

# Efficient Replication Bypass of Size-Expanded DNA Base Pairs in Bacterial Cells\*\*

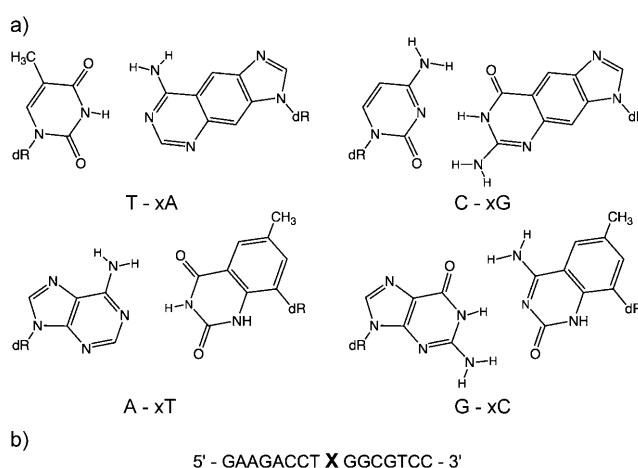
James C. Delaney, Jianmin Gao, Haibo Liu, Nidhi Shrivastav, John M. Essigmann,\* and Eric T. Kool\*

One of the long-standing aims of biomimetic chemistry has been to develop molecules that function as much as possible as the natural ones do.<sup>[1]</sup> Among biopolymers, nucleic acids have long served as a test bed for bioinspired design. The earliest focus of altered structures for DNA was the redesign of the phosphodiester backbone,<sup>[2]</sup> and numerous studies found backbone variants that assembled well into helices. Moreover, in recent years, some altered sugars have also been shown to be substrates for certain polymerase enzymes.<sup>[3]</sup> Such work suggests the possibility of future biological activities associated with altered DNA backbones.

More recently, a number of laboratories have focused on design of replacements for the DNA bases themselves, which encode the chemical information of the cell.<sup>[4]</sup> This is a challenging goal, because biological enzymes have evolved the ability to manipulate these bases and base pairs with extraordinarily high selectivity. The polymerase replication of designed DNA pairs has been studied in a number of laboratories. Significant successes have been reported using varied strategies, including base pairs with altered hydrogen-bonding patterns,<sup>[4a]</sup> pairs that lack hydrogen bonds altogether,<sup>[4b–e,i,j,m–p]</sup> and pairs that have larger-than-natural dimensions.<sup>[4q–s,5]</sup> However, to date such designed base pairs have generally not been tested in living cells. One recent exception is a single isosteric replacement for a natural base that was replicated efficiently.<sup>[6]</sup>

Herein we describe the first tests of a nonnatural DNA base-pair geometry in a living cell. We have inserted single size-expanded DNA (xDNA) bases into phage genomes and measured their replication in *Escherichia coli* cells. Surprisingly, although xDNA base pairs are considerably larger than their natural counterparts, we find that they are bypassed by the cellular replication machinery remarkably well. Indeed, two of the designed pairs possess biological function that is nearly indistinguishable from that of natural base pairs.

The four size-expanded DNA base pairs are shown in Figure 1. The hydrogen-bonded pairs involve benzo homologues opposite complementary natural partners, yielding base pairs 2.4 Å larger than Watson–Crick pairs.<sup>[6]</sup> The



**Figure 1.** a) Structures of the four xDNA bases evaluated in cellular replication, shown with their paired structures. b) DNA sequence context for xDNA bases (at position X) inserted into single-stranded M13 bacteriophage.

concept of a benzo-expanded base was first developed by Leonard and co-workers, who developed *lin*-benzoadenine and -guanine and studied them as ribonucleotide analogues three decades ago.<sup>[7]</sup> We developed analogous benzopyrimidine C-nucleosides and studied the ability of these four deoxynucleoside compounds (xA, xC, xG, xT) to form helices composed of expanded-size pairs. Work to date has shown that fully substituted duplexes of xDNA pairs are highly stable and form right-handed helices with a backbone conformation resembling B-DNA.<sup>[5a,d–f]</sup> However, single xDNA pairs substituted within natural DNA are destabilizing, presumably because of the mismatch in size between the large pair and the naturally sized backbone surrounding it.<sup>[5d]</sup> This mismatch could well present a challenge for a natural DNA polymerase, since such enzymes (especially replicative enzymes) can be highly sensitive to the sizes of base pairs.<sup>[6,8]</sup> However, living cells possess several different specialized classes of polymerases, some of which function to bypass damaged bases that are often larger than those of normal DNA.<sup>[9]</sup> Before completing extensive studies of in vitro replication of xDNA bases by natural and modified polymerases, we decided to test whether the cellular machinery already exists that might process these expanded bases.

[\*] Dr. J. C. Delaney, N. Shrivastav, Prof. Dr. J. M. Essigmann  
Departments of Chemistry and Biological Engineering  
Massachusetts Institute of Technology  
Cambridge, MA 02139 (USA)  
E-mail: jessig@mit.edu

Dr. J. Gao, Dr. H. Liu, Prof. Dr. E. T. Kool  
Department of Chemistry, Stanford University  
Stanford, CA 94305 (USA)  
E-mail: kool@stanford.edu

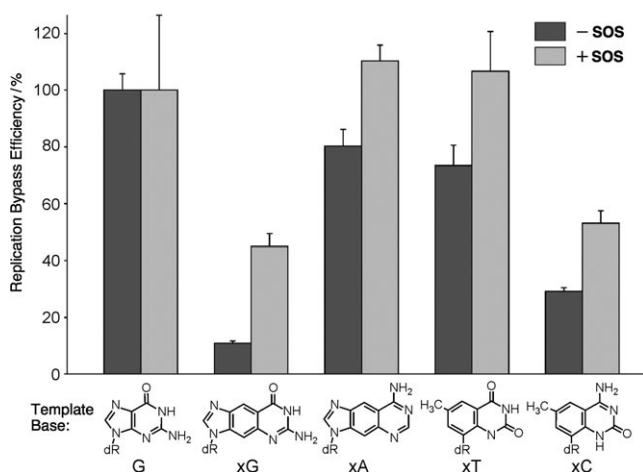
[\*\*] We thank the U.S. National Institutes of Health (CA80024 to J.M.E. and GM63587 to E.T.K.) for support. J.G. and H.L. acknowledge Stanford Graduate Fellowships.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.200805683>.

We measured the information-encoding capability of the single xDNA bases by incorporating them separately into oligonucleotides using standard phosphoramidite solid-phase synthesis.<sup>[5a,c]</sup> The intact incorporation of expanded bases into the DNA strands was confirmed by MALDI-TOF mass spectrometry in all cases (see the Supporting Information).

The oligonucleotides were then ligated into an M13 mp7(L2) single-stranded viral genome.<sup>[10]</sup> These modified genomes were then passaged through *E. coli* to quantify the biological responses. The responses to be derived are replication bypass efficiency (as scored by the amounts of daughter phage produced with respect to an unmodified competitor genome) and replication fidelity (measured by isolation and composition analysis of the daughter sequence, see below). A feature of this system is that there is no complementary strand opposite the unnatural bases in this single-stranded bacteriophage. This situation ensures that outcomes from replication and mutagenesis are derived solely from the initial replicative bypass of the modified bases, rather than from preferential replication of an opposing unmodified strand.

The ability of each of the four xDNA bases to support DNA replication in vivo was addressed using a competitive replication assay in which genomes bearing a nonstandard base are mixed with an unmodified internal standard prior to electroporation into cell.<sup>[11]</sup> The concentration of each xDNA base construct was determined in triplicate, with subsequent normalization and transfection using a 2:1 ratio of xDNA/competitor. The results showed that two of the four size-expanded bases were bypassed highly efficiently (Figure 2). In *E. coli* that were not induced for a damage (SOS) response to translesion DNA replication, the xA and xT bases were bypassed with efficiencies that were, respectively, 80 % and 73 % of the natural guanine control. The xC base was moderately well bypassed, with an efficiency of 29 %, while the signal dropped to 11 % for the xG base. When UV light

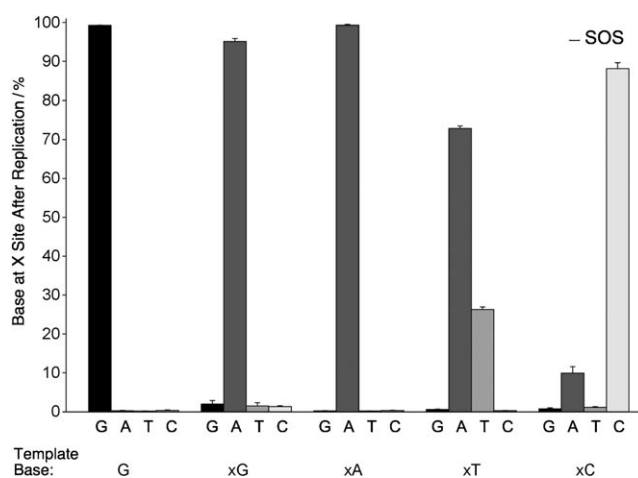


**Figure 2.** The efficiency of replication bypass of single xDNA bases in vivo. Bases were ligated into single-stranded M13 bacteriophage and replicated in *E. coli*. Bypass efficiency was measured by quantifying relative output signals from test and internal standard genomes and normalizing to those from G at the test-site control. Results are shown without (–SOS) and with (+SOS) induced damage response.

was used to induce the SOS response polymerases, bypass was improved by a significant amount in all cases, with xA and xT reaching the level of the unmodified guanine control. The xC base increased to 53 %, while the base that was the strongest block to replication, xG, had the greatest response to the UV-induced SOS DNA polymerases, increasing to 45 % relative to the phage genome containing a natural guanine base at the test site. This result suggests that flexible damage-response enzymes may assist in the synthesis or extension of a large xDNA pair.

Having established that single xDNA base pairs can be processed by the *E. coli* replication machinery, we proceeded to evaluate which natural bases replaced the xDNA analogues in the daughter phage that was recovered, using a published restriction endonuclease/postlabeling assay.<sup>[10]</sup> This assay serves as a quantitative measure of the ability of the bacterial polymerases to accurately read the chemical information encoded in the large-size bases. In the first round of DNA synthesis to produce the (–) strand of the phage DNA, a natural base would be incorporated opposite the xDNA base, thus making a size-expanded pair. In the next round, the all-natural DNA (–) strand would be replicated normally, producing new, all-naturally substituted (+)-strand daughter phage genomes that carry the sequence information encoded by the xDNA base at the original test site.

We found that two of the four xDNA bases, xA and xC, encoded their analogous replacements correctly (Figure 3). The fidelity of replacement of xA by A in the daughter phage



**Figure 3.** The replicative encoding capability of single xDNA bases in *E. coli*. Shown are the relative amounts of each natural base at the test site (X) that replaced the xDNA base shown after phage replication.

is particularly striking, with 99 % of the daughter phage containing adenine at the test site in the genome. This finding establishes that the xDNA-replicating polymerase correctly incorporates T opposite xA despite the large size of this pair. The xC analogue was also read correctly the majority of the time, with replacement by C in 88 % of the cases and by A (implicating T–xC mispairing) in 10 %. The other two xDNA bases were read incorrectly, with T incorporation opposite both xT and xG being dominant over the “correct” xDNA

pairing. Despite the misreading of xG, its coding ambiguity was low, since it coded as A 95 % of the time. We also carried out the same experiments with phage that had been passaged through *E. coli* in which the SOS response had been induced by UV light. The results (see the Supporting Information) showed that this procedure did not significantly change the bases encoded by the large-sized bases.

Although the xA and xC expanded bases are read correctly by the bacterial replication machinery, xT and xG are not, at least in this context. We speculate that the mispairings that are observed with these latter two bases arise from an alternative pairing geometry and from an alternative protonation state (Figure S5 in the Supporting Information). It is possible to pair T opposite xT with a geometry analogous to the T–G wobble, which may be closer to a Watson–Crick geometry. This might explain the observation of “correct” A–xT pairing only 26 % of the time as compared with 73 % T–xT mispairing. As for the xG base, in some contexts it can be deprotonated at neutral and basic pH values.<sup>[5f]</sup> If such deprotonation occurred during replication, xG would present a structure that is more complementary to T than C. We note that although two xDNA bases are incorrectly processed in this context, a substantial information-encoding capability remains. The correct coding of two xDNA pairs involves four different bases, which is, in principle, the same amount of information content as the natural genetic system.

The finding of efficient replication for two of the large-sized pairs is surprising, given that replicative polymerases can be highly sensitive to nucleobase size.<sup>[8]</sup> To examine this aspect in more detail, we carried out preliminary in vitro experiments with DNA polymerase I (Pol I; Klenow fragment (Kf)), the most extensively studied of the *E. coli* polymerases. We used 28 mer DNA templates containing a single xDNA base immediately downstream of a primer. To evaluate enzymatic efficiency and selectivity, we performed steady-state kinetics measurements on the addition of single natural nucleotides opposite each of these large bases. The results are given in Table 1.

The kinetics data confirm that at least one natural enzyme can correctly read sequence information stored in size-expanded bases. Not surprisingly, the Kf enzyme is inefficient in constructing these large base pairs, with  $V_{\max}/K_m$  values approximately 100–1000-fold below those for a natural base pair. Interestingly, this polymerase selectively chose the correct pairing partner with each of the four xDNA bases. Moreover, preliminary experiments on extension of an xDNA pair by this enzyme also show selective bypass of a correct pair over mismatched ones, again with very low efficiency (see the Supporting Information). It appears that enzyme(s) other than Pol I are responsible for the efficient replication of xG and xT in vivo, as these bases exhibited different cellular coding efficiencies (Figure 3). It will be of interest in the future to explore the other classes of polymerases, including types that are functionally more flexible,<sup>[12]</sup> to evaluate which are able to efficiently replicate such large pairs. It will also be important to study the replication in new sequence contexts and to evaluate exonuclease proofreading of such pairs.

Taken together, the results show 1) that a DNA polymerase is able read the chemical information stored in the

**Table 1:** Steady-state kinetics data for insertion of nucleoside triphosphates opposite single xDNA bases by the Klenow fragment of DNA Pol I (exonuclease-free (exo-)).<sup>[a]</sup>

Base	Nucleoside triphosphate	$V_{\max}$ (% min <sup>-1</sup> ) <sup>[b]</sup>	$K_m$ [ $\mu$ M]	Efficiency ( $V_{\max}/K_m$ )	Relative efficiency
xA	dATP	0.0041 (0.0012)	23 (10)	$1.9 \times 10^2$	$2.4 \times 10^{-2}$
xA	dCTP	0.0035 (0.0005)	9.9 (2.3)	$3.9 \times 10^2$	$4.9 \times 10^{-2}$
xA	dGTP	0.0046 (0.0019)	43 (20)	$1.1 \times 10^2$	$1.4 \times 10^{-2}$
xA	dTTP	0.15 (0.01)	18 (3)	$8.0 \times 10^3$	1
xC	dATP	0.40 (0.04)	4.6 (1.5)	$8.7 \times 10^4$	$6.7 \times 10^{-1}$
xC	dCTP	0.065 (0.004)	4.1 (1.5)	$2.0 \times 10^4$	$1.5 \times 10^{-1}$
xC	dGTP	3.5 (0.2)	28 (3)	$1.3 \times 10^5$	1
xC	dTTP	0.19 (0.01)	17 (5)	$1.2 \times 10^4$	$9.2 \times 10^{-2}$
xG	dATP	0.0034 (0.0004)	13 (10)	$3.8 \times 10^2$	$5.1 \times 10^{-3}$
xG	dCTP	0.26 (0.06)	8 (10)	$7.5 \times 10^4$	1
xG	dGTP	0.0046 (0.0011)	20 (9)	$2.7 \times 10^2$	$3.6 \times 10^{-3}$
xG	dTTP	0.0035 (0.0022)	15 (17)	$1.3 \times 10^2$	$1.7 \times 10^{-3}$
xT	dATP	9.0 (0.3)	50 (6)	$1.8 \times 10^5$	1
xT	dCTP	0.19 (0.03)	160 (30)	$1.2 \times 10^3$	$6.7 \times 10^{-3}$
xT	dGTP	0.064 (0.008)	880 (10)	$7.7 \times 10^1$	$4.3 \times 10^{-4}$
xT	dTTP	0.59 (0.03)	17 (2)	$3.6 \times 10^4$	$2.0 \times 10^{-1}$
T	dATP	36 (4)	3.4 (1.4)	$1.1 \times 10^7$	–

[a] Conditions: 200 nM 23 mer/28 mer primer–template duplex and varied polymerase concentrations in a buffer containing 50 mM Tris-HCl (pH 7.5), 10 mM MgCl<sub>2</sub>, 50  $\mu$ g mL<sup>-1</sup> BSA, and 1 mM dithiothreitol, incubated at 37 °C in a reaction volume of 10 mL. Standard deviations ( $n = 3–5$ ) are given in parentheses. [b] Normalized for the lowest enzyme concentration used.

size-expanded bases, and 2) that the full replication machinery of *E. coli* is able to recognize the sequence encoded by two of the xDNA bases correctly and efficiently. This intriguing outcome suggests that it may be possible in the future to incorporate multiple xA or xC bases into phage genomes or to incorporate xDNA pairs into plasmids that encode protein expression. In addition, it would be of interest to explore whether other organisms might also possess the ability to read xDNA pairs. The findings may ultimately lead to new strategies for modifying biological systems in useful ways.

## Experimental Section

**Synthesis:** The deoxynucleoside phosphoramidite derivatives of xA, xT, xC, and xG were prepared as described previously.<sup>[5a–c]</sup> They were incorporated into oligodeoxynucleotides using the published methods, purified by HPLC, and characterized by MALDI-TOF mass spectrometry (see the Supporting Information).

**Enzyme kinetics:** 28-nt oligonucleotides containing single xA, xC, xG, xT residues were prepared along with a 23-nt complementary primer, which was labeled at its 5' end with <sup>32</sup>P. Polymerase reaction conditions were as follows: DNA concentration 5  $\mu$ M in a 37 °C buffer containing 100 mM Tris-HCl (pH 7.5; Tris = tris(hydroxymethyl)aminomethane), 20 mM MgCl<sub>2</sub>, 2 mM dithiothreitol, and 0.1 mg mL<sup>-1</sup> acetylated bovine serum albumin (BSA). Enzyme and nucleotide concentrations were varied. See the Supporting Information for details.

Cellular assays: The competitive replication of adduct bypass (CRAB) assay<sup>[10b]</sup> was used to determine the replication blocking (if any) by the size-expanded bases in M13 phage. Figure S1 in the Supporting Information shows an outline of the assay. Quantification of the modified and control phage sequences was performed on the daughter phage population as described. The restriction endonuclease and postlabeling determination of mutation frequency (REAP) assay<sup>[10b]</sup> was used to quantify the type and amount of mutagenesis at the modified base site by obtaining the base composition at that position after cellular replication (Figure S1 in the Supporting Information). After PCR amplification, products were cleaved, labeled, and enzymatically digested, then analyzed by TLC and quantified by phosphorimager. Experiments were performed in triplicate.

Received: November 20, 2008

Published online: May 14, 2009

**Keywords:** DNA replication · nucleobases · polymerase · steric hindrance

- [1] a) R. Breslow, *Pure Appl. Chem.* **1998**, *70*, 267–270; b) E. T. Kool, M. L. Waters, *Nat. Chem. Biol.* **2007**, *3*, 70–73.
- [2] a) A. Murakami, K. R. Blake, P. S. Miller, *Biochemistry* **1985**, *24*, 4041–4046; b) P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, *Science* **1991**, *254*, 1497–1500; c) P. Nielsen, H. M. Pfundheller, J. Wengel, *Chem. Commun.* **1997**, 825–826; d) E. Lescrinier, R. Esnouf, J. Schraml, R. Busson, H. Heus, C. Hilbers, P. Herdewijn, *Chem. Biol.* **2000**, *7*, 719–731; e) M. Egli, P. S. Pallan, R. Pattanayek, C. J. Wilds, P. Lubini, G. Minasov, M. Dobler, C. J. Leumann, A. Eschenmoser, *J. Am. Chem. Soc.* **2006**, *128*, 10847–10856; f) K. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, A. Eschenmoser, *Science* **2000**, *290*, 1347–1351.
- [3] a) J. C. Chaput, J. W. Szostak, *J. Am. Chem. Soc.* **2003**, *125*, 9274–9275; b) B. R. Shaw, M. Dobrikov, X. Wang, J. Wan, K. He, J. L. Lin, P. Li, V. Rait, Z. Sergueeva, D. Sergueev, *Ann. N. Y. Acad. Sci.* **2003**, *1002*, 12–29; c) S. Pochet, P. A. Kaminski, A. Van Aerschot, P. Herdewijn, P. Marlière, *C. R. Biol.* **2003**, *326*, 1175–1184; d) S. Sinha, P. H. Kim, C. Switzer, *J. Am. Chem. Soc.* **2004**, *126*, 40–41; e) K. H. Jung, A. Marx, *Cell. Mol. Life Sci.* **2005**, *62*, 2080–2091; f) R. N. Veedu, B. Vester, J. Wengel, *Nucleosides Nucleotides Nucleic Acids* **2007**, *26*, 1207–1210.
- [4] a) J. A. Piccirilli, T. Krauch, S. E. Moroney, S. A. Benner, *Nature* **1990**, *343*, 33–37; b) S. Moran, R. X.-F. Ren, S. Rumney, E. T. Kool, *J. Am. Chem. Soc.* **1997**, *119*, 2056–2057; c) T. J. Matray, E. T. Kool, *Nature* **1999**, *399*, 704–708; d) E. L. Tae, Y. Wu, G. Xia, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **2001**, *123*, 7439–7440; e) J. Parsch, J. W. Engels, *Nucl. Acids Res.* **2001**, *20*, 815–818; f) H. Weizman, Y. Tor, *J. Am. Chem. Soc.* **2001**, *123*, 3375–3376; g) C. Beuck, I. Singh, A. Bhattacharya, W. Hecker, V. S. Parmar, O. Seitz, E. Weinhold, *Angew. Chem.* **2003**, *115*, 4088–4091; *Angew. Chem. Int. Ed.* **2003**, *42*, 3958–3960; h) N. Paul, V. C. Nashine, G. Hoops, P. Zhang, J. Zhou, D. E. Bergstrom, V. J. Davisson, *Chem. Biol.* **2003**, *10*, 815–825; i) A. A. Henry, F. E. Romesberg, *Curr. Opin. Chem. Biol.* **2003**, *7*, 727–733; j) I. Hirao, Y. Harada, M. Kimoto, T. Mitsui, T. Fujiwara, S. Yokoyama, *J. Am. Chem. Soc.* **2004**, *126*, 13298–13305; k) X. Zhang, I. Lee, X. Zhou, A. J. Berdis, *J. Am. Chem. Soc.* **2006**, *128*, 143–149; l) C. L. Moore, A. Zivkovic, J. W. Engels, R. D. Kuchta, *Biochemistry* **2004**, *43*, 12367–12374; m) I. Hirao, *Curr. Opin. Chem. Biol.* **2006**, *10*, 622–627; n) A. Zahn, C. J. Leumann, *Bioorg. Med. Chem.* **2006**, *14*, 6174–6188; o) G. T. Hwang, F. E. Romesberg, *J. Am. Chem. Soc.* **2008**, *130*, 14872–14882; p) I. Hirao, T. Mitsui, M. Kimoto, S. Yokoyama, *J. Am. Chem. Soc.* **2006**, *129*, 15549–15555; q) S. Hikishima, N. Minakawa, K. Kuramoto, Y. Fujisawa, M. Ogawa, A. Matsuda, *Angew. Chem.* **2005**, *117*, 602–604; *Angew. Chem. Int. Ed.* **2005**, *44*, 596–598; r) T. R. Battersby, M. Albalos, M. J. Friesenhahn, *Chem. Biol.* **2007**, *14*, 525–531; s) Y. Doi, J. Chiba, T. Morikawa, M. J. Inouye, *J. Am. Chem. Soc.* **2008**, *130*, 8762–8768.
- [5] a) H. Liu, J. Gao, S. R. Lynch, L. Maynard, D. Saito, E. T. Kool, *Science* **2003**, *302*, 868–871; b) H. Liu, J. Gao, L. Maynard, Y. D. Saito, E. T. Kool, *J. Am. Chem. Soc.* **2004**, *126*, 1102–1109; c) J. Gao, H. Liu, E. T. Kool, *Angew. Chem.* **2005**, *117*, 3178–3182; *Angew. Chem. Int. Ed.* **2005**, *44*, 3118–3122; d) J. Gao, H. Liu, E. T. Kool, *J. Am. Chem. Soc.* **2004**, *126*, 11826–11831; e) H. Liu, S. R. Lynch, E. T. Kool, *J. Am. Chem. Soc.* **2004**, *126*, 6900–6905; f) S. R. Lynch, H. Liu, J. Gao, E. T. Kool, *J. Am. Chem. Soc.* **2006**, *128*, 14704–14711.
- [6] T. W. Kim, J. C. Delaney, J. M. Essigmann, E. T. Kool, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15803–15808.
- [7] a) D. I. Scopes, J. R. Barrio, N. J. Leonard, *Science* **1977**, *195*, 296–298; b) N. J. Leonard, *Acc. Chem. Res.* **1982**, *15*, 128–135.
- [8] T. W. Kim, L. G. Briebe, T. Ellenberger, E. T. Kool, *J. Biol. Chem.* **2006**, *281*, 2289–2295.
- [9] a) T. Nohmi, *Annu. Rev. Microbiol.* **2006**, *60*, 231–253; b) D. F. Jarosz, V. G. Godoy, G. C. Walker, *Cell Cycle* **2007**, *6*, 817–822.
- [10] a) J. C. Delaney, J. M. Essigmann, *Chem. Biol.* **1999**, *6*, 743–753; b) J. C. Delaney, J. M. Essigmann, *Methods Enzymol.* **2006**, *408*, 1–15.
- [11] J. C. Delaney, J. M. Essigmann, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 14051–14056.
- [12] S. Mizukami, T. W. Kim, S. A. Helquist, E. T. Kool, *Biochemistry* **2006**, *45*, 2772–2778.