CHEMISTRY
A EUROPEAN JOURNAL

FULL PAPER

DOI: 10.1002/chem.201200816

Expanding the Horizon of the Thymine Isostere Biochemistry: Unique Cyclobutane Dimers Formed by Photoreaction between a Thymine and a Toluene Residue in the Dinucleotide Framework

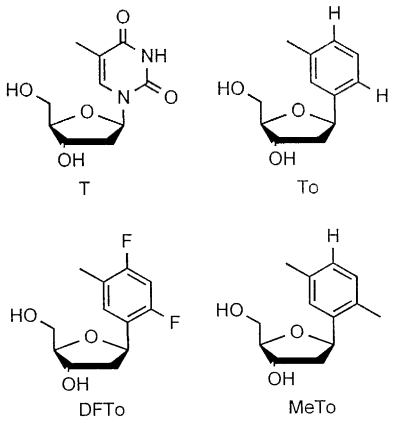
Degang Liu,[a] Yan Zhou,[a] Jingzhi Pu,*[a] and Lei Li*[a, b]

Abstract: Substituted toluenyl groups are considered as close isosteres of the thymine residue. They can be recognized by DNA polymerases as if they were thymine. These toluene derivatives are generally inert toward radical additions, including the [2+2] photocycloadditions, due to the stable structure of the aromatic ring and are usually used as solvents for radical reactions. Surprisingly, after incorporating toluene into the dinucleotide framework, we found that the UV excited thymine residue readily dimerizes with the toluenyl moiety through a [2+2] photoaddition reaction. Furthermore, the reaction site on the toluenyl moiety is not the C5=C6 bond, as commonly observed in cyclobutane pyrimidine dimers, but the C4=C5 or C3=C4 instead. Such a reaction pattern suggests that in the stacked structure, it is one of these bonds, not the C5=C6, that is close to the thymine C5=C6 bond. A similar structural feature is found in DNA duplex with a thymine replaced by a 2,4-difluorotoluene. Our results argue that although the substituted toluenyl moieties closely mimic the size and shape of the thymine residue, their more hydrophobic nature determines that they stack on DNA bases differently from the natural thymine residue and likely cause local conformational changes in duplex DNA.

Keywords: cyclobutane pyrimidine dimers · DNA · photochemistry · stereochemistry · thymine

Introduction

Watson–Crick hydrogen bonding formed between nucleobases AT or GC is widely regarded as one of the basic interactions in life. It enables the formation of the double-stranded oligonucleotide and becomes the foundation for genetic information storage, direct replication, transcription and translation. Kool et al. replaced thymine with a non-polar 2,4-difluorotoluene residue (DFTo, Scheme 1), which possesses a nearly identical size and shape to thymine,[1] and found that DNA polymerases readily incorporate it into DNA as if it were thymine.[1c,2] This result argues that DNA polymerases are governed largely by the steric effects, not the hydrogen bonds between the corresponding nucleobases in making DNA copies.[2c,3] Two recent X-ray crystallographic studies, however, indicate that DFTo still forms hydrogen bonds with adenine,[4] even though the bonding interactions might not be as strong as those between the natural A and T.



Scheme 1. Thymidine and its steric analogues.

Thymine (T) is known for its rich photochemical reactivity and is the most UV sensitive nucleobase.[5] In B-form DNA, thymine dimerizes with a neighboring T; this results in the cyclobutane pyrimidine dimer (CPD, or T<>T) as the major photoproduct.[5] In contrast, toluene is suggested to possess different molecular orbital diagrams from thymine and is rather inert under UV irradiation.[5,6] This is why it is commonly used as the solvent for radical reactions, including many [2+2] photo-cycloadditions.[7] T<>T is considered to form through a [2+2] photoaddition at the singlet excited state of thymine,[8] as implied by its ultrafast

[a] Dr. D. Liu, Dr. Y. Zhou, Prof. Dr. J. Pu, Prof. Dr. L. Li
Department of Chemistry and Chemical Biology
Indiana University–Purdue University Indianapolis (IUPUI)
402 N. Blackford St., Indianapolis, IN 46202 (USA)
Fax: (+1) 317-274-4701
E-mail: lilei@iupui.edu
jpu@iupui.edu

[b] Prof. Dr. L. Li
Department of Biochemistry and Molecular Biology
Indiana University School of Medicine
635 Barnhill Drive, Indianapolis, Indiana 46202 (USA)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.201200816.

formation rate.[9] Our previous research showed that the excited 5′-T behaves like a di-radical and conducts the H atom abstraction during the spore photoproduct (SP) formation.[10] We thus wondered if the excited thymine can induce the [2+2] cycloaddition to a rather inert toluenyl moiety incorporated into the DNA framework.

Toluene is approximately 0.15 Å smaller than thymine. Like DFTo, it is recognized by DNA polymerase as a thymine. The transcription efficiency, however, decreases by approximately 2–3 orders of magnitude due to its smaller size.[11] Compared with DFTo, the synthesis of 2-deoxy-1-(3-toluenyl)-β-D-ribofuranose (To, Scheme 1) is much easier.[12] We thus chose To instead of DFTo for our proof-of-concept photochemical studies. Dinucleotide TpT mimics the photochemistry occurring in duplex DNA due to the similar thymine stacking interaction;[5,13] this results in the *cis–syn* T<>T as the dominating photoproduct. As the toluenyl moiety exhibits an even stronger stacking interaction with DNA bases than thymine due to the enhanced desolvation effect,[14] the dinucleotide analogue TpTo (**1**; Scheme 2) is likely to adopt a stacked structure in aqueous solution. We wondered if such a TpTo compound mimics the rich TpT photochemistry.

Results and Discussion

After radiating **1** (1 mM) under 254 nm UV light in a frozen 1:1 glycerol/water glass in liquid N_2 for 15 min, analysis of the reaction by HPLC revealed formation of two products **2** and **3** (Figure 1a). The reaction yields were found to be 7.8% for **2** and 4.6% for **3**. ESI-MS analysis displayed a $[M+H]^+$ signal of 513.2 for both **2** and **3**, which is the same as that of **1**. Such a result is typically observed in the thymine dimerization reaction,[5] and suggests that a similar reaction might have occurred in the irradiation of **1**. Both **2** and **3** were also isolated in the TpTo photoreaction in H_2O at 298 K (Figure S1 in the Supporting Information).[12b] The yields were 3.2 and 3.3% for **2** and **3**, respectively. These lower yields are likely due to the enhanced thermal motion at ambient temperature, which results in fewer TpTo molecules with the stacked reactive structure.

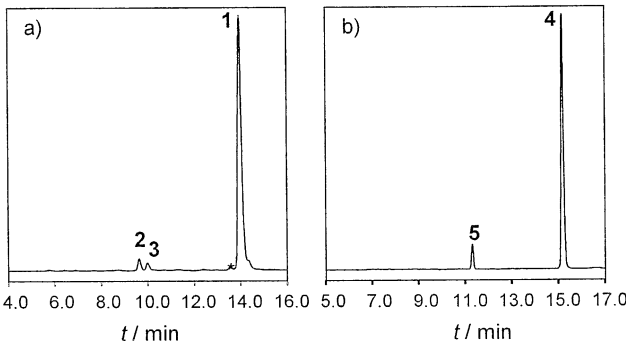
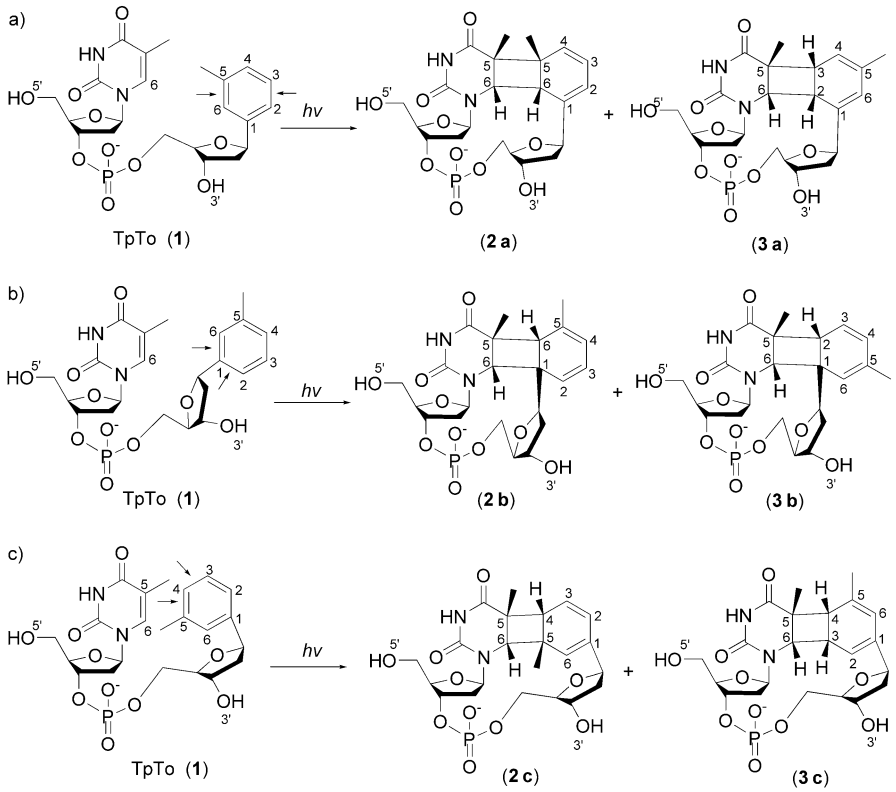


Figure 1. a) The HPLC chromatograph of the CPD analogues **2** and **3** formed from the irradiation of TpTo (**1**) at 77 K. b) The HPLC chromatograph of the CPD analogue **5** formed from the TpMeTo (**4**) photoreaction at 77 K (* denotes an impurity in 1). No other product with the same mass was detected above the basal level in either reaction.



Scheme 2. Photoreaction of **1** could yield six possible *ortho* CPD regioisomers (note: all these CPDs are drawn as the *cis–syn* isomers. The *cis–syn* conformations for products **2** and **3** are later confirmed by NOE spectroscopy, as discussed in the main text).

Compared with the TpT photoreaction, where the C5=C6 is the reaction site for T<>T formation, all six C=C bonds of the toluenyl group can in theory react with the C5=C6 of T should the right conformation be adopted; this results in six possible *ortho* T<>To regioisomers (Scheme 2).[6b] Among these products, the symmetric structure of the

phenyl ring determines that reflected by the C1C4 axis, the six possible isomers can be divided into three groups. It is logical to assume that only one of the three reactions in Scheme 2 occurred in irradiation of **1**. The fact that only two products, **2** and **3**, were isolated further supports this assumption. Which reaction has occurred here?

The ^{1}H NMR spectrum of **2** exhibited two methyl singlet signals at 1.35 and 1.49 ppm; these correspond to CH_3–To and CH_3–T, respectively. These two signals were distinguished as the later signal exhibits a long-range ^{1}H–^{13}C coupling interaction with the thymine C4=O carbon in HMBC NMR spectroscopy.[12b] Three ^{1}H NMR signals were observed in the low-field region; this suggests that three protons are associated with the double bonds. These signals include one singlet (s, 5.55 ppm), one doublet (d, 5.59 ppm) and one dd signal (5.36 ppm; Figure 2a). Among the six possible structures, only **3b** and **2c** support such a coupling pattern. In addition, H6–T exhibits a singlet peak at 3.81 ppm, whereas H4–To (**2c**) or H2–To (**3b**) exhibit a doublet signal at 2.58 ppm. The dd signal at 5.41 ppm was assigned to the H1′ on the thymine 2-deoxyribose.

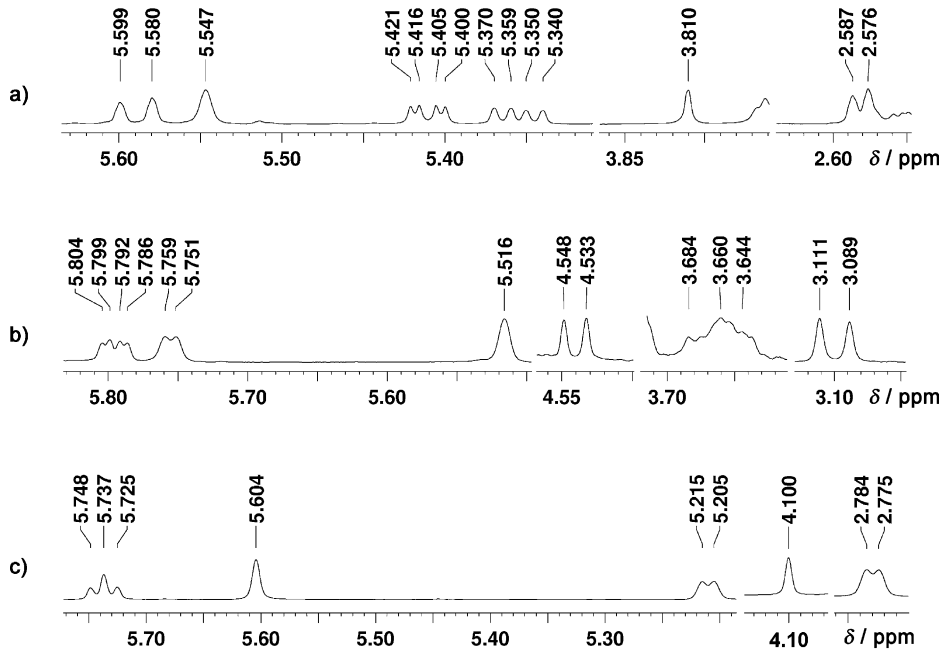


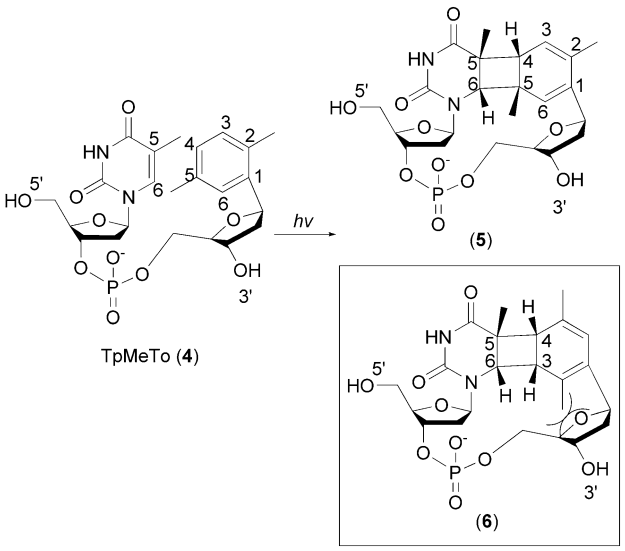
Figure 2. Zoom-in view of the 1H NMR signals associated with the protons on the To or MeTo ring as well as the protons on the cyclobutane ring in the resulting CPD photoproducts. a) Compound **2** in [D_6]DMSO; b) **3** in D_2O; c) **5** in D_2O. The NMR spectroscopy solvents were chosen to achieve the optimal signal resolution, especially for the protons associated with the unsaturated region on the To ring. Full NMR spectra for these CPD species can be found in the Supporting Information.

To distinguish which compound is the true product, we further examined the HMBC NMR spectrum of **2**. Three strong ^{1}H–^{13}C couplings were observed within the three-bond distance between the CH_3–To protons and the three neighboring carbons, respectively. Among these three carbons, one is in the unsaturated region and two in the saturated.[12b] This can only be fulfilled by **2c** as **3b** contains only two carbons associated with the C=C bonds, which are within the three-bond distance to the CH_3–To to enable a strong ^{1}H–^{13}C coupling interaction. Taken together, our data strongly suggest that **2c** is the TpTo photo-dimerization product **2**.

Two methyl signals were observed in the ^{1}H NMR spectra of **3**. The signal at 1.60 ppm was assigned to CH_3–T and that at 1.73 ppm to the CH_3–To. Such assignments were confirmed by the NOSEY spectrum as the CH_3–T showed coupling interaction with the two H atoms on the cyclobutane ring, whereas the CH_3–To did not exhibit any coupling to the ring protons. Two ^{1}H NMR signals were obtained in the low magnetic field (Figure 2b), a doublet at 5.76 ppm and a singlet at 5.52 ppm; this suggests that two protons are associated with the C=C bonds. Among the six possible products in Scheme 2, only **3a** and **3c** are considered possible for such an NMR pattern. The **3a** structure suggests that a dd signal should be observed for the H2–To, however, the signal observed at 3.66 ppm exhibited a multiple splitting pattern, which can only be caused by the H3–To in **3c**. Such a structural assignment is further supported by the COSY spectrum.[12b] The doublet at 3.10 ppm was assigned to H4–To as it exhibited a strong coupling interaction with both methyl groups in the NOSEY spectrum.[12b] The doublet at 4.54 ppm was ascribed to the H6–T. The dd signal at 5.80 ppm was assigned to the H1′ on the 2-deoxyribose of the thymine residue. Compared with those in **2**, some of the ^{1}H NMR signals of **3** shifted toward the low magnetic field; these shifts are likely due to the change of the NMR solvent. The fact that **2** and **3** correspond to products **2c** and **3c**, respectively, suggests that it is the reaction C, not A, which is commonly observed in pyrimidine dimerization reaction, that occurred in the TpTo photoreaction.

Such CPD structures are very rare in pyrimidine photoreactions. To confirm the structural assignments above, we introduced a methyl group at the C2 position of the toluenyl moiety, to yield a dinucleotide analogue TpMeTo (**4**) and studied its photoreaction. MeTo (Scheme 1) was chosen for multiple purposes: 1) to decrease the number of H atoms on the toluenyl ring to simplify the NMR spectroscopy analysis of the resulting T<>MeTo complex. 2) To test if the special CPD configuration identified above is induced by the unique interaction between the

T and the To rings. The new methyl moiety introduced locates at the periphery of the phenyl ring and is expected to have little effect on the ring–ring interaction. 3) To simplify the reaction pattern. As the 2-deoxyribose is connected to the C1 position of the To ring, a methyl group at C2 will not affect the formation of **5**, the analogue of **2** (Scheme 3),



Scheme 3. Photoreaction of **4** only produces CPD analogue **5**. Formation of **6** is inhibited due to the steric hindrance caused by the 2-CH_3 moiety.

should reaction c in Scheme 2 occur in the irradiation of **4**. However, steric hindrance between the methyl group and the 2-deoxyribose should prevent the formation of **6**, the analogue of **3** (Scheme 3). Thus, only **5** is expected in the photoreaction of **4**. Again, the synthesis of MeTo is straightforward, which makes the preparation of **4** much easier.

Indeed, irradiation of **4** generated only one new product with a yield of 16.5% (Figure 1b). NMR spectroscopy confirmed that the product adopts the same structure as **5**. As shown in Figure 2, two ^{1}H signals were revealed at the low magnetic field in the NMR spectrum, a doublet at 5.21 ppm and a singlet at 5.60 ppm; these were assigned to the H3 and H6 atoms of the To group, respectively. The singlet at 4.10 ppm was assigned to the H6–T and the doublet at 2.78 ppm to the H4–To. Such assignments were further supported by the COSY/HSQC/HMBC spectroscopic analyses, the details for which are available in the Supporting Information. Again, the triplet signal at 5.74 ppm was due to the H1′ on the thymine 2-deoxyribose.[15] The fact that the same reaction pattern was obtained in irradiations of both **1** and **4** confirms that the stacking conformation between T and To is mainly determined by the ring–ring interaction; introduction of the methyl substitute has little effect to this interaction.

To confirm that the lack of formation of **6** during the UVC irradiation of **4** is due to the steric hindrance caused by the 2-methyl group on the toluenyl ring, stabilities of **5** and **6** were examined computationally by using density functional theory. These products were optimized by Gaussian 03[16] at the B3LYP/6-31+G(d,p) level[17] by using a TpDFTo conformation previously determined in duplex DNA at an intrahelical position[18] as the initial structural template. Additional models based on an extrahelical conformation found in a ternary complex of DNA–photolyase that contains a flipped CPD-analogue[19] were also constructed. Optimizations initiated from these models yield almost identical geometries for each compound except the orientation of 5′-OH on the thymine 2-deoxyribose due to the different initial orientation of the phosphate group in the two starting structures. Thus, our discussion below focuses on the results of the TpDFTo based models.

The computation shows that at the B3LYP/6-31+G(d,p) level, **5** is 5.47 kcal mol^{-1} more stable than **6** in the gas phase. The cyclobutane ring in **5** is very flat with the dihedral angle C5–C6–C5–C4 being 1.7°; this results in a configuration that allows the T ring and the MeTo ring to overlap very well (Figure 3a). In contrast, the corresponding dihedral angle (C5–C6–C3–C4) is 17.4° in **6**; this suggests that the cyclobutane ring is twisted and the interaction between the T and To rings is weakened (Figure 3b). The ring deformation is due to the steric repulsion from the 2-CH_3 of MeTo; this leads the cyclobutane ring to adopt a twisted structure, making it too unstable to form. A similar steric repulsion is expected in **4** should **6** form, making few molecules of **4** to adopt the conformation favoring **6** formation. This subsequently causes the conformers favoring **5** to increase, as implied by the improved yield of **5** relative to those of **2** and **3**.

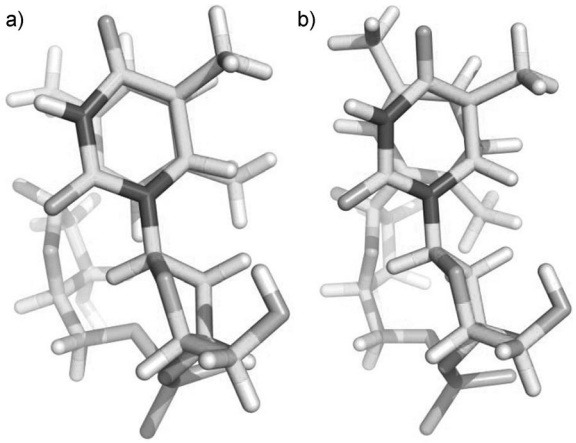


Figure 3. The stacking conformation of the T and MeTo rings in: a) **5**, and b) **6** the structures of which were computationally optimized at the B3LYP/6-31+G(d,p) level by using TpDFTo conformation previously determined in duplex DNA as the initial template. The poor stacking interaction in **6** is likely due to the steric hindrance of the 2-CH_3 moiety on the toluenyl ring.

All these CPD products adopt the *cis–syn* configuration, as revealed by their 2D-NOESY spectra.[12b] As illustrated by the spectrum of **2** in Figure 4, the H6–T and H4–To protons were found to associate with both methyl groups; this

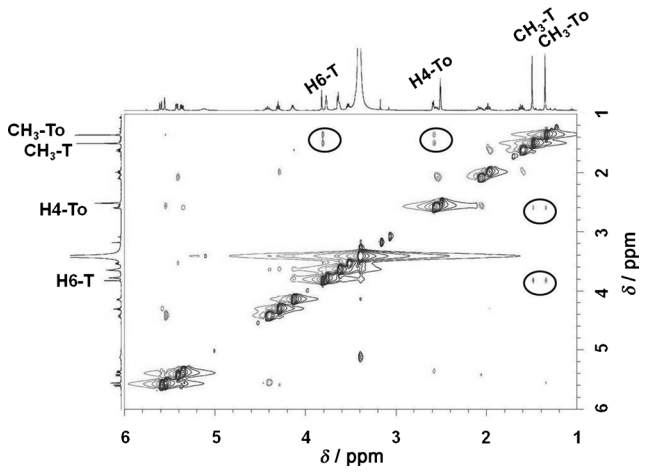


Figure 4. Zoomed-in view of the ROESY spectra of **2** in [D_6]DMSO. The protons associated with the cyclobutane ring are labeled and the interactions among these protons are highlighted in circles.

suggests that these protons and the methyl groups locate on the same side of the cyclobutane plane. This observation also provides evidence that the two CH_3 moieties and the two H protons are at the same side of the plane. The *syn* conformation of these four groups dictates the thymine ring to be *cis* to the toluenyl plane. Similar CH_3–H6 interactions were observed in the NOESY spectrum of *cis–syn* T<>T.[20] In contrast, the CH_3 in the *trans–syn* T<>T only interacts with the H6 proton on the same thymine.[20,21] Taken together, the NMR spectroscopy data prove **2** to be a CPD analogue with a *cis–syn* configuration.

It is worth pointing out that in dinucleotide TpT photochemistry, two *cis–syn* and two *trans–syn* diastereomers, are possible in theory.[20b] The first *cis–syn* diastereomer adopts a structure similar to **2a** (Scheme 2); its formation requires both thymidines to adopt the *anti N*-glycosidic conformations. The second *cis–syn* isomer, the four chiral centers on which the cyclobutane ring adopts opposite stereogeometries, requires the *syn–syn N*-glycosidic conformations instead and is never observed in TpT photochemistry.[20b,22] X-ray structure reveals the DFTo nucleoside to adopt an *anti* glycosidic conformation,[23] the conformation remains once it is incorporated into the duplex framework as revealed by the NMR spectroscopy studies.[18] These observations suggest that the replacement of the natural C–N glycosidic bond by a C–C bond results in little conformational change. For the To residue, the *syn* and *anti* conformation makes little difference energetically due to the lack of substitution at the 2-position. The 2-MeTo is likely to resemble DFTo and adopt the *anti* glycosidic conformation. As the *anti* conformation is dominant in thymidine,[24] the *syn–syn N*-glycosidic conformations are not expected in either compound **1** or **4**. Consequently, although our NMR spectroscopy data do not allow us to completely rule out the possible formation of type II *cis–syn* diastereomer, the resulting CPD species **2**, **3** and **5** are most likely to possess the conformations similar to the type I *cis–syn* species observed in the TpT photochemistry.

The photo-formation of a CPD analogue between a thymine and a toluene residue is surprising. The substituted toluene anions have been extensively studied as computational models to facilitate the understanding of the pyrimidine electronic properties. Neutral molecules like toluene, however, were suggested to possess rather different molecular orbital diagrams due to their higher symmetries, less π electrons, fewer nonbonding electronic transitions as well as lack of tautomeric structure,[5,6] making them rather inert under UV radiation. That is why toluene is often selected as the solvent for photo-cycloaddition reactions.[7b,c] Formations of CPD analogues **2**, **3** and **5** must be due to the photochemically excited thymine residue. In previous mechanistic studies of the spore photoproduct formation, we proved that the excited thymine C5=C6 behaves like a di-radical and conducts typical radical reactions, such as the H atom abstraction.[10] The formation of T<>To is likely due to the radical mediated [2+2] cycloaddition, and further illustrates the di-radical nature of the excited thymine C5=C6 bond.

The fact that toluene is often used as a solvent for photoreaction without causing obvious side reactions[7] suggests that the stacked structure in the dinucleotide TpTo framework must play a key role in minimizing the entropic effect and providing the required template to enable the occurrence of the [2+2] photoaddition reaction. As revealed by Kool et al., the desolvation effect is largely responsible for the strong stacking association of an aromatic ring to the DNA bases.[14] Disturbing this stacking interaction by conducting the TpTo photoreaction in methanol or dry film totally abolished the T<>To formation. Thymine photo-dimerization occurs on the picosecond timescale.[9] This ultrafast nature determines that the original DNA conformation controls the outcome of the photoreaction.[9,13,25] The different CPDs generated by TpT and TpTo photoreactions suggest that they are likely to adopt different stacking structures. In TpT, the two C5=C6 bonds must be close to each other and dimerize to form the CPD commonly observed in pyrimidine photochemistry. In contrast, although in theory all six bonds of the To ring are reactive, that the cross-linking reaction involves the C4=C5 in the formation of **2** implies that this bond is in the vicinity of the thymine C5=C6 bond.

To test this hypothesis, we turned to the duplex DNAs that contain either a TpT step or a TpDFTo sequence. It was suggested that the base stacking interaction is similar in single- and double-stranded DNA contexts; the DNA context does not affect the CPD formation.[9,26] A recent study further proves that the vertical base stacking, not the base-paring interaction, is the dominating force in stabilizing the double helical DNA.[27] Our DFT calculation also suggests that the TpTo conformation remains the same whether it is at intrahelical or extrahelical position. Although the structural information for the dinucleotides is not available, the stacking interaction revealed in duplex DNA could help explain the reaction pattern observed in dinucleotide photochemistry.

Indeed, analyses of the DFTo containing DNA structure indicate that the thymine C5=C6 bond is very close to the C4=C5 bond of the toluenyl moiety in duplex DNA. As revealed by an NMR spectroscopy study,[18] the C5=C6 bond of T aligns perfectly with the C4=C5, not the C5=C6 bond of DFTo, in a duplex DNA dodecamer.[18] In a recent X-ray structure, although the C5=C6 bond of T is no longer parallel with the C4=C5 of DFTo, the distance between these two bonds is still about 0.5 Å shorter than that between the two C5=C6 bonds.[4a] In contrast, the distance between the two thymine C5=C6 bonds is approximately 0.5 Å shorter than that between the C5=C6 of the 5′-T and C3−C4 of the 3′-T in the typical B-form DNA, as revealed by both NMR spectroscopy[28] and X-ray crystallographic studies.[29]

Such results support our findings in the TpTo photochemical studies, and suggest that the thymine substitution by DFTo might cause an additional twist to the DNA double-helical turns. This twist could have been largely overlooked in previous thymine isostere biochemical studies as attention was mainly focused on the maintenance of the duplex structure after the thymine substitution. Little structural difference was suggested after the substitution as compared with thymine and adenine in the same duplex sequence.[1c,18,30] Our results indicate that although the substituted toluene closely mimics the size and shape of thymine, its stronger hydrophobic nature can change the DNA local structure, which can be probed by the structure-sensitive DNA photochemistry.

The unique TpTo configuration discussed above clearly favors the formation of **2**. The To ring might flip and stack on the T ring using the other side of the aromatic plane. Such a conformation brings the C3=C4 of To to the neighborhood of the thymine C5=C6, producing **3** under UV irradiation. Although both TpTo conformers are possible, the lower yield of **3** (4.6% for **3** vs. 7.8% for **2**) suggests that the TpTo conformers favoring the formation of **2** outnumber those favoring that of **3** by twofold at 77 K. Such a difference in TpTo population is ascribed to the hydrophobic nature of the methyl group, as a potential Me–Me interaction is expected in the TpTo conformers before **2** forms. In contrast, both methyl groups are exposed to water in those favoring the formation of **3**.

The methyl–methyl interaction has been suggested to be the major stabilizing force in the TT steps of duplex DNA, which could contribute up to 15% of the total stacking energy.[31] Our data provide the first experimental evidence for the importance of this interaction. Such a stabilization effect is best reflected by reactions at low temperature. As the pyrimidine–pyrimidine stacking interaction was estimated to stabilize DNA by approximately 2 kcal mol^{-1},[32] the stabilization effect by the Me–Me interaction is negligible at 298 K, as reflected by the roughly equal yields of **2** and **3** in the solution reactions (Figure S1 in the Supporting Information).[12b] The sum of the yields for **2** and **3** roughly equals that of **5** (16.5%) for reactions at 77 K; this suggests that the overall populations in TpTo complexes favoring the CPD formation are similar.

It is generally accepted that the DNA conformation controls the outcome of the DNA photoreaction.[9,13,25] However, the exact thymine conformation and factors controlling that conformation to facilitate a specific thymine dimer formation are largely unclear. The DNA tertiary structure could impact the relative positions of the thymines, resulting in the formation of "unusual" CPD product(s). This conclusion can be illustrated by the formation of interstrand *cis–anti* cyclobutane dimer in a 14-mer duplex DNA[33] as well as the *cis–syn*/*cis–anti* thymine dimers formed between Ts located at different positions of G-quadruplex loops.[34] The conformation can also be affected by modifications on either 2-deoxyribose[13a] or nucleobase.[35] A recent study used locked nucleic acids (LNA) to constrain the sugar at the C3′-*endo* conformation, and obtained the *cis–syn* T<>T as the exclusive thymine photoproduct.[13a] Acetylation of the 4-NH_2 moiety in cytosine resulted in a *trans–syn* C<>T isomer with the *syn–anti* conformation in a highly stereoselective manner due to the steric hindrance induced by the acetyl modification.[35] Our report here further reveals that by simply changing the hydrophobicity of the bases, the DNA conformation can also be altered. The stacked TpTo structure is likely approximately 60° away from the TpT conformation; this results in the formation of the regular *cis–syn* T<>T. Such a conformational change alters the outcome of thymine photochemistry, as demonstrated by the formation of new CPD analogues **2**, **3**, and **5** reported here.

Conclusion

In summary, we have synthesized a dinucleotide TpTo and studied its photoreaction under 254 nm UV irradiation. Although toluene is largely inert under UV light, it dimerizes with a thymine residue through a [2+2] photoreaction probably due to the presence of the dinucleotide framework, which minimizes the entropic effect and provides the required reaction template. The TpTo photochemistry yields two *cis–syn* T<>To species as the major photoproducts. More importantly, different from the normal CPDs formed between the two C5=C6 bonds in pyrimidine residues, the CPD species in TpTo forms between the C5=C6 of T and the C3=C4 or C4=C5 bond of To. Considering the ultrafast reaction rate in DNA photoreaction, the dinucleotide conformation must remain during the dimerization reaction. Thus, the stacked structure in TpTo is likely to be 60° away from the photoactive conformation in TpT. These data suggest that although the substituted toluene closely mimics the size and shape of thymine, its stronger hydrophobic nature might result in a subtle DNA conformational change.

The natural DNA adopts a right-handed helical structure, which results in an approximately 30° turn between the adjacent bases. To make the C5=C6 of T close to the C4=C5 bond of To, an additional base rotation that can be as small as $\leq$30° is needed. We believe that such a minor rotational movement induced by the To substitution is reasonable as it is unlikely that the more hydrophobic toluenyl moiety

would result in no structural change at all after it replaces the thymine residue in duplex DNA. At the same time, this minor rotation largely retains the DNA duplex structure, which likely explains why such a conformational change were present in the previous NMR spectroscopy[18] and X-ray structures[29b] of the DFTo containing duplex DNA but was never clearly reported in these previous studies. These structures predict that a similar photoreaction should still occur in the context of duplex DNA. We will repeat the TpTo photochemistry in double-strand DNA to confirm this prediction in our future research. Further work to reveal the mechanistic details of the TpTo photochemistry (for instance, singlet vs. triplet excited state) as well as the correlation between the thymine conformational change due to the base rotation and thymine photochemistry is also in progress.

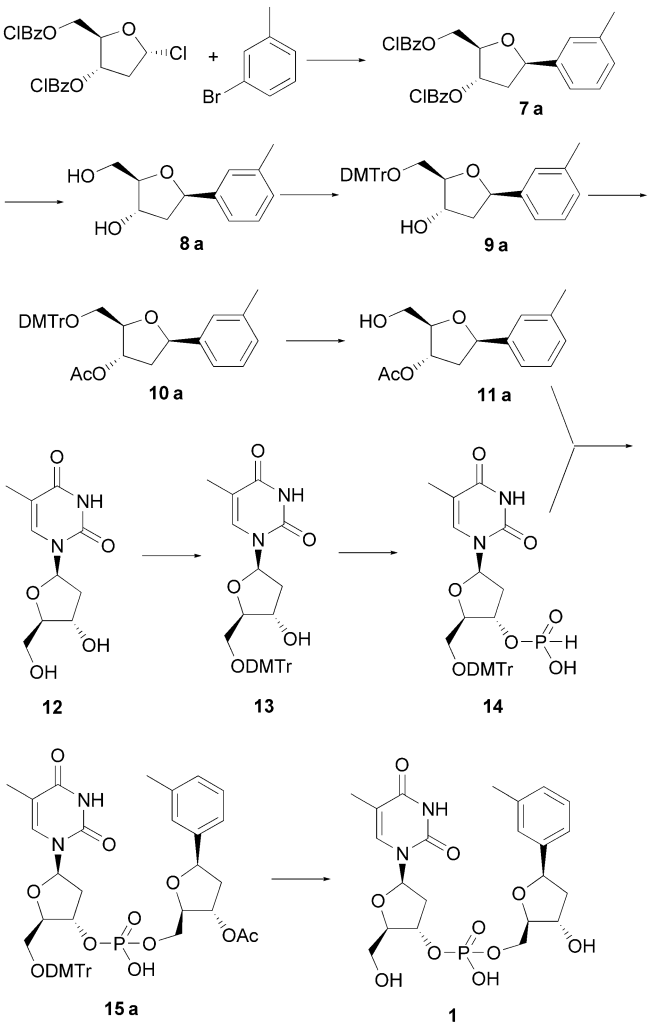
Experimental Section

General methods: All reagent grade chemicals were purchased from Sigma, Fisher, or VWR and were used without further purification except the 3,5-bis(4-chlorobenzoyl)-2-deoxy-α-D-ribofuranosyl chloride, which was purchased from Shanghai Hanhong Chemical Co., Ltd. (Shanghai, P.R. China). All reactions were carried out by using oven- or flame-dried glassware under a nitrogen atmosphere in distilled solvents. Dichloromethane and pyridine were distilled over calcium hydride. Purification of reaction products was carried out by flash chromatography by using silica gel (Dynamic Adsorbents Inc, 32–63 μm). For TLC analysis, precoated plates (w/h F254, Dynamic Adsorbents, Inc., 0.25 mm thick) were used. ^{1}H and ^{13}C NMR spectra were obtained on a Bruker 500 MHz NMR Fourier transform spectrometer. NMR spectra were recorded in sample solutions in deuterated chloroform ($CDCl_3$) with residual chloroform ($\delta = 77.0$ ppm for ^{13}C NMR) and TMS ($\delta = 0$ ppm for ^{1}H NMR), deuterated methanol ($\delta = 3.31$ ppm for ^{1}H NMR and $\delta = 49.0$ ppm for ^{13}C NMR), deuterated DMSO ($\delta = 2.50$ ppm for ^{1}H NMR and $\delta = 39.52$ ppm for ^{13}C NMR). The chemical shifts in the NMR spectra are reported in parts per million (ppm). Mass (MS) analysis was obtained by using Agilent 1100 series LC/MSD system with electrospray ionization (ESI). The TpTo and TpMeTo photoreactions were carried out by using a Spectroline germicidal UV sterilizing lamp (Dual-tube, 15 W, intensity: 1550 uW cm^{-2}) with the samples about 5 inches from the lamp.

DFT calculations on the stability of CPD analogues 5 and 6: The initial structures of the cyclobutane pyrimidine dimer (CPD) were constructed by using residues 5 and 6 of chain A, corresponding to the T<>T in the intrahelical position of a duplex DNA in structure 1BW7 (PDB ID), and residues 7 and 12 of chain I, corresponding to the T<>T in the extrahelical 1TEZ (PDB ID). In 1TEZ, the intradimer contains a formacetal linker instead of the phosphate in the structure. The potential CPD products **5** and **6** were optimized from these two T<>T structures revealed by the PDB files. The structures of **5** and **6** were optimized with G03[16] at the Austin Model 1 (AM1) level first, then at HF/6-31G(d) and B3LYP/6-31+G(d,p) level.[17]

Synthesis of 7a: Magnesium (3.24 g, 126 mmol) was activated by iodine etching in a Schlenk flask while being stirred without solvent. THF (354 mL) and 3-bromotoluene (22.6 g, 123 mmol) were then added and the mixture was stirred at 50 °C until all the magnesium was consumed. The resulting solution was cooled to 0 °C before copper iodide (13.1 g, 68.7 mmol) was added. The reaction mixture was allowed to warm to 20 °C and stirred until all the copper iodide was dissolved (~30 min). The solution was heated to 40 °C, 2-deoxy-α-D-*erythro*-pentofuranosyl chloride bis(4-chloride benzoate) (25 g, 58.1 mmol) was added and the mixture was allowed to stir for 1 h at 40 °C. The reaction was stopped by the addition of a 10% aqueous solution of NH_4Cl followed by extraction with CH_2Cl_2 (30 mL). The combined organic layers were washed with a saturated aqueous solution of $NaHCO_3$ and brine, subsequently dried over anhydrous $MgSO_4$ and the solvent was removed under reduced pressure. The product was purified by silica gel column chromatography (hexane/ethyl acetate = 16:1) to yield an α/β mixture (α/β = 1.6:1) as a colorless oil (26.42 g, 98%).

The resulting isomer mixture was then dissolved in CH_2Cl_2 (529 mL) and benzenesulfonic acid (16.57 g, mmol) as well as TFA (100 mL) were added. The solution was refluxed for 24 h and Et_3N (331 mL) was added to neutralize the acids. The solvents were removed by rotary evaporation and the resulting residue was purified by column chromatography (hexane/ethyl acetate = 16:1) to afford the β isomer compound **7a** (Scheme 4) as a colorless oil (15.39 g, 58%). ^{1}H NMR ($CDCl_3$): $\delta = 2.13$ (m, 1H), 2.21 (s, 3H), 2.31 (s, 3H), 2.56 (dd, $J = 4.55$, 13.45 Hz, 1H), 4.51 (dd, $J = 2.0$, 4.0 Hz, 1H), 4.66 (t, $J = 5.0$ Hz, 2H), 5.22 (dd, $J = 5.0$, 11.0 Hz, 1H), 5.59 (d, $J = 6.5$ Hz, 1H), 7.10 (d, $J = 7.5$ Hz, 1H), 7.24 (m, 4H), 7.40 (d, $J = 8.5$ Hz, 2H), 7.46 (d, $J = 8.6$ Hz, 2H), 8.00 ppm (m, 4H); ^{13}C NMR ($CDCl_3$): $\delta = 21.3$, 41.5, 65.0, 77.6, 80.8, 82.7, 122.9, 126.4, 128.1, 128.2, 128.4, 128.7, 128.8, 131.0, 138.2, 139.6, 139.9, 140.3, 165.2, 165.4 ppm; ESI-MS (positive mode) calcd for $C_{26}H_{23}C_{l2}O_5$: 485.1 [M+H$^+$], found 485.0.



Scheme 4. Synthesis of the dinucleotide TpTo (**1**).

Synthesis of 8a: Compound **7a** (1.92 g, 3.935 mmol) was dissolved in methanol (44 mL) and NaOMe (0.66 g, 12.24 mmol) was added immedi-

ately. The mixture was stirred for 3 h at room temperature and NH_4Cl (0.66 g, 12.24 mmol) was added to quench the reaction. The solvent was removed by rotary evaporation and the residue was purified by column chromatography (CH_2Cl_2/MeOH = 20:1) to afford the product as a colorless oil (0.88 g, 82%). ^1H NMR ($CDCl_3$): δ = 1.94 (m, 1H), 2.16 (m, 1H), 2.32 (s, 3H), 3.67 (d, J = 5.5 Hz, 2H), 3.96 (dd, J = 2.5, 5.5 Hz, 1H), 4.30 (m, 1H), 5.09 (dd, J = 5.5, 10.5 Hz, 1H), 7.08 (m, 3H), 7.20 ppm (t, J = 7.5 Hz, 1H); ^{13}C NMR ($CDCl_3$): δ = 21.4, 43.3, 63.2, 73.5, 80.2, 87.3, 123.1, 126.7, 128.4, 128.5, 138.1, 140.9 ppm; ESI-MS (positive mode) calcd for $C_{12}H_{16}NaO_3$: 231.1 [$M+H^+$], found 231.0.

Synthesis of 9a: Compound **8a** (0.88 g, 4.25 mmol) was dissolved in pyridine (8 mL) and Et_3N (0.77 mL) and DMAP (0.104 g, 0.85 mmol) was added to this solution. After being stirred for 3 min at room temperature, 4,4′-dimethoxytrityl chloride (DMTrCl; 1.87 g, 5.52 mmol) was added at room temperature. After being stirred for 6 h, MeOH (4 mL) was added to quench the reaction and the solvent was removed by rotary evaporation. The resulting residue was purified by column chromatography (hexane/ethyl acetate = 4:1) to afford compound **9a** as a white solid (1.504 g, 70%). ^1H NMR ($CDCl_3$): δ = 2.18 (m, 1H), 2.29 (s, 3H), 2.40 (d, J = 3.5 Hz, 1H), 3.28 (d, J = 5.0 Hz, 1H), 3.31 (d, J = 4.5 Hz, 1H), 4.06 (d, J = 2.5 Hz, 1H), 4.38 (d, J = 2.0 Hz, 1H), 5.13 (dd, J = 5.5, 10.0 Hz, 4H), 6.80 (dd, J = 3.0, 12.0 Hz, 4H), 7.06 (d, J = 7.5 Hz, 1H), 7.20 (m, 3H), 7.25 (m, 3H), 7.36 (m, 4H), 7.47 ppm (d, J = 8.5 Hz, 2H); ^{13}C NMR ($CDCl_3$): δ = 21.3, 43.9, 55.1, 64.5, 74.5, 80.0, 86.1, 86.3, 113.0, 123.1, 126.6, 126.7, 127.7, 128.1, 130.0, 136.0, 137.8, 141.6, 144.8, 158.3 ppm.

Synthesis of 11a: Compound **9a** (1.5 g, 2.95 mmol) was dissolved in pyridine (10 mL) and DMAP (37 mg, 0.30 mmol) and Ac_2O (0.42 mL, 4.43 mmol) were added to this solution at room temperature. After being stirred for 6 h, H_2O (5 mL) was added to quench the reaction and the solvent removed by rotary evaporation. The resulting residue was dissolved in CH_2Cl_2 (50 mL) and washed with NH_4Cl (aq., saturated) and $NaHCO_3$ (aq., saturated). Then the solvent was evaporated and the resulting residue purified by column chromatography to afford compound **10a** as a white solid (1.57 g, 96%). The product was used directly in the next step of synthesis without further purification.

Compound **10a** (1.57 g, 1.09 mmol) was dissolved in a solution of TFA (3%) in dichloromethane (40 mL) at room temperature. After 30 min, the red solution was evaporated to dryness under reduced pressure. The crude material was redissolved in CH_2Cl_2 and purified by silica gel chromatography (hexane/ethyl acetate = 4:1). Compound **11a** was obtained as a white foam after evaporation of the solvent (0.483 g, 68%). ^1H NMR ($CDCl_3$): δ = 2.09 (m, 1H), 2.13 (s, 3H), 2.34 (m, 1H), 2.36 (s, 3H), 3.83 (dd, J = 4.5, 11.5 Hz, 1H), 3.87 (dd, J = 4.0, 12.0 Hz, 1H), 4.08 (m, 1H), 5.06 (dd, J = 5.0, 11.0 Hz, 1H), 5.22 (dd, J = 2.0, 6.0 Hz, 1H), 7.15 (m, 3H), 7.26 ppm (m, 1H); ^{13}C NMR ($CDCl_3$): δ = 21.1, 21.4, 41.0, 63.3, 76.8, 80.4, 85.5, 123.1, 126.7, 128.5, 128.8, 138.2, 140.1, 171.0 ppm; ESI-MS (positive mode) calcd for $C_{14}H_{19}O_4$: 251.1 [$M+H^+$], found 251.1.

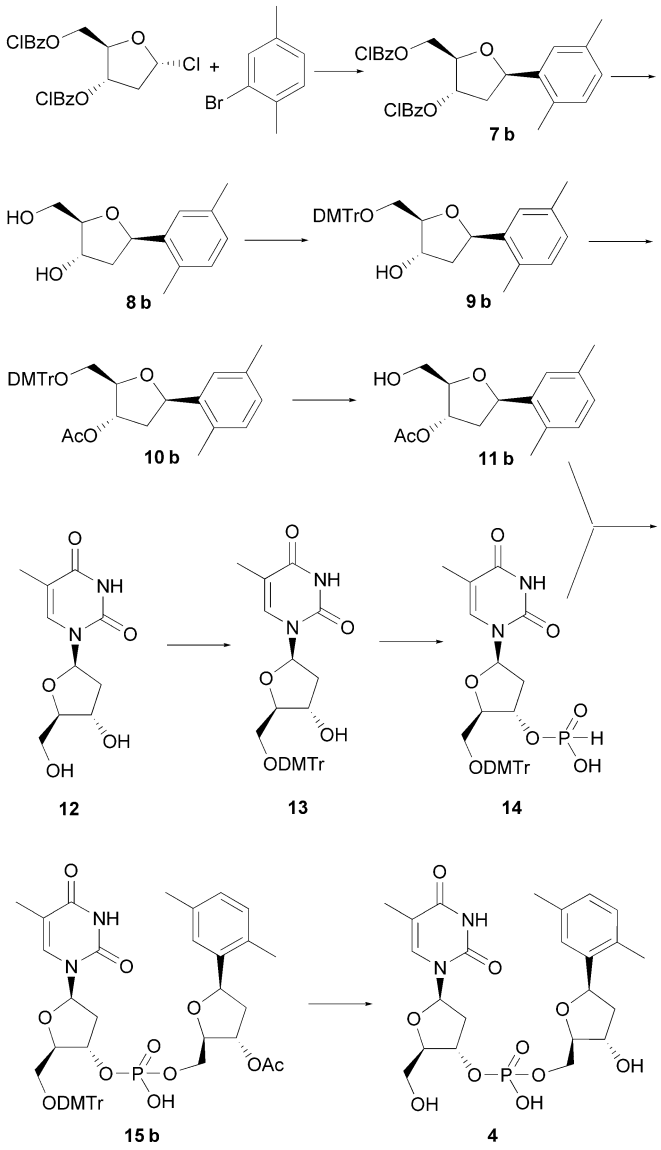
Synthesis of 13: Thymidine (4.00 g, 16.6 mmol; **12**) was dissolved in pyridine (150 mL) and DMTrCl (6.75 g, 20.0 mmol) was added at room temperature. After being stirred for 6 h, MeOH (4.00 g, 125 mmol) was added to quench the reaction and the solvents were evaporated under a reduced pressure. The residue was purified by column chromatography (CH_2Cl_2/MeOH = 40:1) to afford compound **13** as a white solid (7.8 g, 86%).[5] ^1H NMR ($CDCl_3$): δ = 1.45 (s, 3H), 2.31 (m, 1H), 2.43 (m, 1H), 3.36 (dd, J = 3.0, 10.5 Hz, 1H), 3.45 (dd, J = 3.0, 10.5 Hz, 1H), 3.76 (s, 6H), 4.09 (d, J = 2.0 Hz, 1H), 4.57 (s, 1H), 6.44 (t, J = 7.0 Hz, 1H), 6.82 (d, J = 8.5 Hz, 4H), 7.21 (m, 1H), 7.27 (m, 7H), 7.39 (J = 7.5 Hz, 2H), 7.61 ppm (s, 1H); ^{13}C NMR ($CDCl_3$): δ = 11.7, 40.9, 53.4, 55.2, 63.6, 72.4, 84.8, 86.3, 86.8, 111.2, 113.2, 127.0, 127.9, 128.1, 130.0, 135.3, 135.4, 135.7, 144.3, 149.4, 150.7, 158.6, 164.0 ppm.

Synthesis of 14: Diphenyl phosphite (4.00 g, 11.1 mmol) was added to a solution of **13** (4.00 g, 11.1 mmol) in pyridine (5 mL). The reaction was quenched after 15 min by addition of a mixture of water–triethylamine (1:1 v/v, 2 mL) and left standing for 15 min. The solvent was removed by rotary evaporation and the residue was partitioned between CH_2Cl_2 (50 mL) and 5% aq. $NaHCO_3$ (20 mL). The organic layer was washed twice with 5% aq. $NaHCO_3$ (20 mL), dried over Na_2SO_4 and finally evaporated to yield an oil. The product was purified by column chromatography (CH_2Cl_2/Et_3N/MeOH = 30:1.5:1) to afford compound **14** as a white solid (3.84 g, 87%). ^1H NMR ($CDCl_3$): δ = 1.34 (s, 1H), 2.39 (m, 1H), 2.59 (m, 1H), 3.39 (dd, J = 2.5, 10.5 Hz, 1H), 3.48 (dd, J = 2.5, 10.5 Hz, 1H), 3.78 (s, 6H), 4.27 (d, J = 2.0 Hz, 1H), 5.01 (m, 1H), 6.46 (dd, J = 5.5, 8.0 Hz, 1H), 6.81 (s, 2H), 6.83 (s, 2H), 7.22 (m, 1H), 7.28 (m, 6H), 7.40 (d, J = 7.0 Hz, 2H), 7.60 ppm (s, 1H); ^{13}C NMR ($CDCl_3$): δ = 9.2, 45.5, 55.1, 63.4, 74.0, 84.3, 85.1, 86.8, 111.0, 113.1, 120.4, 120.5, 122.9, 126.9, 127.8, 128.1, 129.1, 130.0, 135.2, 135.3, 135.6, 144.2, 150.6, 158.5, 164.0 ppm.

Synthesis of compound 15a: Compounds **11a** (297 mg, 0.488 mmol) and **14** (111 mg, 0.444 mmol) were dried and dissolved in anhydrous pyridine (10 mL); PivCl (0.43 mL, 3.11 mmol) was added dropwise to this solution. The reaction mixture was stirred at room temperature for 20 min, quenched by addition of water (1.3 mL) and oxidized with I_2 (345 mg, 1.36 mmol). The solution was further stirred for 1 h before $Na_2S_2O_4$ was added to quench the reaction. Et_3N (5.5 mL) was then added to facilitate the formation of the triethylammonium salt with the phosphate moiety in **15a**, thus enhancing its solubility in organic solvents. After removing the solvents by rotary evaporation, the resulting residue was dissolved in CH_2Cl_2. The solution was washed with $NaHCO_3$ (aq., saturated) three times, the solvent was evaporated and the resulting compound was purified by column chromatography (CH_2Cl_2/Et_3N/MeOH = 20:1:1) to afford a triethylammonium salt of compound **15a** as a white solid (0.433 g, 56.6%). ^1H NMR ($CDCl_3$): δ = 3.78 (s, 6H), 4.02 (m, 2H), 4.15 (d, J = 1.5 Hz, 1H), 4.32 (d, J = 1.0 Hz, 1H), 5.01 (m, 1H), 5.25 (d, J = 5.5 Hz, 1H), 5.35 (s, 1H), 6.81 (d, J = 8.5 Hz, 4H), 7.05 (s, 1H), 7.15 (m, 3H), 7.24 (m, 7H), 7.37 (d, J = 7.5 Hz, 2H), 7.58 ppm (s, 1H); ^{13}C NMR ($CDCl_3$): δ = 10.9, 20.5, 20.8, 38.8, 40.3, 52.5, 53.1, 54.6, 63.4, 65.1, 76.0, 79.8, 83.0, 83.1, 83.9, 86.3, 110.4, 112.5, 122.5, 126.2, 126.4, 127.3, 127.6, 127.9, 129.5, 134.6, 134.8, 135.2, 137.2, 140.0, 143.7, 149.9, 158.0, 163.3, 169.8 ppm.

Synthesis of 1: The triethylammonium salt of compound **15a** (0.43 g, 0.251 mmol) was dissolved in 20 mL CH_2Cl_2 solution of TFA (3%). After 10 min, the solvent was removed by rotary evaporation. The residue was subsequently dissolved in MeOH (13 mL) and $NH_3{\cdot}H_2O$ (38 mL) was added to this solution. The mixture was stirred for 12 h, the solvents were removed by rotary evaporation and the crude product was stirred with the cation exchange resin in water for 3 h to exchange the corresponding cations to proton. After filtration to remove the resin, the solvents were evaporated and the crude product was purified with RP-HPLC. The eluent was collected and solvents were removed to afford compound **1** as a colorless solid (76 mg, 58.9%). ^1H NMR (CD_3OD): δ = 1.86 (s, 3H), 2.02 (m, 1H), 2.19 (m, 2H), 2.31 (s, 3H), 2.49 (m, 1H), 3.76 (t, J = 2.5 Hz, 2H), 3.98 (dd, J = 5.0, 10.0 Hz, 1H), 4.02 (dd, J = 5.0, 10.0 Hz, 1H), 4.05 (m, 1H), 4.14 (dd, J = 3.0, 5.5 Hz, 1H), 4.42 (dd, J = 1.5, 4.0 Hz, 1H), 4.84 (dd, J = 3.5, 6.5 Hz, 1H), 6.27 (dd, J = 6.0, 7.5 Hz, 1H), 7.04 (d, J = 6.0 Hz, 1H), 7.17 (m, 3H), 7.78 ppm (s, 1H); ^{13}C NMR (CD_3OD): δ = 21.5, 22.7, 40.0, 44.4, 62.9, 64.4, 67.2, 74.4, 76.9, 81.8, 86.3, 87.5, 111.6, 124.4, 128.0, 129.3, 138.2, 139.1, 142.9, 152.4, 166.4, 178.2 ppm; ESI-MS (positive mode) calcd for $C_{22}H_{30}N_2O_{10}P$: 513.2 [$M+H^+$], found 513.2.

Synthesis of 7b: Magnesium (0.98 g, 40.8 mmol) was activated by iodine etching in a Schlenk flask while being stirred without solvent. THF (60 mL) and 3-bromo-4-methyl-toluene (7.400 g, 40 mmol) were then added and the mixture was stirred at 50 °C until all the magnesium was consumed. Then THF (150 mL) was added. The resulting solution was cooled to 0 °C before copper iodide (3.960 g, 20.7 mmol) was added. The reaction mixture was allowed to warm to 20 °C and stirred until all the copper iodide was dissolved (~30 min). The solution was heated to 40 °C, 2-deoxy-α-D-*erythro*-pentofuranosyl chloride bis(4-chloride benzoate) (7.570 mg, 17.6 mmol) was added and the mixture was allowed to stir for 2 h at 40 °C. The reaction was stopped by the addition of a 10% aqueous solution of NH_4Cl followed by extraction with CH_2Cl_2 (30 mL). The combined organic layers were washed with a saturated aqueous solution of $NaHCO_3$ and brine, subsequently dried over anhydrous $MgSO_4$ and solvent was removed under reduced pressure. The product was purified by silica gel column chromatography (hexane/ethyl acetate = 30:1) to yield compound **7b** (Scheme 5) as a colorless oil (5.300 g, 55%). ^1H NMR



Scheme 5. Synthesis of the dinucleotide TpMeTo (**4**).

($CDCl_3$): $\delta = 2.22$ (m, 1H), 2.29 (s, 3H), 2.51 (dd, $J = 4.1$, 14.0 Hz, 1H), 4.50 (m, 1H), 4.72 (m, 2H), 5.41 (m, 1H), 5.60 (t, $J = 2.9$ Hz, 1H), 6.99 (d, $J = 7.7$ Hz, 1H), 7.03 (d, $J = 7.6$ Hz, 1H), 7.33 (s, 1H), 7.39 (d, $J = 8.4$ Hz, 2H), 7.46 (d, $J = 8.5$ Hz, 2H), 8.01 ppm (m, 4H); ^{13}C NMR ($CDCl_3$): $\delta = 18.6$, 21.0, 40.2, 64.9, 77.6, 77.7, 82.4, 125.2, 128.1, 128.2, 128.3, 128.8, 128.9, 130.2, 131.0, 131.2, 135.8, 138.3, 139.6, 139.9, 165.2, 165.4 ppm; ESI-MS (positive mode) calcd for $C_{27}H_{24}Cl_2O_5Na$: 521.1 [$M+Na^+$], found 521.0.

Synthesis of 8b: Compound **7b** (5.300 g, 10.6 mmol) was dissolved in methanol (118 mL) and NaOMe (1.775 g, 32.9 mmol) was added immediately. The mixture was stirred for 3 h at room temperature and NH_4Cl (1.775 g, 32.6 mmol) was added to quench the reaction. The solvent was removed by rotary evaporation and the residue was purified by column chromatography (CH_2Cl_2/MeOH = 20:1) to afford the product as a colorless oil (2.110 g, 90%). ^1H NMR (CD_3OD): $\delta = 1.79$ (m, 1H), 2.21 (m, 1H), 2.26 (s, 3H), 2.28 (s, 3H), 3.70 (m, 2H), 3.92 (m, 1H), 4.31 (m, 1H), 5.30 (dd, $J = 5.4$, 10.5 Hz, 1H), 6.94 (d, $J = 7.8$ Hz, 1H), 6.98 (d, $J = 7.7$ Hz, 1H), 7.34 ppm (s, 1H); ^{13}C NMR (CD_3OD): $\delta = 18.7$, 21.2, 43.4, 64.0, 74.4, 78.3, 88.7, 126.6, 128.8, 131.0, 132.6, 136.6, 140.8 ppm; ESI-MS (positive mode) calcd for $C_{13}H_{19}O_3$: 223.1 [$M+H^+$], found 223.1.

Synthesis of 9b: Compound **8b** (2.110 g, 9.51 mmol) was dissolved in pyridine (18 mL) and Et_3N (1.72 mL) and DMAP (0.233 g, 1.91 mmol) was added to this solution. After being stirred for 3 min at room temperature, 4,4′-dimethoxytrityl chloride (DMTrCl; 4.187 g, 12.36 mmol) was added at room temperature. After being stirred for 6 h, MeOH (9 mL) was added to quench the reaction and the solvent was removed by rotary evaporation. The resulting residue was purified by column chromatography (hexane/ethyl acetate = 4:1) to afford compound **9b** as a yellow solid (4.548 g, 91%). ^1H NMR ($CDCl_3$): $\delta = 1.95$ (m, 1H), 2.23 (s, 3H), 2.25 (m, 1H), 2.27 (s, 3H), 3.36 (m, 2H), 3.77 (s, 6H), 4.03 (d, $J = 3.0$ Hz, 1H), 4.41 (t, $J = 3.0$ Hz, 1H), 5.32 (dd, $J = 5.7$, 9.6 Hz, 1H), 6.82 (d, $J = 8.8$ Hz, 4H), 6.96 (d, $J = 7.7$ Hz, 1H), 7.01 (d, $J = 7.7$ Hz, 1H), 7.20 (t, $J = 7.4$ Hz, 1H), 7.27 (dd, $J = 7.4$, 8.6 Hz, 2H), 7.37 (m, 4H), 7.41 (s, 1H), 7.49 ppm (d, $J = 8.1$ Hz, 2H); ^{13}C NMR ($CDCl_3$): $\delta = 18.7$, 21.1, 42.6, 46.1, 55.2, 64.3, 74.7, 76.8, 85.7, 86.2, 113.1, 125.6, 126.8, 127.7, 127.8, 128.2, 130.0, 130.1, 131.3, 135.5, 136.1, 139.9, 144.9, 158.4 ppm.

Synthesis of 11b: Compound **9b** (4.546 g, 2.95 mmol) was dissolved in pyridine (10 mL) and DMAP (37 mg, 0.30 mmol) and Ac_2O (0.42 mL, 4.43 mmol) were added to this solution at room temperature. After being stirred for 6 h, H_2O (5 mL) was added to quench the reaction and the solvent was removed by rotary evaporation. The resulting residue was dissolved in CH_2Cl_2 (50 mL) and washed with NH_4Cl (aq., saturated) and $NaHCO_3$ (aq., saturated). Then the solvent was evaporated and the resulting residue was purified by column chromatography to afford compound **10b** as a white solid (1.66 g, 99%). The product was used directly in the next step of synthesis without further purification.

Compound **10b** (1.66 g, 2.94 mmol) was dissolved in a solution of TFA (3%) in dichloromethane (40 mL) at room temperature. After 30 min, the red solution was evaporated to dryness under reduced pressure. The crude material was redissolved in CH_2Cl_2 and purified by silica gel chromatography (hexane/ethyl acetate = 4:1). Compound **11b** was obtained as a white solid after evaporation of the solvent (0.565 g, 73%). ^1H NMR ($CDCl_3$): $\delta = 2.00$ (m, 1H), 2.13 (s, 3H), 2.30 (s, 3H), 2.32 (s, 3H), 2.36 (m, 1H), 3.88 (m, 2H), 4.06 (dd, $J = 4.4$, 7.2 Hz, 1H), 5.22 (m, 2H), 7.00 (d, $J = 7.8$ Hz, 1H), 7.04 (d, $J = 7.7$ Hz, 1H), 7.26 ppm (s, 1H); ^{13}C NMR ($CDCl_3$): $\delta = 18.7$, 21.1, 39.7, 63.2, 85.1, 125.2, 128.3, 130.3, 131.8, 135.7, 137.9, 171.1 ppm. ESI-MS (positive mode) calcd for $C_{15}H_{21}O_4$: 265.1 [$M+H^+$], found 265.0.

Synthesis of compound 15b: Compound **11b** (400 mg, 1.52 mmol) and compound **14** (1.010 mg, 1.67 mmol) were dried and dissolved in anhydrous pyridine (34 mL); PivCl (1.48 mL, 10.70 mmol) was added dropwise to this solution. The reaction mixture was stirred at room temperature for 20 min, quenched by addition of water (1.3 mL) and oxidized with I_2 (561 mg, 2.21 mmol). The solution was further stirred for 1 h before $Na_2S_2O_4$ was added to quench the reaction. Et_3N (5.5 mL) was then added to facilitate the formation of the triethylammonium salt with the phosphate moiety in **15b**, thus enhancing its solubility in organic solvents. After removing the solvents by rotary evaporation, the resulting residue was dissolved in CH_2Cl_2. The solution was washed with $NaHCO_3$ (aq., saturated) for three times, the solvent was evaporated and the resulting compound was purified by column chromatography (CH_2Cl_2/Et_3N/MeOH = 20:1:1) to afford a triethylammonium salt of compound **15b** as a white solid (0.412 g, 31%). ^1H NMR ($CDCl_3$): $\delta = 2.07$ (s, 3H), 2.24 (s, 6H), 2.71 (m, 1H), 3.40 (d, $J = 2.4$ Hz, 2H), 3.76 (s, 3H), 3.77 (s, 3H), 4.00 (m, 1H), 4.07 (m, 1H), 4.17 (m, 1H), 4.33 (d, $J = 1.8$ Hz, 1H), 5.06 (t, $J = 5.8$ Hz, 1H), 5.15 (dd, $J = 4.9$, 11.0 Hz, 1H), 5.30 (b, 1H), 6.48 (dd, $J = 5.4$, 8.9 Hz, 1H), 6.80 (d, $J = 8.9$ Hz, 4H), 6.93 (d, $J = 7.6$ Hz, 1H), 6.98 (d, $J = 7.7$ Hz, 1H), 7.25 (m, 8H), 7.37 (d, $J = 7.6$ Hz, 2H), 7.59 ppm (s, 1H); ^{13}C NMR ($CDCl_3$): $\delta = 10.4$, 11.4, 18.6, 21.0, 21.1, 39.3, 45.8, 55.1, 84.5, 86.8, 110.9, 113.1, 125.6, 126.9, 127.8, 127.9, 128.2, 129.9, 130.1, 131.4, 135.2, 135.4, 135.8, 138.4, 144.3, 150.4, 158.5, 163.9, 170.3 ppm.

Synthesis of 4: The triethylammonium salt of compound **15b** (0.412 g, 0.251 mmol) was dissolved in a CH_2Cl_2 (30 mL) solution of TFA (3%). After 10 min, the solvent was removed by rotary evaporation. The residue was subsequently dissolved in MeOH (7 mL) and $NH_3 \cdot H_2O$ (20 mL) was added to this solution. The mixture was stirred for 12 h, the solvents were removed by rotary evaporation and the crude product was stirred with the cation exchange resin in water for 3 h to exchange the corre-

sponding cation to proton. After filtration to remove the resin, the solvents were evaporated and the crude product was purified with RP-HPLC. The eluent was collected and solvents were removed to afford compound **4** as a colorless solid (0.166 mg, 65%). ^{1}H NMR (CD_3OD): δ=1.85 (s, 3H), 1.89 (m, 1H), 2.23 (m, 1H), 2.26 (m, 1H), 2.26 (s, 3H), 2.27 (s, 3H), 2.51 (m, 1H), 3.74 (d, J=2.9 Hz, 2H), 4.07 (d, J=3.1 Hz, 1H), 4.19 (m, 3H), 4.37 (t, J=2.9 Hz, 1H), 4.99 (t, J=6.1 Hz, 1H), 5.32 (dd, J=5.5, 10.3 Hz, 1H), 6.26 (dd, J=5.9, 8.2 Hz, 1H), 6.92 (d, J=7.6 Hz, 1H), 6.97 (d, J=7.7 Hz, 1H), 7.32 (s, 1H), 7.74 ppm (s, 1H); ^{13}C NMR (CD_3OD): δ=12.5, 18.8, 21.3, 39.6, 43.2, 62.6, 68.6, 74.1, 78.6, 79.1, 86.1, 86.4, 87.3, 111.7, 126.6, 128.9, 131.1, 132.8, 136.6, 137.9, 140.4, 452.3, 166.3 ppm; ESI-MS (positive mode) calcd for $C_{23}H_{32}N_2O_{10}P$: 527.2 [M+H$^+$], found 527.1.

Photoreaction of compound 1 and 4: Irradiation of **1** or **4** (1 mM) was conducted in aqueous solution at pH 7 at ambient temperature. The reaction was repeated at 77 K. To ensure the formation of a transparent glass, an equal volume of glycerol was added to the aqueous solution before being frozen. The resulting glass was then exposed for 15 min to the UVC (254 nm) light. The ice was subsequently warmed to room temperature and analyzed by HPLC.

Preparation of 2 and 3 in glycerol/H_2O for NMR spectroscopy: An aqueous solution of **1** (10 mM; 20 mL) was titrated with NaOH (0.1 M) to pH 7.0. After removal of water under vacuum, the resulting sodium salt was dissolved in glycerol/H_2O (1:1, 25 mL, 4.36 mg mL^{-1}). The solution was frozen in a 40×50 cm plate under liquid nitrogen. The frozen solution was then exposed for 15 min to the UVC (254 nm) light. The solid was subsequently warmed to room temperature, frozen again in liquid N_2 and exposed for 15 min to UVC (254 nm) light. This process was repeated six times and the resulting products were purified by reverse phase HPLC in the gradient mode by using ammonium acetate, pH 6.8, and acetonitrile as solvents. The eluted **1** was recycled for another round of photoreaction. Compound **2**: ^{1}H NMR (D_2O): δ=1.47 (s, 3H), 1.63 (s, 3H), 1.96 (m, 1H), 2.19 (m, 1H), 2.52 (m, 1H), 2.63 (m, 1H), 2.87 (d, J=5.5 Hz, 1H), 3.75 (dd, J=6.0, 12.5 Hz, 1H), 3.84 (dd, J=3.5, 12.5 Hz, 1H), 3.93 (m, 1H), 3.97 (dd, J=2.5, 6.0 Hz, 1H), 4.00 (m, 1H), 4.03 (dd, J=2.0, 11.0 Hz, 1H), 4.11 (d, J=0.5 Hz, 1H), 4.47 (m, 2H), 4.56 (t, J=7.5 Hz, 1H), 5.51 (dd, J=5.0, 9.0 Hz, 1H), 5.52 (d, J=5.0 Hz, 1H), 5.74 (t, J=6.0 Hz, 1H), 5.81 ppm (d, J=9.0 Hz, 1H); ^{13}C NMR (D_2O): δ=21.3, 21.6, 24.3, 28.6, 37.7, 39.9, 40.3, 48.7, 48.8, 60.9, 65.8, 66.0, 71.3, 73.9, 79.0, 82.5, 82.6, 84.3, 84.4, 121.9, 123.5, 124.1, 137.0, 152.5, 174.5 ppm; ^{1}H NMR (CD_3OD): δ=1.47 (s, 3H), 1.61 (s, 3H), 1.83 (m, 1H), 2.10 (m, 1H), 2.36 (m, 1H), 2.71 (d, J=5.2 Hz, 1H), 2.85 (m., 1H), 3.77 (m, 1H), 3.83 (m, 2H), 3.93 (dd, J=3.8, 12.5 Hz, 1H), 4.01 (m, 1H), 4.04 (s, 1H), 4.06 (m, 1H), 4.36 (m, 1H), 4.43 (t, J=7.2 Hz, 1H), 4.55 (m, 1H), 5.50 (dd, J=5.2, 9.7 Hz, 1H), 5.66 ppm (m, 3H); ^{1}H NMR ([D_6]DMSO): δ=1.35 (s, 3H), 1.49 (s, 3H), 1.60 (m, 1H), 1.98 (m, 1H), 2.07 (m, 1H), 2.54 (m, 1H), 2.58 (d, J=5.2 Hz, 1H), 3.52 (m, 1H), 3.63 (m, 3H), 3.76 (m, 2H), 3.81 (s, 1H), 4.13 (dd, J=5.2, 10.9 Hz, 1H), 4.29 (t, J=6.8 Hz, 1H), 4.41 (m, 1H), 5.10 (b, 1H), 5.36 (dd, J=5.1, 9.8 Hz, 1H), 5.41 (dd, J=2.7, 7.9 Hz, 1H), 5.55 (s, 1H), 5.59 ppm (d, J=9.8 Hz, 1H); ^{13}C NMR ([D_6]DMSO): δ=26.2, 29.8, 39.6, 39.8, 42.2, 47.4, 48.4, 60.5, 64.5, 65.7, 69.4, 72.2, 77.5, 81.7, 83.6, 84.5, 121.1, 123.0, 123.6, 136.0, 150.1, 170.7 ppm; ESI-MS (positive mode) calcd for $C_{22}H_{30}N_2O_{10}P$: 513.2 [M+H$^+$], found 513.2. Compound **3**: ^{1}H NMR (D_2O): δ=1.60 (s, 3H), 1.73 (s, 3H), 1.95 (m, 1H), 2.27 (m, 1H), 2.52 (m, 1H), 2.57 (m, 1H), 3.10 (d, J=11.0 Hz, 1H), 3.66 (b, 1H), 3.74 (dd, J=5.0, 12.5 Hz, 1H), 3.80 (dd, J=6.5, 11.0 Hz, 1H), 3.90 (dd, J=2.5, 12.5 Hz, 1H), 4.00 (m, 1H), 4.01 (d, J=11.0 Hz, 1H), 4.11 (m, 1H), 4.34 (t, J=7.5 Hz, 1H), 4.42 (dd, J=3.5, 8.0 Hz, 1H), 4.54 (d, J=8.0 Hz, 1H), 4.64 (t, J=7.0 Hz, 1H), 5.52 (s, 1H), 5.76 (d, J=4.0 Hz, 1H), 5.80 ppm (dd, J=2.5, 6.0 Hz, 1H); ^{13}C NMR (D_2O): δ=21.5, 24.8, 37.7, 38.4, 38.7, 45.8, 52.1, 57.6, 60.4, 65.6, 71.1, 72.2, 78.9, 82.5, 83.9, 84.3, 112.0, 118.7, 134.9, 140.2, 152.7, 175.1 ppm; ESI-MS (positive mode) calcd for $C_{22}H_{30}N_2O_{10}P$: 513.2 [M+H$^+$], found 513.2.

Preparation of 5 in glycerol/H_2O for NMR analysis: An aqueous solution of **4** (10 mM; 20 mL) was titrated with NaOH (0.1 M) to pH 7.0. After removal of water under vacuum, the resulting sodium salt was dissolved in glycerol/H_2O (1:1, 25 mL, 4.36 mg mL^{-1}). The solution was frozen in a 40×50 cm plate under liquid nitrogen. The frozen solution was then exposed for 15 min to UVC (254 nm) light. The solid was subsequently warmed to room temperature, frozen again in liquid N_2 and exposed for 15 min to UVC (254 nm) light. This process was repeated six times and the resulting products were purified by reverse phase HPLC in the gradient mode by using ammonium acetate, pH 6.8, and acetonitrile as solvents. The eluted **4** was recycled for another round of photoreaction. Compound **5**: ^{1}H NMR (D_2O): δ=1.44 (s, 3H), 1.58 (s, 3H), 1.76 (s, 3H), 1.95 (dd, J=6.5, 13.4 Hz, 1H), 2.22 (m, 1H), 2.51 (m, 2H), 2.78 (d, J=4.8 Hz, 1H), 3.75 (dd, J=6.1, 12.5 Hz, 1H), 3.83 (dd, J=3.3, 12.6 Hz, 1H), 3.93 (dd, J=4.8, 11.1 Hz, 1H), 3.98 (m, 2H), 4.04 (d, J=11.1 Hz, 1H), 4.10 (s, 1H), 4.47 (m, 2H), 4.89 (t, J=7.5 Hz, 1H), 5.21 (d, J=5.0 Hz, 1H), 5.60 (s, 1H), 5.74 ppm (t, J=5.8 Hz, 1H); ^{13}C NMR (D_2O): δ=18.5, 24.2, 28.4, 37.7, 39.8, 41.4, 48.2, 48.5, 60.9, 65.5, 65.6, 71.2, 73.9, 75.8, 82.7, 84.1, 84.3, 118.7, 123.5, 130.4, 138.5, 152.6, 174.8 ppm; ESI-MS (positive mode) calcd for $C_{23}H_{32}N_2O_{10}P$: 527.2 [M+H$^+$], found 527.1.

Acknowledgements

The authors thank Professors Eric Long and Kyungsoo Oh at IUPUI for helpful discussions. This work was supported by the National Institutes of Health (ES017177 to L.L.), the IUCRG grant (L.L.) and the IUPUI startup fund (L.L. and J.P.). The NMR spectroscopy and MS facilities are supported by the National Science Foundation (CHE-0619254, DBI-0821661).

[1] a) B. A. Schweitzer, E. T. Kool, *J. Am. Chem. Soc.* **1995**, *117*, 1863–1872; b) E. T. Kool, *Acc. Chem. Res.* **2002**, *35*, 936–943; c) E. T. Kool, H. O. Sintim, *Chem. Commun.* **2006**, 3665–3675.

[2] a) S. Moran, R. X. Ren, E. T. Kool, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10506–10511; b) S. Moran, R. X. Ren, S. Rumney, E. T. Kool, *J. Am. Chem. Soc.* **1997**, *119*, 2056–2057; c) E. T. Kool, *Annu. Rev. Biochem.* **2002**, *71*, 191–219.

[3] a) T. A. Kunkel, R. Bebenek, *Annu. Rev. Biochem.* **2000**, *69*, 497–529; b) M. Strerath, J. Cramer, T. Restle, A. Marx, *J. Am. Chem. Soc.* **2002**, *124*, 11230–11231; c) M. F. Goodman, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10493–10495.

[4] a) P. S. Pallan, M. Egli, *J. Am. Chem. Soc.* **2009**, *131*, 12548–12549; b) S. Xia, W. H. Konigsberg, J. Wang, *J. Am. Chem. Soc.* **2011**, *133*, 10003–10005.

[5] J. Cadet, P. Vigny, *Photochemistry and the Nucleic Acids*, Wiley, New York, **1990**, p. 437.

[6] a) P. R. Callis, *Ann. Rev. Phys. Chem.* **1983**, *34*, 329–357; b) U. Streit, C. G. Bochet, *Beilstein J. Org. Chem.* **2011**, *7*, 525–542.

[7] a) G. Litwinienko, A. L. J. Beckwith, K. U. Ingold, *Chem. Soc. Rev.* **2011**, *40*, 2157–2163; b) F. Mueller, J. Mattay, *Chem. Rev.* **1993**, *93*, 99–117; c) J. Svoboda, B. Konig, *Chem. Rev.* **2006**, *106*, 5413–5430.

[8] a) D. Roca-Sanjuàn, G. Olaso-Gonzalez, I. GonzaÌlez-Ramirez, L. Serrano-Andres, M. Merchan, *J. Am. Chem. Soc.* **2008**, *130*, 10768–10779; b) M. Boggio-Pasqua, G. Groenhof, L. V. Schafer, H. Grubmuller, M. A. Robb, *J. Am. Chem. Soc.* **2007**, *129*, 10996–10997; c) L. Blancafort, A. Migani, *J. Am. Chem. Soc.* **2007**, *129*, 14540–14541; d) W. J. Schreier, J. Kubon, N. Regner, K. Haiser, T. E. Schrader, W. Zinth, P. Clivio, P. Gilch, *J. Am. Chem. Soc.* **2009**, *131*, 5038–5039.

[9] W. J. Schreier, T. E. Schrader, F. O. Koller, P. Gilch, C. E. Crespo-Hernandez, V. N. Swaminathan, T. Carell, W. Zinth, B. Kohler, *Science* **2007**, *315*, 625–629.

[10] G. Lin, L. Li, *Angew. Chem.* **2010**, *122*, 10122–10125; *Angew. Chem. Int. Ed.* **2010**, *49*, 9926–9929.

[11] T. W. Kim, J. C. Delaney, J. M. Essigmann, E. T. Kool, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15803–15808.

[12] a) B. A. Schweitzer, E. T. Kool, *J. Org. Chem.* **1994**, *59*, 7238–7242; b) see the Supporting Information.

[13] a) C. Desnous, B. R. Babu, C. Moriou, J. U. O. Mayo, A. Favre, J. Wengel, P. Clivio, *J. Am. Chem. Soc.* **2008**, *130*, 30–31; b) C. Moriou, M. Thomas, M. T. Adeline, M. T. Martin, A. Chiaroni, S. Pochet, J.-L. Fourrey, A. Favre, P. Clivio, *J. Org. Chem.* **2007**, *72*, 43–50.

[14] K. M. Guckian, B. A. Schweitzer, R. X. F. Ren, C. J. Sheils, D. C. Tahmassebi, E. T. Kool, *J. Am. Chem. Soc.* **2000**, *122*, 2213–2222.

[15] Note: the triplet signal observed suggests that in D_2O, the H2′ and H2″ protons become equal. In contast, a dd signal was observed for **5** in D_4-methanol, similar to those exhibited by **2** and **3** for this H1′ (see the Supporting Information).

[16] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, et al. Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT, **2004**.

[17] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623–11627.

[18] K. M. Guckian, T. R. Krugh, E. T. Kool, *Nat. Struct. Biol.* **1998**, *5*, 954–959.

[19] A. Mees, T. Klar, P. Gnau, U. Hennecke, A. P. M. Eker, T. Carell, L.-O. Essen, *Science* **2004**, *306*, 1789–1793.

[20] a) J. Kemmink, R. Boelens, R. Kaptein, *Eur. Biophys. J.* **1987**, *14*, 293–299; b) F. E. Hruska, D. J. Wood, K. K. Ogilvie, J. L. Charlton, *Can. J. Chem.* **1975**, *53*, 1193–1203.

[21] J. L. F. Kao, S. Nadji, J. S. Taylor, *Chem. Res. Toxicol.* **1993**, *6*, 561–567.

[22] T. M. G. Koning, J. J. G. van Soest, R. Kaptein, *Eur. J. Biochem.* **1991**, *195*, 29–40.

[23] K. M. Guckian, E. T. Kool, *Angew. Chem.* **1997**, *109*, 2942–2945; *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 2825–2828.

[24] W. Saenger, *Principles of Nucleic Acid Structure*, Springer, New York, **1984**, p. 556.

[25] T. Ostrowski, J. C. Maurizot, M. T. Adeline, J. L. Fourrey, P. Clivio, *J. Org. Chem.* **2003**, *68*, 6502–6510.

[26] C. E. Crespo-Hernández, B. Cohen, B. Kohler, *Nature* **2005**, *436*, 1141–1144.

[27] P. Yakovchuk, E. Protozanova, M. D. Frank-Kamenetskii, *Nucleic Acids Res.* **2006**, *34*, 564–574.

[28] a) J.-C. Xuan, I. T. Weber, *Nucleic Acids Res.* **1992**, *20*, 5457–5464; b) S.-G. Kim, L.-J. Lin, B. R. Reid, *Biochemistry* **1992**, *31*, 3564–3574.

[29] a) H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, R. E. Dickerson, *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 2179–2183; b) F. D. Lewis, X. Liu, Y. Wu, S. E. Miller, M. R. Wasielewski, R. L. Letsinger, R. Sanishvili, A. Joachimiak, V. Tereshko, M. Egli, *J. Am. Chem. Soc.* **1999**, *121*, 9905–9906; c) X. Shui, L. McFail-Isom, G. G. Hu, L. D. Williams, *Biochemistry* **1998**, *37*, 8341–8355.

[30] K. M. Guckian, T. R. Krugh, E. T. Kool, *J. Am. Chem. Soc.* **2000**, *122*, 6841–6847.

[31] V. R. Cooper, T. Thonhauser, A. Puzder, E. Schroder, B. I. Lundqvist, D. C. Langreth, *J. Am. Chem. Soc.* **2008**, *130*, 1304–1308.

[32] J. Norberg, L. Nilsson, *J. Am. Chem. Soc.* **1995**, *117*, 10832–10840.

[33] D. G. T. Su, J. L. F. Kao, M. L. Gross, J.-S. A. Taylor, *J. Am. Chem. Soc.* **2008**, *130*, 11328–11337.

[34] D. G. T. Su, H. Fang, M. L. Gross, J.-S. A. Taylor, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12861–12866.

[35] J. Yamamoto, K. Nishiguchi, K. Manabe, C. Masutani, F. Hanaoka, S. Iwai, *Nucleic Acids Res.* **2011**, *39*, 1165–1175.

[36] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

Received: March 11, 2012
Published online: May 15, 2012

