RNA Photolabeling Mechanistic Studies: X-ray Crystal Structure of the Photoproduct Formed between 4-Thiothymidine and Adenosine upon Near UV Irradiation

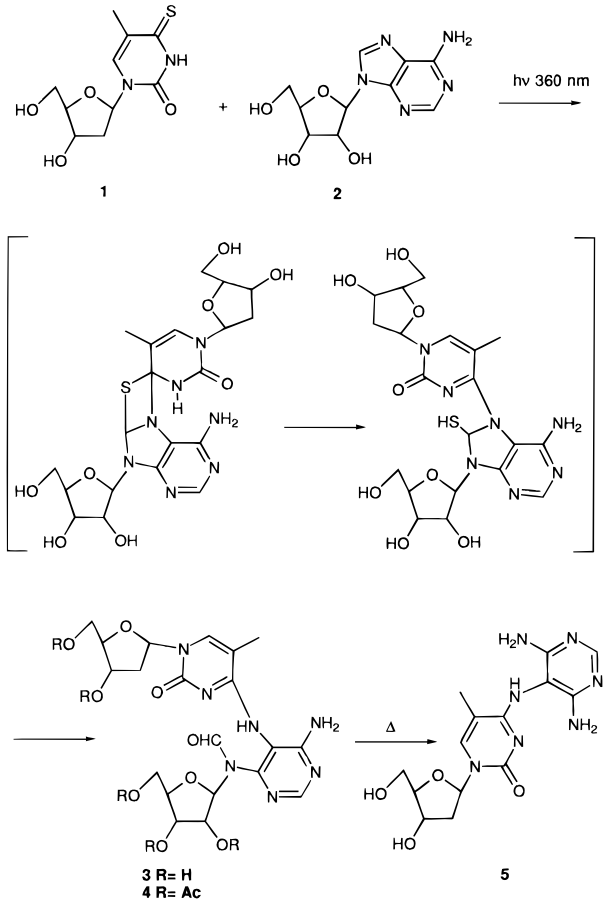
Carole Saintomé,[†,‡] Pascale Clivio,[†] Alain Favre,[‡] Jean-Louis Fourrey,*[,†] and Claude Riche[†]

Institut de Chimie des Substances Naturelles CNRS, 91198 Gif-sur-Yvette Cedex, France Institut Jacques Monod, CNRS-Université Paris VII 2 Place Jussieu, 75251 Paris Cedex O5, France

Received April 22, 1996

In recent years, reliable photolabeling methods have been developed to probe RNA tertiary structure in solution.[1] One of these exploits the remarkable photochemical properties of the sulfur analogs of the current nucleic acid bases 4-thiouracil, 4-thiothymine, and 6-mercaptopurine to gain informative data on tertiary interactions within such biomolecules.[2] Indeed, the latter bases in which an oxygen has been replaced by sulfur are stable in the dark and can be selectively photoactivated to give highly reactive excited states manifesting the capacity to undergo covalent bonding with any nucleic acid base.[3] Hence, when a thio-substituted nucleobase is introduced either ramdomly or at a defined position in such a system (by application of appropriate enzymatic[2b] or chemical[4] methods), subsequent irradiation usually results in the formation of informative cross-links. Obviously, only residues located at bonding distance with the thio-modified nucleobase can be involved in the cross-links which can be mapped by suitable sequencing procedures.[5] Interestingly, the spatial relationships which are evidenced in this manner can serve as constraints for the reconstruction of the three-dimensional structure of the system and the analysis of its conformational flexibility by molecular modeling.[6] Furthermore, the respective orientation of the two residues leading to a cross-link can be more precisely defined when the corresponding photochemical pathways leading to the products have been elucidated. In general, these pathways can be reasonably deduced from the structural analysis of the photoproducts. So far, detailed structural analysis of photoproducts resulting from thio-substituted nucleobases has been accomplished with pyrimidine only.[7] As photo-cross-links involving purine residues are frequently encountered in the analysis of RNA folding,[5] we describe here the X-ray crystal structure of the unique photoproduct[8] which was formed upon irradiation

Scheme 1. Photochemical Reaction between 4-Thiothymidine (**1**) and Adenosine (**2**)



(*ca.* 360 nm, inert atmosphere) of an equimolecular mixture (8.6 mM) of 4-thiothymidine (**1)** and adenosine (**2)** in water.

The reaction, monitored by the disappearance on the UV spectrum of the thiocarbonyl absorption maximum of **1** at 335 nm, yielded **3** (mp 167−168 °C) in 90% yield after purification (based on reacted **2** after *ca.* 70% consumption) (Scheme 1).[9] Structure **3** was supported by the spectroscopic data. Exact mass measurement of the quasimolecular ion (*m/z* 532.1790, M + Na^+) in the positive mode FAB mass spectrum indicated that **3** resulted from the condensation of **1** with **2** and replacement of the sulfur atom by an oxygen atom. The UV absorption peak of **3** (λ_{max} 293 nm, ϵ 13 840 mol^{-1} cm^{-1} dm^3, H_2O, pH 7) was in good agreement with that of a 5-methylcytosine. Key NMR arguments for structure **3** are the following: opening of the imidazole ring of adenosine to give a 6-*N*-formyl-4,5,6-triaminopyrimidine is supported by the presence on the HMBC spectrum of the penta-*O*-acetyl derivative **4**,[10] of a methine carbon at 164.2 ppm that correlates with H1′ of ribose. This long-range C−H coupling (^3J) allowed this carbon to be

† Institut de Chimie des Substances Naturelles.

‡ Institut Jacques Monod.

(1) (a) Branch, A. D.; Benenfeld, B. J.; Paul, C. P.; Robertson H. D. *Methods Enzymol.* **1989**, *180*, 418−442. (b) Jaeger, J. A.; Santa Lucia, J., Jr.; Tinoco, I., Jr. *Annu. Rev. Biochem.* **1993**, *62*, 255−287.

(2) (a) Favre, A. In *Bioorganic Photochemistry: Photochemistry and the Nucleic Acids*; Morrison, H., Ed.; J. Wiley & Sons: New York, 1990; Vol. 1, pp 317−378. (b) Sontheimer, E. J. *Mol. Biol. Rep.* **1994**, *20*, 35−44. (c) Favre, A.; Fourrey, J.-L. *Acc. Chem. Res.* **1995**, *28*, 375−382.

(3) Favre, A.; Lemaigre Dubreuil, Y.; Fourrey, J.-L. *New J. Chem.* **1991**, *15*, 593−599.

(4) (a) Beaucage, S.; Iyer, R. P. *Tetrahedron* **1993**, *49*, 6123−6194. (b) Beaucage, S.; Iyer, R. P. *Ibid.* **1992**, *48*, 2223−2311.

(5) (a) Wollenzien, P.; Expert-Bezançon, A.; Favre, A. *Biochemistry* **1991**, *30*, 1788−1795. (b) Lemaigre Dubreuil, Y.; Expert-Bezançon, A.; Favre, A. *Nucleic Acids Res.* **1991**, *19*, 3653−3660. (c) Dos Santos, V. D.; Vianna, A.; Fourrey, J.-L.; Favre, A. *Nucleic Acids Res.* **1993**, *21*, 201−207. (d) Dos Santos, V. D.; Fourrey, J.-L.; Favre, A. *Biochem. Biophys. Res. Commun.* **1993**, *190*, 377−385. (e) Sontheimer, E. J.; Steitz, J. A. *Science* **1993**, *262*, 1989−1996. (f) Woisard, A.; Fourrey, J.-L.; Favre, A. *J. Mol. Biol.* **1994**, *239*, 366−370. (g) Rinke-Appel, J.; Jünke, N.; Brimacombe, R.; Lavrik, I.; Dokudovskaya, S.; Dontsova, O.; Bogdanov, A. *Nucleic Acids Res.* **1994**, *22*, 3018−3025. (h) Bravo, C.; Lescure, F.; Laugâa, P.; Fourrey, J.-L.; Favre, A. *Nucleic Acids Res.* **1996**, *24*, 1351−1359. (i) Saintomé, C. Unpublished results.

(6) Laugâa, P.; Woisard, A.; Fourrey, J.-L.; Favre, A. *C. R. Acad. Sci., Life Sci.* **1995**, *318*, 307−313.

(7) (a) Fourrey, J.-L.; Gasche, J.; Fontaine, C.; Guittet, E.; Favre, A. *J. Chem. Soc., Chem. Commun.* **1989**, 1334−1336. (b) Clivio, P.; Fourrey, J.-L.; Gasche, J.; Favre, A. *J. Am. Chem. Soc.* **1991**, *113*, 5481−5483. (c) Woisard, A.; Favre, A.; Clivio, P.; Fourrey, J.-L. *J. Am. Chem. Soc.* **1992**, *114*, 10072−10074. (d) Clivio, P.; Favre, A.; Fontaine, C.; Fourrey, J.-L.; Gasche, J.; Guittet, E.; Laugâa, P. *Tetrahedron* **1992**, *48*, 1605−1616. (e) Clivio, P.; Fourrey, J.-L.; Gasche, J.; Favre, A. *Tetrahedron Lett.* **1992**, *33*, 1615−1618. (f) Blazek, E. R.; Alderfer, J. L.; Tabaczynski, W. A.; Stamoudis, V. C.; Churchill, M. E.; Peak, J. G.; Peak, M. J. *Photochem. Photobiol.* **1993**, *57*, 255−265. (g) Clivio, P.; Fourrey, J.-L.; Szabo, T.; Stawinski, J. *J. Org. Chem.* **1994**, *59*, 7273−7283.

(8) From inspection of the ^1H NMR spectrum of the crude irradiated mixture.

(9) Purification was achieved on a medium-pressure (400 mbar) RP 18 column using a gradient of acetonitrile in water. Nucleosides **1** and **2** were recovered in 24% and 33% yield, respectively.

S0002-7863(96)01329-7 CCC: $12.00

identified as being originally C8 of adenine. Its deshielded chemical shift ($\delta\Delta$ 22 ppm), compared to the corresponding one of the parent adenosine (δ 142 ppm), is typical of a formamido group. Analysis of the NMR chemical shifts of the carbons of the former thymine base[11] favors substitution of carbon C4 by a nitrogen (the chemical shifts of C4, C5, and C6 are very similar to the corresponding ones of 5-methylcytosine[12]) rather than by a carbon as in the case of the pyrimidinone of the [6−4] photoproduct of TpT.[13] Upon standing in aqueous solution, photoproduct **3** was slowly transformed into **5**, this transformation being accelerated and brought to completion upon mild heating.[14] The positive FAB mass spectrum of **5** exhibited a pseudomolecular ion at *m/z* 350 ($M + H^+$), suggesting the loss of the ribose unit of **3** and deformylation. This was evident from the examination of the ^{1}H NMR spectrum of **5**, recorded in a D_2O/DMSO-d_6 mixture, which showed only the signals of the deoxyribose moiety together with one methyl group (δ 2.08 ppm) and two vinylic protons (δ 7.86 and 7.77 ppm). The most significant difference between the ^{13}C NMR spectra of **5** and **3** was the shielding of carbon C5 ($\delta\Delta$ *ca.* 17 ppm) of the 4,5,6-triaminopyrimidine unit. This was in accordance with the replacement of the formamido group at position C6 by an amino group, inducing shielding in the ortho position. Interestingly, all the carbon chemical shifts which derived from the adenine part are very close to those of 4,6-diamino-5-(formylamino)pyrimidine (FapyAde),[15] in agreement with N-5-substitution. This interpretation is in accordance with the data derived from the HMBC spectrum recorded in DMSO-d_6. Indeed, this spectrum showed an exchangeable signal at 7.92 ppm integrating for one proton and correlating with C2 (^4J), C4 (^2J), and C5 (^3J) of the 5-methylcytosine unit and C5 (^2J), magnetically equivalent C4 and C6 (^3J) of the 4,5,6-triaminopyrimidine unit of **5**, respectively.

The structure of photoproduct **3** (Figure 1) was definitively established by X-ray diffraction studies carried out on crystals grown from a concentrated aqueous solution (1 M).[16] In this structure, the deoxyribose unit shows a C4′-endo/C3′-exo conformation (South-type puckered conformation) with a pseudorotation phase angle $P = 218.3°$,[17] and a maximum puckering amplitude, $\tau_m = 34.0°$, as in the case of 5-methyl-2′-deoxycytidine.[18] The 5-methylcytosine base retains an anti conformation ($\chi = -178.6°$).[19] On the other hand, the ribose unit adopts a South-type puckered geometry (C2′-endo, $P = 160.3°$ and $\tau_m = 40.8°$) which differs from the one found in the crystalline structure of adenosine (C3′-endo).[20] Finally, in **3** the conformation about the ribose glycosidic bond ($\chi = -108.5°$) is anti as in the case of the parent adenosine. Molecules, in the crystal structure, are linked in a three-dimensional lattice through eight independent intermolecular hydrogen bonds. In particular, this lattice implicates two water molecules which bridge four molecules of photoproduct.

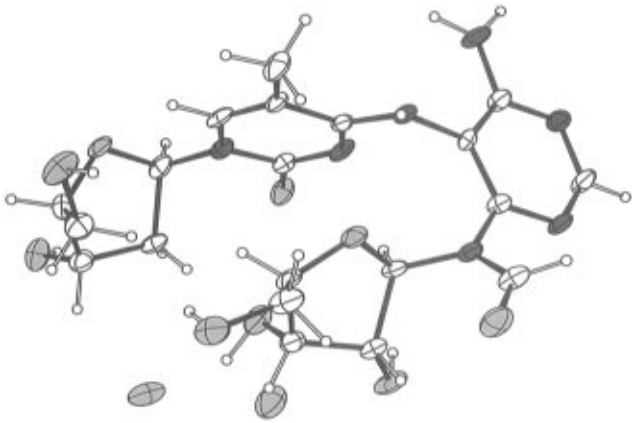


Figure 1. ORTEP drawing of **3**. Ellipsoids are at the 30% level. The two water molecules of crystallization are depicted.

The most probable mechanism to account for the formation of **3** (Scheme 1) is a [2 + 2] cycloaddition in which the thiocarbonyl of **1** adds across the N7−C8 double bond of **2**. This should lead to an unstable four-membered heterocyclic intermediate (thiaazetidine) prone to undergo ring opening to give an imidazoline, yielding **3** after further hydrolysis.[21]

In summary, **3** is the first characterized purine−pyrimidine photoadduct, arising by saturating the N7−C8 double bond of adenine, by means of a remarkably high yielding bimolecular reaction. Indeed, a photoaddition reaction involving the N7−C8 double bond of adenine has already been reported to occur in the case of d(ApA) between the two adjacent adenine bases.[22] But, to the best of our knowledge, such a N7−C8 regioselective reaction has never been observed to be so efficient with a pyrimidine unit.[23] The herein reported X-ray structural data of **3** might hopefully provide important information to be used for the reconstitution of tertiary interactions based on photolabeling experiments. Finally, the facile cleavage of the N-glycosidic bond of the parent adenosine might be potentially used for designing artificial specific endonucleases.

Supporting Information Available: X-ray crystallographic data, ^{1}H NMR spectra of **3**, and ^{13}C NMR chemical shifts of **3**, **4**, and **5** (10 pages). See any current masthead page for ordering and Internet access instructions.

JA961329G

(10) Attempts to obtain correlations between C and H of the *N*-6-formyl-*N*-6-ribosyl-4,5,6-triaminopyrimidine part on the HMBC spectrum of **3**, recorded either in D_2O or in DMSO-d_6 at room temperature or below, failed due to line broadening. Although recording at 90 °C in D_2O led to a well-resolved ^{1}H NMR spectrum of **3** (see the Supporting Information), the transformation of **3** to **5** precluded a HMBC experiment at this temperature. Finally, this was circumvented by inspection of the HMBC spectrum of **4** (*m/z* 742, $M + Na^+$) in CD_3OD whose ^{13}C NMR chemical shifts of the aglycon nuclei are almost identical to those of **3**.

(11) Attributed from the HMBC spectrum.

(12) Shaw, A. A.; Shetlar, M. D. *J. Am. Chem. Soc.* **1990**, *112*, 7736−7742.

(13) Rycyna, R. E.; Alderfer, J. L. *Nucleic Acids Res.* **1985**, *13*, 5949−5863.

(14) Purification was performed by HPLC on a semipreparative NOVA-PAK C18, 6μ radial pak cartridge, using a gradient of acetonitrile in 0.05 M sodium acetate.

(15) Raoul, S.; Bardet, M.; Cadet, J. *Chem. Res. Toxicol.* **1995**, *8*, 924−933.

(16) Crystal data: $C_{20}H_{27}N_7O_9 \cdot 2H_2O$, monoclinic, space group $P2_1$, colorless, $a = 7.381$ (7) Å, $b = 21.298$ (12) Å, $c = 8.414$ (1) Å, $\beta = 108.92$ (8)°, $Z = 2$, $R = 0.067$, $R_w = 0.086$, goodness of fit 1.04, number of observed reflexions 1945.

(17) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1972**, *94*, 8205−8212.

(18) Sato, T. *Acta Crystallogr.* **1988**, *C44*, 870−872.

(19) Saenger, W. In *Principles of Nucleic Acid Structure*; Cantor, C. R., Ed.; Springer-Verlag: New York, 1984.

(20) Lai, T. F.; Marsh, R. E. *Acta Crystallogr.* **1972**, *B28*, 1982−1989.

(21) The preferred N7−C8 imidazole ring fission of adenosine leading to a an *N*-9-formyl derivative is probably governed by the electron deficient nature of the N-7-substituent. This is inferred from analogous ring opening of 9-alkylated adenines undergoing N-7-acylation (Altman, J.; Ben-Ishai, D. *J. Heterocycl. Chem.* **1968**, *5*, 679−682. Leonard, N. J.; McDonald, J. J.; Henderson, R. E. L.; Reichmann, M. E. *Biochemistry* **1971**, *10*, 3335−3342).

(22) Kumar, S.; Joshi, P. C.; Sharma, N. D.; Bose, S. N.; Davies, R. J. H.; Takeda, N.; McCloskey, J. A. *Nucleic Acids Res.* **1991**, *19*, 2841−2847.

(23) For a precedent case see: Paszyc, S.; Skalski, B.; Wenska, G. *Tetrahedron Lett.* **1976**, 449−450.