

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/51701262>

# Encoding Phenotype in Bacteria with an Alternative Genetic Set

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · NOVEMBER 2011

Impact Factor: 12.11 · DOI: 10.1021/ja208025e · Source: PubMed

CITATIONS

24

READS

20

## 5 AUTHORS, INCLUDING:



[Andrew T Krueger](#)

Novartis Institutes for BioMedical Research

11 PUBLICATIONS 320 CITATIONS

[SEE PROFILE](#)



[Larry Peterson](#)

Rhodes College

20 PUBLICATIONS 174 CITATIONS

[SEE PROFILE](#)



[Daniel Jerome Kleinbaum](#)

Emerald Therapeutics

3 PUBLICATIONS 47 CITATIONS

[SEE PROFILE](#)



[Eric T Kool](#)

Stanford University

280 PUBLICATIONS 13,135 CITATIONS

[SEE PROFILE](#)

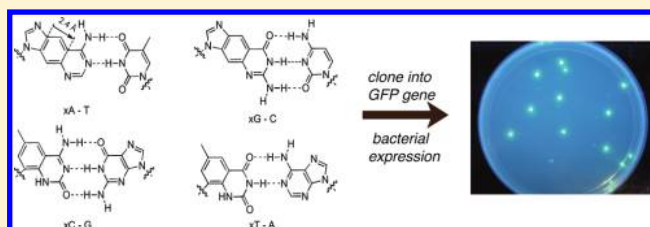
## Encoding Phenotype in Bacteria with an Alternative Genetic Set

Andrew T. Krueger,<sup>†</sup> Larryn W. Peterson,<sup>†</sup> Jijumon Chelliserry, Daniel J. Kleinbaum, and Eric T. Kool\*

Department of Chemistry, Stanford University, Stanford, California 94305-5080, United States

Supporting Information

**ABSTRACT:** An unnatural base-pair architecture with base pairs 2.4 Å larger than the natural DNA-based genetic system (xDNA) is evaluated for its ability to function like DNA, encoding amino acids in the context of living cells. xDNA bases are structurally analogous to natural bases but homologated by the width of a benzene ring, increasing their sizes and resulting in a duplex that is wider than native B-DNA. Plasmids encoding green fluorescent protein were constructed to contain single and multiple xDNA bases (as many as eight) in both strands and were transformed into *Escherichia coli*. Although they yielded fewer colonies than the natural control plasmid, in all cases in which a modified plasmid (containing one, two, three, or four consecutive size-expanded base pairs) was used, the correct codon bases were substituted, yielding green colonies. All four xDNA bases (xA, xC, xG, and xT) were found to encode the correct partners in the replicated plasmid DNA, both alone and in longer segments of xDNA. Controls with mutant cell lines having repair functions deleted were found to express the gene correctly, ruling out repair of xDNA and confirming polymerase reading of the unnatural bases. Preliminary experiments with polymerase deletion mutants suggested combined roles of replicative and lesion-bypass polymerases in inserting correct bases opposite xDNA bases and in bypassing the xDNA segments. These experiments demonstrate a biologically functioning synthetic genetic set with larger-than-natural architecture.



## INTRODUCTION

Studies aimed at mimicking the functional properties of DNA with alternative genetic structures have lent basic insights into why the naturally evolved DNA has its specific structure, and they have offered alternative structures that may retain biophysical and biochemical function. In addition, modified nucleic acids have been useful as basic tools for probing biochemical systems and technological tools for diagnosing disease. Although many studies exist of altered sugar–phosphate backbones for DNA, far fewer studies have been carried out on biochemical and biological function of alternative structures for the DNA bases,<sup>1–3</sup> which are the carriers and encoders of genetic information.

In an ongoing effort to mimic the functional properties of DNA with a new genetic architecture, we have previously explored the synthesis, self-assembly, structure, and enzyme substrate properties of size-expanded DNA (xDNA),<sup>4–8</sup> in which base pairs are made larger than the natural ones by benzohomologation. Results have shown that xDNA forms stable, right-handed, sequence-selective helices,<sup>4–8</sup> and that individual xDNA bases in a template can encode successful formation of base pairs by DNA polymerase enzymes.<sup>9</sup> However, correct function *in vitro* is far from active functioning in a living cell, where many enzymes must operate simultaneously. To date, no unnatural genetic architectures have been used with general success to encode amino acids of a protein in a living system.

A number of laboratories have worked on artificial bases and base pairs to augment the natural DNA genetic system.<sup>10–13</sup> In a different approach, we have investigated whether all base pairs in the DNA framework could be replaced with pairs of larger geometry, comprising an alternative genetic scaffold (Figure 1).

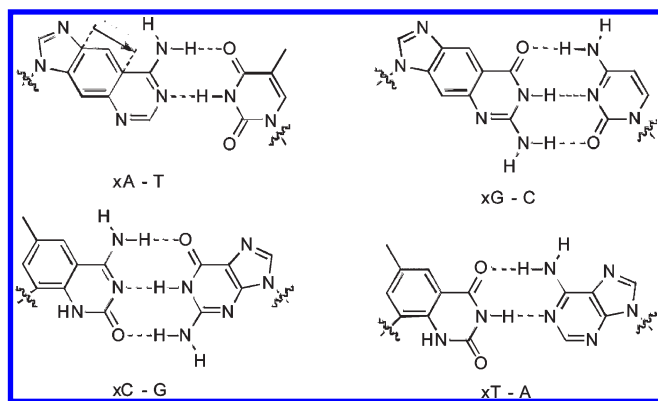
Other research groups have recently adopted an enlarged-base-pair strategy as well.<sup>14–16</sup> In our approach, benzopyrimidines pair with purines and benzopurines<sup>17</sup> with pyrimidines, yielding up to eight letters of information encoding capability.<sup>5,7</sup> The study of this nonnatural genetic set gives basic insight into the chemical, structural, and biophysical properties that govern encoding and transmission of genetic information. Moreover, the unusual properties of xDNA, including high helix stability and inherent fluorescence,<sup>18</sup> suggest applications in the biomedical sciences.

To date, studies on xDNA have been performed primarily *in vitro*, characterizing its pairing, structure, and fluorescence properties. Ultimately, however, to have a fully functional genetic system, it is necessary to have machinery capable of efficiently recognizing and propagating the unnatural base pair architecture during replication. Given that the natural replication enzymes have evolved to function with normally sized DNA, this is a challenging task. Nevertheless, we have begun to explore the ability of existing DNA machinery to tolerate xDNA's size-expanded geometry. Since cells possess several classes of DNA polymerases to carry out replication bypassing a variety of lesions,<sup>19</sup> it is possible that some enzymes may tolerate sterically enlarged xDNA bases with reasonable efficiency.

Recent *in vitro* studies with DNA polymerases suggested that single xDNA bases could be recognized moderately well to make one base pair. In addition, one xDNA base in single-stranded M13 phage was found to be bypassed by the *Escherichia coli* replication machinery with good success for two of the four

Received: August 24, 2011

Published: October 08, 2011



**Figure 1.** Size-expanded architecture of xDNA, in which xA, xC, xT, and xG bases are paired opposite natural bases. To encode phenotype in a cell, the 2.4 Å expanded bases must be read by cellular polymerases accurately and with adequate efficiency.

bases.<sup>1</sup> Encouraged by these results, we decided to evaluate this unnatural genetic set more thoroughly in the context of a biological system by asking whether xDNA could encode information as expressed by a phenotype, and whether multiple xDNA bases could be used in one vector. Here we report the use of multiple xDNA bases to encode biological information as amino acids in a protein, green fluorescent protein (GFP), that confers a visible phenotype.

## EXPERIMENTAL SECTION

**Synthesis and Purification of Oligonucleotides.** Syntheses of dxA, dxG, dxT, and dxC phosphoramidites were carried out as previously reported.<sup>4,7,17</sup> Oligodeoxynucleotides were synthesized on an Applied Biosystems 394 DNA/RNA synthesizer on a 1 μmol scale and possessed a 3'-OH and a 5'-phosphate group (Chemical Phosphorylation II reagent from Glen Research). Coupling employed standard β-cyanoethyl phosphoramidite chemistry, but with extended coupling time (600 s) for nonnatural nucleotides. All oligomers were deprotected in concentrated ammonium hydroxide (55 °C, 16 h), purified by preparative 20% denaturing polyacrylamide gel electrophoresis, and isolated by excision and extraction from the gel, followed by dialysis against water or desalting via PolyPak cartridges. The recovered material was quantified by absorbance at 260 nm, with molar extinction coefficients determined by the nearest-neighbor method. Molar extinction coefficients for unnatural oligomers were estimated by adding the measured value of the molar extinction coefficient of the unnatural nucleoside (at 260 nm) to the calculated value for the natural DNA fragments. Molar extinction coefficients for xDNA nucleosides used were as follows: dxA,  $\epsilon_{260} = 19\,800\text{ M}^{-1}\text{ cm}^{-1}$ ; dxG,  $\epsilon_{260} = 8100\text{ M}^{-1}\text{ cm}^{-1}$ ; dxT,  $\epsilon_{260} = 1200\text{ M}^{-1}\text{ cm}^{-1}$ ; dxC,  $\epsilon_{260} = 5800\text{ M}^{-1}\text{ cm}^{-1}$ . Nonnatural oligomers were characterized by MALDI-TOF mass spectrometry. Their observed and expected masses are listed in Table S1 (Supporting Information, SI).

**Plasmid Digestion.** The pGFPuv vector (Clontech, catalog no. 632312, 3 μL, ~2 μg, see SI for details) was then digested by addition of 3 μL each of the restriction enzymes *Bsr*GI and *Mlu*I (New England Biolabs), 5 μL of NEB buffer 2, 5 μL of 10× BSA, and 31 μL of H<sub>2</sub>O (autoclaved) to an Eppendorf tube. The digestion was allowed to proceed in a thermocycler at 37 °C for 6 h, followed by 65 °C for 20 min to denature the enzymes. After digestion, removal of the 5'-phosphate (to prevent self-ligation of the cut plasmid) was accomplished by addition (to the same Eppendorf tube) of 5 μL of antarctic phosphatase (or calf intestinal phosphatase (NEB) was used with the

corresponding buffer) and 5 μL of 10× antarctic phosphatase buffer and then incubation for 1 h at 37 °C, followed by denaturing at 65 °C for 20 min. The plasmid was purified on a 0.9% agarose (LMP) gel run at 20 V for 12 h. The gel bands were visualized under a UV (365 nm) transilluminator and then cut out (quickly, ~3 s), and the purified DNA was extracted with a plasmid gel extraction kit (Qiagen no. 28704). Concentration was determined by UV absorbance units at 260 nm.

**Ligation of 46 bp xDNA-Containing Segments into Digested pGFPuv Vector.** Ligations of xDNA-containing inserts were carried out at various vector:insert ratios (1:3, 1:5, 1:6, 1:8, and 1:10). Solutions of xDNA inserts were prepared/annealed by adding 2.5 μL of each strand (100 ng/μL), 0.5 μL of 1 M MgCl<sub>2</sub>, and 94.5 μL of H<sub>2</sub>O (autoclaved). This DNA solution was denatured in a thermocycler by heating to 72 °C for 3 min and then cooling slowly to room temperature. Appropriate volumes of the pGFPuv vector (~50 ng/μL) and annealed inserts (~2.5 ng/μL) were combined in an Eppendorf tube with 1 μL of 5× T4 DNA ligase buffer, 1 μL of T4 DNA ligase (New England Biolabs), and H<sub>2</sub>O (autoclaved) to reach a volume of 10 μL. The ligation reaction was carried out at 16 °C for 12 h in a thermocycler.

**Transformation of Modified Vectors into *E. coli*.** Transformation of the ligated inserts into BL21 (DE3) competent cells (Stratagene) was carried out per the transformation protocol described by Stratagene. For each transformation, ~1–10 ng of xDNA-containing insert was transformed into the cells, and after transformation, ~100–150 μL of the cells was plated on LB agar plates containing ampicillin and isopropyl thiogalactoside (IPTG) and allowed to incubate at 37 °C for 20 h. Colonies were briefly visualized under long-wavelength UV (365 nm), and the numbers of white and green colonies were documented. Transformation efficiencies were compared with that of a control ligation reaction under identical conditions using unmodified duplex DNA, which was otherwise identical in sequence to the xDNA-containing segment.

**Colony Picking and Sequencing of Plasmid DNA.** Green colonies (as well as white colonies) from each transformation were picked and grown overnight at 37 °C in LB media, and the plasmid DNA was isolated using a mini prep kit (Qiagen no. 27104). The cloned region within the gene on the isolated DNA was sequenced (Sequetech, Mountain View, CA, or Quintara Biosciences, Berkeley, CA) with the appropriate primers (GFPseq1.24).

**Bacterial Strains.** The bacterial strains used in this study are shown in Table 2; they were purchased from the *E. coli* Genetic Stock Center at Yale University. Bacteria were routinely grown in LB media containing 50 μg/mL kanamycin.

## RESULTS AND DISCUSSION

A high-copy plasmid vector (pGFPuv) containing a gene encoding GFP was used as a biological template in the experiments (see SI for details). A 46 base pair segment within the GFP gene (encoding amino acids 93–107) was excised, and a synthetic insert of identical length containing single or multiple xDNA bases was ligated into the vector in place of the excised segment (Table 1). The synthetic insert contained silent mutations to distinguish it from unmodified vector, and contained an equal number of xDNA bases in both strands so that the replicative enzymes would need to address them regardless of the direction of DNA synthesis. Ultimately, if the cellular machinery were capable of faithfully reading the xDNA bases, the resulting bacterial colonies would fluoresce green under UV illumination.

The xDNA-containing vectors were transformed into *E. coli*. After incubation, green and white colonies were picked, and the DNA was sequenced to determine the fate of encoding by the xDNA bases. It should be noted that xDNAs were only present in



**Table 1. Modified GFP Inserts Containing Various Size-Expanded Base Pairs with a Summary of Sequencing Results for Green Colonies<sup>a</sup>**

Synthetic insert	Sequences thr ile ser phe lys asp asp gly	Seq. Results
xGFP.0 (control)	5'--ACT ATT TCT TTC AAA GAT GAC GGA-- 3'--TGC TAA AGA AAG TTT CTA CTG CCT--	No change
xGFP.1 (one xG-C bp)	5'--ACG ATT TCT TTC AAA GAT GAC GGA-- 3'--TGC TAA AGA AAG TTT CTA CTG CCT--	xG replaced by G
xGFP.2 (one xA-A bp)	5'--ACT ATT TCT TTC AAA GAT GAC GGA-- 3'--TGA TAA AGA AAG TTT CTA CTG CCT--	xA replaced by A
xGFP.3 (one xC-G bp)	5'--ACT ATT TCT TTC AAA GAT GAC GGA-- 3'--TGA TAA AGA AAG TTT CTA CTG CCT--	xC replaced by C
xGFP.4 (one xT-A bp)	5'--ACT ATT TCT TTC AAA GAT GAC GGA-- 3'--TGA TAA AGA AAG TTT CTA CTG CCT--	xT replaced by T
xGFP.5 (two consec. xDNA bps)	5'--ACT ATT TCT TTC AAA GAT GAC GGA-- 3'--TGA TAA AGA AAG TTT CTA CTG CCT--	xCA replaced by CA
xGFP.6 (three consec. xDNA bps)	5'--ACT ATT TCT TTC AAA GAT GAC GGA-- 3'--TGA TAA AGA AAG TTT CTA CTG CCT--	xCAxAXA replaced by CAA
xGFP.7 (four consec. xDNA bps)	5'--ACT ATT TCT TTC AAA GAT GAC GGA-- 3'--TGA TAA AGA AAG TTT CTA CTG CCT--	xCAxAXAXA replaced by CAAA

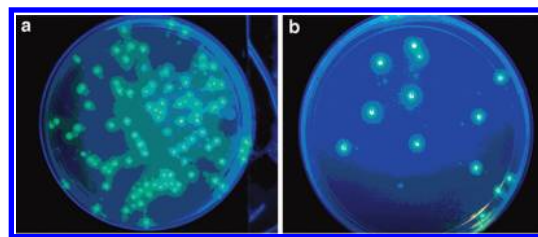
<sup>a</sup> Size-expanded bases in red; silent mutations in blue. Total insert size is 46 nt (Table S1).

the vector template and were not present as deoxynucleoside triphosphates; thus, one round of plasmid replication produces natural DNA in the bacterial colonies.

Significantly, results showed that while efficiency of colony formation was generally reduced relative to the unmodified control (by ca. 10-fold), green colonies were observed in all cases where a single size-expanded base pair was present (xGFP.1–4), demonstrating that the xDNA bases were bypassed by the bacteria's machinery and encoded a correctly functioning amino acid (Figure 2). All green colonies that were sequenced contained the silent mutations (except for a few cases of self-ligated vector missing insert), thus confirming that there was no contamination from natural DNA in these experiments (this was true for all experiments with xDNA-containing inserts). Furthermore, on sequencing the plasmid DNA of the green colonies, we found that in all the synthetic inserts the correct natural base was present opposite the xDNA base in the template (Table 1). Thus, the data suggested that each of the four xDNA bases was read correctly by the replication machinery.

To test the limits of xDNA encoding, we then synthesized inserts containing 2–4 consecutive size-expanded DNA bases (xGFP.5, xGFP.6, and xGFP.7). Note that the insert actually contains double this number of xDNA pairs, 4, 6, and 8 pairs, respectively, since both strands were modified. Extension of unnatural base pairs by polymerases has proved challenging *in vitro*, especially for consecutive substitutions; thus, these inserts represent stringent cases for the replicative machinery. However, although transformation efficiencies remained similar to those with single modifications, green colonies were observed in all cases. Yet more significant was the ability for the information to be replicated faithfully; for example, the vector xGFP.7 (containing xCAxAXAXA) was replaced by CAAA in all green colonies examined that contained the silent mutations (Table 1). This is striking, considering the presence of two 4-bp expanded DNA segments near one another.

A number of white colonies (not successfully expressing GFP, but expressing antibiotic resistance) were sequenced as well. Interestingly, in nearly all cases the xDNA base was replaced by the correct natural base, and the lack of GFP expression was a



**Figure 2.** Example of GFP expression from xDNA-containing gene segment, as shown by green colonies on LB plates after 20 h at 37 °C: (a) control (unmodified plasmid insert) and (b) xGFP.7 insert containing eight xDNA base pairs (four consecutive xDNA bases in each strand).

result of frameshifts (insertions/deletions) that occurred elsewhere in the gene (Table S6). All white colonies that contained the insert also contained the silent mutations. In only one case was the xDNA-encoded pair deleted (one of 10 colonies sequenced from xGFP.3).

Further exploration of the coding limits of xDNA was undertaken by synthesizing more demanding sequences containing six and eight consecutive xDNA pairs, and sequences containing multiple isolated (nonconsecutive) size-expanded base pairs (see Tables S4 and S5). However, not only did experiments with these inserts yield lower expression efficiencies, but none of the observed green colonies contained the silent mutation, even after multiple trials and varying vector:insert ratios. This could occur either from xDNA interference with enzymatic ligation in assembling the plasmid or from failed read-through of these longest segments by polymerases. Overall, the limit in these experiments appeared to be reached with the xGFP.7 vector.

The finding of protein expression from xDNA-containing gene segments suggests that cellular polymerases are able to successfully read the expanded bases as genetic information. However, we considered another possible explanation, namely, that DNA repair enzymes were removing xDNA and filling in the missing information by reading the natural bases opposite. This could occur by one of three different mechanisms: base excision repair (BER), methyl-directed mismatch repair, or nucleotide excision repair (NER).

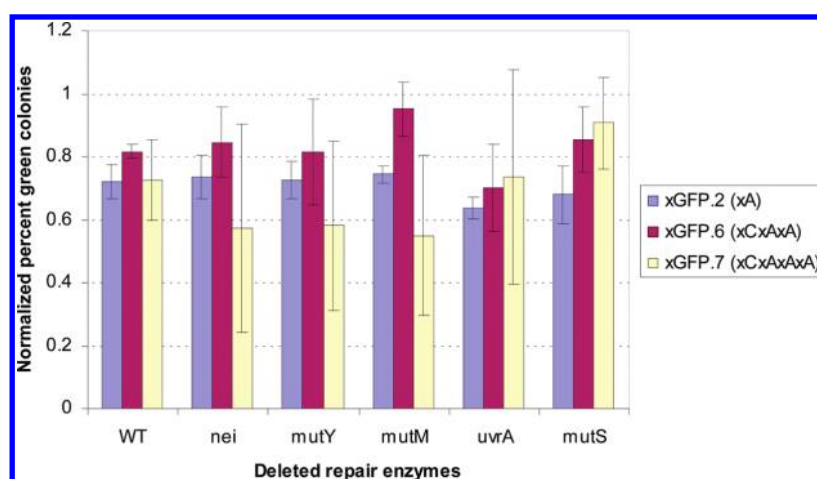
In BER, glycosylases remove damaged bases and then use a lyase mechanism to remove the abasic sugar prior to polymerase filling in the missing nucleotide. BER enzymes in *E. coli* include those encoded by *nei*, *mutY*, and *mutM* genes.<sup>20</sup> In the case of xDNA, the large bases might conceivably be recognized by such enzymes as “damage”; however, the expanded pyrimidines are attached by stable C-glycosidic bonds and so cannot be excised by this mechanism. Thus, full removal and repair of the xDNA is not possible for the inserts xGFP.3 through xGFP.7, all of which contain x-pyrimidines. However, it remained a possibility for inserts xGFP.1 and xGFP.2.

In the absence of data, it is also formally possible that the methyl-directed repair or NER pathways could repair xDNA segments. In these pathways, damage is recognized and excised in segments of DNA. A damaged strand is nicked upstream and downstream of the damage; ultimately the missing information is filled in by polymerases reading the undamaged strand opposite. While such repair would be difficult for the present inserts, where substitutions appear near one another in both strands (see below), it still remained a possibility.

To test the possibility of repair of xDNA segments, we obtained five bacterial strains having known BER, methyl-directed, and NER repair functions (*mutM*, *mutY*, *nei*, *uvrA*, *mutS*) deleted (see Table 2). If GFP expression from the xDNAs arose from a repair process,

Table 2. Bacterial Strain Genotypes Tested for Ability To Replicate xDNA Segments

strain	deleted gene	function/gene product	genotype
BW25113	—		F <sup>−</sup> , $\Delta(\text{araD-araB})\text{S67}$ , $\Delta\text{lacZ4787}::\text{rrnB-3}$ , $\lambda^-$ , <i>rph-1</i> , $\Delta(\text{rhaD-rhaB})\text{S68}$ , <i>hsdR514</i>
JW2703-2	<i>mutS</i>	methyl-mediated mismatch repair	F <sup>−</sup> , $\Delta(\text{araD-araB})\text{S67}$ , $\Delta\text{lacZ4787}::\text{rrnB-3}$ , $\lambda^-$ , $\Delta\text{mutS738}::\text{kan}$ , <i>rph-1</i> , $\Delta(\text{rhaD-rhaB})\text{S68}$ , <i>hsdR514</i>
JW0704-1	<i>nei</i>	endonuclease VII	F <sup>−</sup> , $\Delta(\text{araD-araB})\text{S67}$ , $\Delta\text{lacZ4787}::\text{rrnB-3}$ , $\Delta\text{nei-764}::\text{kan}$ , $\lambda^-$ , <i>rph-1</i> , $\Delta(\text{rhaD-rhaB})\text{S68}$ , <i>hsdR514</i>
JW2928-1	<i>mutY</i>	adenine glycosylase: G-A repair	F <sup>−</sup> , $\Delta(\text{araD-araB})\text{S67}$ , $\Delta\text{lacZ4787}::\text{rrnB-3}$ , $\lambda^-$ , $\Delta\text{mutY736}::\text{kan}$ , <i>rph-1</i> , $\Delta(\text{rhaD-rhaB})\text{S68}$ , <i>hsdR514</i>
JW3610-2	<i>mutM</i> ( <i>fpg</i> )	formamidopyrimidine-DNA glycosylase	F <sup>−</sup> , $\Delta(\text{araD-araB})\text{S67}$ , $\Delta\text{lacZ4787}::\text{rrnB-3}$ , $\lambda^-$ , $\Delta\text{mutM744}::\text{kan}$ , <i>rph-1</i> , $\Delta(\text{rhaD-rhaB})\text{S68}$ , <i>hsdR514</i>
JW4019-2	<i>uvrA</i>	excision nuclease: molecular matchmaker	F <sup>−</sup> , $\Delta(\text{araD-araB})\text{S67}$ , $\Delta\text{lacZ4787}::\text{rrnB-3}$ , $\lambda^-$ , <i>rph-1</i> , $\Delta(\text{rhaD-rhaB})\text{S68}$ , $\Delta\text{uvrA753}::\text{kan}$ , <i>hsdR514</i>
JW0221-1	<i>dinB</i>	Pol IV	F <sup>−</sup> , $\Delta(\text{araD-araB})\text{S67}$ , $\Delta\text{dinB749}::\text{kan}$ , $\Delta\text{lacZ4787}::\text{rrnB-3}$ , $\lambda^-$ , <i>rph-1</i> , $\Delta(\text{rhaD-rhaB})\text{S68}$ , <i>hsdR514</i>
JW1173-1	<i>umuC</i>	Pol V	F <sup>−</sup> , $\Delta(\text{araD-araB})\text{S67}$ , $\Delta\text{lacZ4787}::\text{rrnB-3}$ , $\lambda^-$ , $\Delta\text{umuC773}::\text{kan}$ , <i>rph-1</i> , $\Delta(\text{rhaD-rhaB})\text{S68}$ , <i>hsdR514</i>
JW0059-1	<i>polB</i> ( <i>dinA</i> )	Pol II	F <sup>−</sup> , $\Delta\text{polB770}::\text{kan}$ , $\Delta(\text{araD-araB})\text{S67}$ , $\Delta\text{lacZ4787}::\text{rrnB-3}$ , $\lambda^-$ , <i>rph-1</i> , $\Delta(\text{rhaD-rhaB})\text{S68}$ , <i>hsdR514</i>
JW0205-1	<i>dnaQ</i> ( <i>mutD</i> )	Pol III $\epsilon$ subunit: 3' $\rightarrow$ 5' proofreading	F <sup>−</sup> , $\Delta(\text{araD-araB})\text{S67}$ , $\Delta\text{dnaQ744}::\text{kan}$ , $\Delta\text{lacZ4787}::\text{rrnB-3}$ , $\lambda^-$ , <i>rph-1</i> , $\Delta(\text{rhaD-rhaB})\text{S68}$ , <i>hsdR514</i>



**Figure 3.** Effect of deleted repair functions on xDNA gene expression. Shown are percent green colonies for xDNA-containing inserts xGFP.2, xGFP.6, and xGFP.7 normalized to the unmodified control DNA (GFP.0) in five bacterial strains with known repair functions deleted.

knocking out this gene should result in loss of expression. We transformed each of these strains with three examples of xDNA segments (inserts xGFP.2, xGFP.6, and xGFP.7) and measured colony counts and sequenced as before, with multiple replicates of each case. Results showed no drop in colony counts, GFP expression, or correct encoding for any of these knockout strains (Figure 3 and SI). Considering the mechanisms of possible repair, it is perhaps not surprising that no evidence for repair of xDNA was seen in these experiments. Half of the xDNA nucleosides (the benzopyrimidines) contain unnatural C–C glycosidic bonds, rendering them inert to base excision. Second, if methyl-directed or NER repair were occurring for the xDNA segments, then one might expect equally efficient repair regardless of the length of the xDNA segments. However, our results show decreasing bypass efficiency as the segment is made longer, which is more consistent with polymerase activities rather than repair. Finally, in our segments, this repair would require the nicking to occur in the few nucleotides *between* the two closely spaced segments of xDNA, which would seem unlikely. The more probable nicking outside the xDNA segments would remove one xDNA segment but still leave the other one in the opposite strand and require that it be read through by polymerases. Overall, the results

establish that polymerase-mediated reading of xDNA, rather than repair, is responsible for the GFP expression phenotype.

The above data revealed that up to four consecutive xDNA pairs are read correctly by *E. coli* replication enzymes with surprisingly high fidelity. Previous experiments with bacterial polymerases have been carried out only *in vitro*. Data showed that DNA pol I (Klenow fragment) bypassed single xDNA pairs with low efficiency and fidelity, likely due to the enzyme's steric preference for purine–pyrimidine pairs and its relative rigidity.<sup>1</sup> Interestingly, experiments with Pol IV from *Sulfolobus solfataricus* showed considerably improved extension of xDNA pairs, demonstrating greater flexibility in this Y-family enzyme.<sup>9</sup> This suggests that Y-family enzymes such as Pols II, IV, or V might be involved in the current cellular context. To begin to address this, we performed preliminary experiments with an additional set of bacterial strains having Y-family polymerases Pol II (*polB*), Pol IV (*dinB*), and Pol V (*umuC*) deleted (Table 2). Results measuring GFP expression from xDNA-containing genes showed that in each polymerase deletion mutant, most modified DNAs had little or no drop in expression (Figure S1). Interestingly, single xDNA substitutions showed some apparent reliance on these polymerases,

while multiple substitutions did not, suggesting that the abrupt changes in size at xDNA/DNA junctions may benefit from such a repair enzyme for bypass. No one enzyme deletion yielded full loss of bypass activity, which suggests that either the different Y-family enzymes can substitute for one another in this bypass activity and/or that replicative polymerases I and III also contribute to the replication of xDNA.

To further investigate the involvement of the Y-family polymerases in replication of the xDNA bases, the expression of GFP was examined in cells irradiated with UV light to induce the SOS response and upregulate the expression of Pol II, Pol IV, and Pol V. Induction of the SOS polymerases did not significantly affect the green colony counts and GFP expression (Table S8, Figure S2), suggesting that, rather than repair polymerases alone, the replicative enzymes Pol I and Pol III play a significant role in the xDNA replication. Overall, the results are most consistent with a combination of enzymes processing the xDNA, for example, by replicative enzymes incorporating nucleotides opposite the large bases, followed by extension beyond them by a Y-family enzyme.<sup>9</sup>

An early study tested the ability of isolated xDNA bases in a single-stranded M13 vector to be bypassed in *E. coli* and found some similarities and differences with the current experiments.<sup>1</sup> In the earlier work, xA and xC were found to be replaced correctly by A and C after replication, consistent with the current results. However, in the previous work, xG and xT bases were incorrectly replaced (by A in both cases), indicating frequent mispairing with T, especially in the xG case. In contrast to this are our current findings that single xT and xG bases were correctly replicated in every clone sequenced. We hypothesize that the difference may be due either to the difference in sequence context or to the differences in genome constructs. The current work was carried out within the context of a double-stranded plasmid, while the previous study utilized a single-stranded phage; as the machinery and mechanisms involved in replicating double-stranded plasmid DNA versus a single-stranded genome are not identical, different polymerase enzymes could possibly be involved in the initial nucleotide insertion step.

Taken together, the results show that native enzymes in *E. coli* are capable of accurately reading genetic information encoded in segments of xDNA and expressing this information in a visible phenotype. This is the first example in which an intact biological system has been shown to be capable of reading all components of an unnatural genetic set. The only other examples of biological expression of modified DNA previously involved a modified backbone rather than bases,<sup>21,22</sup> or a simple substitution on a natural base.<sup>2</sup> The results are a precedent for the eventual use of unnatural genetic sets for expression of designed biological activities *in vivo*, a long-term goal of chemical synthetic biology.

## ■ ASSOCIATED CONTENT

**S Supporting Information.** Experimental details, sequencing data, and characterization of synthetic DNAs/xDNAs. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

**Corresponding Author**  
kool@stanford.edu

### Author Contributions

<sup>†</sup>These authors contributed equally to this work.

## ■ ACKNOWLEDGMENT

We thank the National Institutes of Health (GM63587) for support.

## ■ REFERENCES

- (1) Delaney, J. C.; Gao, J.; Liu, H.; Shrivastav, N.; Essigmann, J. M.; Kool, E. T. *Angew. Chem., Int. Ed.* **2009**, *48*, 4524–4527.
- (2) Marlière, P.; Patrouix, J.; Döring, V.; Herdewijn, P.; Tricot, S.; Cruveiller, S.; Bouzon, M.; Mutzel, R. *Angew. Chem., Int. Ed.* **2011**, *50*, 7109–7114.
- (3) Kim, T. W.; Delaney, J. C.; Essigmann, J. M.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15803–15808.
- (4) Liu, H.; Gao, J.; Lynch, S. R.; Saito, Y. D.; Maynard, L.; Kool, E. T. *Science* **2003**, *302*, 868–871.
- (5) Krueger, A.; Lu, H.; Lee, A.; Kool, E. *Acc. Chem. Res.* **2007**, *40*, 141–150.
- (6) Liu, H.; Gao, J.; Kool, E. T. *J. Am. Chem. Soc.* **2005**, *127*, 1396–1402.
- (7) Gao, J.; Liu, H.; Kool, E. T. *Angew. Chem., Int. Ed.* **2005**, *44*, 3118–3122.
- (8) Lynch, S. R.; Liu, H.; Gao, J.; Kool, E. T. *J. Am. Chem. Soc.* **2006**, *128*, 14704–14711.
- (9) Lu, H.; Krueger, A. T.; Gao, J.; Liu, H.; Kool, E. T. *Org. Biomol. Chem.* **2010**, *8*, 2704–2710.
- (10) Henry, A. A.; Romesberg, F. E. *Curr. Opin. Chem. Biol.* **2003**, *7*, 727–733.
- (11) Krueger, A. T.; Kool, E. T. *Chem. Biol.* **2009**, *16*, 242–248.
- (12) Hirao, I. *Curr. Opin. Chem. Biol.* **2006**, *10*, 622–627.
- (13) Krueger, A. T.; Kool, E. T. *Curr. Opin. Chem. Biol.* **2007**, *11*, 588–594.
- (14) Hikishima, S.; Minakawa, N.; Kuramoto, K.; Fujisawa, Y.; Ogawa, M.; Matsuda, A. *Angew. Chem., Int. Ed.* **2005**, *44*, 596–598.
- (15) Battersby, T.; Albalos, M.; Friesenhahn, M. *Chem. Biol.* **2007**, *14*, 525.
- (16) Doi, Y.; Chiba, J.; Morikawa, T.; Inouye, M. *J. Am. Chem. Soc.* **2008**, *130*, 8762–8768.
- (17) Leonard, N. J. *Acc. Chem. Res.* **1982**, *15*, 128–135.
- (18) Krueger, A. T.; Kool, E. T. *J. Am. Chem. Soc.* **2008**, *130*, 3989–3999.
- (19) Nohmi, T. *Annu. Rev. Microbiol.* **2006**, *60*, 231–253.
- (20) David, S. S.; S'Shea, V. L.; Kundu, S. *Nature* **2007**, *447*, 941–950.
- (21) Pochet, S.; Kaminski, P. A.; Van Aerschot, A.; Herdewijn, P.; Marlière, P. C. R. *Biol.* **2003**, *326*, 1175–1184.
- (22) El-Sagheer, A. H.; Sanzone, A. P.; Gao, R.; Tassavoli, A.; Brown, T. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 11338–11343.