Adjacent and non-adjacent dipyrimidine photoproducts as intrinsic probes of DNA secondary and tertiary structure

John-Stephen Taylor

Department of Chemistry, Washington University, St. Louis, MO 63130

Corresponding author email: taylor@wustl.edu.

Contribution for the 50th Anniversary Issue of the American Society for Photobiology (ASP) in Photochemistry and Photobiology

ABSTRACT.

While the photochemistry of duplex DNA has been extensively studied the photochemistry of non-duplex DNA structures is largely unexplored. Because the structure and stereochemistry of DNA photoproducts depend on the secondary structure and conformation of the DNA precursor they can serve as intrinsic probes of DNA structure. This review focuses on the structures and stereoisomers of pyrimidine dimer photoproducts arising from adjacent and non-adjacent pyrimidines in A, B and denatured DNA, bulge loops, G-quadruplexes, and reverse-Hoogsteen hairpins and methods for their detection.

INTRODUCTION

While the photochemistry of DNA has been extensively studied since the discovery of the thymine dimer more than 60 years ago (1-6), it has largely focused on duplex DNA which exists almost exclusively in a B-like conformation. The photochemistry of alternate secondary and tertiary nucleic acid structures such as bulge and hairpin loops, slipped structures, H DNA, cruciforms, and G-quadruplexes which are thought to play roles in mutagenesis, and the regulation of transcription and replication (7-11), has by comparison received very little attention. UV irradiation of duplex DNA under native conditions produces primarily cyclobutane pyrimidine dimers (CPDs) with a specific *cis,syn* stereochemistry that results from the [2+2] cycloaddition of the 5,6-double bonds of adjacent pyrimidines (**Figure 1A**) (4, 12). The *cis,syn* stereochemistry is a direct consequence of the anti-parallel Watson-Crick base-paired right-handed double helical conformation which holds the pyrimidines in a head to head orientation (*syn*) with the Watson Crick faces on the same side and consequently the C5 substituents on the same side (*cis*). The *5R*,*5R* stereochemistry is a result of photodimerization of the bases in an *anti* glycosyl bond conformation. These stereochemical features can be best appreciated from the x-ray crystal structure of a *cis,syn* thymine dimer-containing decamer in comparison to B DNA (13).

Another major direct dipyrimidine photoproduct of duplex DNA is the (6-4) photoproduct (14, 4, 12) which arises from a Paterno-Buchi reaction between a keto group of the 3'-pyrimidine with the 5,6-double bond of the 5'-pyrimidine to form an unstable oxetane (**Figure 1A**). This unstable four-membered ring intermediate opens to leave a bond between the 5'-C6 and 3'-C4 carbons. With a 3'-dC, the photoreaction appears to lead directly to the (6-4) photoproduct with a NH_2 group in place of the OH group (15). The resulting 6*S*,4*P*

stereochemistry is a result of the B DNA conformation, where the *P* designation refers to axial chirality of the pyrimidone ring whose rotation is restricted (16). While another stereocenter is introduced at 5'-C5, it is sufficient to identify the chirality at the 5'-C6 because the photoreaction always adds the pyrimidone ring and heteroatom in a *cis* relationship.

While the relative yields of CPDs and (6-4) products can be modulated by sequence context, transcription factors, nucleosomes, and other proteins which alter the local conformation and dynamics of the duplex DNA through bending and winding, the structure and stereochemistry of the photoproducts remains unchanged (4, 5, 17). A more dramatic change in duplex DNA photochemistry occurs in *Bacillus subtilis* spores where *cis*,*syn* CPD and (6-4) photoproducts are greatly suppressed and the eponymous spore photoproduct (SP) is the major photoproduct (**Figure 1**) (18). Almost exclusive formation of SP in spores has been attributed to the drastically reduced intracellular water content and binding of the small acid-soluble protein (SASP) which changes the B DNA conformation to A DNA (19-21), as well as to the presence of dipicolinic acid (22, 23). The reduction in CPD and (6-4) photoproduct formation in spores has been proposed to be a result of the highly restricted mobility of the DNA which prevents the bases from attaining the appropriate alignment of the bases for photoreaction (18). On the other hand, the A conformation is optimal for radical abstraction of the 3'-C5-methyl hydrogen by a 5'-C6 radical in the excited state and subsequent bond formation between the resulting C5 methylene radical and the C5 of the 5'-pyrimidine to produce SP with a *5R* configuration (24-26). In 80% aqueous ethanol, where A DNA is conformationally flexible, *cis,syn* CPD and (6-4) photoproducts are produced at only a slightly lower the rate than in B DNA, along with the SP product (27, 28). What is puzzling, however, is that no photoproducts were observed when two adjacent T's were embedded in an RNA hairpin, which adopts an A

form (29) suggesting that there may be conformational or photophysical differences between A form DNA and RNA.

Unlike A and B form DNAs which constrain photoproduct formation to adjacent dipyrimidines in a head to head orientation with *cis* stereochemistry, other types of DNA secondary and tertiary structures can lead to adjacent and non-adjacent photoproducts with different structures and stereochemistries (**Figure 1B**). Denaturation of DNA removes constraints on the glycosyl bond conformations, while folding of DNA into bulge loops and G-quadruplexes can bring otherwise non-adjacent pyrimidines into close proximity in both head to head and head to tail orientations. Dipyrimidine photoproducts formed from these other structures would be expected to have their own unique biological effects, and could also serve as intrinsic probes to identify and locate these unusual secondary and tertiary structures in vivo. To understand the extent to which dipyrimidine photoproducts can be used as probes of DNA structure we need to first understand the possible structures and stereochemistries of DNA photoproducts and how they can be identified and located.

We will first focus on cyclobutane pyrimidine dimers (CPDs) and later on the (6-4) and SP photoproducts. To simplify the discussion we will illustrate all the possible regio- and stereochemistries of CPDs with those resulting in thymidine dimers, recognizing that 5-methyl dC (d^mC) within CPDs readily deaminates to T (**Figure 2, Supplementary Figures S1 & S2**) (30, 31). Likewise, dC in dC=T, T=dC and dC=dC CPDs readily deaminate to dU to result in dU=T, T=dU, dU=dU CPDs respectively (**Supplementary Figures S3-S8**). The regiochemistry of CPDs can be categorized as to whether they are head to head (*syn*) (**Figure 2B**) or head to tail (*anti*) (**Figure 2C**). Secondly, the individual regioisomers can be differentiated by the relative stereochemistry of the C5-methyl or H5 on one pyrimidine as being either *cis* or *trans* to the C5-

methyl or H5 on the other pyrimidine. Such an analysis reveals that there are eight possible CPD isomers when embedded in a nucleic acid structure, two each of *cis,syn*, *trans,syn*, *cis,anti* and *trans,anti* isomers. These isomers can then be further differentiated by assignment of the absolute stereochemistry at the C5-carbons of the CPDs as *R* or *S* according to the Cahn-Ingold-Prelog rules as implemented in a useful online app (https://chemapps.stolaf.edu/jmol/jsmol/cip.htm) (32) and by the roman numerals I and II.

Structural assignment and mapping of CPDs. The large number of CPD isomers poses challenges for identifying and locating them in genomic DNA. While CPDs of the bases, nucleosides and short oligonucleotides can be structurally characterized by crystallography (33, 34, 13) and by NMR (35-38), it is much more difficult to determine the structure and location of CPDs in genomic DNA. The one exception being the *cis,syn*-I CPD, which can be mapped by enzymatic cleavage with *cis,syn* CPD-specific pyrimidine dimer glycosylases followed by sequence analysis of the cleavage sites (39-42). Most of the information about the formation of the other CPD isomers comes from acid hydrolysis of the DNA which releases dipyrimidine CPDs (**Figure 2**), or by enzymatic degradation which releases di-, tri- or tetranucleotides depending on the enzymes used (**Figure 3**).

CPDs produced by acid hydrolysis of DNA. Acid hydrolysis of irradiated DNA containing all eight possible T=T CPDs would release six Thy=Thy isomers due to symmetry elements that arise in the *cis,syn* and *trans,anti* CPDs when the are stripped of their sugar phosphate backbone as shown in **Figure 2** and **Supplementary Figures S1 & S2** (43). There would be one *cis,syn* and two enantiomeric *trans,syn* CPDs resulting from a head to head arrangement, and two

enantiomeric *cis,anti* CPDs and one *trans,anti* CPDs from a head to tail arrangement (43). There would be no way of knowing, however, if a Thy=Thy CPD arose from T=T, or a d^mC-containing CPD following deamination of the d^mC. The same six isomers would also be released from all eight possible dC=dC CPDs following deamination to Ura=Ura CPDs (**Supplemental Figures S7&S8**). Because the sugar phosphate backbone is released from the CPDs by acid hydrolysis there is also no way to determine whether a Pyr=Pyr CPD arose from an adjacent or non-adjacent pair of pyrimidines. One cannot also determine whether a *cis,syn*-Thy=Thy or Ura=Ura CPD arose from a *cis,syn*-I or *cis,syn*-II CPD because the acid hydrolysis products are identical. Furthermore only four Thy=Thy or Ura=Ura CPDs would be separable by achiral chromatography because the *trans,syn* and *cis,anti* products exist as enantiomers. The four separable products can be identified by comparison to authentic samples prepared by photodimerization of the parent pyrimidines (44, 38, 45). Acid hydrolysis of T=dC and dC=T CPDs would also produce eight products (**Supplementary Figures S3-S6**), two each of enantiomeric *cis,syn*, *trans,syn*, *cis,anti* and *trans,anti*-Thy=Ura CPDs which would appear as four separable products on an achiral column. The *cis,syn*-Thy=Ura CPD arising from a *cis,syn*-I-T=pdC CPD would be the enantiomer of the one arising from a *cis,syn*-I-dC=pT CPD and cannot be distinguished on an achiral column, and can also not be distinguished from *cis,syn*-Thy=Ura and Ura=Thy arising from the *cis,syn*-II CPDs. One can also not distinguish CPD products arising from C=T and $C=d^mC$ following hydrolysis because both lead to Ura=Thy. **Tables 1** and **2** give the stereochemical relationships between all the Thy=Ura and Ura=Thy CPDs.

CPDs produced by enzymatic degradation of DNA. Unlike the CPD products resulting from acidic hydrolysis of DNA, the CPD photoproducts released by enzymatic degradation contain information about whether the CPD formed from adjacent or non-adjacent pyrimidines as illustrated for T=T CPDs (**Figure 3**). The two principal enzymes used for this purpose are nuclease P1 (NP1) and snake venom phosphodiesterase (SVP) (46-48, 28). The degradation products produced by these enzymes depend on whether the enzyme cleaves to the 3'-side (NP1) or 5'-side (SVP) of an undamaged nucleotide. Degradation of DNA containing an adjacent CPD by NP1 produces a trinucleotide of the form pT=pTpN which is converted to the trinucleotide pTpTpN upon photoreversal of the CPD with 254 nm light because the two thymidines remain linked by a phosphodiester bond (**Figure 3A**). NP1 degradation of a non-adjacent CPD, however, produces a tetranucleotide of the form pT(pN)=pTpN which is a photodimer of two dinucleotides (**Figure 3B**). This non-adjacent CPD-containing tetramer can be distinguished on a gel from a linear tetranucleotide that might have resulted from incomplete enzymatic digestion by photoreversal with 254 nm light to produce two dinucleotides pTpN. It can also be distinguished by mass spectrometry from a tetranucleotide of the same composition by having a 18 amu higher mass due to what amounts to be a hydrolyzed phosphodiester linkage between the second and third nucleotides. LC-MS/MS analysis of NP1 cleavage products can therefore be used to distinguish between adjacent and non-adjacent CPDs and provide additional sequence context information from identification of the 3'-attached nucleotides (46, 49, 38, 45). Another advantage of having an intact nucleotide in the degraded product is that it will have a high absorbance at 260 nm making it more sensitive to UV detection at 254 nm than pyrimidine dimers which have their absorption maxima at much lower wavelengths which are subject to more interference (50).

Degradation with snake venom photophodiesterase (SVP) also produces trinucleotides from adjacent CPDs and tetranucleotides from non-adjacent CPDs, but with the intact nucleotide attached to the 5'-end of the CPD, rather than the 3'-end as with NP1. The intact undamaged nucleotide on the 5'-end of the trinucleotide degradation product of adjacent CPDs enables these products to be detected by postlabeling with [γ-^{32}P]-ATP and kinase following removal of the 5'-phosphate with a phosphatase (**Figure 3A**) (51, 52), unlike the degradation products from NP1 digestion that have photodamaged 5'-ends. The tetramer degradation products from non-adjacent CPDs can also be detected by this postlabeling method and can be distinguished from incompletely degraded DNA tetramers by photoreversal to dinucleotides (**Figure 3B**) (Gutierrez et al, unpublished results).

When irradiated DNA is subjected to both NP1, SVP and alkaline phosphatase, adjacent T=T CPDs are degraded to T=pT with a phosphate linking the two nucleosides, whereas non-adjacent CPDs are degraded to T=T which lacks the internucleotide phosphate. These three enzymes have been used in combination with LC-M/MS to afford an extremely sensitive analytical method for quantifying adjacent and non-adjacent CPD formation in genomic DNA (48, 28). In principle, ten different CPD products of enzymatic degradation can be distinguished by this method. All four adjacent *syn*-T=pT CPDs stereoisomers can be distinguished because none are identical or mirror image isomers (enantiomers) and are thus diastereomers and can be separated by achiral HPLC. Because the substituents X and Y become identical in the enzymatic degradation products of non-adjacent T=T CPDs, there is only one *cis,syn*-T=T CPD, whereas there are two diastereomeric *trans,syn* T=T CPDs (35) (**Figure 2B**). Likewise, there is only one *trans,anti*-TT CPD and two *cis,anti* T=T CPDs (35) (**Figure 2C**). All four possible stereoisomers of adjacent T=pdU, dU=pT and dU=pdU can also be distinguished, but because of

symmetry elements many of the non-adjacent CPDs cannot be distinguished (**Tables 1 & 2, Supplementary Figures S3-S8**). For example, the non-adjacent *cis,syn*-I and –II products of TC and CT sites would not be distinguishable following enzymatic or acid hydrolysis by achiral chromatography because they would form enantiomeric *cis,syn*-T=dU and dU=T products. Because deamination of $d^{m}C$ in a CPD converts it to T, it is also impossible to distinguish a Y=T CPD as having arisen from a Y=T or Y=$d^{m}C$ CPD (**Figure 2**).

Of some concern, however, is a report that the *cis,syn*-pT=pTpT CPD produced by NP1 degradation is subject to intradimer phosphodiester cleavage by the same enzyme (52). Subsequent degradation by SVP and alkaline phosphatase would then produce the same *cis,syn*-T=T CPD with a cleaved internucleotide phosphodiester linkage that is produced from degradation of non-adjacent *cis,syn*-T=T CPDs. Furthermore, it was reported by the same group that nucleotide ex*cis*ion repair products of adjacent CPDs from human cells contain a subst*anti*al fraction of CPDs with cleaved intradimer phosphodiester linkages (53). Adjacent CPDs with cleaved intradimer phosphates would then be indistinguishable from non-adjacent CPDs when using NP1 to degrade the DNA. Other studies utilizing NP1 to analyze CPD formation in native DNA, however, have not reported such intradimer cleaved products (46, 54, 47, 48).

Adjacent *trans,syn* and *cis,syn*-II CPD formation in DNA. In A or B form DNA only the adjacent *cis,syn*-I CPD stereoisomer can form as a consequence of the Watson-Crick base pairing which holds adjacent nucleotides in an *anti* glycosyl conformation with the pyrimidines in a *syn*, head to head arrangement with a *cis* orientation of the methyl groups (**Figure 1A**). Adjacent *trans,syn*-T=T CPDs cannot form in A or B DNA because one of the T's must be in a *syn* glycosyl conformation and Watson-Crick base pairing requires the glycosyl conformations to be

in the *anti* conformation (**Figure 4A**). If one thymine were to flip into the *syn* glycosyl conformation within B DNA it would also be greatly inhibited from photoreacting with the other T because the distances between adjacent 5,6-double bonds would be lengthened, more so when the 3'-T is in a *syn* conformation. Furthermore, the *syn* glycosyl conformation is much less favorable than the *anti* conformation for pyrimidines. As a result, *trans,syn* CPDs have only been detected in very low yields (<0.2%) relative to the *cis,syn* CPD (94%) in UVC irradiated genomic DNA at 0^{o}C in water (**Table 3**) (55).

In single strand DNA pyrimidines are not constrained to be in the *anti* glycosyl conformation and are free to assume *syn* conformations. If the 5'-thymidine of a TpT site was in a *syn* conformation a *trans,syn*-I CPD would result, whereas a *trans,syn*-II CPD would result if the 3'-T was in the *syn* conformation (**Figure 4A**). As a result, the frequency of the *trans,syn* CPD greatly increases upon heat denaturation to 14% (**Table 3**) (55). The *trans,syn*-I CPD is the only adjacent *trans,syn* CPD to be detected in irradiated genomic DNA and in the dinucleotide TpT (56), while the adjacent *trans,syn*-II-T=T CPD has only been found to form in a modified dinucleotide derivative (57). In contrast, irradiation of the dinucleotide TpdC produces both *trans,syn*-I and *trans,syn*–II T=pdU CPDs (56). There is no evidence at this time for formation of the adjacent *cis,syn*-II CPD in a dinucleotide or in longer DNA sequences most likely because it would that require both pyrimidines to be in unfavorable *syn* glycosyl conformations. An enzyme with specificity for both adjacent *cis,syn*-I and *trans,syn*-II-TT CPDs has been reported (58) which could possibly be used to locate *trans,syn*-II CPDs in genomic DNA following enzymatic photoreversal with a *cis,syn*-I CPD specific DNA photolyase.

Non-adjacent *cis,syn* and *trans,syn* CPD formation in DNA. Quite some time ago it was proposed that UV induced frame-shift mutations might be the result of polymerase bypass of non-adjacent CPDs formed by photodimerization of bulge loop structures. It was discovered that UVC irradiation of poly d(GT)● polyd(CT) (59-61) or poly d(AT) (62) resulted in the formation of non-adjacent CPDs as determined by chromatographic analysis of the Thy=Thy CPDs produced by acid hydrolysis. Non-adjacent CPD formation was proposed to occur from a structure in which an intervening nucleotide becomes extrahelical in a bulge loop allowing the two non-adjacent pyrimidines to stack upon each other and form a *cis,syn* CPD (**Figure 4B**). In the case of poly-d(AT) only *cis,syn* Thy=Thy CPDs were detected, presumably because of base pairing interactions between the photoreacting T's and the A's in the complementary strand. In the case of poly-d(GT) the *cis,syn* stereochemistry predominated but *trans,syn* and *cis,anti* Thy=Thy CPDs were also detected indicating the presence of a variety of folded structures. A site-specific non-adjacent *cis,syn* T(pdC)=T CPD has been prepared by UVB irradiation of a 13-mer with a central TpdCT sequence base paired to a complementary 12-mer sequence with a central d(AA) sequence to form a bulge loop structure (**Figure 4C**) (37). The *cis,syn*-I stereochemistry was determined by 2D NMR studies of the duplex (**Figure 4D**), and must have resulted from a conformation in which the intervening dC was extra-helical thereby allowing the two T's stack upon each other. The purified non-adjacent CPD containing 13-mer was ligated into a longer template and primer extension by exonuclease-free Klenow fragment was shown to lead to both -1 and -4 deletion mutations by a proposed misalignment-realignment mediated mechanism. Non-adjacent *trans,syn* CPDs have only been detected in very low yields (1%) relative to other stereoisomers (**Table 3**).

Non-adjacent *anti*-CPD formation in DNA under non-native conditions. When poly-d(GT) was irradiated with UVC in the presence of 80% ethanol or Mn^{2+} a small amount of *anti* Thy=Thy CPD was detected (59). The formation of an *anti* CPD suggested that it arose from photodimerization of thymines from opposing strands which would also explain photocrosslink formation between separate strands that was also observed (59). A later study of UVC irradiated ethanolic or desiccated genomic DNA by an enzyme-coupled mass spectrometry assay found that a significant proportion of the photoproducts formed were non-adjacent CPDs with the *cis*,*anti* or *trans,anti* stereochemistry (**Table 3**) (28). UVC irradiation of freeze dried DNA produced about 8-12% each of the non-adjacent *cis,syn*, *cis,anti*, and *trans,anti* non-adjacent T=T CPDs compared to 32% of the adjacent *cis,syn* T=T CPD. The relative proportion of non-adjacent T=T CPDs in 80% ethanolic solution increased to 16-21% each for non-adjacent *cis,syn*-, *cis,anti*-, and *trans,anti*-T=T CPDs compared to about 25% of adjacent *cis,syn*-T=T CPDs. These distributions can be compared to only produced 0.3% of the non-adjacent *cis,syn*-T=T CPD and 0.05 % of the non-adjacent *trans,anti* CPD compared to the adjacent *cis,syn*-T=T CPD in UVC irradiated aqueous solutions of calf thymus DNA. When compared to the photoproducts of thymidine under similar conditions, it is apparent that *trans,syn* CPDs are greatly suppressed in UVC irradiated native, freeze dried and ethanolic genomic DNA. The DNA structure leading to the non-adjacent CPDs formed in freeze dried and alcoholic solutions is unknown but cannot be A DNA because the reacting double bonds are too far apart.

A non-adjacent *cis,anti* CPD was also unexpectedly discovered to form in 40% yield between T2 and T7 in d(G**T**ATCA**T**GAGGTGC) when it was UVB irradiated at pH 5 but not at pH 7 (**Figure 4E**) (38). The formation of a CPD between T2 and T7 was determined by digestion with nuclease P1 which produced the tetranucleotide pT(pA)=pTpG and by MS

analysis of the termination products of 5'-3' and 3'-5' exonuclease digestion with exonucleases III and IV, respectively. The *cis,anti* stereochemistry was established by acid hydrolysis to give a *cis,anti*-Thy=Thy CPD that was correlated with the authentic product produced by irradiation of thymine. The stereochemistry and location of the CPD was further confirmed by 2D NMR NOE analysis of the purified CPD-containing 14-mer. The formation of the *cis,anti*-T2=T7 CPD was highly stereoselective and sequence dependent suggesting that it arises from some as yet unknown folded or aggregated DNA structure that brings the two T's into a head to tail alignment. Examination of the sequence does not reveal any obvious secondary structure that might explain the formation and location of the non-adjacent CPD. Mutating nucleotides 3,4,5,6, & 8 caused a drastic reduction or elimination of *anti* CPD formation, with the exception of an A6G mutation. Likewise truncation of the 3'-end of the sequence by even two nucleotides was enough to suppress *anti* CPD formation suggesting that they play an important role.

***Anti*-CPD formation in G-quadruplex forming human telomeric sequences under native conditions.** The discovery of *anti* CPD formation in a single stranded ODN prompted an investigation of *anti* CPD formation in single stranded human telomeric sequences $d(GGGTTA)_n$ which fold into various types of G quadruplex structures which bring non-adjacent T's into close proximity (**Figure 5**). G-quadruplexes are four stranded structures that contain G-quartets formed from a cyclic array of four Hoogsteen base paired G's (**Figure 6**) (63-65). The type of G-quadruplex formed depends on the glycosyl bond conformations of the G's. When both G's in a Hoogsteen base pair are in an *anti* glycosyl conformation, they would form parallel strands. When one of the G's in a Hoogsteen base pair is in a *syn* conformation and the other is in an *anti* conformation, antiparallel strands would result. The relative orientation of strands that result

from different combinations of glycosyl bonds for both Hoogsteen and reverse Hoogsteen base pairs is shown in the table in **Figure 6**.

It was initially speculated that *anti*-TT CPDs would form between the T's in adjacent lateral TTA loops in antiparallel G-quadruplexes such as the basket and chair forms due to their proximity (dashed arrows), but would not be able to form in the parallel or hybrid forms (**Figure 5**). The sequence Tel22 d(AGGG(TTAGGG)$_3$) (**Figure 7A**) was initially chosen for study because it was shown to adopt an antiparallel basket-type quadruplex structure in NaCl solution (66). In the basket structure the lateral loops 1 and 3 are adjacent to each other which would allow the T's of one loop to photodimerize with a T in the other loop to form an *anti* CPD. In KCl solution, however, Tel22 was found to adopt a hybrid-1 or hybrid-2 structure, which is composed of two non-adjacent lateral loops and one propeller loop, and was not expected to form *anti* CPDs (67). Unexpectedly, UVB irradiation of Tel22 produced *anti*-CPDs in much greater yields in KCl solution which favors hybrid structures than in NaCl solution which favors the basket structure. This is biologically significant because K^+ ion is the major intracellular ion in all types of cells.

CPD formation in the G-quadruplexes was assayed by an enzyme coupled mass spectrometry assay in which UVB irradiated DNA was treated with NP1 to degrade adjacent CPDs to trinucleotides (46) and non-adjacent CPDs to tetranucleotides (38) which were then characterized by ESI MS/MS. In all, eight different *anti* CPD-containing tetramer NP1 digestion products were detected, five *anti* pTp(A)=pTpA CPDs (three *cis,anti*, one *trans,anti* and one uncharacterized, though in principle only two *cis,anti* CPDs are possible) and three pT(pT)=pTpA *anti* CPDs (one *cis,anti*, one *trans,anti*, and one uncharacterized). To determine which T's were involved in *anti* CPD formation, individual T's in loops 1-3 were replaced with

dU which is identical to T except for the absence of the C5 methyl group. The absence of the methyl group reduces the elution time of a tetramer product on HPLC and reduces the mass by 14 which makes it easily identifiable by LC-MS. Based on uracil substitution experiments, the *anti* CPDs were determined to have formed between T1 or T2 of loop 1 and T2 of loop 3, though some CPDs involving loop 2 were also detected. To account for the preferential formation of *anti* CPDs in KCl solution but not NaCl solution, it was proposed that the *anti* CPDs form from a photoreactive form 3 basket G-quadruplex that is in equilibrium with non-photoreactive hybrid structures. The form 3 G-quadruplex structure is a two tetrad basket type quadruplex in which the 5'-GGG slips by one nucleotide to form a four-nucleotide GTTA lateral loop (68) (**Figure 8C**). It was proposed that while Tel22 forms a three tetrad basket structure in NaCl solution, the two TTA loops are too short for the T's to overlap and form an *anti* CPD (**Figure 8B**). In contrast, the T's in the form 3 basket structure in KCl could form an *anti* CPD because loop 1 was longer by one nucleotide.

It was also possible, however, that the *anti* photoproducts were arising from a partially unfolded structure like a triplex (**Figure 5B**) (69). This hypothesis was tested with Tel19, a truncated version of Tel22 lacking the 5'-AGG, and Tel22-A3 in which the 5'-most GGG sequence was replaced with AAA. Both sequences still produced a mixture of *anti*-CPDs but in much lower yield, and both had a diminished amount of the *trans,anti* product between the T2s in loops 1 and 3, T2(L1)=T2(L3). Interestingly, the Tel26 sequence with A's on both the 5'- and 3'-ends that had been found to form a hybrid-I structure in KCl solution (67) gave primarily the *trans,anti* product. Theoretical calculations supported the proposal that the *trans,anti*-T2(L1)=T2(L3) CPD could be produced in the form 3 basket conformation in KCl and not the three G-quartet basket conformation in NaCl solution (70). Interestingly it was also concluded

that a non-adjacent *cis,syn*-T2(L1)=T2(L3) CPD could also be produced which was not reported in the original experimental study and may have been misidentified as some other product.

The hypothesis that the form 3 basket structure could explain the formation of the *trans,anti*-T2(L1)=T2(L3) CPD was subsequently tested by UVB irradiating a variety of mutants of Tel22 designed to stabilize the form 3 structure as well as irradiating the sequence used to establish the form 3 structure (71). None of the mutants of the form 3 sequence resulted in an increased yield of the *trans,anti*-T2=T2 CPD, and most mutations diminished its yield. The question was then whether some other structure in equilibrium with the quadruplex structures was responsible for the formation of the *trans,anti*-T2=T2 CPD.

***Trans,anti* CPD formation in reverse Hoogsteen hairpins.** To determine if G quadruplexes were involved in the formation of the *trans,anti*-T2=T2 CPD, the effect of LiCl on photoproduct formation was studied next (72). G-quadruplexes are known not to form in lithium ion solutions because the lithium cation has a small ionic radius and is strongly hydrated making it unavailable to coordinate and stabilize the G-quartet structures (73-76). Most surprisingly, the *trans,anti*-T2=T2 CPD was produced most selectively and in the highest yield in LiCl solution compared to NaCl and KCl. What structure then, was present in LiCl solution that caused the highly stereospecific formation of the *trans*,*anti*-T2=T2 CPD? A hint as to a possible structure came from an NMR study of d(GGGGTTTTGGGG) which determined that the sequence formed a reverse Hoogsteen hairpin in LiCl solution (77). A reverse Hoogsteen hairpin is expected to be more stable than the alternative Hoogsteen hairpin because the glycosyl bonds of all of the G's are in the lower energy *anti* form. In contrast, a Hoogsteen hairpin requires that one of the G's in each base pair is in the higher energy *syn* conformation (**Figure 6**). The reverse Hoogsteen

hairpin form for Tel26 would have a TTA hairpin loop 2, and a interior loop consisting of TTA loops 1 and 3 which would bring T2(L1) and T2(L3) in contact with each other in the head to tail orientation required to form an *anti* CPD (**Figure 8D**). To obtain the *trans* stereochemistry, one of the photoreacting T's would have to be in a *syn* glycosyl orientation. A reverse Hoogsteen hairpin would also explain why Tel26 produces the *trans,anti* CPD with greater selectivity and yield than Tel22 because the two extra A's at both ends of the sequence could form additional reverse Hoogsteen base pairs and increase the stability of the hairpin (vide infra).

The idea that a reverse Hoogsteen hairpin was the conformation responsible for *trans*,*anti*-T2(L1)=T2(L3) CPD formation was first tested by systematically replacing putative G●G base pairs with G●I base pairs (72). Replacement of G with I destabilizes G-quartet structures by eliminating C2-NH2 H-bonding and forces the Hoogsteen or reverse Hoogsteen acceptor to be the I (**Figure 6)**. In spite of having disrupted G quadruplex formation by the base substitution, the *trans,anti*-T2=T2 CPD was still the major product and was produced in high yield in KCl solution in Tel21-3'con, 3'-alt, and 3'-mix sequences (**Figure 7B**) supporting the involvement of a Hoogsteen or reverse Hoogsteen hairpin. Most interestingly, however, the *trans,anti*-T2=T2 CPD now formed in high yield in NaCl solution, presumably because the three-quartet basket structure that inhibits *anti* CPD formation was no longer able to form. The important role of a Hoogsteen or reverse Hoogsteen base pairing in both the 5'- and 3'-stem portions of the hairpin was demonstrated by the absence of *trans,anti*-T2=T2 CPD formation upon replacing two to three base pairs with Watson Crick base pairs (Tel21WC and HP15WC).

Because one cannot distinguish between a Hoogsteen or reverse Hoogsteen hairpin using inosine substitutions, G's were systematically replaced with A's. A's can only form reverse Hoogsteen base pairs and therefore cannot form cyclic quartets required for formation of G-

quadruplex-type structures. UVB irradiation of the A-substituted sequences Tel-21 GAG, GGA and AAG (**Figure 7B**) also produced the *trans*,*anti*-T2=T2 CPDs in high yield in both NaCl and KCl solution, thereby confirming the involvement of the reverse Hoogsteen hairpin. When the effect of the various adenosine and inosine substituted sequences were taken into account it was concluded that the reverse Hoogsteen conformation for Tel22 and Tel26 leading to the *trans,anti* T2=T2 CPD could have the G's in the stem adjacent to the hairpin loop 2 in either reverse Hoogsteen orientation. On the other hand, the highest yield of the trans,anti-T2=T2 CPD was obtained when the G's in the stem on the other end had the reverse Hoogsteen base pairing shown in **Figure 8A**. The increased yield and selectivity for the *trans,anti*-T2(L1)=T2(L3) CPD in Tel26 compared to Tel22 in KCl solution could then be explained by the formation of additional A-A reverse Hoogsteen basepairs which selectively stabilize the reverse Hoogsteen hairpin. To account for the observed cation effects on CPD yields, it was proposed that Tel22 and Tel26 adopt a three tetrad G-quadruplex basket in NaCl solution (**Figure 8B**) that is refractory to *anti*-CPD formation. In KCl solution, however, the sequences are proposed to equilibrate between the hybrid-I and -II quadruplexes, the form 3 basket (**Figure 8C**), and the reverse Hoogsteen hairpin (**Figure 8D**), the latter of which is favored in LiCl solution. While the form 3 basket can lead to a variety of *anti* CPDs, the reverse Hoogsteen hairpin preferentially produces the *trans,anti*-T2(L1)=T2(L3) CPD.

***Anti*-CPD formation in human promoter sequences.** Analysis of the human genome has revealed that there are millions of potential G-quadruplex forming sequences which are largely localized in promoter regions (78-81). To determine whether or not *anti*-CPD formation is unique to human telomeric sequences or might also form in G-quadruplex forming sequences in

promoters, a study was undertaken to identify sequences within promoters that had the potential for forming *anti* CPDs (82). To identify potential sequences the EuQuad database of G-quadruplex forming sequences was queried for sequences within 500 bp of transcription start sites that contained three G-quartets with three or more pyrimidines in loops 1 and 3, at least one of which was thymidine. Such sequences were expected to be able to fold into basket or chair type G-quadruplexes that would form non-adjacent CPDs between loops 1 and 3 when irradiated with UVB light. From over 300 sequences that met the criteria, 15 putative G-quadruplex forming sequences 21-28 nucleotides in length were selected for irradiation in both NaCl and KCl solution and analyzed by NP1-coupled mass spectrometry. Three of these sequences taken from CALU, SISD3 and STARD3NL promoters showed significant amounts of non-adjacent CPD formation in both NaCl and KCl solution. In addition to detecting non-adjacent CPDs of TT, a non-adjacent T=dU CPD was found to form in STARD2NL in KCl solution, which would have resulted from the deamination product of a non-adjacent T=dC CPD. The formation of non-adjacent CPDs in these sequences is indicative of their ability to adopt basket (**Figure 9A**), chair and/or reverse Hoogsteen hairpin structures. These results suggest that non-adjacent CPDs could be used to detect G-quadruplex and other non-B DNA forming sequences in promoter DNA that are capable of folding into structures with appropriately positioned pyrimidines in loops 1 and 3.

Non-adjacent photoproduct formation in other DNA quadruplex sequences. A review of the literature reveals that there was a report in 1989 of efficient intrastrand photocrosslinking of Oxytrichia (Oxy-4) and Tetrahymena (Tet-4) telomere sequences by 254 nm light in NaCl solution (73). Analysis of hot pyrrolidine cleavage suggested that a photocrosslink had formed

between T11 and T27 in Oxy-4, $(TTTTGGGG)_4$, which was proposed to result from a chair conformation in which T11 was in loop 1 and T27 was in loop 3 of a two G-quartet chair conformation (**Figure 9B**). In Tet-4, $(TTGGGG)_4$, a photocrosslink was mapped between T1 and T13 which was proposed to arise from T1 in the 5'-tail and T13 in loop 2 of a chair G-quadruplex, suggesting another way that non-adjacent photoproducts could be formed in G-quadruplexes. It is also possible that the photocrosslinking occurred in a basket conformation, but because photocrosslinking was inefficient in LiCl solution it is unlikely to have involved a Hoogsteen or reverse Hoogsteen hairpin conformation. In neither quadruplex was the actual photoproduct identified, and could be a CPD or possibly a (6-4) photoproduct (**Figure 10A**) based on the fact that strand cleavage was induced by hot pyrrolidine (42, 83). Because DNA cleavage of adjacent (6-4) photoproducts only occurs on the pyrimidone side (84), however, DNA cleavage at the pyrimidine of a non-adjacent (6-4) product would not be expected to occur. A (5-4) product arising from a head to tail photoreaction (**Figure 10B**), however, might be expected to cleave at the pyrimidine side in addition to the pyrimidone side because of the presence of a C6 hemi-aminal which is why DNA is cleavable at 5,6-dihydroxythymidine sites by base (85). Alternatively, *anti* CPDs may have a greater propensity for glycosyl bond hydrolysis than their *syn* isomers but this remains to be investigated.

Inter-strand photocrosslinking has been observed to occur between TTTTGGGGT strands and between TGGGGTTTTGGGGT strands and was used as evidence for intermolecular G-quadruplex formation (86). In another study, primer extension opposite the HSP1 G-quadruplex by human mitochondrial DNA polymerase was found to be inhibited in a UV dose dependent manner that was proposed to arise from photocrosslinking between the pyrimidine-rich loops, though it may have been due to the formation of adjacent photoproducts (87).

Interestingly the HSP1 G-quadruplex is a site of UV-induced deletions in mitochondrial DNA in human skin.

Adjacent and non-adjacent (6-4) and spore photoproduct formation.

One can imagine that if the regio- and stereochemistry of CPDs depends on DNA structure, so should the regio- and stereochemistry of other types of dipyrimidine photoproducts such as the (6-4) and spore photoproducts (**Figure 1A & 10A**). (6-4) photoproducts can photoisomerize to their Dewar valence isomers which introduces another stereocenter at 3'-C6 but leaves the stereochemistry of the 5'-C5 & C6 centers unchanged and therefore does not contribute any additional information about the precursor DNA (88, 89). Adjacent and non-adjacent (6-4) and SP photoproducts can be assayed by the same enzyme-coupled LC-MS/MS methods used for CPDs (54, 49, 46, 28). In addition, the pyrimidone ring of (6-4) as well as Dewar photoproducts can been mapped at the sequence level by hot piperidine cleavage (42, 83, 90). In principle, there are four possible adjacent (6-4) photoproducts that could arise from different combinations of *syn* and *anti*-glycosyl conformations. There are also four possible adjacent (4-6) regioisomeric photoproducts that could arise from the alternate photoreaction of the 5'-pyrimidine with the 5,6-double bond of the 3'-pyrimidine. In spite of these possibilities only the (*6S,4P*) (6-4) product of adjacent pyrimidines have ever been isolated and characterized. The chirality of the pyrimidone ring and whether a (6-4) or (4-6) product was involved, however, becomes lost in the enzymatic degradation products of non-adjacent photoproducts because of free rotation due to the absence of a constraining internucleotide linkage. As a result, only two enzymatic degradation products of non-adjacent (6-4) products are produced and have the *6S* or *6R* chirality. The structures of the enzymatic digestion of adjacent and non-adjacent (6-4)

photoproducts arising from of all combinations of T and dC is given in **Supplementary Figure S9**.

In principle, there are four possible adjacent spore photoproducts (24). Two stereoisomers would arise from bond formation between the 3'-C5 methyl group and the 5'-C5 with either 5*R* or 5*S* stereochemistry, while the other two would arise from a regioisomer in which bond formation occurs between the 5'-C5 methyl group and the 3'-C5 with either 5*R* or 5*S* stereochemistry. The only adjacent SP product that has been isolated from irradiation of TpT, A DNA, or spores, however, is the (5'-5*R*)-SP product **Figure 1A & 10A** (24). As the case with enzymatic degradation products of non-adjacent (6-4) products, there are only two possible non-adjacent SP products with 5*S* and 5*R* stereochemistry.

Non-adjacent SP and T(6-4)T photoproducts with both 6*S* and 6*R* configurations have been detected by enzyme-coupled LC-MS/MS methods in UVC-irradiated freeze-dried calf thymus DNA and 80% ethanol (**Table 3**) (28). They were identified by comparison to authentic products produced from irradiation of thymidine, though the ratios of the two stereoisomers were not reported. Because head to head and head to tail orientations would produce the same SP products it is not possible to deduce the tertiary structure of the DNA precursor. On the other hand, because the non-adjacent 6*S* and 6*R* (6-4) photoproducts retain the C6-C4 linkage, they must have been produced from a head to head orientation and may arise from bulge loop intermediates. The spore product was the major non-adjacent product in freeze dried DNA (15%) whereas the non-adjacent T-T (6-4) was the least frequent (1%). Both non-adjacent products were minor products in 80% ethanol. In contrast, the non-adjacent spore photoproduct was the major photoproduct of thymidine in frozen aqueous, lyophilized, and air dried ethanolic conditions whereas the non-adjacent (6-4) was a minor photoproduct. It still remains to be

determined which DNA structures give rise to non-adjacent SP and (6-4) photoproducts. It is also be possible, however, that a head to tail (5-4) photoproduct is also produced in some DNA structures such as Oxy-4 and Tet-4 as discussed above. Irradiation of thymine has been reported to produce a compound whose spectroscopic data is consistent with a (5-4) adduct (91). There is also evidence for the formation of (5-4) product from cytidine which eliminates water to produce a pyrimidine-pyrimidone product (**Supplementary Figure S9**) (92).

Conclusion. The structure and stereochemistry of dipyrimidine photoproducts are complex and reveal information about the secondary and tertiary structures of the precursor DNAs. Non-adjacent dipyrimidine photoproducts are of particular interest because they provide information on how DNA may be folded. New analytical methods will be required, however, to map non-adjacent photoproducts in genomic DNA so that they could be used to reveal the presence and location of unusual structures such as G-quadruplexes. Non-adjacent photoproducts are also expected to be much more disruptive to DNA structure than adjacent photoproducts and may be playing an under recognized role in mutagenesis, especially with regard to frameshift and deletion mutations.

Acknowledgements—This material is based upon work supported by the National Science Foundation under NSF Grant No. 2003688. Robert Hanson is also thanked for help in assigning chirality of various photoproducts.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Figure S1. Head to head adjacent and non-adjacent TT CPDs

Figure S2. Head to tail non-adjacent TT CPDs

Figure S3. Head to head adjacent and non-adjacent TdU CPDs

Figure S4. Head to tail non-adjacent TdU CPDs

Figure S5. Head to head adjacent and non-adjacent dUT CPDs

Figure S6. Head to tail non-adjacent dUT CPDs

Figure S7. Head to head dUdU CPDs

Figure S8. Head to tail non-adjacent dUdU CPDs

Figure S9. Adjacent and non-adjacent head to head (6-4) and non-adjacent head to tail (5-4) photoproducts

REFERENCES

1. Beukers, R., A. P. Eker and P. H. Lohman (2008) 50 years thymine dimer. *DNA Repair (Amst)* **7,** 530-543.

2. Cadet, J., S. Mouret, J. L. Ravanat and T. Douki (2012) Photoinduced damage to cellular DNA: direct and photosensitized reactions. *Photochem. Photobiol.* **88,** 1048-1065.

3. Schreier, W. J., P. Gilch and W. Zinth (2015) Early events of DNA photodamage. *Annu. Rev. Phys. Chem.* **66,** 497-519.

4. Cadet, J., A. Grand and T. Douki (2015) Solar UV Radiation-Induced DNA Bipyrimidine Photoproducts: Formation and Mechanistic Insights. In *Top Curr Chem*. (Edited by M. Barbatti, A. C. Borin and S. Ullrich), pp. 249-275. Springer International Publishing, Cham.

5. Mao, P., J. J. Wyrick, S. A. Roberts and M. J. Smerdon (2017) UV-Induced DNA Damage and Mutagenesis in Chromatin. *Photochem. Photobiol.* **93,** 216-228.

6. Improta, R. and T. Douki (2022) DNA photodamage: From light absorption to cellular responses and skin cancer. In *Comprehensive Series in Photochemical and Photobiological Sciences*. (Edited by L. Rhodes, E. Sage and M. Trotta), pp. 346. The Royal Chemical Society, Croydon.

7. Sinden, R. R. (2007) Slipped strand DNA structures. *Front. Biosci.* **12,** 4788.

8. Choi, J. and T. Majima (2011) Conformational changes of non-B DNA. *Chem. Soc. Rev.* **40,** 5893-5909.

9. Saini, N., Y. Zhang, K. Usdin and K. S. Lobachev (2013) When secondary comes first - the importance of non-canonical DNA structures. *Biochimie* **95,** 117-123.

10. Ganser, L. R., M. L. Kelly, D. Herschlag and H. M. Al-Hashimi (2019) The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell. Biol.* **20,** 474-489.

11. Batey, R. T., R. P. Rambo and J. A. Doudna (1999) Tertiary Motifs in RNA Structure and Folding. *Angew. Chem. Int. Ed.* **38,** 2326-2343.

12. Cadet, J. and T. Douki (2018) Formation of UV-induced DNA damage contributing to skin cancer development. *Photochem. Photobiol. Sci.* **17,** 1816-1841.

13. Park, H., K. Zhang, Y. Ren, S. Nadji, N. Sinha, J. S. Taylor and C. Kang (2002) Crystal structure of a DNA decamer containing a cis-syn thymine dimer. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 15965-15970.

14. Yokoyama, H. and R. Mizutani (2014) Structural Biology of DNA (6-4) Photoproducts Formed by Ultraviolet Radiation and Interactions with Their Binding Proteins. *Int. J. Mol. Sci.* **15**.

15. Giussani, A., L. Serrano-Andrés, M. Merchán, D. Roca-Sanjuán and M. Garavelli (2013) Photoinduced Formation Mechanism of the Thymine–Thymine (6–4) Adduct. *J. Phys. Chem. B* **117,** 1999-2004.

16. Favre, H. A. and W. H. Pwell (2014) CHAPTER P-9 Specification of Configuration and Conformation. In *Nomenclature of Organic Chemistry: IUPAC Recommendations and Preferred Names 2013*. pp. 1156-1292. The Royal Society of Chemistry.

17. Lu, C., N. E. Gutierrez-Bayona and J. S. Taylor (2021) The effect of flanking bases on direct and triplet sensitized cyclobutane pyrimidine dimer formation in DNA depends on the dipyrimidine, wavelength and the photosensitizer. *Nucleic Acids Res.* **49,** 4266-4280.

18. Setlow, P. and L. Li (2015) Photochemistry and Photobiology of the Spore Photoproduct: A 50-Year Journey. *Photochem. Photobiol.* **91,** 1263-1290.

19. Setlow, B., A. R. Hand and P. Setlow (1991) Synthesis of a Bacillus subtilis small, acid-soluble spore protein in Escherichia coli causes cell DNA to assume some characteristics of spore DNA. *J. Bacteriol.* **173,** 1642-1653.

20. Mohr, S. C., N. V. H. A. Sokolov, C. M. He and P. Setlow (1991) Binding of Small Acid-Soluble Spore Proteins from Bacillus-Subtilis Changes the Conformation of DNA from B to A. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 77-81.

21. Douki, T., B. Setlow and P. Setlow (2005) Effects of the binding of alpha/beta-type small, acid-soluble spore proteins on the photochemistry of DNA in spores of Bacillus subtilis and in vitro. *Photochem. Photobiol.* **81,** 163-169.

22. Setlow, B. and P. Setlow (1993) Dipicolinic Acid Greatly Enhances Production of Spore Photoproduct in Bacterial Spores upon UV Irradiation. *Appl. Environ. Microbiol.* **59,** 640-643.

23. Douki, T., B. Setlow and P. Setlow (2005) Photosensitization of DNA by dipicolinic acid, a major component of spores of Bacillus species. *Photochem. Photobiol. Sci.* **4,** 591-597.

24. Mantel, C., A. Chandor, D. Gasparutto, T. Douki, M. Atta, M. Fontecave, P.-A. Bayle, J.-M. Mouesca and M. Bardet (2008) Combined NMR and DFT Studies for the Absolute Configuration Elucidation of the Spore Photoproduct, a UV-Induced DNA Lesion. *J. Am. Chem. Soc.* **130,** 16978-16984.

25. Lin, G. and L. Li (2010) Elucidation of spore-photoproduct formation by isotope labeling. *Angew. Chem. Int. Ed.* **49,** 9926-9929.

26. Singh, I., Y. J. Jian, L. Li and M. M. Georgiadis (2014) The structure of an authentic spore photoproduct lesion in DNA suggests a basis for recognition. *Acta. Crystallogr. D* **70,** 752-759.

27. Patrick, M. H. and D. M. Gray (1976) Independence of photoproduct formation on DNA conformation. *Photochem. Photobiol.* **24,** 507-513.

28. Douki, T., G. Laporte and J. Cadet (2003) Inter-strand photoproducts are produced in high yield within A-DNA exposed to UVC radiation. *Nucleic Acids Res.* **31,** 3134-3142.

29. Kundu, L. M., U. Linne, M. Marahiel and T. Carell (2004) RNA is more UV resistant than DNA: The formation of UV-induced DNA lesions is strongly sequence and conformation dependent. *Chem. Eur. J.* **10,** 5697-5705.

30. Cannistraro, V. J. and J. S. Taylor (2009) Acceleration of 5-methylcytosine deamination in cyclobutane dimers by G and its implications for UV-induced C-to-T mutation hotspots. *J. Mol. Biol.* **392,** 1145-1157.

31. Taylor, J.-S. (2022) Deamination of C-containing Cyclobutane Pyrimidine Dimers and Its Role in C to T and CC to TT Signature Mutations Caused by UV Light. In *DNA*

Photodamage: From Light Absorption to Cellular Responses and Skin Cancer. pp. 164-196. The Royal Society of Chemistry.

32. Hanson, R. M., S. Musacchio, J. W. Mayfield, M. J. Vainio, A. Yerin and D. Redkin (2018) Algorithmic Analysis of Cahn–Ingold–Prelog Rules of Stereochemistry: Proposals for Revised Rules and a Guide for Machine Implementation. *J. Chem. Info. Model.* **58,** 1755-1765.
33. Camerman, N. and S. C. Nyburg (1969) The crystal and molecular structure of a thymine photodimer (dimer `E'). *Acta Crystallogr. B* **25,** 388-394.
34. Kemmink, J., R. Boelens and R. Kaptein (1987) Two-dimensional 1H NMR study of two cyclobutane type photodimers of thymidylyl-(3′→5′)-thymidine. *Eur. Biophys. J.* **14,** 293-299.
35. Cadet, J., L. Voituriez, F. E. Hruska, L. S. Kan, F. A. A. M. Deleeuw and C. Altona (1985) Characterization of Thymidine Ultraviolet Photoproducts - Cyclobutane Dimers and 5,6-Dihydrothymidines. *Can. J. Chem.* **63,** 2861-2868.
36. Taylor, J. S., D. S. Garrett, I. R. Brockie, D. L. Svoboda and J. Telser (1990) 1H NMR assignment and melting temperature study of cis-syn and trans-syn thymine dimer containing duplexes of d(CGTATTATGC).d(GCATAATACG). *Biochemistry* **29,** 8858-8866.
37. Lingbeck, J. M. and J. S. Taylor (1999) Preparation and characterization of DNA containing a site-specific nonadjacent cyclobutane thymine dimer of the type implicated in UV-induced -1 frameshift mutagenesis. *Biochemistry* **38,** 13717-13724.

38. Su, D. G., J. L. Kao, M. L. Gross and J. S. Taylor (2008) Structure determination of an interstrand-type cis-anti cyclobutane thymine dimer produced in high yield by UVB light in an oligodeoxynucleotide at acidic pH. *J. Am. Chem. Soc.* **130,** 11328-11337.

39. Yasuda, S. and M. Sekiguchi (1970) T4 endonuclease involved in repair of DNA. *Proc. Natl. Acad. Sci. U.S.A.* **67,** 1839-1845.

40. Bourre, F., G. Renault and A. Sarasin (1987) Sequence effect on alkali-sensitive sites in UV-irradiated SV40 DNA. *Nucleic Acids Res.* **15,** 8861-8875.

41. Haseltine, W. A., L. K. Gordon, C. P. Lindan, R. H. Grafstrom, N. L. Shaper and L. Grossman (1980) Cleavage of pyrimidine dimers in specific DNA sequences by a pyrimidine dimer DNA-glycosylase of M. luteus. *Nature* **285,** 634-641.

42. Lippke, J. A., L. K. Gordon, D. E. Brash and W. A. Haseltine (1981) Distribution of UV light-induced damage in a defined sequence of human DNA: detection of alkaline-sensitive lesions at pyrimidine nucleoside-cytidine sequences. *Proc. Natl. Acad. Sci. U.S.A.* **78,** 3388-3392.

43. Weinblum, D. and H. E. Johns (1966) Isolation and properties of isomeric thymine dimers. *Biochim. Biophys. Acta* **114,** 450-459.

44. Douki, T. (2013) The variety of UV-induced pyrimidine dimeric photoproducts in DNA as shown by chromatographic quantification methods. *Photochem. Photobiol. Sci.* **12,** 1286-1302.

45. Su, D. G., H. Fang, M. L. Gross and J. S. Taylor (2009) Photocrosslinking of human telomeric G-quadruplex loops by anti cyclobutane thymine dimer formation. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 12861-12866.

46. Wang, Y., J. S. Taylor and M. L. Gross (1999) Nuclease P1 digestion combined with tandem mass spectrometry for the structure determination of DNA photoproducts. *Chem. Res. Toxicol.* **12,** 1077-1082.

47. Wang, Y., D. L. Rempel, J. S. Taylor and M. L. Gross (2001) A method for quantification from composite spectra: application to the determination of isomeric DNA photoproducts by tandem mass spectrometry. *Anal. Chem.* **73,** 185-191.

48. Douki, T. and J. Cadet (2001) Individual Determination of the Yield of the Main UV-Induced Dimeric Pyrimidine Photoproducts in DNA Suggests a High Mutagenicity of CC Photolesions. *Biochemistry* **40,** 2495-2501.

49. Wang, Y., M. L. Gross and J. S. Taylor (2001) Use of a combined enzymatic digestion/ESI mass spectrometry assay to study the effect of TATA-binding protein on photoproduct formation in a TATA box. *Biochemistry* **40,** 11785-11793.

50. Herbert, M. A., J. C. Leblanc, D. Weinblum and H. E. Johns (1969) Properties of thymine dimers. *Photochem. Photobiol.* **9,** 33-43.

51. Weinfeld, M., M. Liuzzi and M. C. Paterson (1989) Enzymatic analysis of isomeric trithymidylates containing ultraviolet light-induced cyclobutane pyrimidine dimers. II. Phosphorylation by phage T4 polynucleotide kinase. *J. Biol. Chem.* **264,** 6364-6370.

52. Liuzzi, M., M. Weinfeld and M. C. Paterson (1989) Enzymatic analysis of isomeric trithymidylates containing ultraviolet light-induced cyclobutane pyrimidine dimers. I. Nuclease P1-mediated hydrolysis of the intradimer phosphodiester linkage. *J. Biol. Chem.* **264,** 6355-6363.

53. Galloway, A. M., M. Liuzzi and M. C. Paterson (1994) Metabolic processing of cyclobutyl pyrimidine dimers and (6-4) photoproducts in UV-treated human cells. Evidence for distinct excision-repair pathways. *J. Biol. Chem.* **269,** 974-980.

54. Douki, T., M. Court, S. Sauvaigo, F. Odin and J. Cadet (2000) Formation of the main UV-induced thymine dimeric lesions within isolated and cellular DNA as measured by high performance liquid chromatography-tandem mass spectrometry. *J. Biol. Chem.* **275,** 11678-11685.

55. Douki, T. (2006) Effect of denaturation on the photochemistry of pyrimidine bases in isolated DNA. *J. Photochem. Photobiol. B* **82,** 45-52.

56. Liu, F.-T. and N. C. Yang (1978) Photochemistry of cytosine derivatives. 1. Photochemistry of thymidylyl-(3' →5')-deoxycytidine. *Biochemistry* **17,** 4865-4876.

57. Kao, J. L., S. Nadji and J. S. Taylor (1993) Identification and structure determination of a third cyclobutane photodimer of thymidylyl-(3'-->5')-thymidine: the trans-syn-II product. *Chem. Res. Toxicol.* **6,** 561-567.

58. McCullough, A. K., M. T. Romberg, S. Nyaga, Y. Wei, T. G. Wood, J. S. Taylor, J. L. Van Etten, M. L. Dodson and R. S. Lloyd (1998) Characterization of a novel cis-syn and trans-syn-II pyrimidine dimer glycosylase/AP lyase from a eukaryotic algal virus, Paramecium bursaria chlorella virus-1. *J. Biol. Chem.* **273,** 13136-13142.

59. Love, J. D., H. T. Nguyen, A. Or, A. K. Attri and K. W. Minton (1986) UV-induced interstrand cross-linking of d(GT)n.d(CA)n is facilitated by a structural transition. *J. Biol. Chem.* **261,** 10051-10057.

60. Nguyen, H. T. and K. W. Minton (1989) Extensive photodimerization of non-adjacent pyrimidines. *J. Mol. Biol.* **210,** 869-874.

61. Nguyen, H. T. and K. W. Minton (1988) Ultraviolet-induced dimerization of non-adjacent pyrimidines. A potential mechanism for the targeted -1 frameshift mutation. *J. Mol. Biol.* **200,** 681-693.

62. Love, J. D. and K. W. Minton (1992) Ultraviolet-induced dimerization of non-adjacent pyrimidines in poly[d(A-T)]. *J. Biol. Chem.* **267,** 24953-24959.

63. Karsisiotis, A. I., C. O'Kane and M. Webba da Silva (2013) DNA quadruplex folding formalism--a tutorial on quadruplex topologies. *Methods* **64,** 28-35.

64. Harkness, R. W. t. and A. K. Mittermaier (2017) G-quadruplex dynamics. *Biochim. Biophys. Acta Proteins Proteom.* **1865,** 1544-1554.

65. Dolinnaya, N. G., A. M. Ogloblina and M. G. Yakubovskaya (2016) Structure, Properties, and Biological Relevance of the DNA and RNA G-Quadruplexes: Overview 50 Years after Their Discovery. *Biochemistry (Mosc)* **81,** 1602-1649.

66. Wang, Y. and D. J. Patel (1993) Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure* **1,** 263-282.

67. Ambrus, A., D. Chen, J. Dai, T. Bialis, R. A. Jones and D. Yang (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.* **34,** 2723-2735.

68. Lim, K. W., S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J. Patel, K. N. Luu and A. T. Phan (2009) Structure of the human telomere in K+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J. Am. Chem. Soc.* **131,** 4301-4309.

69. Hou, X. M., Y. B. Fu, W. Q. Wu, L. Wang, F. Y. Teng, P. Xie, P. Y. Wang and X. G. Xi (2017) Involvement of G-triplex and G-hairpin in the multi-pathway folding of human telomeric G-quadruplex. *Nucleic Acids Res.* **45,** 11401-11412.

70. Lee, W. and S. Matsika (2017) Conformational and electronic effects on the formation of anti cyclobutane pyrimidine dimers in G-quadruplex structures. *Phys. Chem. Chem. Phys.* **19,** 3325-3336.

71. Smith, J. E., C. Lu and J. S. Taylor (2014) Effect of sequence and metal ions on UVB-induced anti cyclobutane pyrimidine dimer formation in human telomeric DNA sequences. *Nucleic Acids Res.* **42,** 5007-5019.

72. Lu, C., J. E. Smith-Carpenter and J. A. Taylor (2018) Evidence for Reverse Hoogsteen Hairpin Intermediates in the Photocrosslinking of Human Telomeric DNA Sequences. *Photochem. Photobiol.* **94,** 685-697.

73. Williamson, J. R., M. K. Raghuraman and T. R. Cech (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* **59,** 871-880.

74. Hardin, C. C., T. Watson, M. Corregan and C. Bailey (1992) Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d(CGCG3GCG). *Biochemistry* **31,** 833-841.

75. Sen, D. and W. Gilbert (1990) A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* **344,** 410-414.

76. Bhattacharyya, D., G. Mirihana Arachchilage and S. Basu (2016) Metal Cations in G-Quadruplex Folding and Stability. *Front. Chem.* **4,** 38.

77. Choi, K. H. and B. S. Choi (1994) Formation of a hairpin structure by telomere 3' overhang. *Biochim. Biophys. Acta* **1217,** 341-344.

78. Huppert, J. L. and S. Balasubramanian (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33,** 2908-2916.

79. Huppert, J. L. and S. Balasubramanian (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **35,** 406-413.

80. Rigo, R., M. Palumbo and C. Sissi (2017) G-quadruplexes in human promoters: A challenge for therapeutic applications. *Biochim. Biophys. Acta* **1861,** 1399-1413.

81. Gong, J.-y., C.-j. Wen, M.-l. Tang, R.-f. Duan, J.-n. Chen, J.-y. Zhang, K.-w. Zheng, Y.-d. He, Y.-h. Hao, Q. Yu, S.-p. Ren and Z. Tan (2021) G-quadruplex structural variations in human genome associated with single-nucleotide variations and their impact on gene activity. *Proc. Natl. Acad. Sci. U.S.A.* **118,** e2013230118.

82. Smith-Carpenter, J. E. and J. S. Taylor (2019) Photocrosslinking of G-Quadruplex-Forming Sequences found in Human Promoters. *Photochem. Photobiol.* **95,** 252-266.

83. Brash, D. E. and W. A. Haseltine (1982) UV-induced mutation hotspots occur at DNA damage hotspots. *Nature* **298,** 189-192.

84. Arichi, N., A. Inase, S. Eto, T. Mizukoshi, J. Yamamoto and S. Iwai (2012) Mechanism of the alkali degradation of (6-4) photoproduct-containing DNA. *Org. Biomol. Chem.* **10,** 2318-2325.

85. Palecek, E. (1992) Probing DNA structure with osmium tetroxide complexes in vitro. *Methods Enzymol.* **212,** 139-155.

86. Roberts, C., J. C. Chaput and C. Switzer (1997) Beyond guanine quartets: cation-induced formation of homogenous and chimeric DNA tetraplexes incorporating iso-guanine and guanine. *Chem. Biol.* **4,** 899-908.

87. Sullivan, E. D., M. J. Longley and W. C. Copeland (2020) Polymerase γ efficiently replicates through many natural template barriers but stalls at the HSP1 quadruplex. *J. Biol. Chem.* **295,** 17802-17815.

88. Taylor, J. S. and M. P. Cohrs (1987) DNA, Light, and Dewar Pyrimidinones - the Structure and Biological Significance of Tpt3. *J. Am. Chem. Soc.* **109,** 2834-2835.

89. Douki, T. and E. Sage (2016) Dewar valence isomers, the third type of environmentally relevant DNA photoproducts induced by solar radiation. *Photochem. Photobiol. Sci.* **15,** 24-30.

90. Mitchell, D. L., D. E. Brash and R. S. Nairn (1990) Rapid repair kinetics of pyrimidine(6-4)pyrimidone photoproducts in human cells are due to excision rather than conformational change. *Nucleic Acids Res.* **18,** 963-971.

91. Shetlar, M. D. and V. J. Basus (2013) The Photochemistry of Thymine in Frozen Aqueous Solution: Trimeric and Minor Dimeric Products. *Photochem. Photobiol.* **89,** 631-639.

92. Rhoades, D. F. and S. Y. Wang (1971) Photochemistry of polycytidylic acid, deoxycytidine, and cytidine. *Biochemistry* **10,** 4603-4611.

TABLES

Table 1. Stereochemical relationships of adjacent head to head T=dU and dU=T, and Thy=Ura and Ura=Thy CPDs. Relationships in () apply only to Pyr=Pyr products. Where not otherwise indicated, the products are diastereomers and would be separable by achiral chromatography.

dU=T & Ura=Thy	**T=dU & Thy=Ura**			
	R,S-cis,*syn*-I	*S,R-cis*,*syn*-II	*S,S-trans*,*syn*-I	*S,S-trans*,*syn*-II
R,S-cis,*syn*-I	(enantiomers)	identical		
S,R-cis,*syn*-II	identical	(enantiomers)		
S,S-trans,*syn*-I			identical	(enantiomers)
R,R-tran,*syn*-II			(enantiomers)	identical

Table 2. Stereochemical relationships of non-adjacent head to tail T=dU and dU=T, and Thy=Ura and Ura=Thy CPDs. Relationships in () apply only to Pyr=Pyr products. Where not otherwise indicated, the products are diastereomers and separable by achiral chromatography.

dU=T & Ura=Thy	**T=dU & Thy=Ura**			
	S,S-cis,*anti*-I	*R,R-cis*,*anti*-II	*R,*S-*trans*,*anti*-I	*S,R-trans*,*anti*-II
S,S-cis,*anti*-I	identical	(enantiomers)		
R,R-cis,*anti*-II	(enantiomers)	identical		
R,S-trans,*anti*-I			(enantiomers)	identical
S,R-trans,*anti*-II			identical	(enantiomers)

Table 3. Relative percentages of dithymidine photoproducts in genomic DNA and thymidine under various conditions.

	Type	*cis-syn*-I	*trans-syn*[1]	*cis-anti*[1]	*trans,anti*	SP[1]	(6-4)[1]
Genomic DNA							
native 0°C[2]	adjacent	94.2	0.2	-	-	-	5.7
native 90°C[2]	adjacent	77	14	-	-	-	9
freeze dried DNA[3]	adjacent	32	4	-	-	11	6
	non-adjacent	12	1	10	8	15	1
80% aq. ethanol[3]	adjacent	25	-	-	-	18	4
	non-adjacent	18	1	13	16	2	3
Thymidine							
frozen aqueous[3]	non-adjacent	10	12	24	3	17/21	13
lyophilized[3]	non-adjacent	12	22	28	7	14/14	4
air dried ethanolic[3]	non-adjacent	5	6	8	1	4/78	2
aqueous acetone[4]	non-adjacent	38	5/6	9/8	34	-	-

[1] Yields for adjacent photoproducts are presumed to be for *trans,syn*-I, (5*R*)-SP, and (6*S*,4*P*)-(6-4), while those for non-adjacent photoproducts are the sum of the two possible stereoisomers if not specified. [2] UVC (55). [3] UVC, adjacent *cis*,*syn* and (6-4) values based on statements and data in the text and referenced data (28). [4] UVB sensitized, based on isolated yields (35).

FIGURES.

Figure 1. Dipyrimidine photoproducts of various types of DNA secondary structure. A) *Cis,syn* thymidine dimer, (6-4) and spore photoproducts arising from adjacent T's in A and B DNA. The structure and stereochemistry of the photoproducts results from relative positioning and orientation of the precursor thymidines in a right handed Watson-Crick base paired antiparallel double helix (bolded base on top) and from the *anti* glycosyl conformations of the thymidines. B) Adjacent and non-adjacent head to head (*syn*) and head to tail (*anti*) dipyrimidine photoproducts that result from various types of DNA secondary structures

Figure 2. Structure and stereochemistry of adjacent and non-adjacent thymidine CPDs formed in DNA. A) Structures of the precursor nucleotides, and structures of B) adjacent and non-adjacent head to head and C) non-adjacent head to tail T=T CPDs following deamination and enzymatic degradation with SVP, NP1 and BAP (ED), and Thy=Thy CPDs following acid hydrolysis (H^+). T refers to thymidine, d^mC to 5-methyldeoxycytidine which readily deaminates to T in a CPD, Thy to thymine, dR to deoxyribose, and dRpdR to the sugar phosphate backbone of a dinucleotide. An asterisk indicates that the CPD is an enantiomer of the CPD to the other side of a mirror plane shown as a dotted line. CPDs indicated to be *meso* can been seen to be identical when X=Y by rotating 180^o about the axis shown in the plane of the paper or perpendicular to the plane as indicated by a dot. The *R/S* designations refers to the chirality of the C5 carbon in order of the first and second pyrimidine. Detailed structures are given in **Supplementary Figures S1 & S2**.

Figure 3. Products of enzymatic degradation of adjacent and non-adjacent CPDs formed in DNA. A) *Syn*-T=T CPD-containing trinucleotides and dinucleotides from enzymatic degradation of adjacent *cis,syn*- and *trans,syn*-T=T CPDs. The trinucleotide produced from SVP can be dephosphorylated and rephosphorylated for post-labeling assays. B) *Syn*- and *anti*-T=T CPD-containing tetranucleotides pT(pN)=pTpN and pNpT=(pN)pT, and dinucleotides T=T from enzymatic degradation of non-adjacent CPDs. The tetranucleotide produced from SVP can also be dephosphorylated and rephosphorylated for postlabeling assays (Gutierrez et al., unpublished results). Enzymes used are: NP1, nuclease P1; SVP, snake venom phosphodiesterase; and BAP, bacterial alkaline phosphatase.

Figure 4. DNA structures that produce adjacent *syn*-T=T CPDs and non-adjacent *cis*-T=T CPDs. A) Formation of adjacent *trans,syn*-I and –II CPDs requires one pyrimidine to be in a *syn* glycosyl conformation and the other *anti*. The *cis,syn*-II-CPD would require both T's to be *syn*. The bolded base is on top. B) Formation of a non-adjacent *cis,syn*-T=T CPD in an alternating polymer of A and T via a bulge loop structure; C) formation of a *cis,syn*-I-T=T CPD in a bulge loop duplex DNA; D) structure of a non-adjacent *cis,syn*-I-T=T CPD; and E) an oligodeoxynucleotide with no known folded structure that produces a high yield of a *cis,anti*-I-T=T CPD.

Figure 5. Various types of intramolecular G-quadruplex structures and folding/unfolding intermediates. A) The parallel quadruplex with all strands parallel and propeller loops P that prevent CPD formation between loops. B) A triplex intermediate between hybrid-II and chair forms with two lateral loops L. CPD formation can occur between lateral loop 1 and the

disordered loop 3. C) The hybrid-2 conformation with two antiparallel strands, two lateral loops and one propeller loop. CPD formation cannot take place between any of the loops. D) The basket conformation with two antiparallel strands, one diagonal loop D and two lateral loops 1 and 3 between which CPD formation can occur. E) The chair conformation with two antiparallel strands and three lateral loops in which CPD formation could take place between loops 1 and 3. F) A long hairpin loop that is in equilibrium with the chair form in which CPD formation could take place between loops 1 and 3. Figure adapted from Figure 1 of Smith-Carpenter and Taylor (82).

Figure 6. Hoogsteen and reverse Hoogsteen base pairing and its relationship to relative strand orientation and G-quartet formation. The G quartets of G quadruplexes arise from the cyclic arrangement of four Hoogsteen base-paired G's which have negative electrostatic potential at the center due to the remaining non-H bonded lone pairs on the four oxygens that are stabilized by metal ions. The strands formed between a Hoogsteen base paired G-G can either be parallel or antiparallel depending on the glycosyl conformations which are indicated by a bold line for *anti*, and dashed line for *syn*. In the lowest energy *anti* glycosyl conformation Hoogsteen base paired G's form parallel strands whereas reverse Hoogsteen paired G's form *anti*-parallel strands. While A-T can base pair by either Hoogsteen or reverse Hoogsteen base pairing, A-A can only form reverse Hoogsteen base pairs that form antiparallel strands when all the A's are in the lowest energy *anti* glycosyl conformation. Inosine (I) can only serve as a Hoogsteen H-bond acceptor due to the absence of a C2-NH_2 group. The table shows the relative strand orientations for various combinations of *anti* and *syn* glycosyl conformations for a Hoogsteen or reverse Hoogsteen base pair.

Figure 7. Non-adjacent *trans,anti* CPD formation in G-quadruplex and reverse-Hoogsteen forming sequences. A) Sequences derived from human telomeric DNA that formed *trans,anti* CPDs in KCl solution but not in NaCl solution. B) Sequences used to investigate the involvement of a reverse Hoogsteen hairpin in the formation of the *trans,anti*-CPDs in LiCl, NaCl and KCl solutions, where I is inosine. A filled dot represents Hoogsteen or reverse Hoogsteen base pairing, an empty dot represents only reverse Hoogsteen base pairing. An arrow represents the direction from the Watson Crick face of one base to the Hoogsteen face of the other base when there is only one possible base pairing orientation, and a solid line represents Watson-Crick base pairing. TA base pairs can adopt Watson-Crick, Hoogsteen, and reverse Hoogsteen base pairing. Other possible types of base pairing are not shown. The yields are for the *trans,anti* T2=T2 CPD which is the major non-adjacent CPD product in all cases.

Figure 8. Proposed scheme to explain the favored formation of the *trans,anti* T2(L1)=T2(L3) CPD in Tel26. A) Human telomeric sequences are proposed to adopt a non-adjacent CPD prohibiting basket structure in NaCl solution. In KCl solution, the sequences adopt the non-adjacent CPD prohibiting hybrid-I and II conformations, and the form 3 basket structure which can lead to various non-adjacent CPDs. In LiCl solution, G-quadruplexes are disfavored and the reverse Hoogsteen hairpin becomes favored which leads to the *trans,anti*-T2(L1)=T2(L3) CPD. The extra A's at the 5'- and 3'-ends of Tel26 further stabilize the reverse Hoogsteen hairpin in NaCl, KCl, and LiCl solutions favoring formation of the *trans,anti* CPD. B) Possible structure of basket structure with in plane Na^+ ions that prevents non-adjacent CPD formation. C) Possible structure of two-quartet form 3 basket structure with a single intercalated K^+ ion

enlarged loop 1 that facilitates *trans,anti* T2(L1)=T2(L3) CPD formation. D) Proposed reverse Hoogsteen hairpin that also facilitates *trans,anti* T2(L1)=T2(L3) CPD formation. Figure adapted from Figure 10 of Lu et al. (72).

Figure 9. G quadruplex structures that could explain the formation of non-adjacent CPDs in human promoter sequences and photocrosslinks in telomeric sequences from other organisms. A) Chair structures consistent with identification of pT(pA)=pTpG, pT(pG)=pTpG, pU(pG)=pTpG non-adjacent CPDs from NP1 degradation of UVB irradiated G-quadruplex forming sequences found in human promoter sequences. These products could either have arisen from interloop photoreactions indicated by thick dotted lines or intraloop photoreactions indicated by thin dotted lines. The photoproduct site in CALU was confirmed by substitution of T's with dU's. Shaded sites were found to be more reactive to dimethylsulfate. B) Sites of photocrosslinking of Tet-4 and Oxy-4 (X=T_4) as mapped onto the originally proposed chair conformations. The crosslinking site in Oxy-4, however, might be better explained by the basket structure shown. Photocrosslinking was not observed to be very efficient in lithium ion solution, suggesting that Hoogsteen or reverse Hoogsteen hairpins are not involved. Panel A adapted from Figure 11 of Smith-Carpenter and Taylor (82).

Figure 10. Structures of adjacent and non-adjacent (6-4) and spore photoproducts of DNA following enzymatic digestion. A) Stereochemistry of adjacent head to head (6-4) photoproducts formed in A and B DNA and spore photoproducts formed in A DNA, and non-adjacent products formed from other structures following enzymatic digestion with SVP, NP1 and BAP. B)

Possible head to tail regioisomer of the (6-4) photoproduct which is a (5-4) photoproduct. Non-adjacent head to tail SP products are indistinguishable from head to head products.

Graphic Abstract. While irradiation of duplex DNA produces adjacent *cis,syn* cyclobutane pyrimidine dimers (CPDs), irradiation of other structures can produce other isomers, such as the non-adjacent *trans,anti* CPD formed in G quadruplexes. This review investigates the various types of adjacent and non-adjacent dipyrimidine photoproducts and methods for their detection that could be used to identify the presence various types of unusual structures in DNA.

Figure 1.

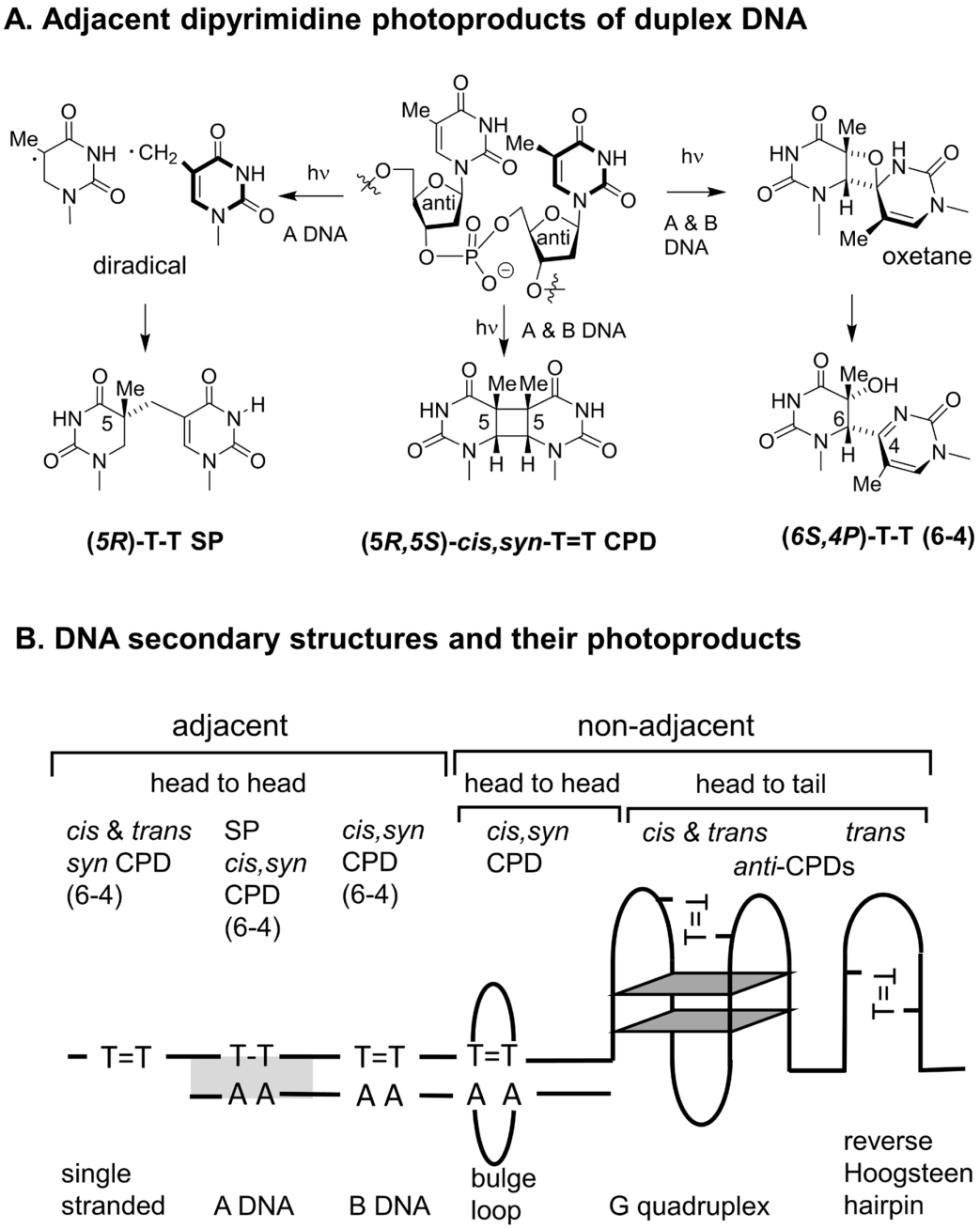

A. Adjacent dipyrimidine photoproducts of duplex DNA
diradical
hν
A DNA
anti
anti
hν
A & B DNA
hν
A & B DNA
oxetane
(5R)-T-T SP
(5R,5S)-cis,syn-T=T CPD
(6S,4P)-T-T (6-4)
B. DNA secondary structures and their photoproducts
adjacent
non-adjacent
head to head
head to head
head to tail
cis & trans syn CPD (6-4)
SP cis,syn CPD (6-4)
cis,syn CPD (6-4)
cis,syn CPD
cis & trans anti-CPDs
trans
T=T
T-T
T=T
T=T
A A
A A
A A
single stranded
A DNA
B DNA
bulge loop
G quadruplex
reverse Hoogsteen hairpin

Figure 2.

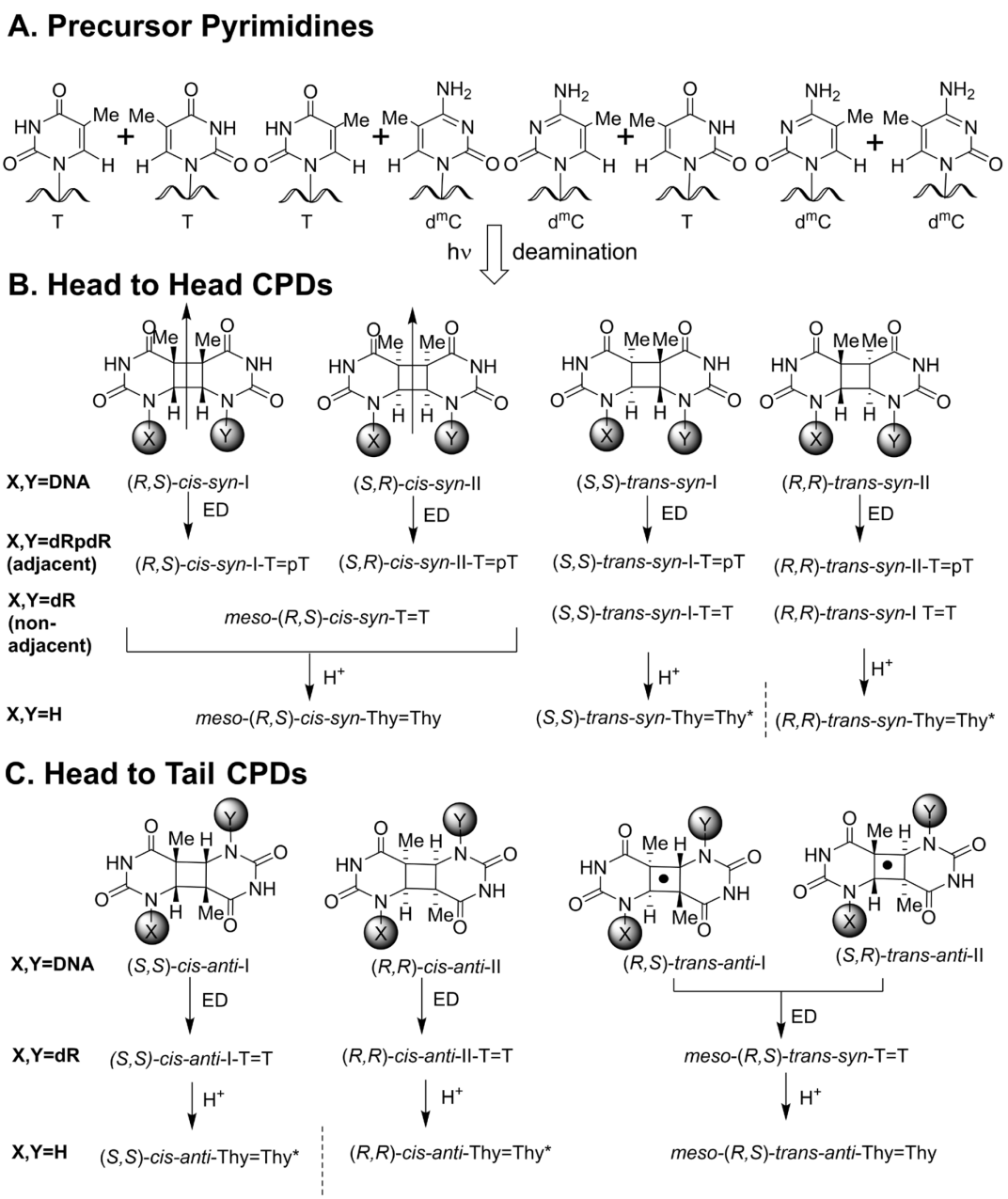

A. Precursor Pyrimidines
T
T
T
d^mC
d^mC
T
d^mC
d^mC
hν
deamination
B. Head to Head CPDs
X,Y=DNA
(R,S)-cis-syn-I
(S,R)-cis-syn-II
(S,S)-trans-syn-I
(R,R)-trans-syn-II
ED
X,Y=dRpdR (adjacent)
(R,S)-cis-syn-I-T=pT
(S,R)-cis-syn-II-T=pT
(S,S)-trans-syn-I-T=pT
(R,R)-trans-syn-II-T=pT
X,Y=dR (non-adjacent)
meso-(R,S)-cis-syn-T=T
(S,S)-trans-syn-I-T=T
(R,R)-trans-syn-I T=T
H+
X,Y=H
meso-(R,S)-cis-syn-Thy=Thy
(S,S)-trans-syn-Thy=Thy*
(R,R)-trans-syn-Thy=Thy*
C. Head to Tail CPDs
X,Y=DNA
(S,S)-cis-anti-I
(R,R)-cis-anti-II
(R,S)-trans-anti-I
(S,R)-trans-anti-II
ED
X,Y=dR
(S,S)-cis-anti-I-T=T
(R,R)-cis-anti-II-T=T
meso-(R,S)-trans-syn-T=T
H+
X,Y=H
(S,S)-cis-anti-Thy=Thy*
(R,R)-cis-anti-Thy=Thy*
meso-(R,S)-trans-anti-Thy=Thy

Figure 3.

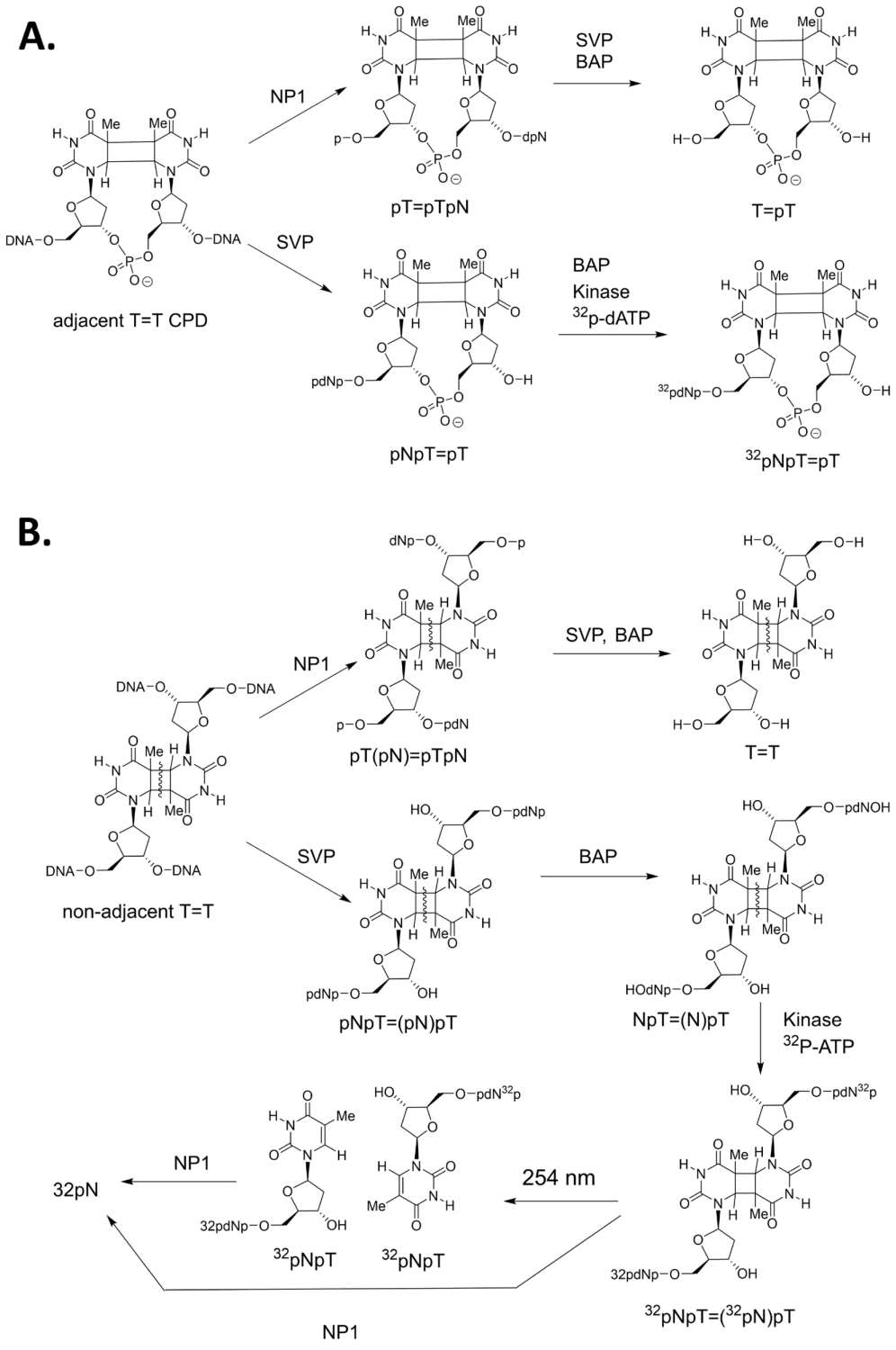

A.
NP1
SVP
BAP
pT=pTpN
T=pT
adjacent T=T CPD
SVP
BAP
Kinase
32p-dATP
pNpT=pT
32pNpT=pT
B.
NP1
SVP, BAP
pT(pN)=pTpN
T=T
non-adjacent T=T
SVP
BAP
pNpT=(pN)pT
NpT=(N)pT
Kinase
32P-ATP
254 nm
NP1
32pN
32pNpT
32pNpT
32pNpT=(32pN)pT
NP1

Figure 4.

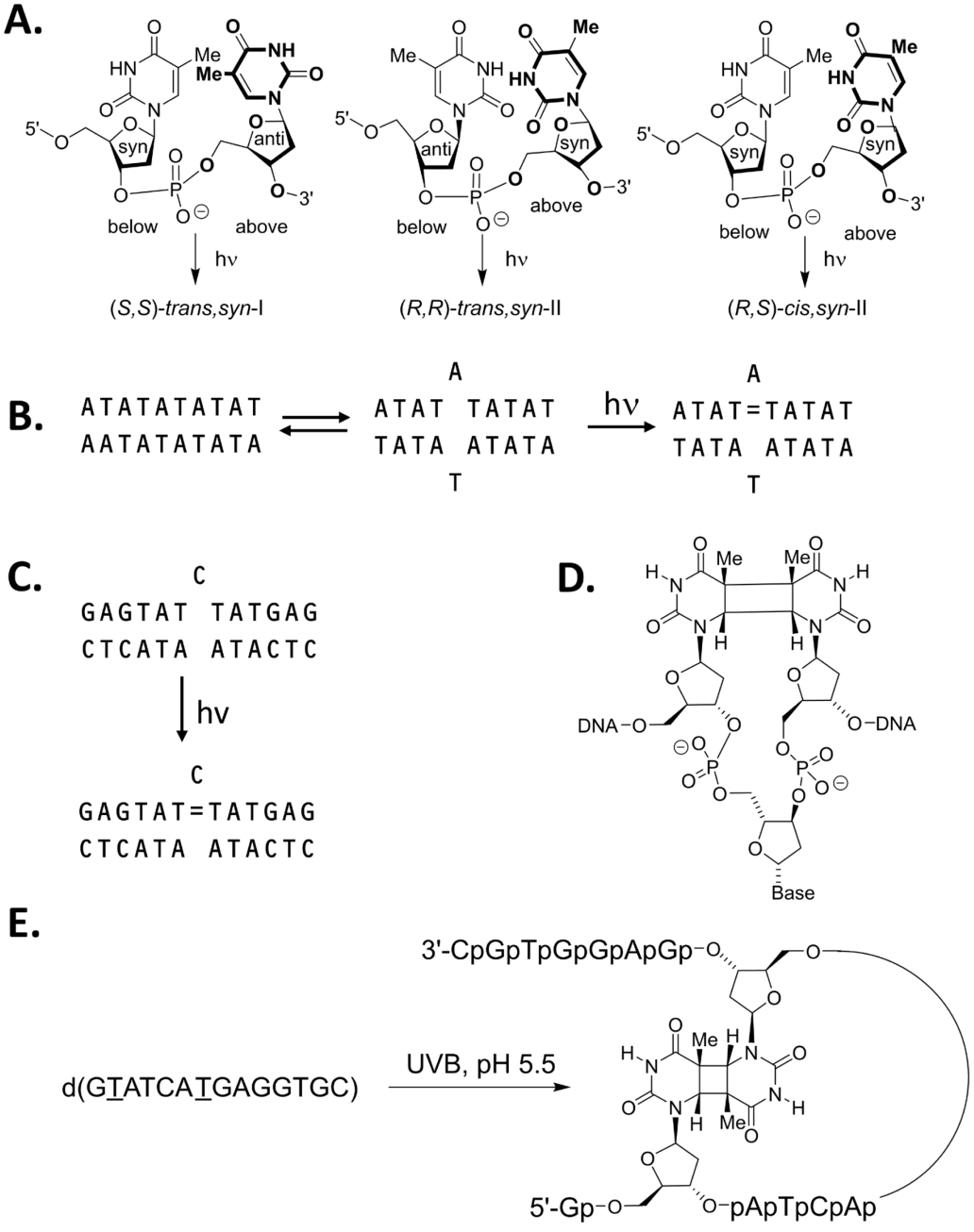

A.
(S,S)-trans,syn-I
(R,R)-trans,syn-II
(R,S)-cis,syn-II
B.
ATATATATAT
AATATATATA
ATAT TATAT
TATA ATATA
ATAT=TATAT
TATA ATATA
C.
GAGTAT TATGAG
CTCATA ATACTC
hν
GAGTAT=TATGAG
CTCATA ATACTC
D.
E.
3'-CpGpTpGpGpApGp
d(GTATCATGAGGTGC)
UVB, pH 5.5
5'-Gp
pApTpCpAp

Figure 5.

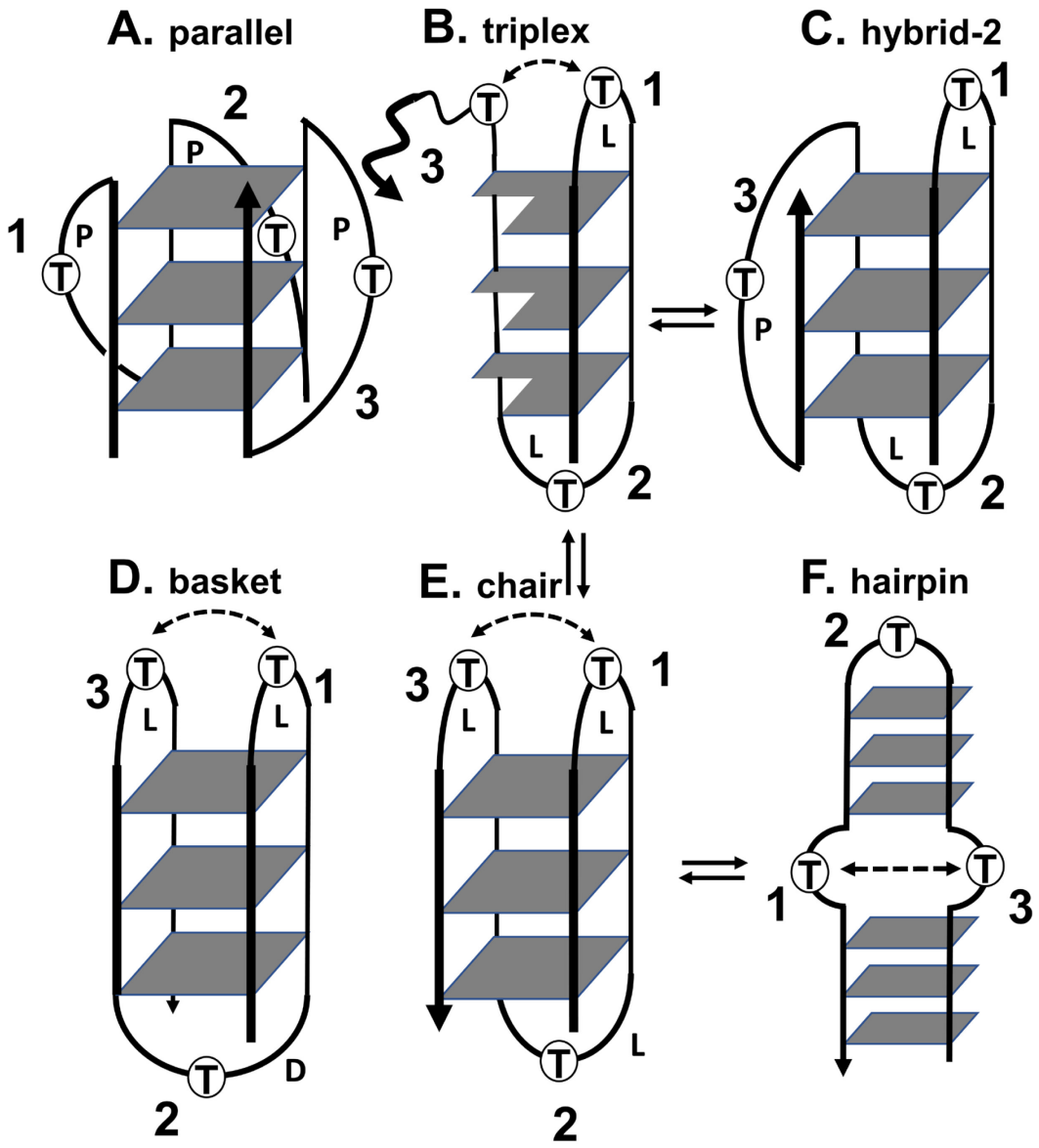
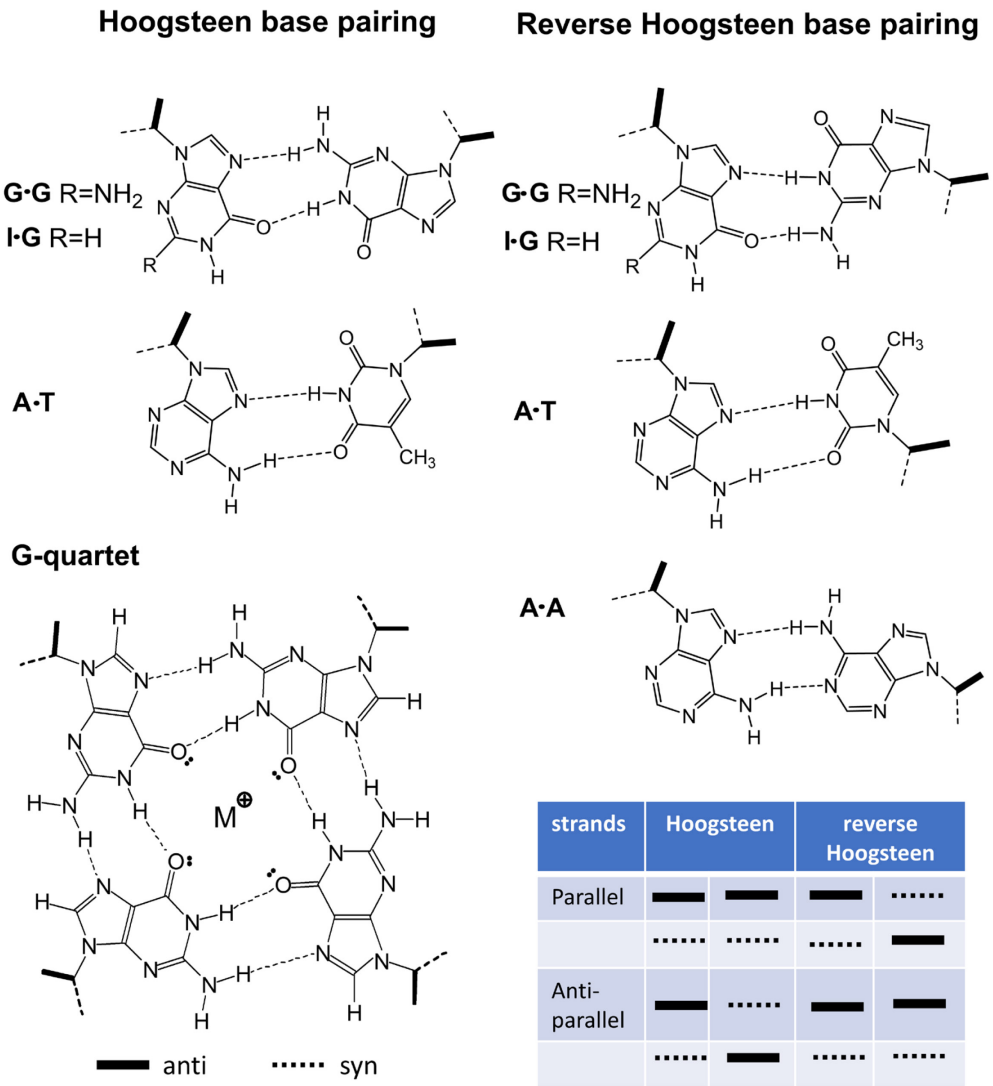

A. parallel
B. triplex
C. hybrid-2
D. basket
E. chair
F. hairpin
T
P
L
D
1
2
3

Figure 6.



strands	Hoogsteen		reverse Hoogsteen	
Parallel	▬	▬	▬	·····
	·····	·····	·····	▬
Anti-parallel	▬	·····	▬	▬
	·····	▬	·····	·····

Figure 7.

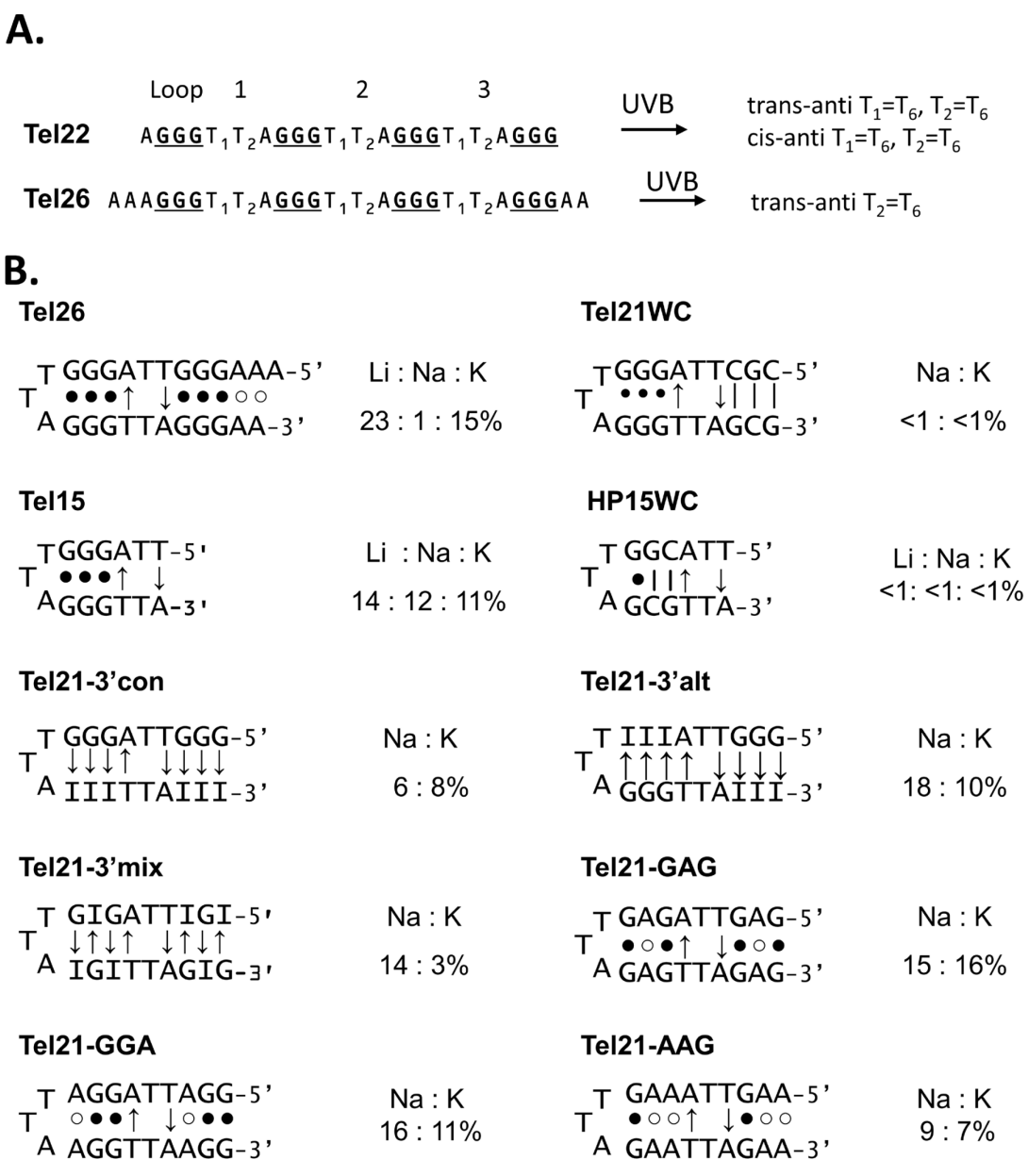

A.
Loop 1 2 3
Tel22
AGGGT1T2AGGGT1T2AGGGT1T2AGGG
UVB
trans-anti T1=T6, T2=T6
cis-anti T1=T6, T2=T6
Tel26
AAAGGGT1T2AGGGT1T2AGGGT1T2AGGGAA
UVB
trans-anti T2=T6
B.
Tel26
T GGGATTGGGAAA-5'
T ●●●↑ ↓●●●○○
A GGGTTAGGGAA-3'
Li : Na : K
23 : 1 : 15%
Tel21WC
T GGGATTCGC-5'
T ●●●↑ ↓| | |
A GGGTTAGCG-3'
Na : K
<1 : <1%
Tel15
T GGGATT-5'
T ●●●↑ ↓
A GGGTTA-3'
Li : Na : K
14 : 12 : 11%
HP15WC
T GGCATT-5'
T ●||↑ ↓
A GCGTTA-3'
Li : Na : K
<1: <1: <1%
Tel21-3'con
T GGGATTGGG-5'
T ↓↓↓↑ ↓↓↓↓
A IIITTAIII-3'
Na : K
6 : 8%
Tel21-3'alt
T IIIATTGGG-5'
T ↑↑↑↑ ↓↓↓↓
A GGGTTAIII-3'
Na : K
18 : 10%
Tel21-3'mix
T GIGATTIGI-5'
T ↓↑↓↑ ↓↑↓↑
A IGITTAGIG-3'
Na : K
14 : 3%
Tel21-GAG
T GAGATTGAG-5'
T ●○●↑ ↓●○●
A GAGTTAGAG-3'
Na : K
15 : 16%
Tel21-GGA
T AGGATTAGG-5'
T ○●●↑ ↓○●●
A AGGTTAAGG-3'
Na : K
16 : 11%
Tel21-AAG
T GAAATTGAA-5'
T ●○○↑ ↓●○○
A GAATTAGAA-3'
Na : K
9 : 7%

Figure 8.

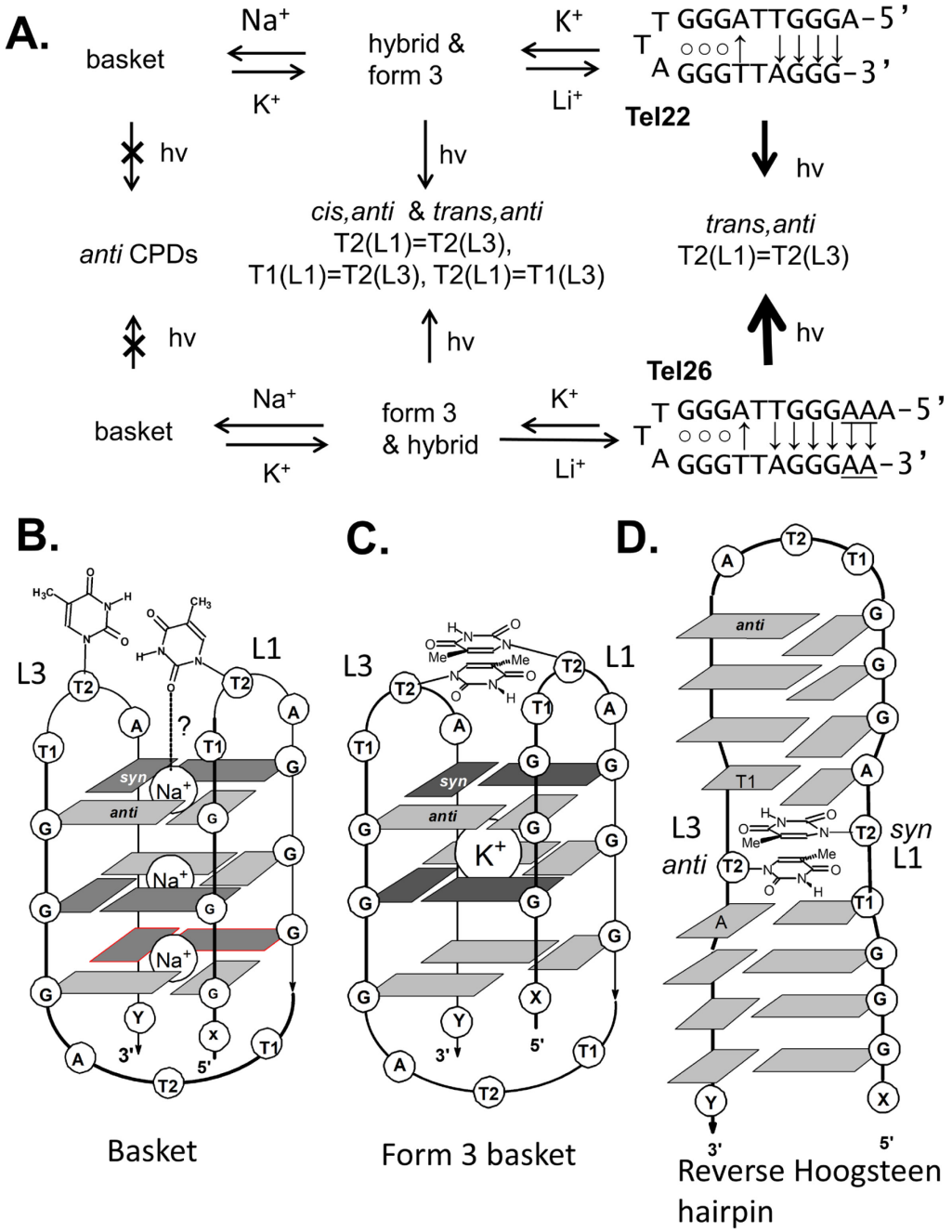

A.
basket
Na+
K+
hybrid &
form 3
K+
Li+
T GGGATTGGGA-5'
A GGGTTAGGG-3'
Tel22
hv
hv
hv
anti CPDs
cis,anti & trans,anti
T2(L1)=T2(L3),
T1(L1)=T2(L3), T2(L1)=T1(L3)
trans,anti
T2(L1)=T2(L3)
hv
hv
hv
Tel26
basket
Na+
K+
form 3
& hybrid
K+
Li+
T GGGATTGGGAAA-5'
A GGGTTAGGGAA-3'
B.
C.
D.
L3
L1
syn
anti
Na+
K+
3'
5'
Basket
Form 3 basket
Reverse Hoogsteen
hairpin

Figure 9.

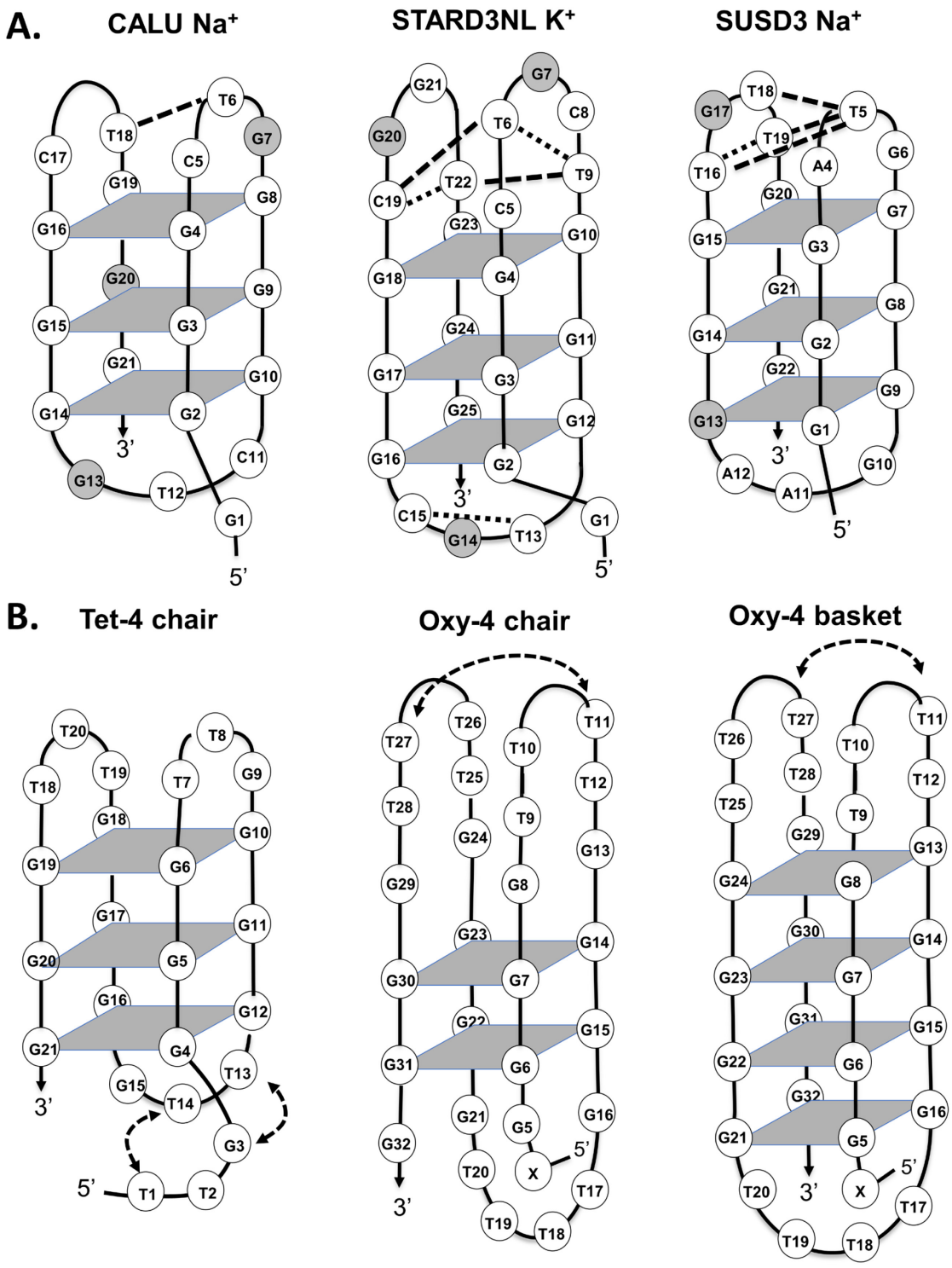
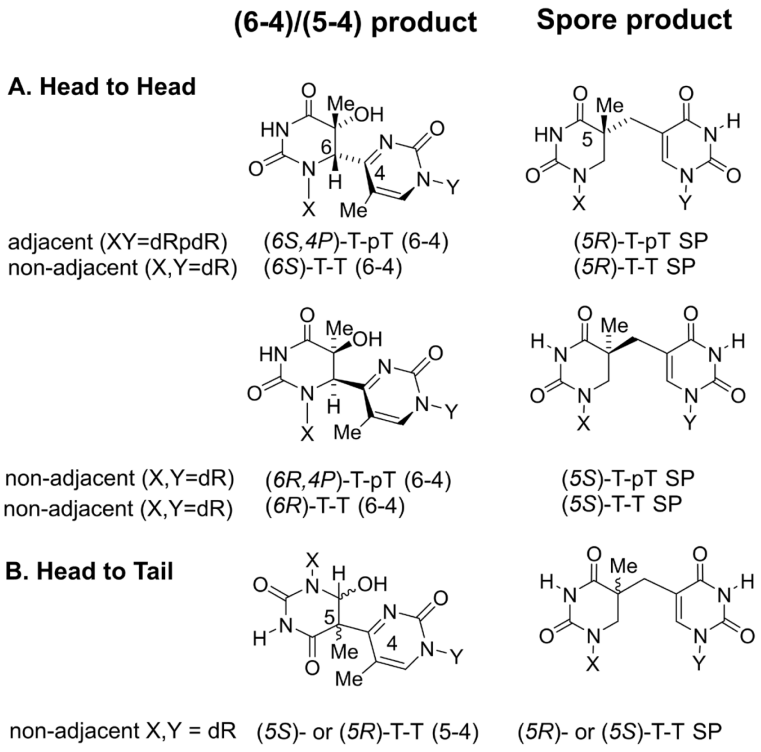

A.
CALU Na+
STARD3NL K+
SUSD3 Na+
B.
Tet-4 chair
Oxy-4 chair
Oxy-4 basket

Figure 10.


(6-4)/(5-4) product
Spore product
A. Head to Head
adjacent (XY=dRpdR)
non-adjacent (X,Y=dR)
(6S,4P)-T-pT (6-4)
(6S)-T-T (6-4)
(5R)-T-pT SP
(5R)-T-T SP
non-adjacent (X,Y=dR)
non-adjacent (X,Y=dR)
(6R,4P)-T-pT (6-4)
(6R)-T-T (6-4)
(5S)-T-pT SP
(5S)-T-T SP
B. Head to Tail
non-adjacent X,Y = dR
(5S)- or (5R)-T-T (5-4)
(5R)- or (5S)-T-T SP