Published in final edited form as:

Chemistry. 2011 August 22; 17(35): 9658–9668. doi:10.1002/chem.201101821.

Chemical Synthesis, Crystal Structure and Enzymatic Evaluation of a Dinucleotide Spore Photoproduct Analogue Containing a Formacetal Linker

Dr. Gengjie Lin^[a], Dr. Chun-Hsing Chen^[b], Dr. Maren Pink^[b], Prof. Dr Jingzhi Pu^[a], and Prof. Dr Lei Li^{[a],[c],*}

^[a]Department of Chemistry and Chemical Biology, Indiana University, Purdue University Indianapolis (IUPUI), 402 N. Blackford St., Indianapolis, IN 46202 (USA)

^[b]Indiana University Molecular Structure Center, Chemistry, A421, Indiana University, 800 E. Kirkwood Avenue, Bloomington, IN 47405 (USA)

^[c]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, 635 Barnhill Drive, Indianapolis, IN 46202 (USA)

Abstract

Spore photoproduct (SP) is the exclusive DNA photodamage product found in bacterial endospores. Its photoformation and repair by a metal-loenzyme spore photoproduct lyase (SPL) composes the unique SP biochemistry. Despite the fact that the SP was discovered almost 50 years ago, its crystal structure is still unknown and the lack of structural information greatly hinders the study of SP biochemistry. Employing a formacetal linker and organic synthesis, we successfully prepared a dinucleotide SP isostere 5*R*-CH₂SP, which contains a neutral CH₂ moiety between the two thymine residues instead of a phosphate. The neutral linker dramatically facilitates the crystallization process, allowing us to obtain the crystal structure for this intriguing thymine dimer half a century after its discovery. Further ROESY spectroscopic, DFT computational, and enzymatic studies of this 5*R*-CH₂SP compound prove that it possesses similar properties with the 5*R*-SP species, suggesting that the revealed structure truly reflects that of SP generated in Nature.

Keywords

DNA damage; enzymes; oligonucleotides; photochemistry; UV irradiation

Introduction

Endospore-forming bacterial strains are responsible for a number of human diseases such as anthrax and tetanus. Due to the unique spore DNA photochemistry coupled to efficient damage repair in germinated spores, these spores are extremely resistant to UV radiation.^[1] Contrasting to the normal cells where the UV radiation induces thymine dimerization in genomic DNA to generate cyclobutane pyrimidine dimers (CPD) and pyrimidine (6-4) photoproducts, the exclusive DNA photoproduct in endospores is a special thymine dimer 5-thyminy-5,6-dihydrothymine, also called spore photoproduct or SP.^[2] Spores contain a special enzyme, the spore photoproduct lyase (SPL), to effectively reverse SP formation at

the early germination phase, resuming their normal life cycle (Scheme 1).^[1a,2] However, despite the fact that SP was discovered nearly half a century ago^[3] and the strong interest of the scientific community to understand its formation and repair,^[1a,4] the process is not completely understood.

Our group is interested in the mechanistic elucidation of both aspects of the SP biochemistry. By selectively labeling a dinucleotide TpT with deuterium and following the deuterium migration during the photoreaction, we demonstrated that one of the three methyl hydrogen atoms migrates to the *pro-S* position of the C6 carbon in resulting SP.^[5a] Furthermore, utilizing the deuterium-labeled SP substrate generated via photochemistry, we proved that the hydrogen atom abstracted by SPL in the initial step of the enzyme repair process was not returned to the resulting TpT product, a protein residue to be identified instead serves as the H-atom donor.^[6] Our data suggest that the currently hypothesized enzyme mechanism, which excludes the involvement of protein residue in catalysis, needs to be modified;^[1f,7] the SPL reaction is much more complex than currently thought.^[6]

Although being discovered almost half a century ago, the structure of SP remains unknown. In fact, among the three thymine dimers, CPD, (6-4) photoproduct and SP, only CPD is structurally characterized in duplex DNA^[8] as well as in dinucleotide fragment.^[9] The lack of structural information has drastically hindered our understanding of the SP biochemistry. As SP contains a chiral center in C5, both *R* and *S* isomers can be obtained should the thymine residues react freely.^[10] Such a free rotation does not happen due to the restriction of the double helical structure in genomic DNA, which determines that only one of two SP isomers can possibly form in Nature. There was a heated debate in the past several years on the steric configuration of SP in bacterial DNA.^[11] Although recent NMR spectroscopic studies established SP to adopt a 5*R* configuration, a crystal structure which allows a direct observation of the atom arrangement in SP is clearly needed.

Dinucleotide is the smallest DNA fragment to accommodate a SP dimer. Such a SP possesses the size of a small organic molecule; it can thus provide the precise structural details by the single-crystal X-ray diffraction methods. We rationalized that the negatively charged phosphate moiety in natural DNA may hinder the crystallization process, as reflected by the fact that all the CPD dimers crystallized to date contain either no or neutral linker between the two T residues.^[9,12] A formacetal is the simplest and smallest non-chiral isostere for a phosphate moiety and has proved very useful in oligonucleotide biochemical studies.^[9a,13] It was used to successfully crystallize a dinucleotide CPD analogue formed between two uracil residues.^[9a] It was included between the thymine residues of a CPD which was incorporated into the middle of a duplex DNA strand to facilitate the crystallization of the photolyase–DNA complex as well.^[14]

We therefore decided to incorporate a formacetal linker into the SP dinucleotide to examine if the resulting compound would mimic the properties of natural phosphate containing SP. Moreover, the neutral methylene linker was expected to facilitate the crystallization process, allowing us to gain structural insight into this important thymine photo-damage product.

Results and Discussion

Chemical synthesis of 5*R*-CH₂SP

As summarized in Scheme 2, the formacetal containing SP analog was synthesized using a 15-step chemical reaction via a modified procedure employed in the synthesis of phosphate containing di-nucleotide SP adopted by Begley and co-worker.^[11d] However, our synthesis is not a simple repetition of the previous work. For instance, SEM ([2-(trimethylsilyl)ethoxy]methyl) was used to protect the thymine N3 in Begley's original

approach; such a protecting reagent proved futile in our synthesis as it is incompatible with the formacetal linkage in the SnCl_4 -promoted de-protection step to obtain compound **11**. Changing to other de-protection reagents such as TBAF/DMPU, or TBAF/HMPA^[15] led to a complete decomposition of SEM-protected **11** as well. We therefore chose the benzyloxymethyl group (CH_2OBn) for N3 protection, which as a bonus increased the diastereoselectivity of the 5*R* over 5*S* isomer to 2.6:1 after the coupling step to yield compound **5**, contrasting to the non-selective 1:1 mixture obtained in the previous work.^[11d] (Note: The steric configurations of the compounds involved were assigned in consistence to the SP assignment in the previous NMR studies.^[4]) To remove the CH_2OBn from compound **10**, the traditional hydrogenation reaction did not work probably due to the intoxication of the Pd/C catalyst by the thymine nucleobase.^[16] Fortunately, the oxidative de-benzoylation strategy using the 2,3-di-chloro-5,6-dicyano-1,4-benzoquinone (DDQ) proved effective;^[17] the 5*R*- CH_2SP was successfully obtained after a standard desilylation reaction via TBAF (tetrabutylammonium fluoride) treatment.

Chemical synthesis and photolysis of $\text{T}_{\text{CH}_2\text{T}}$

The synthesis of thymine dinucleotide with a formacetal linker was achieved using a modified procedure of the $\text{U}_{\text{CH}_2\text{U}}$ synthesis,^[9a] as shown in Scheme 3. The $\text{T}_{\text{CH}_2\text{T}}$ was readily obtained after the standard deprotection step via ammonium hydroxide treatment.

However, UV irradiation of the resulting $\text{T}_{\text{CH}_2\text{T}}$ in either dry film or ice at 77 K under the established photochemical procedure^[5,18] led to no formation of the SP analog 5*R*- CH_2SP , as suggested by the HPLC analysis using the synthesized 5*R*- CH_2SP as the standard. This finding is in sharp contrast to what was observed in the TpT photolysis, which readily produced 5*R*-SP in ~1% yield in the dry film reaction as shown in our previous studies.^[5a] As the outcome of DNA photoreaction is determined by the DNA conformation at the instant of photoexcitation,^[19] such an observation suggests that the $\text{T}_{\text{CH}_2\text{T}}$ adopts a conformation different from that of TpT at the solid state.

Although the solid state is found to promote SP formation under UV photolysis probably due to the unique thymine stacking interaction, the exact thymine conformation under such a state is unclear. Our research to date found that this photoreaction can be easily disturbed by small modifications on the nucleotide residues. For instance, disturbing the methyl–methyl interaction between the thymine residues could dramatically alter the outcome of SP photochemistry. Due to the lower zero-point energy, deuterium is slightly smaller than protium even though it is the heavier isotope.^[20] However, substituting the CH_3 of 5'-T by a smaller CD_3 moiety while leaving the CH_3 group of the 3'-T which is involved in the H-transfer process intact decreases the yield of SP by 2.5-fold after photolysis of TpT.^[5a] Moreover, the di-nucleotide UpT which does not contain such a Me–Me interaction totally stopped the SP formation in the dry film reaction^[5b] although the mechanism suggests a SP analogue to be produced should the UpT adopt a similar conformation as TpT.

The highly sensitive DNA conformation favoring SP formation indicates that changing the negatively charged phosphate moiety to a neutral and smaller formacetal linker is likely to alter the thymine conformation in the dry state, making it no longer support the SP formation upon UV excitation. The observation suggests that to obtain the formacetal containing SP, the lengthy chemical synthesis shown above becomes the only choice, which is in sharp contrast to the cyclobutane pyrimidine dimer formation in the $\text{U}_{\text{CH}_2\text{U}}$ photolysis conducted by Carell and co-workers.^[9a] In their studies, the *cis-syn* CPD product was isolated as the major photoproduct with a 44% yield, making the CPD dimer preparation much easier.

Crystal structure of the 5R-CH₂SP

The resulting 5R-CH₂SP is very soluble in water due to the hydrophilic nature of the thymidine residue. More importantly, as expected, this neutral linker dramatically improved the crystallization property of SP. Dissolving the 5R-CH₂SP in water and allowing the solvent to slowly evaporate in two days afforded a colorless block crystal. The X-ray analysis of the crystal revealed the structure of SP half a century after it was discovered.^[3]

As shown in Figure 1, the most noteworthy feature in the SP structure is the methylene bridge between the two C5 carbons of the thymine residues. The C5 carbon on the 5'-T is chiral as the consequence of the methyl addition; and our structure clearly shows that it adopts an *R* configuration, confirming the assignment by the previous NMR spectroscopic studies.^[4] (Note: in the structure shown in Figure 1, C5 was used to label the methylene bridge between the thymine bases in the PDB file; while C6 was used to label the C5 position of thymine, which adopts an *R* configuration.). The 3'-thymine ring in SP is still planar; in contrast, the 5'-thymine is distorted because of the loss of C5=C6 bond and the subsequent loss of the ring aromaticity.

Both deoxyribose adopt the C3'-*endo* conformation in the structure, contrasting to the C2'-*endo* claimed in the previous NMR studies.^[4] Re-examining the NMR studies revealed that the C2'-*endo* conformation was reached as the values of ($J_{1'2'} + J_{1'2''}$) for the two deoxyribose rings were found to be 11.7 and 15.1 Hz, respectively. The ($J_{1'2'} + J_{1'2''}$) value ranging from 7.1 to 16.1 Hz was suggested to be characteristic for the C2'-*endo* conformation, which we believe is not correct. As the C2'-*endo* and C3'-*endo* conformers are in equilibrium in solution in most cases, a ($J_{1'2'} + J_{1'2''}$) value of 7.1 Hz roughly corresponds to a 100% C3'-*endo* conformer; while a ($J_{1'2'} + J_{1'2''}$) value of 16.1 Hz roughly corresponds to a 100% C2'-*endo* conformer.^[21] Therefore, what was observed in the previous NMR studies also indicates a mixture of C2'- and C3'-*endo* conformers, and the question thus is which conformer dominates?

We therefore re-visited the 1D-NMR studies of both 5R-SP and 5R-CH₂SP species and obtained the coupling constants for the ribose protons. As shown in Table 1, all coupling constants for the deoxyribose attached to the 5'-T can be readily obtained; however, the $J_{3'4'}$ for the deoxyribose at the 3'-end was difficult to determine owing to the poor resolution. The ($J_{1'2'} + J_{1'2''}$) values for the 5R-SP species were found to be 12.1 and 14.2 Hz, respectively, which are comparable to the values determined in the previous NMR studies.^[4] Comparing with the ($J_{1'2'} + J_{1'2''}$) value, the values of $J_{1'2'}/(J_{1'2'} + J_{3'4'})$ and $J_{3'4'}/(J_{1'2'} + J_{3'4'})$ are more commonly used to calculate the percentages of C2'- and C3'-*endo* conformers in the mixture.^[21–22] As shown in Table 1, the calculation shows that the majority of the deoxyriboses at the 5'-end of SP adopt the C3'-*endo* conformation in both 5R-SP and 5R-CH₂SP species. The large $J_{1'2'}$ (7.1 Hz) in the deoxyriboses at the 3'-end of 5R-SP suggests that the C2'-*endo* conformer could be the dominant species as suggested by the previous NMR studies,^[4] although significant amount of C2'-*endo* conformers may still exist in the 3'-end ribose of the 5R-CH₂SP as suggested by the relatively small $J_{1'2'}$ (5.9 Hz) value. However, the absolute percentages cannot be obtained due to the lack of $J_{3'4'}$ value for both SP species. Once crystallized, the C3'-*endo* conformation prevails for both ribose rings, suggesting such a conformation is more stable in the solid state.

The slightly higher percentages of C3'-*endo* conformer for both deoxyriboses in 5R-CH₂SP than its phosphate counter-part seem to suggest that the formacetal linker may induce the ribose to adopt the C3'-*endo* conformation in solution. However, the formacetal linker proved not to change the C2'-*endo* deoxyribose conformation in regular duplex DNA as revealed by a previous NMR spectroscopic studies.^[13b] Moreover, the deoxyriboses in a U_{CH₂}U CPD dimer with a formacetal linker exhibit the C2'-*endo* conformation as well.^[9a]

These results suggest that the linker alone is not responsible for the ribose conformation; such a conformation is more likely due to the intrinsic property of the SP compound.

C3'-*endo* is typically observed in A-form DNA while the C2'-*endo* is the feature of B-form DNA.^[23] It is widely accepted that thymine dimerization in B-form DNA produces CPD as the dominant species; while DNA conformation has to change to an A-like DNA form before SP product can be generated upon UV radiation.^[1a,24] The deoxyribose conformation observed in our structure thus provides a direct evidence to support this assumption. The C–O bonds associated with the formacetal linker are found to be 1.412 Å long, which is about 0.2 Å shorter than the typical P–O bond in DNA.^[23] However, such a bond shortening is suggested to be readily compensated by small changes of the backbone torsional angles and is not projected to result in big DNA conformational change.^[9a]

As shown in our previous deuterium labeling studies, the $H_{6a-pro-S}$ atom of SP originates from the methyl group of the 3'-T in dinucleotide TpT.^[5a] The distance between the $H_{6a-pro-S}$ and the bridging methylene carbon is found to be 2.63 Å in our SP structure (distance indicated by the double-head blue arrow in Figure 1), only 1 Å longer than a typical C–H bond. Should our structure reflect that of natural SP and a direct H atom abstraction occur during SP formation as proposed by our mechanism,^[5a] only relatively small conformational change is needed from the reactive TpT to the resultant SP. In contrast, the distance between the $H_{6a-pro-R}$, which is the retained H atom at 5'-T during SP formation, and the methylene carbon, is 0.7 Å longer at 3.36 Å (distance indicated by the double-head red arrow in Figure 1). This observation supports our findings in previous labeling studies that one of the methyl H atoms migrates to the $H_{6a-pro-S}$ position during SP photo-formation; relatively minimal structural changes are required as the consequence of this H atom migration even in dinucleotide TpT.

Structural comparison of 5R-SP and 5R-CH₂SP via 2D ROESY spectroscopy

To confirm that the formacetal linkage does not alter the base interaction in SP and the 5R-CH₂SP truly reflects the conformation of 5R-SP formed in Nature, we first compared the solution structure of these two species using 2D-ROESY NMR spectroscopy which reflects the H–H correlations through space. As summarized in Figure 2d, strong cross-peaks were observed among protons located in front of the thymine ring ($H_{6a-pro-R}$, CH₃, H_S) or the protons behind this ring ($H_{6a-pro-S}$, H_{6b}, H_R) in ROESY spectrum of 5R-SP.^[4–5] Between the two C6 protons on the 5'-T of 5R-SP, Only the $H_{6a-pro-S}$ exhibits a strong interaction with the deoxyribose protons H_{2'A/2''A} and H_{3'A}. Moreover, although both protons interact with the -CH₃ group, the coupling with $H_{6a-pro-S}$ is weaker than that with $H_{6a-pro-R}$. All these interactions have been observed in the ROESY spectrum of the 5R-CH₂SP species except that the coupling between $H_{6a-pro-S}$ and -CH₃ is also strong. In the two ROESY spectra exhibited in Figure 2, only the coupling interactions among the protons associated with the 5'-T ring are highlighted for clarity purpose. Furthermore, H–H coupling can be typically observed for two protons 3–4 Å away from each other in the ROESY spectrum. Comparing the H–H distances predicted by the ROESY spectrum with those found in the X-ray structure reveals that these two agree with each other fairly well, suggesting that the SP crystal structure obtained in solid state largely reflects its solution structure.

Structural comparison of 5R-SP and 5R-CH₂SP via DFT calculation

Theoretical calculations further support that the formacetal linkage does not change the overall structure of the SP molecule. The 5R-CH₂SP and 5R-SP structures were optimized using density functional theory at the B3LYP/6-31+G(d,p) level and basis set (Figure 3).^[25] The root-mean-square-deviation (RMSD) of common heavy atoms between the optimized 5R-CH₂SP and the crystal structure is 0.290 Å. The RMSD between the optimized 5R-

CH₂SP and 5*R*-SP is 0.372 Å. In 5*R*-CH₂SP, the C–O bond distances associated with the formacetal linkage were calculated to be 1.413 and 1.405 Å, respectively, which are comparable to those revealed by the X-ray structure but shorter than the corresponding P–O bonds of 1.689 Å found in 5*R*-SP, suggesting a slightly tighter linkage between the two deoxyribose rings in 5*R*-CH₂SP. However, such a difference at the ribose interface does not alter the overall structural features of SP, with only negligible differences bond lengths and bond angles obtained in other part of these two species.^[25]

In addition to these essentially identical geometric parameters, a few changes were observed for the stacking configuration between two thymine residues in these two SPs: the dihedral angles between the two base planes are 49.9 and 50.0°, in 5*R*-CH₂SP and 5*R*-SP, respectively. The conservation of angular features in SPs upon PO₂/CH₂ replacement may be ascribed to the geometric restraint imposed by the methylene bridge between two rigid thymine bases, which results a minimum structural impact of the PO₂/CH₂ replacement in the more flexible ribose part. Population analysis of the SPs further suggests that the formacetal linkage does not change the atomic charge distribution in the nucleobase, with the atomic charges in the two SPs essentially the same.^[25]

Structural comparison of 5*R*-SP and 5*R*-CH₂SP via enzymatic reaction

In germinated spores, the SP dimer is repaired by a radical SAM enzyme spore photoproduct lyase (SPL).^[1d–g,6,26] SPL can tolerate a number of substrates, ranging from SP containing plasmid DNA^[1d,f,7] to a 6-mer oligonucleotide.^[1g] SPL even repairs the dinucleotide SP, the minimal substrate possible, to TpT.^[1e,6,27] The phosphate linker in SP proved important to maintain its conformation, enabling its fitting into the enzyme binding pocket. The presence of phosphate results a 50-fold rate increase in enzyme reaction compared with a SP substrate with no such a linker.^[11b,c] Comparing the SPL reactions using SP substrates within DNA strand of different length; the dinucleotide 5*R*-SP results in a relatively slow repair rate due to its weak binding affinity to SPL. Despite that, our recent work proves that it still exhibits all the important substrate properties during SPL catalysis. If 5*R*-CH₂SP adopts a similar conformation as the 5*R*-SP, as proved above, it should support a reasonable SPL activity as well.

We therefore incubated the 5*R*-CH₂SP with the SPL enzyme using the experimental conditions described in our previous studies.^[6] The substrate concentration was kept at ~150-fold of the enzyme concentration; decreasing the concentration by half did not change the reaction rate. This observation suggests that the reaction rate determined very likely reflects the $V_{\max}$. As shown in Figure 4, the TCH₂T species was produced as the consequence of the enzyme repair reaction; plotting the amount of TCH₂T isolated versus the reaction time leads to a $V_{\max}$ of 0.17 – 0.01 min^{–1}, which is roughly half of the $V_{\max}$ determined using 5*R*-SP as the enzyme substrate.^[6] The slightly decreased reaction rate suggests that the phosphate linker in natural SP may be involved in substrate recognition. Of the 5*R* and 5*S* SP isomers, SPL is proved to only take one as its substrate.^[11b,c] The significant repair rate exhibited by 5*R*-CH₂SP thus demonstrates that it is a SPL substrate and its structure revealed in 5*R*-CH₂SP must be very similar to the naturally occurring 5*R*-SP.

Conclusion

In summary, by employing a neutral formacetal linker and synthetic chemistry, a dinucleotide SP isostere was successfully prepared. Such a compound can only be obtained via this lengthy synthetic route. Photolysis of the dinucleotide TCH₂T did not result in any SP analogue formation, indicating that the loss of the negative charge in the phosphate after the linker replacement and/or the smaller size of the formacetal linker change the relative

positions of the thymine residues, making them to no longer support the SP photochemistry. Such a neutral linker greatly facilitates the crystallization process, allowing the SP to be crystallized and its structure determined 50 years after its discovery. ROESY spectroscopy, DFT calculations as well as enzymatic activity studies all verify that the resulting 5*R*-CH₂SP possesses a similar structure to the phosphate-containing 5*R*-SP, suggesting that different from the T_{CH₂}T residue, in SP the methylene bridge formed between the methyl moiety of the 3'-T and the C5 carbon of the 5'-T in the rigid nucleobase region dictates the overall SP structure, making the influence of such a linker change trivial.

Such a simple structure of dinucleotide SP isostere still provides several valuable mechanistic insights regarding to the SP photo-formation. The shorter distance between the H_{6a-pro-S} and the methylene bridge between the two thymine bases supports our mechanism that both the H_{6a-pro-S} and the methylene bridge originate from the methyl group of the 3'-T.^[5a] While the C3'-*endo* conformation, which is typically observed in A-form DNA, for both deoxyribose in 5*R*-CH₂SP supports the assumption that the spore genomic DNA has to change to the A-like conformation due to the binding of the small acid soluble proteins before SP can be produced upon UV irradiation.^[24] Our data further suggests that after SP formation, the methylene bridge between the thymine bases becomes the dominant factor, forcing the two deoxyribose to maintain the A-like conformation even though the DNA binding proteins are absent.

The fact that 5*R*-CH₂SP is still an enzyme substrate suggests that the phosphate moiety between the two thymine residues in SP most likely plays a structural role to hold the two thymine residues in the right positions so that SP can fit into the enzyme binding pocket. Such a linker may also be recognized by some protein residues due to its negative charge as the linker change to a neutral formacetal moiety results a 50% rate decrease in the enzyme reaction. The formacetal linker has proved to be highly useful in studying the DNA repairing enzymes. For instance, the formacetal containing CPD dimer was incorporated into a duplex DNA, enabling the co-crystallization of such a DNA sequence with the CPD repair enzyme - DNA photolyase.^[14] Using this 5*R*-CH₂SP to explore the reaction mechanism of SPL is currently in progress.^[33]

Experimental Section

General methods

Unless otherwise stated all solvents and chemicals used were of commercially available analytical grade. They were purchased from Sigma, Fisher, or VWR and used without further purification. All reactions were carried out 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 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. The ¹H and ¹³C NMR spectra were obtained on a Bruker 500 MHz NMR Fourier transform spectrometer. NMR spectra were recorded in the following solvents: CDCl₃, with residual chloroform (δ 77.2 ppm for ¹³C NMR) and TMS (δ 0 ppm for ¹H NMR), CD₃OH (δ 3.31 ppm for ¹H NMR and δ 49.1 ppm for ¹³C NMR) or [D₆]DMSO, with residual methyl sulfoxide (δ 2.50 ppm for ¹H NMR and δ 39.5 ppm for ¹³C NMR) taken as the standard. MS analysis was obtained using Agilent 1100 series LC/MSD system with Electrospray Ionization (ESI). The T_{CH₂}T photoreactions under UVC light were carried out using a Spectroline germicidal UV sterilizing lamp at 254 nm (Dual-tube, 15 W, intensity: 1550 μWcm⁻²).

All DNA-modifying enzymes and reagents were purchased from Fermentas Life Sciences (Glen Burnie, MD). *B. subtilis* strain 168 chromosomal DNA was purchased from the ATCC (ATCC 23857D). Oligonucleotide primers were obtained from Integrated DNA Technologies (Coralville, IA). *E. coli* BL21(DE3) and expression vector pET-28a were purchased from Novagen (Madison, WI). The construct containing the SPL gene was co-expressed with plasmid pDB1282, which was a generous gift from Professor Squire Booker at the Pennsylvania State University. 5'-deoxyadenosine (5'-dA) and S-adenosylmethionine (SAM) was purchased from Aldrich and used directly without further purification. 5 Prime Perfect-pro* Nickel nitrilotriacetic acid (Ni-NTA) resin was purchased from Fisher Scientific. All other buffers and chemicals were of the highest grade available.

The protein purification as well as the enzyme reactions were carried out under an inert atmosphere using a CoyLab anaerobic chamber (Grass Lake, MI) with the H₂ concentration around 3%. Product analysis was conducted using a Waters (Milford, MA) breeze HPLC system with a 2489 UV/Visible detector and an Agilent 1100 LC-MS system under both a positive and negative ionization mode. Sonic disruption of *E. coli* cells was conducted by a Fisher Scientific* Model 500 Digital Sonic Dismembrator. DNA sequencing was performed by the Indiana University Nucleic Acid Core Facility at IU School of Medicine.

T_{CH₂}T photochemical studies

The photolysis of T_{CH₂}T was conducted using the same procedure in our previous studies.^[5a] Briefly, a 5 mL 1 mM aqueous solution of T_{CH₂}T was prepared by titration with 0.3 M NaOH to pH 7.0. After removal of water under vacuum, the resulting T_{CH₂}T sodium salt was dissolved in MeOH (5 mLs). The solution was transferred to a 10 cm petri dish and methanol evaporation afforded a nice thin film. The film was then exposed for 45 min to the UVC (254 nm) light. The products were subsequently dissolved in water and analyzed by reverse phase HPLC in the gradient mode using triethyl ammonium acetate, pH 6.5, and acetonitrile as solvents.

In another set of experiments, the resulting T_{CH₂}T sodium salt was re-dissolved in a glycerol/water solution (1:1, 5mL total) and flash freeze in liquid N₂ to obtain a transparent glass. The glass was subsequently irradiated for 45 min under 254 nm UV light in liquid N₂. The glass was allowed to warm up to ambient temperature and the UV reaction products in the glycerol/water solution analyzed directly by HPLC.

DFT calculation

The computational procedures are described below. To obtain an initial geometry of 5R-SP, we started from the crystal structure of 5R-CH₂SP, where we replaced the carbon atom in the CH₂ linkage by a phosphorous atom. Then the two oxygen atoms in PO₂⁻ were placed at initial positions that are 1.6 Å away from the phosphorous atom along the crystallographic C-H bond directions. We noticed that there is a crystal water molecule captured in the X-ray structure, where inclusion of the crystal water in the models introduced additional computational cost for optimizing the intermolecular interaction between SPs and the solvent. Our preliminary calculations showed that the presence of the crystal water did not change the essential features of the SPs in geometry and charge distributions. Therefore, we omitted the crystal water from our models in the subsequent calculations. The 5R-CH₂SP crystal structure and the 5R-SP structure generated with above procedures were then optimized with density function theory (DFT) calculations at the B3LYP/6-31+G(d,p) level and basis set, by using the electronic structure package Gaussian03.^[28]

The computationally optimized structures were superimposed with the original X-ray structure based on common heavy atoms using the macro-molecular modeling package

CHARMM,^[29] where for the comparison of 5*R*-CH₂SP and 5*R*-SP, the O/P atoms involved in the CH₂/PO₂⁻ replacement were included in the superposition, but the two oxygen atoms in PO₂⁻ were excluded. The dihedral angles between the two thymine planes were calculated by the least square fitting plane utility (LSQP) of the COOR module in CHARMM, where 12 atoms are included to define the 3'-thymine ring, and 10 atoms are included to define the 5'-thymine ring.

Chemical synthesis of 5*R*-CH₂SP

As shown in Scheme 2, the synthesis of the 5*R*-CH₂SP was achieved via a modified procedure used in the synthesis of 5*R*-SP originally adopted by Begley and co-workers.^[30]

N³-Benzyloxymethyl-3',5'-O-di(*tert*-butyldimethylsilyl)-5-bromomethyl-2'-deoxyuridine (2)

A mixture of **1** (10.0 g, 18.2 mmol), *i*Pr₂NEt (11.1 mL, 63.7 mmol) and BnOCH₂Cl (6.35 mL, 45.5 mmol) in anhydrous CH₂Cl₂ (25 mL) was stirred for 1 d.^[31] The reaction solution was diluted with CH₂Cl₂ (200 mL), washed with saturated aqueous sodium bicarbonate (3×20 mL) and dried over anhydrous sodium sulfate. After filtration and concentration under vacuum, the residue was purified by flash chromatography (hexane/EtOAc 6:1) to afford **2** as a yellow oil (4.40 g, 36%). ¹H NMR (CDCl₃): δ = 0.08 (s, 3H), 0.09 (s, 3H), 0.12 (s, 3H), 0.13 (s, 3H), 0.90 (s, 9H), 0.94 (s, 9H), 1.96 (ddd, *J*=5.9, 7.7, 13.5 Hz, 1H), 2.33 (ddd, *J*=2.6, 5.8, 13.5 Hz, 1H), 3.77 (dd, *J*=2.6, 11.4 Hz, 1H), 3.87 (dd, *J*=2.8, 11.4 Hz, 1H), 3.96–4.00 (m, 1H), 4.32 (d, *J*=12.0 Hz, 1H), 4.36–4.41 (m, 1H), 4.39 (d, *J*=12.0 Hz, 1H), 4.71 (s, 2H), 5.50 (s, 2H), 6.29 (dd, *J*=5.8, 7.7 Hz, 1H), 7.24–7.28 (m, 1H), 7.29–7.35 (m, 2H), 7.35–7.39 (m, 2H), 7.82 ppm (s, 1H); ¹³C NMR (CDCl₃): δ = -5.3 (2C), -4.7, -4.5, 18.1, 18.5, 25.8, 26.1, 39.5, 41.9, 63.1, 70.7, 72.4, 72.5, 86.3, 88.3, 110.7, 127.8, 128.4, 137.9, 138.1, 150.5, 161.6 ppm; ESI-MS (positive ion): *m/z*: calcd for C₃₀H₅₀BrN₂O₆Si₂⁺: 669.2 [*M*+H]⁺, found 669.2.

N³-Benzyloxymethyl-3',5'-O-di(triethylsilyl)-5,6-dihydrothymidine (4)

A mixture of **3** (3.40 g, 7.19 mmol), *i*Pr₂NEt (5.66 mL, 32.5 mmol) and BnOCH₂Cl (3.00 mL, 21.7 mmol) in anhydrous CH₂Cl₂ (15 mL) was stirred for 36 h.^[11d] The reaction solution was diluted with CH₂Cl₂ (150 mL), washed with saturated aqueous sodium bicarbonate (3× 10 mL) and dried over anhydrous sodium sulfate. After filtration and concentration under vacuum, the residue was purified by flash chromatography (hexane/EtOAc 6:1) to afford **4** as a colorless oil (3.27 g, 77%). ¹H NMR (CDCl₃): δ = 0.61 (q, *J* = 7.9 Hz, 12H), 0.96 (t, *J* = 7.9 Hz, 9H), 0.97 (t, *J*=7.9 Hz, 9H), 1.21 (d, *J*=7.1 Hz, 3H), 1.92–1.98 (m, 2H), 2.54 (ddq, *J*=5.2, 8.8, 7.1 Hz, 1H), 3.08 (dd, *J*=8.8, 12.9 Hz, 1H), 3.17 (dd, *J*=5.2, 12.9 Hz, 1H), 3.67–3.71 (m, 2H), 3.80–3.84 (m, 1H), 4.32–4.36 (m, 1H), 4.66 (s, 2H), 5.33 (d, *J*=9.9 Hz, 1H), 5.36 (d, *J*=9.9 Hz, 1H), 6.36 (dd, *J*=6.3, 8.1 Hz, 1H), 7.23–7.27 (m, 1H), 7.27–7.33 (m, 2H), 7.33–7.37 ppm (m, 2H); ¹³C NMR (CDCl₃): δ = 4.3, 4.8, 6.8, 13.4, 35.8, 38.1, 40.5, 62.9, 70.4, 71.8, 72.2, 84.6, 86.7, 127.5, 127.6, 128.3, 138.6, 153.0, 172.8 ppm; ESI-MS (positive ion): *m/z*: calcd for C₃₀H₅₃N₂O₆Si₂⁺: 593.3 [*M*+H]⁺, found 593.3.

Synthesis of 5

LDA solution (3.20 mL, 2.00M in THF) was added to a solution of dihydrothymidine **4** (2.50 g, 4.22 mmol) in anhydrous THF (25 mL) at -78°C. The reaction mixture was stirred at -78°C for 3 h before the addition of **2** (2.90 g, 4.33 mmol, 0.25M in THF). The reaction mixture was stirred at -78°C for 3 h and allowed to warm to room temperature for overnight. The reaction was quenched by saturated aqueous sodium bicarbonate (40 mL) and the aqueous phase was extracted with CH₂Cl₂ (3×100 mL). The collected extracts were dried over anhydrous sodium sulfate. After filtration and concentration under vacuum, the

residue was purified by flash chromatography (hexane/EtOAc 6:1) to afford isomer **5R-5** as a pale yellow oil (1.95 g, 39%). Further purification by flash chromatography (hexane/CH₂Cl₂/EtOAc 2:1:0.3) afforded isomer **5S-5** as a pale yellow oil (0.74 g, 15%). **5R-5**: ¹H NMR (CDCl₃): δ = 0.08 (s, 3H), 0.09 (s, 3H), 0.12 (s, 6H), 0.60 (q, *J*=7.9 Hz, 6H), 0.65 (q, *J*=8.0 Hz, 6H), 0.89 (s, 9H), 0.93 (s, 9H), 0.96 (t, *J*=8.0 Hz, 9H), 0.98 (t, *J*=7.9 Hz, 9H), 1.18 (s, 3H), 1.85–1.98 (m, 3H), 2.10–2.18 (m, 1H), 2.62 (d, *J*=14.1 Hz, 1H), 2.68 (d, *J*=14.1 Hz, 1H), 2.91 (d, *J*=13.0 Hz, 1H), 3.11 (d, *J*=13.0 Hz, 1H), 3.60 (dd, *J*=5.2, 10.4 Hz, 1H), 3.68 (dd, *J*=3.3, 10.7 Hz, 1H), 3.74 (dd, *J*=5.0, 10.7 Hz, 1H), 3.77–3.83 (m, 2H), 3.86–3.92 (m, 1H), 4.29–4.33 (m, 1H), 4.39–4.44 (m, 1H), 4.55 (d, *J*=12.4 Hz, 1H), 4.61 (d, *J*=12.4 Hz, 1H), 4.66 (d, *J*=12.1 Hz, 1H), 4.69 (d, *J*=12.1 Hz, 1H), 5.32 (s, 2H), 5.47 (s, 2H), 6.30 (t, *J*=6.6 Hz, 1H), 6.33 (t, *J*=7.1 Hz, 1H), 7.20–7.27 (m, 6H), 7.27–7.32 (m, 2H), 7.32–7.37 (m, 2H), 7.49 ppm (s, 1H); ¹³C NMR (CDCl₃): δ = −5.3, −5.2, −4.8, −4.6, 4.4, 4.8, 6.8, 6.9, 18.0, 18.5, 22.1, 25.8, 26.1, 32.4, 36.9, 40.2, 43.0, 44.2, 63.0, 63.1, 70.9, 71.9, 72.0, 72.2, 72.3, 84.4, 85.4, 86.4, 87.6, 108.6, 127.5, 127.6, 127.7, 128.3, 128.4, 138.1, 138.5, 139.3, 150.8, 152.4, 163.6, 173.7 ppm; ESI-MS (positive ion): *m/z*: calcd for C₆₀H₁₀₁N₄O₁₂Si₄⁺: 1181.7 [*M*+H]⁺, found 1181.7; **5S-5**: ¹H NMR (CDCl₃): δ = 0.09 (s, 12H), 0.58–0.66 (m, 12H), 0.90 (s, 18H), 0.97 (t, *J*=7.9 Hz, 18H), 1.11 (s, 3H), 1.88–2.00 (m, 3H), 2.28 (dd, *J*=4.7, 12.7 Hz, 1H), 2.43 (d, *J*=14.2 Hz, 1H), 2.78 (d, *J*=13.0 Hz, 1H), 2.79 (d, *J*=14.2 Hz, 1H), 3.18 (d, *J*=13.0 Hz, 1H), 3.59 (dd, *J*=4.9, 10.4 Hz, 1H), 3.65–3.80 (m, 3H), 3.81–3.87 (m, 1H), 3.90–3.96 (m, 1H), 4.30–4.35 (m, 1H), 4.35–4.39 (m, 1H), 4.61 (s, 2H), 4.65 (s, 2H), 5.25 (d, *J*=9.9 Hz, 1H), 5.37 (d, *J*=9.9 Hz, 1H), 5.44 (s, 2H), 6.23 (t, *J*=6.7 Hz, 1H), 6.37 (t, *J*=7.0 Hz, 1H), 7.20–7.33 (m, 8H), 7.33–7.38 (m, 2H), 7.41 ppm (s, 1H); ¹³C NMR (CDCl₃): δ = −5.4, −5.2, −4.7, −4.6, 4.4, 4.8, 6.8, 6.9, 18.0, 18.4, 20.8, 25.8, 26.0, 32.5, 37.2, 40.8, 42.3, 44.7, 63.4, 63.5, 70.7, 71.0, 71.9, 72.2, 72.8, 72.9, 85.0, 86.3, 86.8, 88.1, 108.6, 127.4, 127.5, 127.6, 127.7, 128.3, 128.4, 138.0, 138.1, 138.9, 150.7, 152.5, 163.4, 173.5 ppm; ESI-MS (positive ion): *m/z*: calcd for C₆₀H₁₀₁N₄O₁₂Si₄⁺: 1181.7 [*M*+H]⁺, found 1181.7.

Selective deprotection of **5R-5** to yield **6**

Compound **5R-5** (2.00 g, 1.69 mmol) was dissolved in CH₃CN (30 mL) and cooled to −20 °C. A HF-pyridine solution (1.70M in acetonitrile, 20.3 mmol, 11.9 mL) was added and the reaction mixture was stirred at −20 °C for 4 h. The reaction was quenched by saturated aqueous NaHCO₃ (40 mL). The mixture was extracted with EtOAc (3×150 mL). The collected extracts were dried over anhydrous sodium sulfate. After filtration and concentration under vacuum, the residue was purified by flash chromatography (hexane/EtOAc/MeOH 1:1:0.2) to afford **6** as a colorless oil (1.25 g, 78%). ¹H NMR (CDCl₃): δ = 0.09 (s, 6H), 0.12 (s, 6H), 0.90 (s, 9H), 0.93 (s, 9H), 1.11 (s, 3H), 1.95 (ddd, *J*=5.9, 7.9, 13.4 Hz, 1H), 2.00 (ddd, *J*=3.3, 6.4, 13.6 Hz, 1H), 2.08–2.15 (m, 1H), 2.20 (ddd, *J*=2.6, 5.8, 13.4 Hz, 1H), 2.68 (d, *J*=14.3 Hz, 1H), 2.69 (brs, 1H), 2.74 (d, *J*=14.3 Hz, 1H), 2.93 (brs, 1H), 2.94 (d, *J*=13.1 Hz, 1H), 3.35 (d, *J*=13.1 Hz, 1H), 3.43–3.65 (m, 1H), 3.58–3.65 (m, 1H), 3.73–3.82 (m, 3H), 3.92–3.97 (m, 1H), 4.26–4.31 (m, 1H), 4.39–4.43 (m, 1H), 4.58 (d, *J*=12.4 Hz, 1H), 4.62 (d, *J*=12.4 Hz, 1H), 4.68 (d, *J*=11.8 Hz, 1H), 4.71 (d, *J*=11.8 Hz, 1H), 5.28 (d, *J*=9.9 Hz, 1H), 5.35 (d, *J*=9.9 Hz, 1H), 5.41 (d, *J*=9.4 Hz, 1H), 5.52 (d, *J*=9.4 Hz, 1H), 6.19 (dd, *J*₁ = *J*₂ = 6.8 Hz, 1H), 6.33 (dd, *J*=5.9, 7.9 Hz, 1H), 7.22–7.29 (m, 6H), 7.30–7.34 (m, 2H), 7.34–7.38 (m, 2H), 7.54 ppm (s, 1H); ¹³C NMR (CDCl₃): δ = −5.3, −5.2, −4.7, −4.6, 18.0, 18.5, 21.0, 25.8, 26.1, 32.0, 37.5, 40.7, 42.3, 44.8, 62.5, 63.2, 70.8, 70.9, 71.9, 72.0, 72.4, 72.5, 85.5, 85.8, 86.1, 87.9, 108.7, 127.5, 127.7, 127.8, 128.1, 128.4, 128.5, 137.6, 138.5, 139.1, 150.7, 152.3, 163.7, 174.3 ppm; ESI-MS (positive ion): *m/z*: calcd for C₄₈H₇₃N₄O₁₂Si₂⁺: 953.5 [*M*+H]⁺, found 953.5.

Selective protection of **6** to yield **7**

A solution of **6** (1.06 g, 1.11 mmol), *tert*-butyl diphenylchlorosilane (TBDPSCI) (0.31 mL, 1.20 mmol) and imidazole (152 mg, 2.23 mmol) in DMF (5 mL) was stirred at 0 °C for 16 h. The reaction was allowed to warm to room temperature and stirred for another 20 h. The reaction was quenched by water (30 mL). The mixture was extracted with Et₂O (5×50 mL). The collected extracts were washed with brine (20 mL) and dried over anhydrous sodium sulfate. After filtration and concentration under vacuum, the residue was purified by flash chromatography (hexane/EtOAc 1:1) to afford **7** as a colorless oil (990 mg, 75%). ¹H NMR (CDCl₃): δ = 0.09 (s, 3H), 0.10 (s, 3H), 0.11 (s, 3H), 0.12 (s, 3H), 0.90 (s, 9H), 0.92 (s, 9H), 1.06 (s, 3H), 1.08 (s, 9H), 1.90–1.97 (m, 1H), 2.00–2.06 (m, 2H), 2.16 (ddd, *J*=3.2, 6.1, 13.4 Hz, 1H), 2.44 (d, *J*=3.9 Hz, 1H), 2.53 (d, *J*=14.2 Hz, 1H), 2.60 (d, *J*=14.2 Hz, 1H), 2.92 (d, *J*=13.1 Hz, 1H), 3.04 (d, *J*=13.1 Hz, 1H), 3.72 (dd, *J*=5.4, 11.1 Hz, 1H), 3.75–3.85 (m, 4H), 3.88–3.92 (m, 1H), 4.36–4.44 (m, 2H), 4.54 (d, *J*=12.4 Hz, 1H), 4.60 (d, *J*=12.4 Hz, 1H), 4.64 (d, *J*=12.2 Hz, 1H), 4.67 (d, *J*=12.2 Hz, 1H), 5.29 (d, *J*=10.0 Hz, 1H), 5.33 (d, *J*=10.0 Hz, 1H), 5.43 (s, 2H), 6.29 (dd, *J*₁ = *J*₂ = 7.2 Hz, 1H), 6.34 (dd, *J*₁ = *J*₂ = 7.0 Hz, 1H), 7.17–7.25 (m, 6H), 7.25–7.31 (m, 2H), 7.32–7.35 (m, 2H), 7.36–7.43 (m, 6H), 7.44 (s, 1H), 7.66–7.71 ppm (m, 4H); ¹³C NMR (CDCl₃): δ = −5.3 (2C), −4.8, −4.7, 18.0, 18.4, 19.3, 21.9, 25.8, 26.0, 27.0, 32.4, 36.5, 40.2, 42.7, 44.7, 62.9, 64.3, 70.8, 70.9, 71.9, 72.1, 72.2, 83.9, 84.6, 85.4, 87.5, 108.4, 127.4, 127.6, 127.7, 127.9 (2C), 128.3 (2C), 129.9, 130.0, 132.9, 133.2, 135.5, 135.6, 137.9, 138.4, 139.1, 150.7, 152.4, 163.5, 173.4 ppm; ESI-MS (positive ion): *m/z*: calcd for C₆₄H₉₁N₄O₁₂Si₃⁺: 1191.6 [*M*+H]⁺, found 1191.7.

Synthesis of **8**

A solution of **7** (910 mg, 0.76 mmol), Ac₂O (3.60 mL) and AcOH (1.15 mL) in DMSO (5 mL) was stirred at room temperature for 50 h. The reaction was quenched by addition of an aq. saturated NaHCO₃ solution (20 mL) at 0 °C and stirred for another 1 h. The aqueous phase was extracted with Et₂O (4×30 mL). The combined organic layers were washed with water (5×50 mL) and dried with Na₂SO₄. After filtration and concentration under vacuum, the residue was purified by flash chromatography (hexane/EtOAc 4:1) to afford **8** as a white solid (830 mg, 87%). ¹H NMR (CDCl₃): δ = 0.08 (s, 3H), 0.09 (s, 3H), 0.10 (s, 3H), 0.11 (s, 3H), 0.90 (s, 9H), 0.92 (s, 9H), 1.09 (s, 3H), 1.10 (s, 9H), 1.90–1.98 (m, 2H), 2.04 (ddd, *J*=2.3, 6.0, 13.8 Hz, 1H), 2.10 (s, 3H), 2.16 (ddd, *J*=3.1, 6.0, 13.4 Hz, 1H), 2.55 (d, *J*=14.2 Hz, 1H), 2.61 (d, *J*=14.2 Hz, 1H), 2.93 (d, *J*=13.0 Hz, 1H), 3.07 (d, *J*=13.0 Hz, 1H), 3.71–3.81 (m, 4H), 3.88–3.91 (m, 1H), 3.92–3.95 (m, 1H), 4.40–4.43 (m, 1H), 4.47–4.51 (m, 1H), 4.56 (d, *J*=12.4 Hz, 1H), 4.58 (d, *J*=11.6 Hz, 1H), 4.61 (d, *J*=12.4 Hz, 1H), 4.62 (d, *J*=11.6 Hz, 1H), 4.64 (d, *J*=12.5 Hz, 1H), 4.67 (d, *J*=12.5 Hz, 1H), 5.30 (d, *J*=10.0 Hz, 1H), 5.34 (d, *J*=10.0 Hz, 1H), 5.44 (s, 2H), 6.28 (dd, *J*=4.0, 5.5 Hz, 1H), 6.30 (t, *J*=5.7 Hz, 1H), 7.18–7.27 (m, 6H), 7.27–7.32 (m, 2H), 7.32–7.36 (m, 2H), 7.37–7.43 (m, 6H), 7.44 (s, 1H), 7.67–7.72 ppm (m, 4H); ¹³C NMR (CDCl₃): δ = −5.3, −5.2, −4.8, −4.7, 13.8, 18.0, 18.5, 19.3, 21.9, 25.8, 26.0, 27.0, 32.4, 33.5, 40.2, 42.8, 44.5, 62.9, 63.8, 70.7, 70.9, 71.9 (2C), 72.2, 73.5, 76.0, 83.1, 84.3, 85.4, 87.5, 108.5, 127.4, 127.6 (2C), 127.7, 127.9, 128.3 (2C), 129.9 (2C), 133.0, 133.2, 135.6 (2C), 138.0, 138.4, 139.1, 150.7, 152.4, 163.4, 173.5 ppm; ESI-MS (positive ion): *m/z*: calcd for C₆₆H₉₅N₄O₁₂SSi₃⁺: 1251.6 [*M*+H]⁺, found 1251.6.

Mono-deprotection of **8** to yield **9**

A solution containing **8** (1.10 g, 0.88 mmol) and pyridinium *p*-toluenesulfonate (0.90 g, 3.58 mmol) in MeOH (30 mL) was stirred at 0 °C for 1 h. The reaction mixture was allowed to warm to room temperature and stirred for overnight. After quenched by ice-cooled saturated aqueous sodium bicarbonate (20 mL), the reaction mixture was diluted with CH₂Cl₂ (200 mL), washed with saturated aqueous sodium bicarbonate (3×10 mL) and dried over anhydrous sodium sulfate. After filtration and concentration under vacuum, the residue was purified by flash chromatography (hexane/EtOAc 2:1) to afford **9** as a colorless oil (0.61 g,

62%). ^1H NMR (CDCl_3): δ = 0.09 (s, 6H), 0.91 (s, 9H), 1.12 (s, 12H), 1.95 (ddd, J =2.5, 7.0, 13.6 Hz, 1H), 2.05 (ddd, J =1.9, 5.7, 13.6 Hz, 1H), 2.09–2.16 (m, 1H), 2.13 (s, 3H), 2.33 (ddd, J =2.8, 6.2, 13.6 Hz, 1H), 2.51 (d, J =14.2 Hz, 1H), 2.70 (d, J =14.2 Hz, 1H), 2.87 (d, J =13.1 Hz, 1H), 3.11 (d, J =13.1 Hz, 1H), 3.15 (t, J =4.9 Hz, 1H), 3.70 (ddd, J =3.1, 4.9, 12.0 Hz, 1H), 3.76 (dd, J =5.6, 11.0 Hz, 1H), 3.81 (dd, J =4.3, 11.0 Hz, 1H), 3.94 (ddd, J =2.4, 4.8, 12.0 Hz, 1H), 3.94–3.99 (m, 1H), 4.00–4.03 (m, 1H), 4.41–4.45 (m, 1H), 4.50–4.55 (m, 1H), 4.59 (d, J =12.5 Hz, 1H), 4.61 (d, J =11.6 Hz, 1H), 4.65 (d, J =11.6 Hz, 1H), 4.68 (d, J =12.5 Hz, 1H), 4.71 (s, 2H), 5.32 (s, 2H), 5.49 (d, J =10.2 Hz, 1H), 5.51 (d, J =10.2 Hz, 1H), 6.16 (dd, $J_1 = J_2 = 6.7$ Hz, 1H), 6.30 (dd, J = 5.7, 9.0 Hz, 1H), 7.21–7.35 (m, 8H), 7.35–7.39 (m, 2H), 7.39–7.47 (m, 6H), 7.71–7.75 (m, 4H), 7.80 ppm (s, 1H); ^{13}C NMR (CDCl_3): δ = -4.8, -4.7, 13.9, 18.0, 19.3, 21.1, 21.7, 25.8, 27.0, 32.1, 33.5, 41.8, 42.8, 44.6, 62.1, 63.9, 70.7, 71.0, 72.2, 72.3, 72.5, 73.5, 76.1, 83.2, 84.3, 87.7, 88.6, 107.6, 127.6 (2C), 127.7 (2C), 127.9, 128.2, 128.3, 129.9 (2C), 133.0, 133.1, 135.6 (2C), 138.0, 138.2, 139.7, 150.6, 152.3, 163.6, 174.2 ppm; ESI-MS (positive ion): m/z : calcd for $\text{C}_{60}\text{H}_{81}\text{N}_4\text{O}_{12}\text{SSi}_2^+$: 1137.5 [$M + \text{H}$] $^+$, found 1137.5.

Preparation of 10

To a cooled (0°C) solution of **9** (420 mg, 0.37 mmol) and molecular sieve in anhydrous 1,2-dichloroethane (8 mL) was added a freshly mixed solution of *N*-iodosuccinimide (NIS) (90 mg, 0.40 mmol) and TfOH (5.00 μL , 55.5 μmol) in 1,2-dichloroethane/diethyl ether 1:1 (5 mL). After 30 min, the reaction mixture was filtered, diluted with CH_2Cl_2 (50 mL), washed with aq. saturated $\text{Na}_2\text{S}_2\text{O}_3/\text{NaHCO}_3$ solution (5 mL). The organic layer was dried over anhydrous sodium sulfate. After filtration and concentration under vacuum, the residue was purified by flash chromatography (hexane/EtOAc=2/1) to afford **10** as a white solid (170 mg, 42%). ^1H NMR (CDCl_3): δ = 0.10 (s, 3H), 0.11 (s, 3H), 0.91 (s, 9H), 1.05 (s, 9H), 1.09 (s, 3H), 1.98 (ddd, $J_1 = J_2 = 7.4$ Hz, $J_3 = 13.0$ Hz, 1H), 2.23 (ddd, J =5.2, 7.5, 13.0 Hz, 1H), 2.32 (d, J =14.3 Hz, 1H), 2.37–2.46 (m, 2H), 2.80 (d, J =14.3 Hz, 1H), 2.90 (d, J =13.0 Hz, 1H), 3.07 (d, J =13.0 Hz, 1H), 3.48 (dd, J =2.0, 11.0 Hz, 1H), 3.74–3.82 (m, 2H), 3.91–3.98 (m, 2H), 4.13–4.20 (m, 2H), 4.52–4.56 (m, 1H), 4.60–4.67 (m, 4H), 4.72 (d, J =7.2 Hz, 1H), 4.77 (d, J =7.2 Hz, 1H), 5.28 (d, J =9.8 Hz, 1H), 5.32 (d, J =9.8 Hz, 1H), 5.39 (s, 2H), 6.07 (dd, J =5.7, 8.5 Hz, 1H), 6.33 (dd, J =4.9, 7.9 Hz, 1H), 7.20–7.25 (m, 3H), 7.25–7.30 (m 5H), 7.32–7.38 (m, 9H), 7.62–7.67 ppm (m, 4H); ^{13}C NMR (CDCl_3): δ = -4.8, -4.6, 18.0, 19.4, 22.7, 25.8, 27.1, 36.3, 41.0, 41.1, 45.3, 62.6, 65.8, 70.7, 71.0, 71.1, 71.9, 72.3, 77.8, 82.1, 83.2, 84.9, 85.3, 96.4, 109.3, 127.5, 127.6, 127.7, 127.9 (2C), 128.3 (2C), 130.0 (2C), 132.9, 133.2, 135.5, 135.6, 136.5, 138.1, 138.5, 150.6, 153.8, 162.8, 173.2 ppm; ESI-MS (positive ion): m/z : calcd for $\text{C}_{59}\text{H}_{77}\text{N}_4\text{O}_{12}\text{Si}_2^+$: 1089.5 [$M + \text{H}$] $^+$, found 1089.5.

De-benzyloxylation of 10 to yield 11

A mixture of **10** (140 mg, 0.13 mmol) and DDQ (325 mg, 1.43 mmol) in wet 1,2-dichloroethane (40 mL) was stirred at 55°C for 1 day. The reaction solution was diluted with CH_2Cl_2 (150 mL), and washed with aq. saturated NaHCO_3 solution (3 $\times$ 10 mL). The organic layer was dried over anhydrous sodium sulfate. After filtration and concentration under vacuum, the residue was purified by flash chromatography (hexane/EtOAc/MeOH 1:1:0.1). The collected crude product was dissolved in a $\text{CH}_2\text{Cl}_2/i\text{Pr}_2\text{NEt}$ solution (20:1, 10 mL) and refluxed overnight. After concentration under vacuum, the residue was purified by flash chromatography (hexane/EtOAc/MeOH 1:1:0.1) to afford **11** as a white solid (92 mg, 84%). ^1H NMR (CDCl_3): δ = 0.10 (s, 3H), 0.11 (s, 3H), 0.90 (s, 9H), 1.05 (s, 9H), 1.18 (s, 3H), 2.13 (ddd, $J_1 = J_2 = 7.9$ Hz, $J_3 = 13.5$ Hz, 1H), 2.19 (d, J = 14.4 Hz, 1H), 2.26–2.38 (m, 3H), 2.92 (d, J =13.2 Hz, 1H), 3.11 (d, J =14.4 Hz, 1H), 3.30 (d, J = 13.2 Hz, 1H), 3.50 (d, J =10.9 Hz, 1H), 3.66–3.72 (m, 1H), 3.77 (dd, J = 3.2, 11.7 Hz, 1H), 3.82–3.86 (m, 1H), 3.96 (dd, J =2.3, 11.7 Hz, 1H), 4.08–4.15 (m, 1H), 4.14 (d, J =10.9 Hz, 1H), 4.55–4.61 (m, 1H), 4.73 (d, J =7.3 Hz, 1H), 4.77 (d, J =7.3 Hz, 1H), 6.28 (t, J =7.2 Hz, 1H), 6.32 (dd, J =4.3, 8.0

Hz, 1H), 7.19 (s, 1H), 7.34–7.39 (m, 4H), 7.39–7.43 (m, 2H), 7.63–7.68 (m, 4H), 9.11 (s, 1H), 10.50 ppm (s, 1H); ^{13}C NMR (CDCl_3): δ = -4.8, -4.6, 18.0, 19.4, 24.1, 25.7, 27.0, 36.4, 40.0, 40.5, 46.5, 62.0, 64.7, 69.8, 77.6, 81.6, 82.1, 82.7, 84.3, 96.3, 110.9, 127.7, 127.8, 130.0 (2C), 133.1, 133.2, 135.5, 135.6, 137.5, 150.2, 154.0, 164.2, 173.0 ppm; ESI-MS (positive ion): m/z : calcd for $\text{C}_{43}\text{H}_{61}\text{N}_4\text{O}_{10}\text{Si}_2^+$: 849.4 $[M+H]^+$, found 849.4.

De-protection of **11** to yield **5R-CH₂SP**

To a solution of **11** (85 mg, 0.10 mmol) in THF (10 mL) was added TBAF (1.0M in THF, 0.30 mL) and stirred for overnight. After the evaporation of the solvent under vacuum, the residue was then purified by reverse phase HPLC in the gradient mode using water/ CH_3CN as solvent. The compound **5R-CH₂SP** was isolated as colorless micro-crystals (45 mg, 90%). ^1H NMR ($[\text{D}_6]\text{DMSO}$): δ = 1.09 (s, 3H), 2.09–2.22 (m, 3H), 2.30 (ddd, J = 3.4, 7.8, 14.2 Hz, 1H), 2.55 (d, J = 14.9 Hz, 1H), 2.59 (d, J = 14.9 Hz, 1H), 3.13 (d, J = 12.9 Hz, 1H), 3.18 (d, J = 12.9 Hz, 1H), 3.40 (dd, J = 4.1, 12.1 Hz, 1H), 3.49 (ddd, J = 2.9, 4.1, 7.3 Hz, 1H), 3.54 (dd, J = 3.7, 10.9 Hz, 1H), 3.58 (dd, J = 2.9, 12.1 Hz, 1H), 3.81–3.87 (m, 2H), 3.99–4.06 (m, 1H), 4.32–4.39 (m, 1H), 4.72 (d, J = 7.0 Hz, 1H), 4.74 (brs, 1H), 4.75 (d, J = 7.0 Hz, 1H), 5.31 (brs, 1H), 5.95 (dd, J = 5.9, 8.6 Hz, 1H), 5.99 (dd, J = 3.3, 8.6 Hz, 1H), 7.59 (s, 1H), 10.03 (s, 1H), 11.28 ppm (brs, 1H); ^{13}C NMR ($[\text{D}_6]\text{DMSO}$): δ = 23.2, 34.8, 35.2, 38.7, 40.0, 45.1, 60.3, 66.3, 69.4, 76.6, 82.0, 82.1, 83.6, 84.6, 95.4, 109.7, 138.2, 150.6, 153.3, 163.4, 174.3 ppm; ESI-MS (positive ion): m/z : calcd for $\text{C}_{21}\text{H}_{29}\text{N}_4\text{O}_{10}^+$: 497.2 $[M+H]^+$, found 497.2.

Synthesis of **T_{CH₂T}**

As shown in Scheme 3, the dinucleotide **T_{CH₂T}** was synthesized via a modified procedure used in the synthesis of **U_{CH₂U}** originally adopted by Carell and co-workers.^[9a]

5'-O-Pivaloylthymidine (**13**)

To a solution of thymidine (**12**, 5.00 g, 20.7 mmol) in pyridine (70 mL) was added trimethylacetyl chloride (PivCl) (2.26 mL, 24.8 mmol) at 0°C. The reaction was allowed to warm to room temperature and stirred overnight. After quenching with MeOH (5 mL), the solvent was removed under vacuum. The residue was purified by flash chromatography (hexane/EtOAc/MeOH 1:1:0.2) to afford compound **13** as a colorless oil (5.30 g, 78%). ^1H NMR (CDCl_3): δ = 1.23 (s, 9H), 1.91 (s, 3H), 2.08 (ddd, $J_1 = J_2 = 6.6$ Hz, $J_3 = 13.8$ Hz, 1H), 2.52 (ddd, J = 2.6, 5.7, 13.8 Hz, 1H), 4.22–4.29 (m, 2H), 4.33 (brs, 1H), 4.37–4.44 (m, 2H), 6.33 (dd, J = 6.0, 7.7 Hz, 1H), 7.33 (s, 1H), 10.32 ppm (s, 1H); ^{13}C NMR (CDCl_3): δ = 12.5, 27.2, 38.9, 40.5, 64.1, 71.5, 84.7, 85.1, 111.3, 135.2, 150.8, 164.4, 178.5 ppm; ESI-MS (positive ion): m/z : calcd for $\text{C}_{15}\text{H}_{23}\text{N}_2\text{O}_6^+$: 327.2 $[M+H]^+$, found 327.2.

5'-O-Pivaloyl-3'-O-methylthiomethylthymidine (**14**)

Compound **13** (6.50 g, 19.9 mmol) was dissolved in DMSO (40 mL), Ac_2O (27 mL) and AcOH (8.50 mL) and stirred for 3 d. After the completion of the reaction, ice cold saturated sodium bicarbonate solution (100 mL) was added and the mixture was stirred for 1 h. The aqueous solution was extracted with Et_2O (3×80 mL). The combined organic phases were washed with water (4×100 mL) and dried with Na_2SO_4 . After concentration under vacuum, the residue was purified by flash chromatography (hexane/EtOAc 1:1) to afford **14** as a white solid (6.15 g, 80%). ^1H NMR (CDCl_3): δ = 1.25 (s, 9H), 1.93 (s, 3H), 2.05 (ddd, J = 6.6, 6.7, 14.0 Hz, 1H), 2.15 (s, 3H), 2.50 (ddd, J = 2.6, 5.9, 14.0 Hz, 1H), 4.22–4.29 (m, 2H), 4.34–4.40 (m, 1H), 4.42–4.47 (m, 1H), 4.62 (d, J = 11.8 Hz, 1H), 4.70 (d, J = 11.8 Hz, 1H), 6.26 (dd, J = 5.9, 7.9 Hz, 1H), 7.28 (s, 1H), 10.09 ppm (s, 1H); ^{13}C NMR (CDCl_3): δ = 12.5, 13.7, 27.2, 37.4, 38.8, 63.6, 73.7, 75.5, 82.0, 85.1, 111.2, 134.7, 150.5, 164.1, 178.0 ppm; ESI-MS (positive ion): m/z : calcd for $\text{C}_{17}\text{H}_{27}\text{N}_2\text{O}_6\text{S}^+$: 387.2 $[M+H]^+$, found 387.2.

Preparation of 16

To a mixture of **14** (1.50 g, 3.88 mmol), **15** (1.10 g, 3.88 mmol) and molecular sieve in dichloromethane (80 mL) was added TfOH (51 μ L, 0.58 mmol) and NIS (921 mg, 4.11 mmol) at room temperature.^[32] After 5 min, the reaction mixture was filter, diluted with dichloromethane (100 mL), washed with saturated aq. Na₂S₂O₃ (10 mL) and saturated aq. NaHCO₃ (10 mL). The organic phase was dried with Na₂SO₄. After concentration under vacuum, the residue was purified by flash chromatography (hexane/EtOAc/MeOH 1:1:0.2) to afford compound **16** as a white solid (1.82 g, 77%). ¹H NMR (CDCl₃): δ = 1.23 (s, 9H), 1.92 (s, 6H), 2.01–2.09 (m, 1H), 2.12 (s, 3H), 2.22–2.30 (m, 1H), 2.43 (ddd, J =1.9, 6.0, 13.4 Hz, 1H), 2.52–2.59 (m, 1H), 3.82 (dd, J =2.7, 10.9 Hz, 1H), 3.92 (dd, J =3.4, 10.9 Hz, 1H), 4.16–4.20 (m, 1H), 4.25–4.34 (m, 3H), 4.37–4.43 (m, 1H), 4.82 (d, J =7.1 Hz, 1H), 4.86 (d, J =7.1 Hz, 1H), 5.24–5.31 (m, 1H), 6.24 (dd, J =6.0, 7.9 Hz, 1H), 6.36 (dd, J =6.1, 8.1 Hz, 1H), 7.28 (s, 1H), 7.41 (s, 1H), 10.20 (s, 1H), 10.23 ppm (s, 1H); ¹³C NMR (CDCl₃): δ = 12.4, 12.7, 20.9, 27.1, 37.3, 38.0, 38.7, 63.9, 68.2, 74.2, 82.6, 83.2, 84.6, 85.2, 94.8, 111.2, 111.3, 134.8, 135.1, 150.5, 150.7, 164.0, 164.1, 170.5, 177.9 ppm; ESI-MS (positive ion): m/z : calcd for C₂₈H₃₉N₄O₁₂⁺: 623.3 [M +H]⁺, found 623.2.

Deprotection of 16 to yield T_{CH₂T}

Compound **16** (830 mg, 1.37 mmol) was stirred with aqueous NH₃ solution (25%, 20 mL) at 80 °C for overnight. The solvent was removed in vacuum. The residue was then purified by reverse phase HPLC in the gradient mode using water/CH₃CN as solvent. The compound T_{CH₂T} was isolated as a white solid (440 mg, 65%). ¹H NMR ([D₄]MeOH): 1.86 (s, 3H), 1.87 (s, 3H), 2.20–2.33 (m, 3H), 2.43 (ddd, J =2.4, 5.6, 13.5 Hz, 1H), 3.76 (dd, J =3.3, 11.9 Hz, 1H), 3.78–3.87 (m, 3H), 4.01–4.06 (m, 1H), 4.06–4.09 (m, 1H), 4.40–4.47 (m, 2H), 4.84 (d, J =7.1 Hz, 1H), 4.88 (d, J =7.1 Hz, 1H), 6.23 (t, J =6.9 Hz, 1H), 6.38 (t, J =6.6 Hz, 1H), 7.65 (s, 1H), 7.77 ppm (s, 1H); ¹³C NMR ([D₄]MeOH): δ = 12.5, 12.7, 38.8, 40.8, 62.7, 68.9, 72.2, 78.6, 86.2, 86.3, 86.7, 86.8, 95.9, 111.5, 111.6, 137.8, 138.0, 152.2 (2C), 166.2 ppm (2C); ESI-MS (positive ion): m/z : calcd for C₂₁H₂₉N₄O₁₀⁺: 497.2 [M +H]⁺, found 497.2.

Synthesis of 5R-SP

The 5R-SP was chemically synthesized according to the published procedure.^[5,11d]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the National Institute of Environmental Health Sciences (NIH; R00ES017177, L.L.), IUPUI startup fund (L.L. and J.P.), IUPUI RSFG fund (L.L.) as well as IUCRG grant (L.L.) for financial support. The NMR and MS facilities are supported by the National Science Foundation MRI grants CHE-0619254 and DBI-0821661, respectively.

References

- a) Desnoux CL, Guillaume D, Clivio P. Chem Rev. 2010; 110:1213–1232. [PubMed: 19891426] b) Douki T, Setlow B, Setlow P. Photochem Photobiol Sci. 2005; 4:591–597. [PubMed: 16052264] c) Douki T, Setlow B, Setlow P. Photochem Photobiol. 2005; 81:163–169. [PubMed: 15458366] d) Rebeil R, Nicholson WL. Proc Natl Acad Sci USA. 2001; 98:9038–9043. [PubMed: 11470912] e) Chandor A, Berteau O, Douki T, Gasparutto D, Sanakis Y, Ollagnier-De-Choudens S, Atta M, Fontecave M. J Biol Chem. 2006; 281:26922–26931. [PubMed: 16829676] f) Buis J, Cheek J, Kalliri E, Broderick J. J Biol Chem. 2006; 281:25994–26003. [PubMed: 16829680] g) Pieck J,

- Hennecke U, Pierik A, Friedel M, Carell T. *J Biol Chem*. 2006; 281:36317–36326. [PubMed: 16968710]
2. Setlow P. *Environ Mol Mutagen*. 2001; 38:97–104. [PubMed: 11746741]
3. Donnellan JE Jr, Setlow RB. *Science*. 1965; 149:308–310. [PubMed: 17838105]
4. Mantel C, Chandor A, Gasparutto D, Douki T, Atta M, Fontecave M, Bayle PA, Mouesca JM, Bardet M. *J Am Chem Soc*. 2008; 130:16978–16984. [PubMed: 19012397]
5. a) Lin G, Li L. *Angew Chem*. 2010; 122:10122–10125. *Angew Chem Int Ed*. 2010; 49:9926–9929. b) Liu D, Li L. unpublished results.
6. Yang L, Lin G, Liu D, Dria KJ, Telser J, Li L. *J Am Chem Soc*. 2011; 133:10434–10447. [PubMed: 21671623]
7. Cheek J, Broderick J. *J Am Chem Soc*. 2002; 124:2860–2861. [PubMed: 11902862]
8. Park H, Zhang K, Ren Y, Nadji S, Sinha N, Taylor JS, Kang C. *Proc Natl Acad Sci USA*. 2002; 99:15965–15970. [PubMed: 12456887]
9. a) Butenandt J, Eker APM, Carell T. *Chem Eur J*. 1998; 4:642–654. b) Hruska FE, Voituriez L, Grand A, Cadet J. *Biopolymers*. 1986; 25:1399–1417.
10. Cadet, J.; Vigny, P. *Photochemistry and the Nucleic Acids*. Vol. 1. Wiley; New York: 1990. *Bioorganic Photochemistry*; p. 1
11. a) Friedel M, Pieck J, Klages J, Dauth C, Kessler H, Carell T. *Chem Eur J*. 2006; 12:6081–6094. b) Chandra T, Silver SC, Zilinskas E, Shepard EM, Broderick WE, Broderick JB. *J Am Chem Soc*. 2009; 131:2420–2421. [PubMed: 19178276] c) Friedel M, Berteau O, Pieck J, Atta M, Ollagnier-de-Choudens S, Fontecave M, Carell T. *Chem Commun*. 2006:445–447. d) Kim SJ, Lester C, Begley TP. *J Org Chem*. 1995; 60:6256–6257.
12. a) Camerman N, Camerman A. *J Am Chem Soc*. 1970; 92:2523–2527. [PubMed: 5438023] b) Leonard NJ, Golankiewicz K, McCredie RS, Johnson SM, Paul IC. *J Am Chem Soc*. 1969; 91:5855–5862.
13. a) Matteucci M. *Tetrahedron Lett*. 1990; 31:2385–2388. b) Gao X, Brown FK, Jeffs P, Bischofberger N, Lin KY, Pipe AJ, Noble SA. *Biochemistry*. 1992; 31:6228–6236. [PubMed: 1320930] c) Jones RJ, Lin KY, Milligan JF, Wadwani S, Matteucci MD. *J Org Chem*. 1993; 58:2983–2991. d) Rozners E, Strömberg R. *J Org Chem*. 1997; 62:1846–1850.
14. Mees A, Klar T, Gnau P, Hennecke U, Eker APM, Carell T, Essen LO. *Science*. 2004; 306:1789–1793. [PubMed: 15576622]
15. Lipshutz BH, Miller TA. *Tetrahedron Lett*. 1989; 30:7149–7152.
16. Johnson DC 2nd, Widlanski TS. *Org Lett*. 2004; 6:4643–4646. [PubMed: 15575650]
17. Ikemoto N, Schreiber SL. *J Am Chem Soc*. 1992; 114:2524–2536.
18. a) Varghese AJ. *Biochem Biophys Res Commun*. 1970; 38:484–490. [PubMed: 5462695] b) Varghese AJ. *Biochemistry*. 1970; 9:4781–4787. [PubMed: 5476256] c) Varghese AJ. *Photochem Photobiol*. 1971; 13:357–364.
19. Schreier WJ, Schrader TE, Koller FO, Gilch P, Crespo-Hernandez CE, Swaminathan VN, Carell T, Zinth W, Kohler B. *Science*. 2007; 315:625–629. [PubMed: 17272716]
20. a) Dunitz JD, Ibberson RM. *Angew Chem*. 2008; 120:4276–4278. *Angew Chem Int Ed*. 2008; 47:4208–4210. b) Bartell LS, Roskos RR. *J Chem Phys*. 1966; 44:457–463. c) Bartell LS, Kuchitsu K, deNeui RJ. *J Chem Phys*. 1961; 35:1211–1218.
21. a) Altona C, Sundaralingam M. *J Am Chem Soc*. 1972; 94:8205–8212. [PubMed: 5079964] b) Cheng DM, Sarma RH. *J Am Chem Soc*. 1977; 99:7333–7348. [PubMed: 915155]
22. a) Altona C, Sundaralingam M. *J Am Chem Soc*. 1973; 95:2333–2344. [PubMed: 4709237] b) Ostrowski T, Maurizot J-C, Adeline M-Trs, Fourrey J-L, Clivio P. *J Org Chem*. 2003; 68:6502–6510. [PubMed: 12919010]
23. Saenger, W. *Principles of Nucleic Acid Structure*. Springer; New York: 1984. p. 556
24. a) Mohr SC, Sokolov NV, He CM, Setlow P. *Proc Natl Acad Sci USA*. 1991; 88:77–81. [PubMed: 1898779] b) Nicholson WL, Setlow B, Setlow P. *Proc Natl Acad Sci USA*. 1991; 88:8288–8292. [PubMed: 1924287] c) Lee KS, Bumbaca D, Kosman J, Setlow P, Jedrzejewski MJ. *Proc Natl Acad Sci USA*. 2008; 105:2806–2811. [PubMed: 18287075]
25. See the Supporting Information.

26. Chandor-Proust A, Berteau O, Douki T, Gasparutto D, Ollagnier-de-Choudens S, Fontecave M, Atta M. *J Biol Chem*. 2008; 283:36361–36368. [PubMed: 18957420]
27. Silver S, Chandra T, Zilinskas E, Ghose S, Broderick W, Broderick J. *J Biol Inorg Chem*. 2010; 15:943–955. [PubMed: 20405152]
28. Frisch, MJ.; Trucks, GW.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Montgomery, JA., Jr; Vreven, T.; Kudin, KN.; Burant, JC.; Millam, JM.; Iyengar, SS.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, GA.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, JE.; Hratchian, HP.; Cross, JB.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, RE.; Yazyev, O.; Austin, AJ.; Cammi, R.; Pomelli, C.; Ochterski, JW.; Ayala, PY.; Morokuma, K.; Voth, GA.; Salvador, P.; Dannenberg, JJ.; Zakrzewski, VG.; Dapprich, S.; Daniels, AD.; Strain, MC.; Farkas, O.; Malick, DK.; Rabuck, AD.; Raghavachari, K.; Foresman, JB.; Ortiz, JV.; Cui, Q.; Baboul, AG.; Clifford, S.; Cioslowski, J.; Stefanov, BB.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, RL.; Fox, DJ.; Keith, T.; Al-Laham, MA.; Peng, CY.; Nanayakkara, A.; Challacombe, M.; Gill, PMW.; Johnson, B.; Chen, W.; Wong, MW.; Gonzalez, C.; Pople, JA. *Gaussian 03*. Gaussian, Inc; Wallingford CT: 2004.
29. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. *J Comput Chem*. 2009; 30:1545–1614. [PubMed: 19444816]
30. Kim ST, Malhotra K, Smith CA, Taylor JS, Sancar A. *J Biol Chem*. 1994; 269:8535–8540. [PubMed: 8132580]
31. Bornemann B, Marx A. *Bioorg Med Chem*. 2006; 14:6235–6238. [PubMed: 16777418]
32. Veeneman GH, Van Der Marel GA, Van Den Elst H, Van Boom JH. *Tetrahedron*. 1991; 47:1547–1562.
33. Carell and co-workers have independently crystallized the SP using a different approach: Heil K, Kneutinger AC, Schneider S, Lischke U, Carell T. *Chem Eur J*. 2011; 17:9651–9657.

Figure 1.

X-ray structure of the 5*R*-CH₂SP with the nitrogen and oxygen atoms labeled. For clarity purpose, only the bridging methylene carbon between the two thymine bases (C5), the newly formed chiral center at the 5'-T (C6) and the formacetal carbon (C16) are labeled among the 21 carbon atoms in the structure. The structure clearly shows that the chiral center (C6) adopts an *R* configuration. The distance between the H_{6a-pro-S} and the C5 carbon (shown in blue arrow) is found to be 2.63 Å and that between the H_{6a-pro-R} and the C5 carbon (shown in red arrow) is found to be 3.36 Å. Details about the bond lengths and bond angles are available in Supporting Information.

Figure 2.

Top) Zoom-in view of ^1H -ROESY NMR spectra for $5R\text{-CH}_2\text{SP}$ and $5R\text{-SP}$, respectively in DMSO. The key H-H correlations from protons associated with the 5'-thymine ring are highlighted by the green arrows in these spectra (Note: in the ROESY of $5R\text{-CH}_2\text{SP}$, the spots pointed by two arrows are actually composed by two dots, which can be separated in a further zoom-in view). Bottom) Key NOESY cross-peaks determined from the NMR spectra of these SP species. The almost identical coupling patterns between these two compounds suggest them to possess very similar steric conformation in the nucleobase structure. Full 1D- and 2D-NMR spectra for these SP species can be found in SI.

Figure 3. Superposition of the 5R-CH₂SP and 5R-SP structures computationally optimized at B3LYP/6-31+G(d,p) level. Carbon atoms in 5R-CH₂SP and 5R-SP are colored in dark-gray and green, respectively; the phosphorus atom in 5R-SP is colored in orange and oxygen atoms are in red. These two structures are found to overlap perfectly for most part of the molecules except for the linkage section.

Figure 4.

A) HPLC chromatograph of the 5R-CH₂SP repair mediated by SPL. 5'-dA was eluted at 16.8 min, 5R-CH₂SP at 18.5 min and T_{CH₂}T at 28.2 min. B) Enzyme reaction rates were calculated via plotting the amount of T_{CH₂}T or TpT isolated versus reaction time.

Scheme 1.

Scheme 2.

Scheme 3.

Table 1

$J_{1'2'}$ and $J_{3'4'}$ coupling constants (Hz) and population distribution of conformers in the deoxyribose in 5*R*-SP and 5*R*-CH₂SP.

	5 <i>R</i> -SP		5 <i>R</i> -CH ₂ SP	
	5'-end	3'-end	5'-end	3'-end
$J_{1'2'}$	4.0	7.1	3.3	5.9
$J_{1'2''}$	8.1	7.1	8.6	8.6
$J_{3'4'}$	6.9	n/d	7.3	n/d
% of C2'-endo ^[a]	37		31	
% of C3'-endo ^[b]	63		69	

^[a] Calculated from $J_{1'2'}/(J_{1'2'} + J_{3'4'})$.

^[b] Calculated from $J_{3'4'}/(J_{1'2'} + J_{3'4'})$.