studied by UV spectroscopy, ESI mass spectrometry and proton NMR. The UV spectrum showed spectral maxima at λ_{max} at 223 and 283 nm. The ESI mass spectrum displayed peaks at 256.4 $[M + H^+]$ and 278.5 $[M + Na^+]$. The former mass corresponds to the mass of protonated 5-formyl-2′-deoxycytidine with the molecular formula $C_{10}H_{13}N_3O_5$. Analysis of the NMR spectrum yielded chemical shifts and spin–spin coupling parameters nearly identical to those previously published for 5-formyl-2′-deoxycytidine (14).

Speculations concerning the mechanism of (α-4) photoproduct formation

The mechanism of conversion of, say, thymidine to its corresponding (α-4) product is an intriguing topic to consider. Studies made by Shaw and Cadet (16) on the mechanism of formation of the Thd "spore" photoproduct (**VIII**) are, perhaps, relevant to this problem. In this study, it was shown that this reaction likely occurs by a concerted process when Thd is irradiated with γ-rays in frozen solution. (Because of the chiral center at C5 in the saturated portion of the product (denoted by an asterisk in **VIII**), the spore product is produced in two diastereomeric forms.) As evidence of the concerted nature of this reactive process, it was shown that reaction of

Table 1. Proton chemical shifts (δ) and coupling constants (J) for the (α-4) product (IIc) of thymidine.

Proton	5Me	CH2	H6	H6′	Ring	1′	2′	2″	3′	4′	5′	5′
δ	2.17	3.75	7.68	8.21	A	6.31	2.43	2.37	4.43	4.03	3.78	3.69
					B	6.22	2.34	2.60	4.44	4.15	3.91	3.80

J	H1′-H2′	H1′-H2″	H2′-H2″	H2′-H3′	H2″-H3′	H3′-H4′	H4′-H5′	H4′-H5″	H5′-H5″
Ring A	6.7	6.6	13.9	4.2	6.0	4.3	3.6	4.9	12.3
Ring B	5.8	6.5	14.2	6.1	4.5	4.1	3.7	4.6	12.4

Proton NMR spectra were run in D_2O, using TSP as a reference compound. Total correlation spectroscopy (TOCSY) was used to partition the sugar protons of the (α-4) product into two spin systems, designated as "A" and "B." Also, based on the TOCSY study, H6′ (8.21) and 5Me (2.17) were found to be coupled; thus, the corresponding protons occur in the same ring system. The pyrimidine ring systems with which Rings A and B are associated have not been determined. The chemical shifts in ppm (δ) (upper table) and magnitudes of the coupling constants in Hz (J; lower table) for the sugar protons were estimated with SPINWORKS (Version 3.1, 2009, Kirt Marat, University of Minnesota, Minneapolis, MN), using the simulation module, NUMMRIT, with default parameters. (NUMMRIT algorithm reference: [21]. The simulated proton NMR spectrum closely matched the experimental spectrum. Integrated areas were in reasonable agreement with expectation; in cases where signal overlap occurred, the sums of the appropriate areas agreed with expected values.

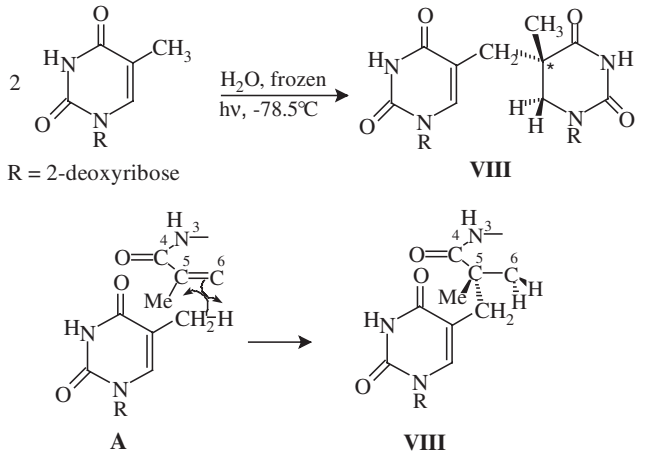
Thd, fully deuterated on the C5 methyl group, led to transfer of deuterium from one of the methyl groups to the 6-position in the resulting spore product, making an alternative radical recombination process unlikely. Both the α-methylene group and the second methyl group were fully deuterated in the final products. (It should be noted that there was evidence for production of **VIII** by a radical combination process when Thd, as a solid film, was irradiated with γ-rays (16).) Possibly, the simplest concerted mechanism that can be visualized is that shown in the bottom panel of Scheme 4, in which the photoreactive pair (A), presumably suitably oriented by the surrounding ice matrix, takes part in a hydrogen transfer process as shown. However, other more complex (and probably more likely) mechanisms can also be considered, one of which is shown in Fig. S8 in the Supporting Information.

A similar simple mechanism can be drawn for the photoinduced conversion of Thd and related compounds to the (α-4) product **II** (see Scheme 5). Again it is assumed that the ice matrix suitably constrains a reactive pair, labeled B in the bottom part of Scheme 5, to a configuration in which hydrogen transfer from the methyl group of one reactant occurs to the carbonyl of the second; this results in the intermediate Int-1. Loss of water then leads to **II**, the observed (α-4) conjugate.

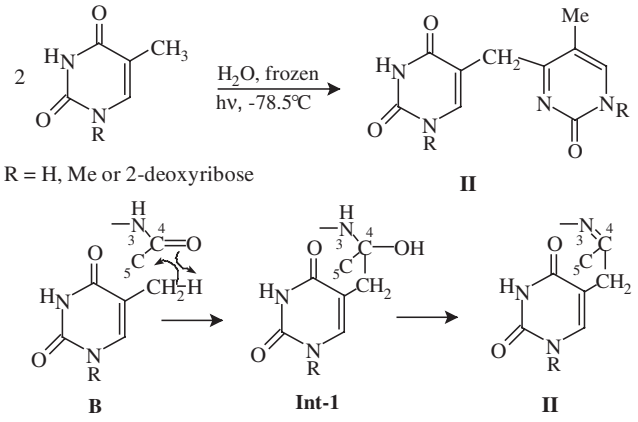
An analogous mechanism can be proposed for formation of (α-4) conjugates of 5-MeC (**Ia**), 1,5-diMeC (**Ib**) and 5-MedCyd (**Ic**). In each of these cases, it would be postulated that the photoexcited state exists as a tautomer with an imine group at C4, rather than an amino group, and this group reacts analogously to carbonyl in the mechanism shown in Scheme 5. Such an imine functionality is likely to be involved in the formation of the (6-4) photoproduct occurring when the dinucleotide 2′-thymidylyl-(3′-5′)-2′-deoxycytidine is irradiated in aqueous solution (22). Such products contain an amino group at the 5-position of the 5,6-dihydropyrimidine portion of the adduct; this probably results from a ring-opening reaction of a four-membered azetadine ring system produced by the putative photoinduced cycloaddition of imine at the 4-position of the cytosine ring to the alkene linkage in the thymine ring.

Unfortunately, in the mechanism outlined elsewhere in the text, deuteration of the methyl group would not lead to useful information as to the validity of this mechanism; any transferred deuterium would be lost in the reaction of Int-1 to form final product. (However, observation of deuterium transfer from a deuterated methyl group to another position [say, a ring carbon] in high yield in the final product would tend to provide evidence against the mechanism outlined in Scheme 5.)

While the above mechanism has the virtue of simplicity, there may be other mechanisms that can be proposed to explain the conversion of, say, Thd to **IIc**. Definitive exper-



Scheme 4.



Scheme 5.

imental elucidation of the actual mechanism could provide a challenge for future experimentalists.

It is unlikely that (α-4) conjugates, such as described elsewhere in the text, would be formed in photoreactions of double stranded DNA in the B form. However, it would be interesting to determine if such products can be formed in the context of denatured DNA or DNA in structural forms alternative to the B form. In such cases, relaxation of structural constraints could allow, say, T residues to take up spatial relationships conducive to formation of (α-4) products. Such relaxed constraints on reactivity have been observed, for example, in the photoreactions of T in A form DNA (23) or human telomeric G-quadruplex DNA (24) to form CBDs with an *anti* configuration.

Acknowledgements—This work was supported by the National Science Foundation (CHE-0131203). We gratefully acknowledge the Bioorganic, Biomedical Mass Spectrometry Resource (A. L. Burlingame, Director), supported by the NIH Division of Research Resources Grant RR 01614 and, especially, David Maltby, who provided needed mass spectra in a timely manner.

SUPPORTING INFORMATION

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

Appendix S1. Experimental details of the preparation and purification protocols for various photoproducts and descriptions of the partial characterizations of certain photoproducts of 5-methyl-2′-deoxycytidine.

Figure S1. The UV spectra of the "dh" products of 1,5-dimethylcytosine (**Ib**) and 1-methylthymine (**IIb**).

Figure S2. The UV spectra of three other photoproducts of 1,5-dimethylcytosine.

Figure S3. The UV spectra of the "dh" products of 5-methyl-2′-deoxycytidine (putatively **Ic**) and thymidine (**IIc**).

Figure S4. The proton NMR spectrum of the "dh" product of Thd (**IIc**).

Figure S5. The chromatogram resulting from HPLC of 5-MedCyd, irradiated at 254 nm in frozen aqueous solution.

Figure S6. The proton NMR spectrum of the "dh" product of 5-MedCyd.

Figure S7. The UV spectra of three 5-MedCyd photoproducts with either tentatively assigned or unassigned structures.

Figure S8. A possible mechanism leading to formation of the "spore" photoproducts of Thd.

Table S1. Correlations between compound designations (*e.g.* PH_{11}), structural assignments (*e.g.* **IIc**) and compound type (*e.g.* Thd (α-4) conjugate).

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. Fisher, G. J. and H. E. Johns (1976) Pyrimidine photodimers. In *Photochemistry and Photobiology of Nucleic Acids*, Vol. I (Edited by S. Y. Wang), pp. 225–294. Academic Press, New York.
2. Fisher, G. J. and H. E. Johns (1976) Pyrimidine photohydrates. In *Photochemistry and Photobiology of Nucleic Acids*, Vol. I (Edited by S. Y. Wang), pp. 169–224. Academic Press, New York.
3. Wang, S. Y. (1976) Pyrimidine bimolecular photoproducts. In *Photochemistry and Photobiology of Nucleic Acids*, Vol. I (Edited by S. Y. Wang), pp. 295–356. Academic Press, New York.
4. Cadet, J. and P. Vigny (1990) The photochemistry of nucleic acids. In *Bioorganic Photochemistry*, Vol. I (Edited by H. Morrison), pp. 1–272. John Wiley and Sons, New York.
5. Davies, R. J. H. (1995) Ultraviolet radiation damage in DNA. *Biochem. Soc. Trans.* **23**, 407–418.
6. 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.
7. 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.
8. 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.
9. Cadet, J., E. Sage and T. Douki (2005) Ultraviolet radiation-mediated damage to cellular DNA. *Mutat. Res.* **571**, 3–17.
10. Friedberg, E. C., G. C. Walker, W. Seide, R. D. Wood, R. A. Schultz and T. Ellenberger (2005) *DNA Repair and Mutagenesis*, 2nd edn. ASM Press, Washington, DC.
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. I (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. La Francois, C. J., Y. H. Jang, T. Cagin, W. A. Goddard III and L. C. Sowers (2000) Conformation and proton configuration of pyrimidine deoxynucleoside oxidation damage products in water. *Chem. Res. Toxicol.* **13**, 462–470.
15. Noll, S., M. Karlj, L. Suman, H. Stephan and I. Piantanida (2009) Synthesis of modified pyrimidine bases and positive impact of chemically reactive substituents on their *in vitro* antiproliferative activity. *Eur. J. Med. Chem.* **44**, 1172–1179.
16. Shaw, A. A. and J. Cadet (1990) Radical combination processes under the direct effects of gamma radiation on thymidine. *J. Chem. Soc. Perkin Trans. 1* **2**, 2063–2070.
17. Khoda, K., I. Kobayahi, K. Itano, S. Asano and Y. Kawazoe (1993) Synthesis and properties of *N*-aminopyrimidines. *Tetrahedron* **49**, 3947–3958.
18. Shetlar, M. D., V. J. Basus, A. M. Falick and A. Mujeeb (2004) The cyclobutane dimers of 5-methylcytosine and their deamination products. *Photochem. Photobiol. Sci.* **3**, 968–979.
19. Shetlar, M. D. and J. Chung (2011) The (5-4) and (6-4) adducts of 1-methylthymine and their Dewar valence isomers. *Photochem. Photobiol.* **87**, 802–817.
20. Stepien, E., R. Lisewski and K. L. Wierzchowski (1973) Cyclobutane dimers of 1-methylthymine: isolation, identification and properties. *Acta Biochim. Pol.* **20**, 297–311.
21. Martin, J. S. and A. R. Quirt (1971) NUMMRIT. *J. Magn. Reson.* **5**, 318.
22. Franklin, W. A., P. W. Doetsch and W. A. Haseltine (1985) Structural determination of the ultraviolet light-induced thymine-cytosine pyrimidine-pyrimidone (6-4) photoproduct. *Nucleic Acids Res.* **13**, 2317–2325.
23. Douki, T., G. Laporte and J. Cadet (2003) Inter-strand photoproducts are produced in high yield with A-DNA exposed to UVC radiation. *Nucleic Acids Res.* **31**, 3134–3142.
24. Su, D. G., H. Fang, M. Gross and J.-S. A. Taylor (2009) Photocrosslinking of human telomeric G-quadruplex loops by *anti* cyclobutane thymine dimer formation. *Proc. Natl Acad. Sci. USA* **106**, 12861–12866.