

# Molecular Requirements of High-Fidelity Replication-Competent DNA Backbones for Orthogonal Chemical Ligation

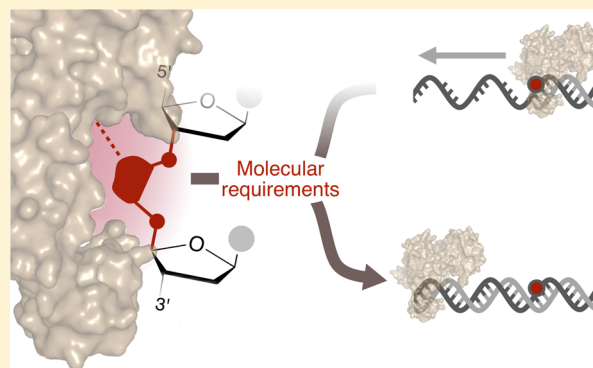
Arun Shivalingam,<sup>†,§</sup> Agnes E. S. Tyburn,<sup>†,§</sup> Afaf H. El-Sagheer,<sup>\*,†,‡</sup> and Tom Brown<sup>\*,†</sup>

<sup>†</sup>Department of Chemistry, Chemistry Research Laboratory, University of Oxford, 12 Mansfield Road, Oxford OX1 3TA, U.K.

<sup>‡</sup>Chemistry Branch, Department of Science and Mathematics, Faculty of Petroleum and Mining Engineering, Suez University, Suez 43721, Egypt

## Supporting Information

**ABSTRACT:** The molecular properties of the phosphodiester backbone that made it the evolutionary choice for the enzymatic replication of genetic information are not well understood. To address this, and to develop new chemical ligation strategies for assembly of biocompatible modified DNA, we have synthesized oligonucleotides containing several structurally and electronically varied artificial linkages. This has yielded a new highly promising ligation method based on amide backbone formation that is chemically orthogonal to CuAAC “click” ligation. A study of kinetics and fidelity of replication through these artificial linkages by primer extension, PCR, and deep sequencing reveals that a subtle interplay between backbone flexibility, steric factors, and ability to hydrogen bond to the polymerase modulates rapid and accurate information decoding. Even minor phosphorothioate modifications can impair the copying process, yet some radical triazole and amide DNA backbones perform surprisingly well, indicating that the phosphate group is not essential. These findings have implications in the field of synthetic biology.



## INTRODUCTION

Over evolutionary time, the molecular structure of DNA has become intricately linked with the enzymatic tools that propagate it. Despite this constraint, scientists have been able to expand the nucleobase alphabet,<sup>1–4</sup> evolve polymerase enzymes to replicate base- and sugar-modified nucleic acids,<sup>5–8</sup> and identify modified nucleic acids that are catalytically functional.<sup>9</sup> Interestingly, the key constant in these transformative studies is the phosphodiester backbone. Phosphate was abundant on prebiotic earth,<sup>10</sup> it has good reaction buffering capacity, it can act as a catalyst,<sup>11</sup> and it forms a stable phosphodiester linkage in DNA. However, it is possible that phosphate is not essential and that other carbon-, nitrogen-, or oxygen-based backbone variants could have evolved if different prebiotic conditions had prevailed. These thoughts led us to study the properties of artificial DNA backbone linkages. Relatively few studies have focused on replication or transcription through artificial DNA backbones, despite many analogues being synthesized for therapeutic applications.<sup>12–18</sup> Some basic information is available: minor phosphodiester modifications are accepted by polymerases,<sup>19–21</sup> an amide variant was imperfectly bypassed in primer–template experiments,<sup>22</sup> certain triazole-based backbones are tolerated by polymerases *in vitro*,<sup>23–25</sup> and one is even functional in bacterial and mammalian cells.<sup>26,27</sup>

In order to expand the boundaries of biologically functional DNA synthesis by chemical methods, we sought a greater

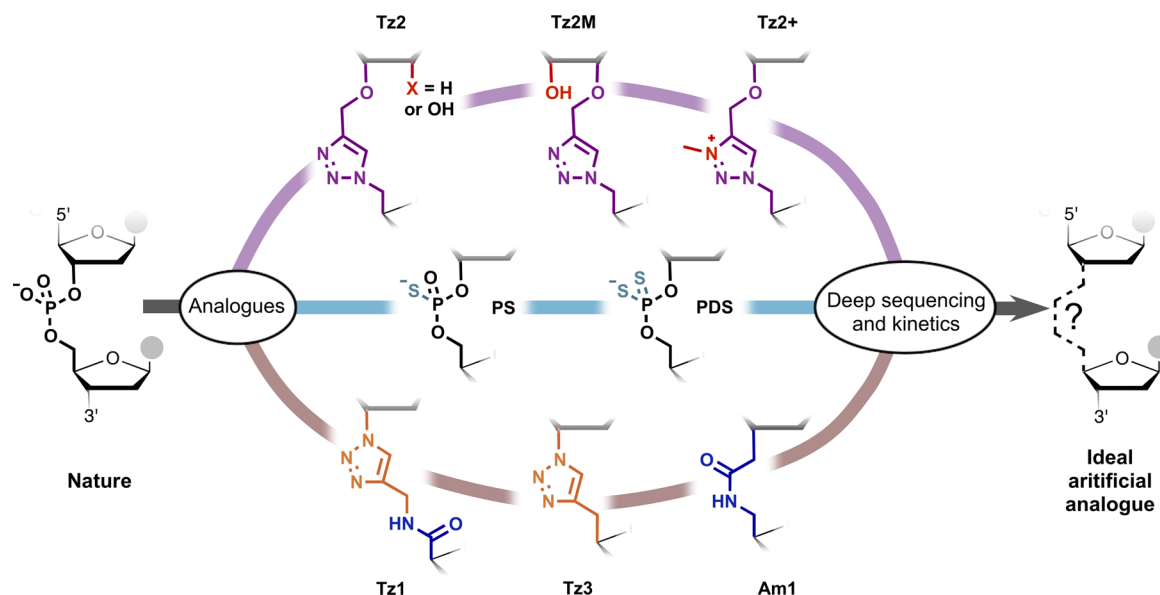
understanding of the molecular requirements for a high-fidelity replication-competent artificial DNA backbone. This might facilitate the design of novel analogues with useful properties. Indeed, nucleic acid linkages that are formed by efficient chemical ligation are beginning to find use in diverse fields ranging from chemical biology to nanotechnology,<sup>13,28–33</sup> and are necessary to produce dense site-specifically modified DNA, a demand that is likely to grow due to recent advances in genome editing, epigenetics, and synthetic biology. Here we study structurally and electronically varied artificial linkages (Figure 1) to identify the molecular requirements for an effective DNA backbone analogue. We also describe a new chemical ligation strategy to produce an amide DNA backbone linkage with excellent polymerase compatibility and read-through kinetics and show surprisingly that much-utilized phosphorothioate/dithioate modifications can impair replication fidelity.

## RESULTS AND DISCUSSION

**Design and Synthesis of Backbone Analogues.** The triazole backbones (Tz1<sup>23</sup> and Tz2, X = H,<sup>24</sup> Figure 1) separate 5' and 3' sugar rings by seven bonds, whereas the natural phosphodiester backbone does so by five bonds. To investigate the significance of this for DNA replication, a five bond triazole

Received: November 6, 2016

Published: January 18, 2017



**Figure 1.** Overview of modified DNA backbones, their structural relationships (color-coded lines), and strategy for identifying the requirements of an ideal analogue.

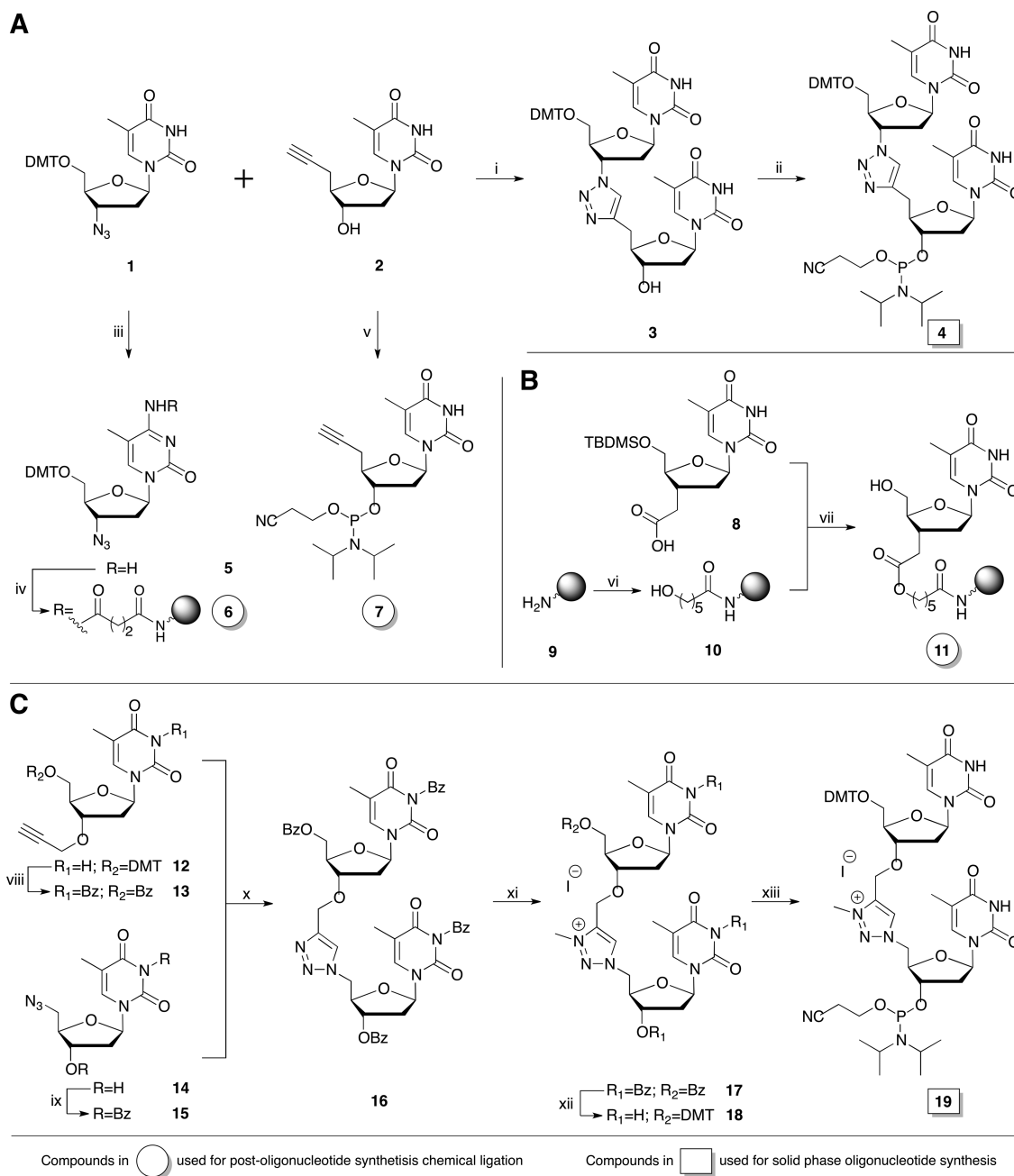
linkage was prepared (Tz3, Figure 1), which at full backbone substitution is duplex-stabilizing.<sup>34</sup> This structure retains the triazole of Tz1 but lacks the amide moiety. Thymidine triazole dimer 3 was synthesized from azide 1<sup>35</sup> and alkyne 2<sup>34</sup> and then converted to phosphoramidite 4 for use in oligonucleotide synthesis (Scheme 1A). An amide-containing thymidine dimer phosphoramidite was then synthesized (Am1, Figure 1),<sup>36</sup> which also separates 5' and 3' sugar rings by five bonds. Duplex NMR studies indicate that the amide oxygen mimics a phosphate oxygen, potentially allowing it to form hydrogen bonds with polymerase enzymes.<sup>37</sup>

To evaluate the formation of Am1 and Tz3 backbones by controlled oligonucleotide ligation, DNA sequences bearing terminal 3'-azide or -carboxylic acids and 5'-alkynes or primary amines were required. For 3'-azide oligonucleotides, despite the risk of deleterious Staudinger reduction,<sup>38</sup> previous reports indicate that resin-bound azides are compatible with the phosphorus(III) monomers used in solid-phase oligonucleotide synthesis.<sup>39,40</sup> Therefore, 3'-thymidine azide derivative (1) was converted to its 5-methylcytosine analogue (5) and attached via the C4 amine to a solid support to give 6, which performed well in oligonucleotide synthesis (Scheme 1A). 5'-Alkyne oligonucleotides were prepared using monomer 7 and ligated to 3'-azide oligonucleotides to form Tz3 in near quantitative yields using aqueous CuAAC conditions (Supplementary Methods and Supplementary Figure 1). For 3'-carboxylic acid oligonucleotides, thymidine analogue 8<sup>36</sup> was esterified onto a hydroxyl-functionalized solid support (10) and deprotected to provide resin 11 for use in oligonucleotide synthesis (Scheme 1B). Cleavage of the oligonucleotide from the solid support required sodium hydroxide, as the conventional concentrated ammonium hydroxide resulted in some amide formation. The required 5'-amino oligonucleotides were obtained using a commercially available phosphoramidite, and duplex-templated oligonucleotide ligation using EDC/NHS proceeded in near quantitative yield (Supplementary Methods and Supplementary Figure 2 and 3). Importantly this amide-based ligation strategy is chemically orthogonal to CuAAC ligation, providing a wider

choice of chemistries for use in technologies such as DNA-encoded library generation.<sup>30</sup>

Next, we focused on Tz2 analogues to study the effect of local base pair stability on polymerase read-through. We employed known chemical ligation methods<sup>24,41</sup> to vary the sequence around the triazole linkage (TT, <sup>m</sup>CU, <sup>m</sup>CC, <sup>m</sup>CG-clamp<sup>42</sup>). NMR structural studies on double stranded DNA indicate that the N3 atom of triazole Tz2 occupies the same space as one of the phosphate oxygen atoms.<sup>43</sup> Consequently, triazole N3 could have an important function as a hydrogen bond acceptor in DNA-polymerase interactions. To test this hypothesis, a methylated (N-blocked) Tz2 analogue (Tz2+) was prepared as a phosphoramidite dimer for use in oligonucleotide synthesis (19, Scheme 1C). Synthesis of the dimer required benzoyl protection of alcohols and thymine N3 (12 → 13 and 14 → 15)<sup>33,44</sup> to prevent per-alkylation. The desired site-specific N3 triazole methylation was confirmed by HMBC-NMR (Supplementary Figure 4). Duplex NMR studies also revealed that the Tz2 linkage disturbs natural sugar positioning,<sup>43</sup> so to change the sugar pucker and triazole backbone trajectory, Tz2 was synthesized with ribonucleotides at the 5'-side of the triazole. Commercially available 3'-alkyne/2'-hydroxyl C, G, and U solid supports were used to prepare the required alkyne oligonucleotides. In addition, a natural phosphodiester backbone template containing 2'-OMe groups was prepared. These templates were used to evaluate the extent to which replication fidelity is linked to sugar conformation. Finally, an unnatural 2'-ribo-5'-deoxyribo-triazole linkage was synthesized to provide an extreme example of DNA backbone distortion (Tz2M, Figure 1).

With the exception of the commercially available 2'-OMe, phosphorothioate (PS) and phosphorodithioate (PDS) modifications, all backbones were introduced either as the appropriate phosphoramidite dimer or by chemical ligation of oligonucleotides (Figure 2, Supplementary Table 1 and 2). Templates contain a fixed 60-base region in which the modification is located, with randomized 18-base primer regions of fixed ACGT content that act as sequencing barcodes for the modification (Supplementary Table 3 and 4). In the

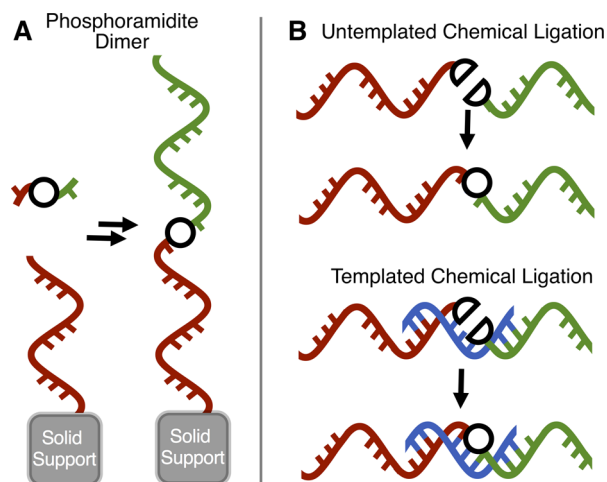
Scheme 1. Synthetic Routes to Phosphoramidites and Solid Supports Required for (A) Tz3, (B) Am1, and (C) Tz2+<sup>a</sup>

<sup>a</sup>Reagents and conditions: (A) (i)  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ , sodium ascorbate, tris(benzyltriazolylmethyl)amine, aq. dimethylformamide, rt, 2 h, 79%; (ii) 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite, *N,N*-diisopropylethylamine,  $\text{CH}_2\text{Cl}_2$ , rt, 2 h, 60%; (iii)  $\text{POCl}_3$ , *N*-methylimidazole, pyridine, 4 °C to rt, then aq.  $\text{NH}_3$ , rt, 17 h, 65%; (iv) succinylated resin, DIC, HOBt, pyridine, rt, 20 h; (v) 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite, *N,N*-diisopropylethylamine, THF, rt, 73%; (B) (vi) 6-*O*-DMT-hexanoic acid, DIC, HOBt, pyridine, rt, 20 h then acetic anhydride, *N*-methylimidazole, THF, pyridine, rt, 1 h then 3% TCA,  $\text{CH}_2\text{Cl}_2$ , rt; (vii) EDC-HCl, DMAP,  $\text{NEt}_3$ , pyridine, rt, 20 h then acetic anhydride, *N*-methylimidazole, THF, pyridine, rt, 1 h, then tetrabutylammonium fluoride, THF, rt, 4 h; (C) (viii) 3% TCA,  $\text{CH}_2\text{Cl}_2$ , rt, 1 h, then  $\text{BzCl}$ , pyridine, 4 °C to rt, 48 h, 92%; (ix)  $\text{BzCl}$ , pyridine, rt, 15 h, 94%; (x)  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ , sodium ascorbate, tris(benzyltriazolylmethyl)amine, aq. dimethylformamide, rt, 3 h, 97%; (xi) MeI, dimethylformamide, rt to 55 °C, 15 h, 92%; (xii) aq.  $\text{NH}_3$ ,  $\text{CH}_2\text{Cl}_2$ , rt, 20 h then DMT-Cl, pyridine, rt, 3 h, 26%; (xiii) 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite, *N,N*-diisopropylethylamine, dimethylformamide,  $\text{CH}_2\text{Cl}_2$ , rt, 3 h, 85%.

case of templates with identical primer regions, a nine base tailed PCR primer was used for barcoding in sequencing studies (Supplementary Table 5).

**Linear Copying of Modified DNA Templates.** Having prepared the DNA templates, we performed primer extension to investigate the read-through properties of the unnatural linkages. For Klenow polymerase at 37 °C, backbone modified

templates were either almost completely replicated (<5 min; Am1, PS, PDS, Tz3, Tz2<sup>m</sup>CC variant), displayed a time-dependent increase in full-length product formation (5–240 min; Tz2 TT and rUT variant), or consisted mainly of truncated products at or before the modification site (Tz2M, Tz2+, Tz1), as determined by denaturing gel electrophoresis and mass spectroscopy (Supplementary Figure 5 and



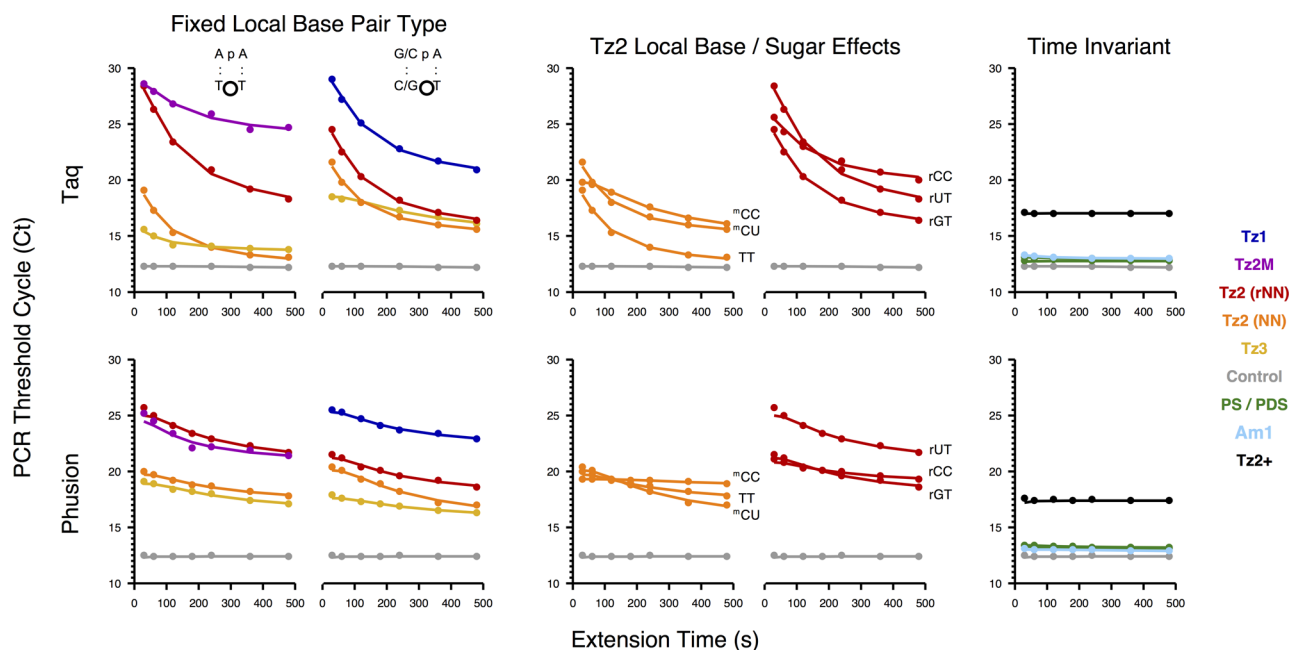
**Figure 2.** Backbone modifications were introduced either (A) as the preformed phosphoramidite dimer in oligonucleotide synthesis or (B) via postsynthetic chemical oligonucleotide ligation (templated or untemplated).

Supplementary Table 6). A significant deletion mutation was observed when replicating the Tz3 TT template (Supplementary Table 6), and this was confirmed by PCR and sequencing (discussed later). For thermostable Taq and Phusion polymerases at 60 °C (Supplementary Figure 6), primer extension from the unmodified control template was significantly slowed down by the high concentration of DNA, a known phenomenon.<sup>45</sup> This inhibition does not occur in the early cycles of PCR where template concentrations are very low (discussed in the next section). Nevertheless, the modified backbones that were replicated fastest by Klenow (Am1, PS, and PDS) remained faster with Taq and Phusion polymerases with the notable exception of Tz3. Most strikingly, Am1 is tolerated by polymerases to a remarkable extent, with Phusion

reading though it more efficiently than even the minor PS and PDS phosphodiester modifications.

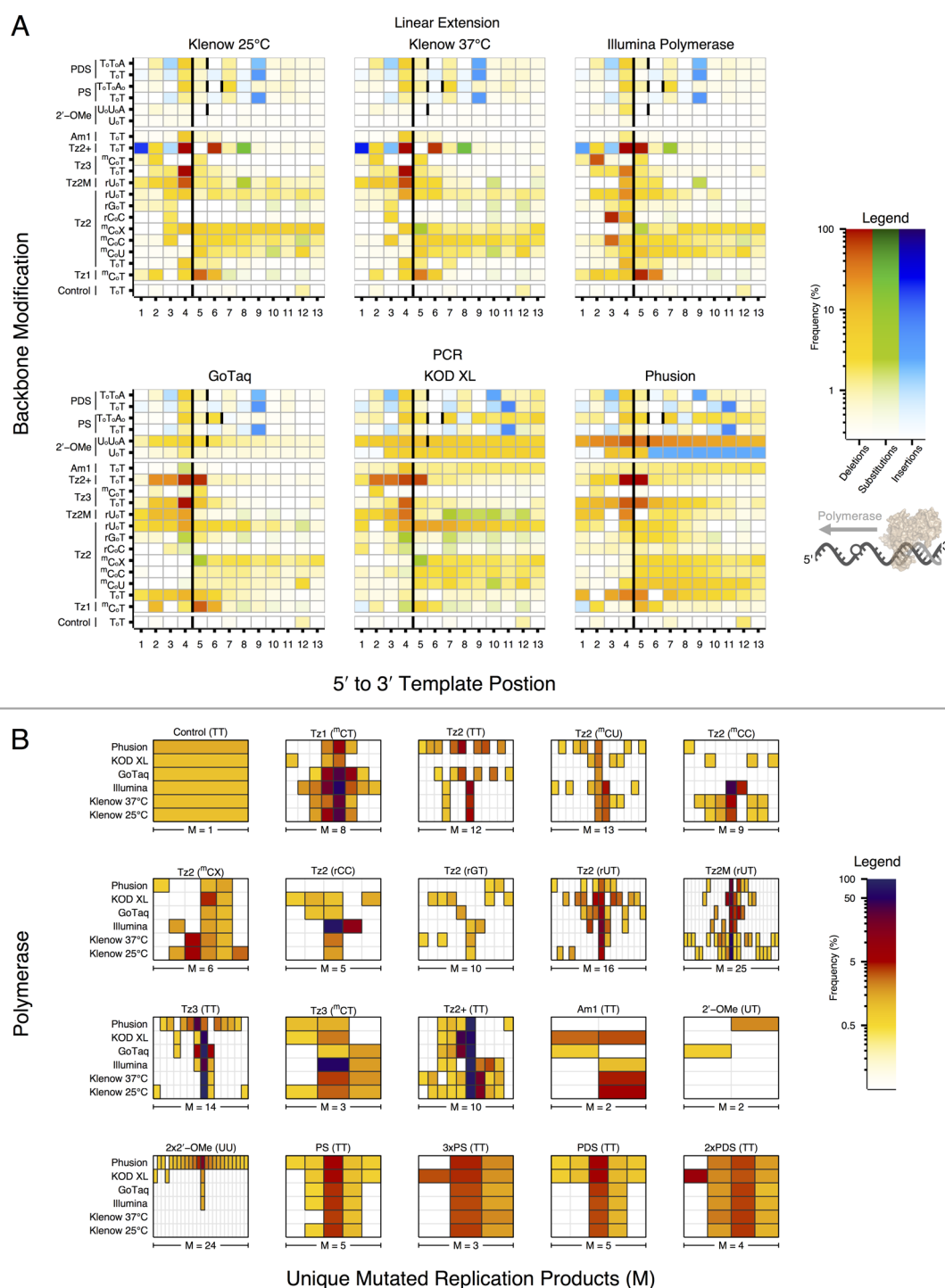
**PCR Amplification.** Linear copying indicates that read-through of some artificial DNA backbone linkages is rate limiting. As many of the applications of modified DNA templates are likely to involve PCR, this process was studied in detail. qPCR was performed as a function of extension time for each template, the premise being that after rate limited generation of the unmodified complementary DNA strand in the first cycle of PCR, amplification efficiency in subsequent cycles will revert to normal. Hence minor differences in artificial backbone read-through will be revealed by monitoring PCR efficiency. Importantly, the dynamic changes in component concentrations that will occur in PCR-based applications will be accounted for in a way that is not possible by linear copying. In this assay, if read-through of the artificial backbone is rate limiting, the PCR cycle at which products can be detected ( $C_t$ ) will decrease as extension times are increased. If false,  $C_t$  values will appear independent of extension times, indicating that the modified template is completely replicated within the shortest time frame assayed (30 s). In addition, comparison of  $C_t$  values for unmodified control vs modified templates will indicate which backbone modifications are read-through fastest by the polymerase. Different primers were used for each template to avoid PCR contamination issues and all primers had identical ACGT content to minimize efficiency-derived artifacts.

Using a hot-start Taq polymerase (Figure 3), amplification of the completely unmodified control template is naturally unaffected by variations in extension time. In contrast,  $C_t$  values for Tz1, Tz2, Tz2M, and Tz3 templates decrease (i.e., to earlier cycles) as extension time is increased, thus indicating that read-through of the unnatural linkage is rate limiting. Keeping the base pairs on either side of the artificial linkage constant gives rise to a similar order of backbone read-through for different base combinations (slowest to fastest, Tz2M < Tz1  $\ll$  Tz2  $\leq$  Tz3). This demonstrates that speed of replication is



**Figure 3.** Kinetics of backbone read-through. Templates were amplified by PCR and extension times varied (0.5–8 min) using hot-start Taq (exo−) and Phusion (exo+) polymerases. Templates are color-coded. Lines of best fit are for trend visualization. The hollow circle represents the site of the modified linkage. Full amplification curves and duplex melting curves confirming product formation can be found in Supplementary Figure 7.





**Figure 4.** Profile of mutations generated by DNA polymerases upon replication of modified backbones using Illumina deep sequencing. Median reads = 158 147; lower/upper quartile = 35 746/290 696. (A) Frequency of mutation at each position of the template as the polymerase approaches the site of modification. Obtained by global alignment of sequences to a master template, with all templates aligned relative to the backbone modification (black line) in the image. Note some templates have multiple backbone modifications (black lines). The main error type at each position is assigned as a substitution (green), insertion (blue), or deletion (red), where the color intensity corresponds to the frequency of observation. (B) Correlation between the unique mutated sequences observed ( $M$ , frequency  $>0.005$ ,  $x$ -axis) and polymerases ( $y$ -axis), with the frequency of observation color-coded. For example, Tz2+ (TT) displays one unique mutated sequence at high frequency for all polymerases (dark blue column). However, of the remaining nine unique mutated sequences, one of them is only observed for Phusion and at low frequency (last column, single pale yellow square). Full sequence profiles for panel A and unique mutated sequences for panel B with their associated frequencies are in [Supplementary Files 1 and 2](#), respectively. PCR and linear copying conditions are described in [Supplementary Methods](#). Conformation of PCR and linear copying products used for library preparation can be found in [Supplementary Figure 8](#).  $mCX = mCG$ -clamp.<sup>42</sup>

dependent on backbone structures and correlates well with their respective steric demands: Tz2M potentially generates the

greatest structural perturbation; Tz1 is long and rigid; Tz2 is longer but more flexible than Tz3. However, the local

nucleobases are also influential; comparing Tz2 base pair variants shows a negative correlation between base pair thermal stability and read-through rates ( $T_m = 50.1, 55.8, 59.2$  °C for TT,  ${}^m$ CU, and  ${}^m$ CC respectively; [Supplementary Table 7](#)). This trend is slightly distorted by uracil containing templates, of which DNA polymerases are less tolerant. Introduction of ribonucleotides 5' of Tz2 impairs extension kinetics, consistent with the effects of the ribo-modification and pyrimidine methylation state on duplex stability ( $\Delta T_m$ , Tz2 TT – rUT = +1.5 °C; Tz2  ${}^m$ CC – rCC = +2.7 °C; [Supplementary Table 7](#)).

Consistent with linear copying assays,  $C_t$  values for Am1, PS, and PDS show no extension time dependence and are within one cycle of the control, suggesting initial input template is completely read-through within the shortest extension time. On the other hand, for Tz2+, which also displays time-independent extension, the  $C_t$  values are notably higher than the control ( $\Delta C_t \approx 6$ ). These two observations suggest that a thermodynamic barrier exists to replication of Tz2+, which is overcome during PCR at some point between extension and denaturation (60 to 95 °C at 3.3 °C/s). The time spent in the temperature range was too short for the polymerase to bypass Tz2+, severely reducing the yield of the PCR amplicon. This phenomenon was observed only for the cationic Tz2+ backbone, suggesting an electrostatic origin, most likely repulsion by positively charged amino acid side chains in the polymerase DNA template recognition site.

Use of proofreading exo+ hot-start Phusion polymerase ([Figure 3](#)) gave  $C_t$  values comparable to exo– Taq polymerase for the control and time-invariant backbones. Moreover, the order of slowest to fastest time-dependent backbone read-through remained consistent with that of Taq. However, there was a lower local base pair bias, and Phusion generally replicated artificial linkages more efficiently than Taq at lower extension times. As these times were lengthened, product yields did not increase as significantly as for Taq, the reason for which will be discussed later.

**Replication Fidelity: Backbone Generated Mismatches, Insertions, and Deletions.** Our previous studies of Tz1 and Tz2 using Sanger sequencing indicated that the latter is accurately replicated by GoTaq polymerase.<sup>23,24</sup> However, the number ( $N$ ) of unique products sequenced (“reads”) was limited, thereby permitting only semiquantitative interpretation of the data. Here we use Illumina next-generation sequencing to significantly increase the number of reads per backbone modification per polymerase to generate a quantitative mutational profile ([Figure 4](#)).

Intuitively, “deep” sequencing could impair visualization of backbone-derived errors if those originating from Illumina platform-specific base calling and oligonucleotide synthesis are not taken into consideration. For the former, established postsequencing recalibration is not possible as the expected variants are unknown.<sup>46</sup> To mitigate this issue, reads were fully sequenced from both directions to obtain a merged higher-quality consensus sequence. For the latter, two independent but complementary analyses of the data account for the error differently.

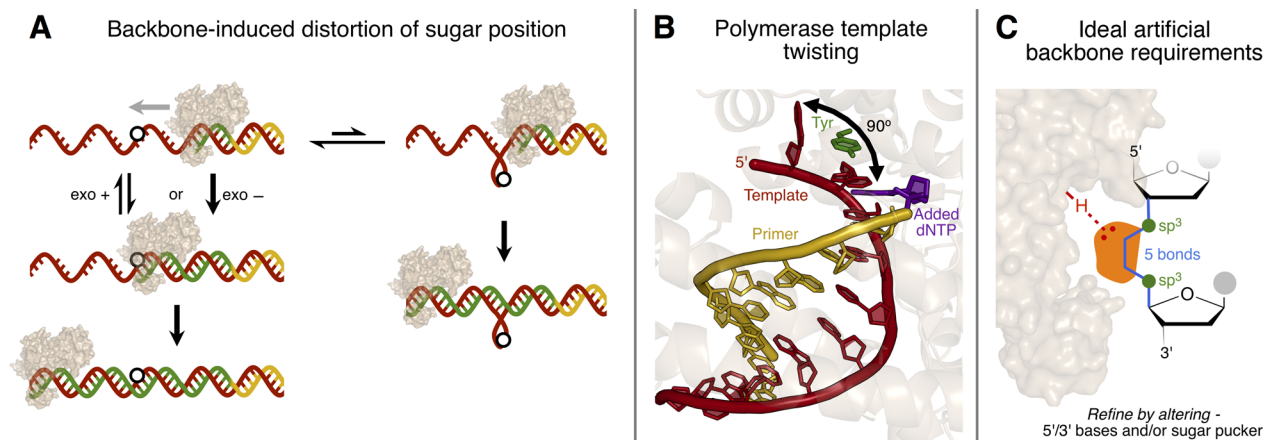
In the first approach, all reads were globally pairwise aligned to a single master template (see [Methods](#) for details). Next, at each position of the aligned sequence, bases observed below a statistically significant level (negative binomial test,  $p > 0.005$ ) or a frequency of 0.0009 (ca. 2–3 $\times$  oligonucleotide synthesis error rates per base<sup>47</sup>) were masked. To simplify visualization,

all templates were aligned relative to the backbone modification with the main error type at each position color-coded ([Figure 4A](#) and [Supplementary File 1](#)). By accumulating the errors from multiple replication events, this approach gives an averaged error “footprint” as the polymerases pass the unnatural linkage. In the second approach, unique sequences observed at a frequency greater than 0.005 (typically 1–25) were counted and pairwise aligned to the expected template, and the region of interest was identified (4/8 bases to 5'-/3'-sides of the modification, respectively). For visualization, the frequency (color-coded) of each unique mutated sequence ( $M$ ,  $x$ -axis columns) is correlated between different polymerases ( $y$ -axis rows) by backbone modification ([Figure 4B](#)). This enables easier identification of mutations that are common to specific polymerases (e.g., linear extension vs PCR polymerases) as well as identifying odd and highly frequent unique errors, the sequences for which are listed in [Supplementary File 2](#). This analysis reasonably assumes that oligonucleotide synthesis errors are randomly distributed while backbone-derived errors are not and places a higher threshold for error detection than the first approach. Moreover, it gives the exact error identities.

Modifications that generate low-level multibase deletions become progressively more intense toward the linkage site, suggesting they are polymerase-generated. This is to be expected when polymerases pass over the modification from the 3' to 5' side of the template ([Figure 4A](#), legend) since the linkage may form nonoptimal interactions with the enzyme. More surprisingly, this phenomenon also occurs 3' of the template modification where the polymerase has apparently not yet encountered the modification.<sup>48</sup> Only GoTaq, which lacks proofreading activity, exhibits significantly reduced deletions to the 3' side of the template modification. Therefore, we propose exo+ polymerase extension continues to the modification site, where it sometimes stalls and passes the primer terminus to the 3'  $\rightarrow$  5' exonuclease site, which arbitrarily digests the extended-primer.<sup>50</sup> This extension and digestion process may continue until either the modification is naturally passed or more rarely the polymerase loops the modified backbone out of the template to enable its unimpeded extension. For exo– polymerases, the looping mechanism is only accessed once, thus reducing the possibility of multibase deletions ([Figure 5A](#)). This postulate correlates well with the generally lower gains in modified backbone read-through efficiency as a function of extension time for Phusion (exo+) vs Taq (exo–); Phusion can iteratively attempt to access the looping mechanism even at lower extension times. Interestingly, this phenomenon appears to be linked to sugar distortions; supplementing the phosphodiester backbone with conformationally altered sugars (via 2'-OMe groups) gives similar deletion footprints to Tz2, which slightly displaces the sugar 3' to the linkage.<sup>43</sup>

Conversely, Am1 does not significantly perturb sugar placement or conformation,<sup>37</sup> which is consistent with the lower-level multibase deletions it generates (cf. Tz2). However, Am1 twists the glycosidic bond of the nucleoside 5' to the linkage (unlike Tz2). The resulting imperfect base orientation may be the source of the minor substitution mutations observed when using proofreading-deficient GoTaq.

It is noteworthy that 2'-OMe sugar-related deletions are negligible for Klenow and Illumina polymerases, suggesting that conformationally altered sugars are more easily addressed by these polymerases than displaced sugars. Surprisingly, ribonucleotides 5' to Tz2 generally mitigate multibase deletions; perhaps the sugar pucker of the ribonucleotide partially restores



**Figure 5.** Proposed features that the ideal artificial backbone must avoid or fulfill based on kinetic and sequencing data reported herein. (A) The mechanism by which backbone-induced distortion of the sugar position may generate multibase deletions around the linkage. (B) The 90° twist in template geometry induced by polymerases to enable dNTP addition (using PDB 2BDP<sup>49</sup>). (C) The requirements of an artificial backbone to enable the mechanism in panel B and thereby allow accurate recognition of the adjacent 5'/3' bases.

deoxyribose positioning 3' to the linkage, that is, two nonoptimal modifications negate the negative impact each has on fidelity. However, ribonucleotide-containing Tz2 templates do display mild substitutions in the absence of strong proofreading activity, suggesting the modification causes base misalignment in the enzyme active site.

Although Tz1 and Tz3 do not display as many multibase deletions, they can generate strong single-point deletions immediately adjacent to the linkage; for TT bases around the linkage, Tz1 always gives a single T deletion (GoTaq, Sanger sequencing,<sup>23</sup>  $N = 65$ ), whereas Tz3 gives 3–97% single T deletion (GoTaq = 87%, Figure 4B and Supplementary File 2). To rationalize this observation, it is known that polymerases twist the backbone of the template by ~90° immediately 5' to the site of dNTP addition (Figure 5B).<sup>48,51</sup> To facilitate this mechanism, backbone flexibility is likely to be crucial. Indeed, the point deletions around the artificial linkage (Tz1 > Tz3 >> Tz2 ≈ Am1, Figure 1) correlate well with the hybridization of the atom immediately adjacent to the 5'-/3'-side nucleosides ( $sp^2$  vs  $sp^3$ ), and poorly with internucleoside bond separation or the backbone functional group (triazole, amide, or triazole-amide). In support of this hypothesis, <sup>m</sup>CT variants of Tz1 and Tz3 display differing rates of 3'-side T deletion (Tz1 = 3–52%, Tz3 = 0–2%; Figure 4B and Supplementary File 2), yet both contain identical triazole motifs connected by a  $sp^2$ -hybridized center to the 5'-side nucleoside. Crucially Tz3 has a more flexible  $sp^3$ -hybridized  $CH_2$  connected to the 3'-side nucleoside (cf. Tz1  $sp^2$ -hybridized amide), thus reducing point deletions. As the base 5' to the modification becomes the site for dNTP addition, the artificial backbone is no longer twisted, but backbone flexibility is required for tyrosine stacking upon the templating base (Figure 5B).<sup>49,52</sup> Despite identical triazole motifs, longer Tz1 may facilitate this interaction better than Tz3, thus explaining its lower level of 5'-side <sup>m</sup>C deletions (Tz1 = 1–11%, Tz3 = 1–54%; Figure 4B and Supplementary File 2). Again  $sp^3$ -hybridization of the atom linked to the 5'-side nucleoside appears to offset this problem since Am1 displays minimal point deletions compared to Tz3 (both have an internucleoside bond separation of 5). It should be noted that the <sup>m</sup>C deletion appears to occur two bases distal to the 5'-side of template modification due to alignment issues; sequencing alone provides insufficient information to determine which

cytosine from a run of 2+ cytosines is deleted, and alignment biases the deletion to one site.

The general reduction in Tz1/Tz3 point deletions for <sup>m</sup>CT compared to TT to local bases may be the result of stronger C–Tyr  $\pi$ – $\pi$  interactions (dipole moment and base stacking in duplex DNA:  $C > T$ ).<sup>53</sup> This corroborates polymerase mutational studies,<sup>54,55</sup> which suggest that only aromatic residues can replace tyrosine and that the tyrosine phenyl ring chaperones templating base orientation. Overall, point deletion rates are highly dependent on the polymerase, suggesting that it may be possible to artificially evolve polymerases to perfectly accommodate modified DNA linkages.

Blocking the ability of the triazole to form hydrogen bonds by *N*-methylation (Tz2+) is catastrophic; a pair of bases is ignored by the polymerase during PCR, either around the linkage or two bases 5' to the modification site (Supplementary File 2), with Klenow and Illumina polymerases giving even odder error footprints. Although the triazole methyl group may create unwanted steric interactions, the structurally more intrusive Tz1/Tz2M modifications fail to generate similar error footprints and behave very differently kinetically. These striking observations demonstrate, for the first time, that hydrogen bonding acceptor capacity is a minimal requirement for polymerase backbone recognition.

Interestingly, the duplex-stabilizing G-clamp of Tz2 <sup>m</sup>CX is mutagenic via its T-like tautomer, despite it forming a very stable base pair with G. This corroborates the hypothesis that polymerases recognize base pairs by shape complementarity<sup>56,57</sup> and underlines the importance of deep sequencing; this error was not previously detected (<http://www.glenresearch.com/GlenReports/GR19-25.html> and ref 41).

Surprisingly, despite being far more similar to Nature's phosphodiester backbone than the other modifications studied, PS and PDS linkages show significant insertions, the length and position of which depend upon the polymerase (Figure 4A). Strong interactions between the sulfur atom and the polymerase, as previously observed for PS antisense oligonucleotides with serum proteins,<sup>58</sup> may inhibit polymerase passage through the template, thereby promoting multiple dNTP additions. This unexpected mutagenesis is highly important given that PS modifications are used as 3' primer modifications in stringent applications such as next-generation sequencing library



preparation and gene synthesis.<sup>59,60</sup> Moreover, it raises the fascinating possibility that the PS modifications identified in bacterial DNA<sup>61</sup> could be a means of genetically diversifying and evolving the population. Further studies are needed to investigate the sequence dependence of insertion mutations.

## CONCLUSION

When designing a polymerase-compatible artificial DNA backbone, certain features are essential. It should be able to accept hydrogen bonds from the enzyme and conform to specific structural and steric demands; a five bond separation between surrounding nucleosides is optimum for faster polymerase read-through but not necessarily fidelity, whereas greater rotational freedom around the modified backbone improves the copying rate and inhibits point deletions. If the linkage disturbs natural sugar conformation or positioning, low level multibase deletions can occur, which can be minimized by further sugar modifications. Overall, triazoles are good DNA backbone analogues with either Tz2 or Tz3 offering optimal performance depending upon the polymerase (e.g., KOD XL with Tz2 TT and GoTaq with Tz2<sup>m</sup>CC).

It is apparent that replication fidelity is not strictly related to replication speed; hence selection of a modified backbone should be application-dependent. For *in vitro* systems lacking selection pressures, fidelity can be optimized at the expense of speed of polymerase read-through, with backbones that facilitate efficient controlled oligonucleotide ligation being particularly useful.<sup>28,29</sup> For *in vivo* applications, speed is likely to be as important as fidelity, since replication must keep pace with other native processes during the cell cycle.

In this respect, the amide analogue (Am1) is very promising; it offers fast read-through and good fidelity, and importantly we show here that it can be formed with high efficiency. However, further studies are required to establish the utility of the Am1 backbone in living systems. In this regard, the triazole linkages are at present more thoroughly validated.<sup>24,26,27</sup>

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications Web site. (PDF) (txt) (txt) The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b11530.

- All methods, supplementary tables, and figures (PDF)
- Supplementary file 1 of sequencing alignment profiles (TXT)
- Supplementary file 2 of unique enriched sequences (TXT)

## AUTHOR INFORMATION

### Corresponding Authors

\*tom.brown@chem.ox.ac.uk

\*afaf.el-sagheer@chem.ox.ac.uk

### ORCID

Tom Brown: 0000-0002-6538-3036

### Author Contributions

<sup>§</sup>A.S. and A.E.S.T. contributed equally.

### Notes

The authors declare the following competing financial interest(s): A.H.E.S. and T.B. are co-inventors on US patent 8,846,883 B2 Oligonucleotide Ligation.

## ACKNOWLEDGMENTS

This work was supported by UK BBSRC Grants BB/M025624/1 (Next-generation DNA synthesis) and BB/J001694/2 (Extending the boundaries of nucleic acid chemistry). We also thank ATDBio Ltd. for a studentship to A.E.S.T.

## REFERENCES

- (1) Kimoto, M.; Kawai, R.; Mitsui, T.; Yokoyama, S.; Hirao, I. *Nucleic Acids Res.* **2009**, *37* (2), e14.
- (2) Malyshev, D. A.; Dhami, K.; Quach, H. T.; Lavergne, T.; Ordoukhanian, P.; Torkamani, A.; Romesberg, F. E. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (30), 12005.
- (3) Yang, Z.; Chen, F.; Alvarado, J. B.; Benner, S. A. *J. Am. Chem. Soc.* **2011**, *133* (38), 15105.
- (4) Malyshev, D. A.; Dhami, K.; Lavergne, T.; Chen, T.; Dai, N.; Foster, J. M.; Correa, I. R.; Romesberg, F. E. *Nature* **2014**, *509* (7500), 385.
- (5) Burmeister, P. E.; Lewis, S. D.; Silva, R. F.; Preiss, J. R.; Horwitz, L. R.; Pendergrast, P. S.; McCauley, T. G.; Kurz, J. C.; Epstein, D. M.; Wilson, C.; Keefe, A. D. *Chem. Biol.* **2005**, *12* (1), 25.
- (6) Chen, T.; Hongdilokkul, N.; Liu, Z.; Adhikary, R.; Tsuen, S. S.; Romesberg, F. E. *Nat. Chem.* **2016**, *8*, 556.
- (7) Pinheiro, V. B.; Taylor, A. I.; Cozens, C.; Abramov, M.; Renders, M.; Zhang, S.; Chaput, J. C.; Wengel, J.; Peak-Chew, S.-Y.; McLaughlin, S. H.; Herdewijn, P.; Holliger, P. *Science* **2012**, *336* (6079), 341.
- (8) Malyshev, D. A.; Romesberg, F. E. *Angew. Chem., Int. Ed.* **2015**, *54* (41), 11930.
- (9) Taylor, A. I.; Pinheiro, V. B.; Smola, M. J.; Morgunov, A. S.; Peak-Chew, S.; Cozens, C.; Weeks, K. M.; Herdewijn, P.; Holliger, P. *Nature* **2015**, *518* (7539), 427.
- (10) Pasek, M. A.; Lauretta, D. S. *Astrobiology* **2005**, *5* (4), 515.
- (11) Sutherland, J. D. *Angew. Chem., Int. Ed.* **2016**, *55* (1), 104.
- (12) Miller, P. S.; Yano, J.; Yano, E.; Carroll, C.; Jayaraman, K.; Ts'o, P. O. P. *Biochemistry* **1979**, *18* (23), 5134.
- (13) Kalinowski, M.; Haug, R.; Said, H.; Piasecka, S.; Kramer, M.; Richert, C. *ChemBioChem* **2016**, *17* (12), 1150.
- (14) Eckstein, F. *Antisense Nucleic Acid Drug Dev.* **2000**, *10* (2), 117.
- (15) De Mesmaeker, A.; Waldner, A.; Lebreton, J.; Hoffmann, P.; Fritsch, V.; Wolf, R. M.; Freier, S. M. *Angew. Chem., Int. Ed. Engl.* **1994**, *33* (2), 226.
- (16) Patzke, V.; McCaskill, J. S.; von Kiedrowski, G. *Angew. Chem., Int. Ed.* **2014**, *53* (16), 4222.
- (17) Blaskó, A.; Dempcy, R. O.; Minyat, E. E.; Bruice, T. C. *J. Am. Chem. Soc.* **1996**, *118* (34), 7892.
- (18) Waldner, A.; De Mesmaeker, A.; Lebreton, J.; Fritsch, V.; Wolf, R. M. *Synlett* **1994**, *1994* (1), 57.
- (19) Clafiré, S. A.; Rinaldi, M.; Gasparini, P.; Seripa, D.; Bisceglia, L.; Zelante, L.; Farace, M. G.; Fazio, V. M. *Nucleic Acids Res.* **1995**, *23* (20), 4134.
- (20) Shaw, B. R.; Dobrikov, M.; Wang, X. I. N.; Wan, J.; He, K.; Lin, J.-L.; Li, P.; Rait, V.; Sergueeva, Z. A.; Sergueev, D. *Ann. N. Y. Acad. Sci.* **2003**, *1002* (1), 12.
- (21) Adelfinskaya, O.; Terrazas, M.; Froeyen, M.; Marlière, P.; Nauwelaerts, K.; Herdewijn, P. *Nucleic Acids Res.* **2007**, *35* (15), 5060.
- (22) Kuwahara, M.; Takeshima, H.; Nagashima, J.; Minezaki, S.; Ozaki, H.; Sawai, H. *Bioorg. Med. Chem.* **2009**, *17* (11), 3782.
- (23) El-Sagheer, A. H.; Brown, T. *J. Am. Chem. Soc.* **2009**, *131* (11), 3958.
- (24) El-Sagheer, A. H.; Sanzone, A. P.; Gao, R.; Tavassoli, A.; Brown, T. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (28), 11338.
- (25) El-Sagheer, A. H.; Brown, T. *Chem. Commun.* **2011**, *47* (44), 12057.
- (26) Birts, C. N.; Sanzone, A. P.; El-Sagheer, A. H.; Blaydes, J. P.; Brown, T.; Tavassoli, A. *Angew. Chem., Int. Ed.* **2014**, *53* (9), 2362.
- (27) Sanzone, A. P.; El-Sagheer, A. H.; Brown, T.; Tavassoli, A. *Nucleic Acids Res.* **2012**, *40* (20), 10567.



- (28) Meng, W.; Muscat, R. A.; McKee, M. L.; Milnes, P. J.; El-Sagheer, A. H.; Bath, J.; Davis, B. G.; Brown, T.; O'Reilly, R. K.; Turberfield, A. J. *Nat. Chem.* **2016**, *8* (6), 542.
- (29) Litovchick, A.; Dumelin, C. E.; Habeshian, S.; Gikunju, D.; Guié, M.-A.; Centrella, P.; Zhang, Y.; Sigel, E. A.; Cuozzo, J. W.; Keefe, A. D.; Clark, M. A. *Sci. Rep.* **2015**, *5*, 10916.
- (30) Keefe, A. D.; Clark, M. A.; Hupp, C. D.; Litovchick, A.; Zhang, Y. *Curr. Opin. Chem. Biol.* **2015**, *26*, 80.
- (31) Walczak, S.; Nowicka, A.; Kubacka, D.; Fac, K.; Wanat, P.; Mroczek, S.; Kowalska, J.; Jemielity, J. *Chem. Sci.* **2017**, *8*, 260.
- (32) Hansen, D. J.; Manuguerra, I.; Kjelstrup, M. B.; Gothelf, K. V. *Angew. Chem., Int. Ed.* **2014**, *53* (52), 14415.
- (33) El-Sagheer, A. H.; Brown, T. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (35), 15329.
- (34) Isobe, H.; Fujino, T.; Yamazaki, N.; Guillot-Nieckowski, M.; Nakamura, E. *Org. Lett.* **2008**, *10* (17), 3729.
- (35) Gogoi, K.; Gunjal, A. D.; Phalgune, U. D.; Kumar, V. A. *Org. Lett.* **2007**, *9* (14), 2697.
- (36) Luo, P.; Leitzel, J. C.; Zhan, Z.-Y. J.; Lynn, D. G. *J. Am. Chem. Soc.* **1998**, *120* (13), 3019.
- (37) Selvam, C.; Thomas, S.; Abbott, J.; Kennedy, S. D.; Rozners, E. *Angew. Chem., Int. Ed.* **2011**, *50* (9), 2068.
- (38) Staudinger, H.; Meyer, J. *Helv. Chim. Acta* **1919**, *2* (1), 635.
- (39) Pourceau, G.; Meyer, A.; Vasseur, J.-J.; Morvan, F. *J. Org. Chem.* **2009**, *74* (17), 6837.
- (40) Steger, J.; Graber, D.; Moroder, H.; Geiermann, A.-S.; Aigner, M.; Micura, R. *Angew. Chem., Int. Ed.* **2010**, *49* (41), 7470.
- (41) El-Sagheer, A. H.; Brown, T. *Chem. Sci.* **2014**, *5* (1), 253.
- (42) Lin, K.-Y.; Matteucci, M. D. *J. Am. Chem. Soc.* **1998**, *120* (33), 8531.
- (43) Dallmann, A.; El-Sagheer, A. H.; Dehmel, L.; Mügge, C.; Griesinger, C.; Ernsting, N. P.; Brown, T. *Chem. - Eur. J.* **2011**, *17* (52), 14714.
- (44) Cui, H.; Carrero-Lérida, J.; Silva, A. P. G.; Whittingham, J. L.; Brannigan, J. A.; Ruiz-Pérez, L. M.; Read, K. D.; Wilson, K. S.; González-Pacanoska, D.; Gilbert, I. H. *J. Med. Chem.* **2012**, *55* (24), 10948.
- (45) Kainz, P. *Biochim. Biophys. Acta, Gene Struct. Expression* **2000**, *1494* (1–2), 23.
- (46) McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytzky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; DePristo, M. A. *Genome Res.* **2010**, *20* (9), 1297.
- (47) Schwartz, J. J.; Lee, C.; Shendure, J. *Nat. Methods* **2012**, *9* (9), 913.
- (48) Kunkel, T. A. *J. Biol. Chem.* **2004**, *279* (17), 16895.
- (49) Kiefer, J. R.; Mao, C.; Braman, J. C.; Beese, L. S. *Nature* **1998**, *391* (6664), 304.
- (50) Morales, J. C.; Kool, E. T. *Biochemistry* **2000**, *39* (10), 2626.
- (51) Doublíé, S.; Sawaya, M. R.; Ellenberger, T. *Structure* **1999**, *7* (2), R31.
- (52) Eom, S. H.; Wang, J.; Steitz, T. A. *Nature* **1996**, *382* (6588), 278.
- (53) Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D. *Nucleic Acids Res.* **2006**, *34* (2), 564.
- (54) Minnick, D. T.; Astatke, M.; Joyce, C. M.; Kunkel, T. A. *J. Biol. Chem.* **1996**, *271* (40), 24954.
- (55) Carroll, S. S.; Cowart, M.; Benkovic, S. J. *Biochemistry* **1991**, *30* (3), 804.
- (56) Moran, S.; Ren, R. X.-F.; Kool, E. T. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (20), 10506.
- (57) Moran, S.; Ren, R. X.-F.; Rumney, S.; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119* (8), 2056.
- (58) Geary, R. S.; Norris, D.; Yu, R.; Bennett, C. F. *Adv. Drug Delivery Rev.* **2015**, *87*, 46.
- (59) Blanus, M.; Schenk, A.; Sadeghi, H.; Marienhagen, J.; Schwaneberg, U. *Anal. Biochem.* **2010**, *406* (2), 141.
- (60) Quail, M. A.; Kozarewa, I.; Smith, F.; Scally, A.; Stephens, P. J.; Durbin, R.; Swerdlow, H.; Turner, D. J. *Nat. Methods* **2008**, *5* (12), 1005.
- (61) Wang, L.; Chen, S.; Xu, T.; Taghizadeh, K.; Wishnok, J. S.; Zhou, X.; You, D.; Deng, Z.; Dedon, P. C. *Nat. Chem. Biol.* **2007**, *3* (11), 709.