

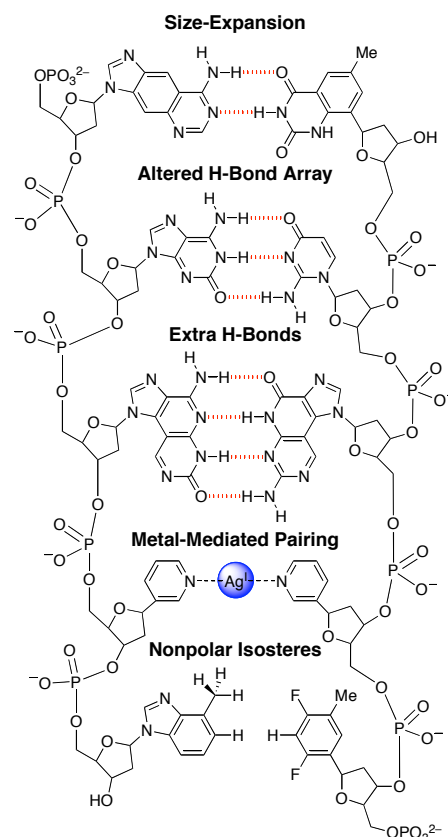
NON-NATURAL NUCLEOBASES: DEVELOPMENT AND RECENT ADVANCES

Reported by Chun-Ho Wong

December 1, 2008

INTRODUCTION

The discovery of the double helical structure of DNA by Watson and Crick in 1953¹ started a new era of molecular biology resulting in the development of a tremendous number of modified nucleobases. The main motivation at the early stage was the potential therapeutic applications of the modified bases. For example, azidothymidine (AZT) became the first FDA-approved drug to be used in the treatment of AIDS.² This chemically modified thymidine analog inhibits reverse transcriptase and terminates viral DNA synthesis by mimicking the natural thymidine nucleoside. In the 1980s, chemists began to study the structure and function of nucleic acids by synthesizing non-natural nucleobases. With the advent of methodology for oligonucleotide synthesis,³ the preparation of non-natural oligonucleotides using non-natural nucleobases also became possible in the 1990s.⁴ Today, non-natural bases and oligonucleotides are used not only for probing biological systems and functions but also as functional materials in the realm of nanotechnology.⁵ Herein, four major strategies for designing novel nucleobases are discussed: (1) size-expanded bases, (2) alternative hydrogen bonding arrays, (3) metal-mediated base pairing, and (4) nonpolar nucleoside isosteres. Applications based on non-natural nucleobases are also presented.



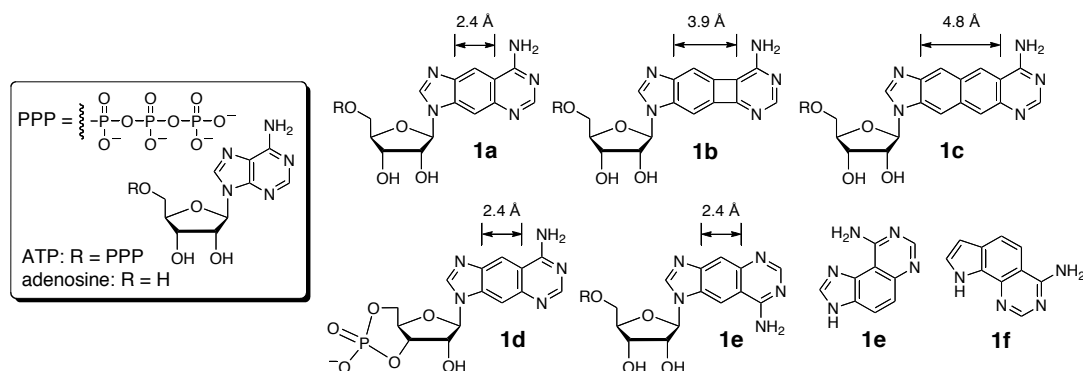
DESIGN AND APPLICATIONS OF NON-NATURAL NUCLEOBASES

Size-Expanded DNA Based On Benzo-Fused Nucleobases

The first systematic studies on enzyme-coenzyme binding sites, including several ATP-dependent enzymes,⁶ cAMP-dependent protein kinase,⁷ xanthine oxidase,⁸ firefly luciferase,⁹ and ribonucleotide reductase,¹⁰ using a series of “stretched-out” adenosine and guanosine ribonucleotides dates to the late 1970s. At that time, the ability to understand the structural details of binding sites was impeded by the lack of high-resolution X-ray structures of enzyme-ATP and enzyme-cAMP. Leonard

developed a series of “dimensional probes” to assess the size of the space available in the enzymes for the adenine moiety in ATP and cAMP (Chart 1).¹¹

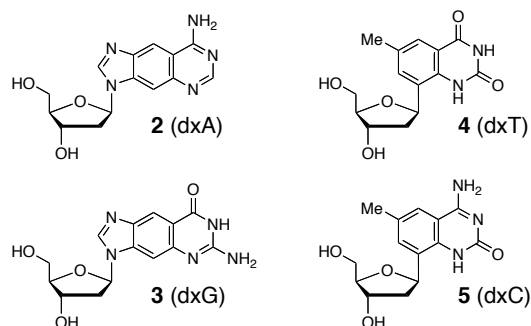
Chart 1. Representative Examples of Dimensional Probes for Enzyme-Coenzyme Binding Studies.



The aim of synthesizing these dimensional probes was to study the enzyme-coenzyme binding and kinetics relative to the natural substrates. For example, it was found that, in the presence of adenosine deaminase, compound **1a** (R = H) underwent deamination at a rate comparable to that of the natural substrate adenosine **1** (R = H), however compound **1e** (R = H) was not deaminated to any detectable quantity. These results suggest that the binding pocket of the deaminase can accommodate a substrate of larger size (i.e., 2.4 Å) than adenosine and the correct hydrogen bonding donor/acceptor array and orientation may be important for deamination.¹² This was a remarkable example of using synthetic chemistry to study biological systems. Current studies of complex biological systems are aided with sophisticated techniques such as high resolution X-ray crystallography and multi-dimensional NMR spectroscopy.

Analogous to Leonard’s “stretched-out” systems, Kool and coworkers have taken the expanded-nucleotide concept a step further by developing expanded DNA (xDNA) based on the benzo-fused deoxynucleotides shown in Chart 2.¹³ These non-natural bases were designed with the aim of developing non-natural DNA with higher stability and enhanced fluorescence properties. This, in turn, may facilitate the generation of a new genetic system that could be useful for chemical biology and nanoscale engineering. One of the factors that contributes to the high fidelity of natural DNA replication is the near absence of other tautomeric forms of the natural bases. The expanded versions of bases, however, have different electronic properties due to the presence of an extra “benzene” ring such that different tautomers may arise and hence, lower the fidelity of the system. For example, on the basis of semiempirical calculations of the heats of formation, it was found that xC has a second tautomeric form (i.e., imino form) which has only slightly higher energy (+1.3 kcal/mol) relative to the most stable form (i.e., the desired form) indicating that the imino form may exist in appreciable amounts in the system

Chart 2. Benzo-Fused Deoxynucleotides for xDNAs.



(Figure 1). Melting temperature (T_m) experiments, however, showed that dxC (**5**) is highly selective in pairing with dG suggesting the imino tautomer does not interfere with the pairing properties of dxC (**5**) even if it exists. In addition, one of the reasons that the high selectivity is achieved could be that there is simply no counterpart with a correct hydrogen-bonding array complementary to the imino tautomer. For other expanded bases, the desired tautomers are the most stable by at least 6–7 kcal/mol and T_m experiments support the theoretical predictions.¹⁴ Based on the T_m studies, Kool and coworkers also found that xDNAs have higher stabilities when compared to the natural sequences which can be ascribed to enhanced π -stacking properties of expanded bases.¹⁵ In addition, fluorescence properties are enhanced due to the extended conjugation.¹⁶ Since the expanded versions of oligonucleotides are readily synthesized by an automated DNA synthesizer, they are expected to be useful tools for probing DNA/RNA hybridization and protein/DNA interactions.¹³

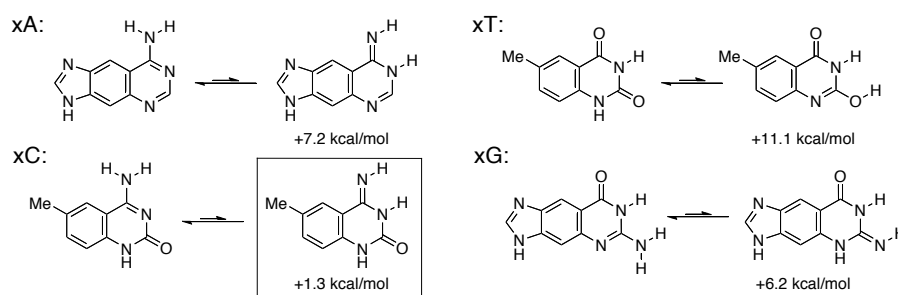


Figure 1. Semiempirical calculations (AM1) of relative tautomeric preferences in the gas phase. Only the tautomer with the closest energy to the most stable one is shown.

Alternative Hydrogen Bonding Arrays

Besides the standard Watson–Crick base pairing that is present in the natural DNA, one can envision that an alternative scheme can be designed in such a way that it is orthogonal to the natural one with the aim of expanding the genetic alphabet. In fact, Benner and coworkers proposed a set of non-natural bases based on altered hydrogen bonding arrays in 1990.¹⁷ This system is known as the Artificially Expanded Genetic Information System (AEGIS).¹⁸ All the bases contain three hydrogen

bonds because, in general, the stability of a non-natural base pair correlates with the number of hydrogen bonds. This system contains six mutually exclusive hydrogen-bonding arrays out of a total of $2^3 = 8$ possibilities. In these six arrays, one of them is the original C•G base pairing (Figure 2). This system has had many practical applications. The most notable one is the application in branched DNA assays that are used to monitor the viral RNA level in approximately 400,000 patients with HIV or hepatitis C annually.¹⁷

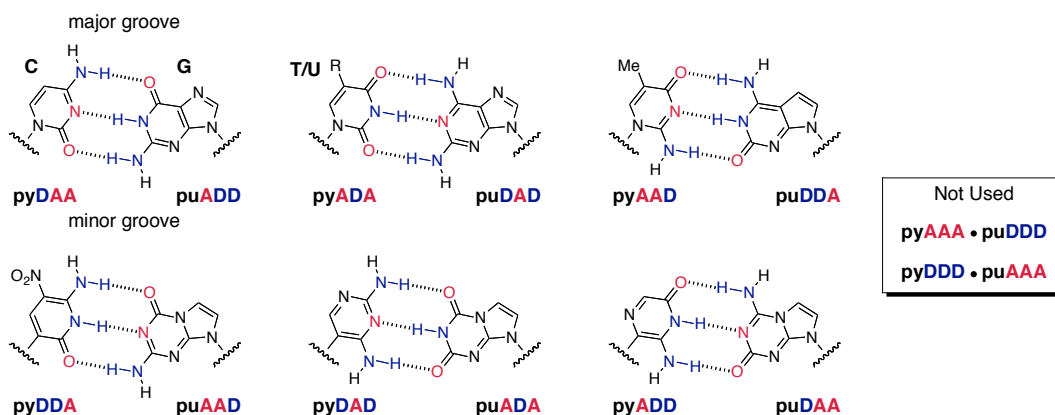


Figure 2. Six hydrogen-bonding arrays pairing 12 different nucleobases in AEGIS. Notation is as follows: pu = a fused ring system; py = a six-membered ring. The array of acceptor (A) and donor (D) groups from major to minor groove is indicated.

Another way to expand the genetic alphabet is to increase the number of hydrogen bonds from three to four and this approach was developed by Matsuda and coworkers recently.¹⁹ This approach represents a combination of Kool's xDNA system and Benner's AEGIS because an additional aromatic ring will be required for the incorporation of the extra hydrogen bond and to maintain the planar geometry of the base (Figure 3). The first quadruple hydrogen-bonding system consists of four different bases, which are all imidazopyridopyrimidine-based tricyclic nucleosides. It was originally anticipated that the replacement of one natural base pair in a given oligonucleotide by one of the quadruple hydrogen-bonding pairs would increase the melting temperature due to the extra hydrogen bonding and π -stacking interactions. In fact, melting temperature studies show a decrease in stability (ΔT_m ranges from -2 to -7 °C) suggesting a distortion of the backbone might have occurred (Table 1).²⁰ On the other hand, the second generation base pairs show significant stabilization ($\Delta T_m \approx +9$ °C) relative to the A•T pair under a more stringent conditions (1 mM vs. 100 mM NaCl). This observation is attributed to the minimized unfavorable backbone distortion due to the similar C1'-C1' distances of the non-natural pairs and the natural base pairs (~ 10.7 Å). No applications based on the quadruple hydrogen-bonding nucleobases are reported, but potential biological applications similar to the Benner's system can be envisioned.²⁰

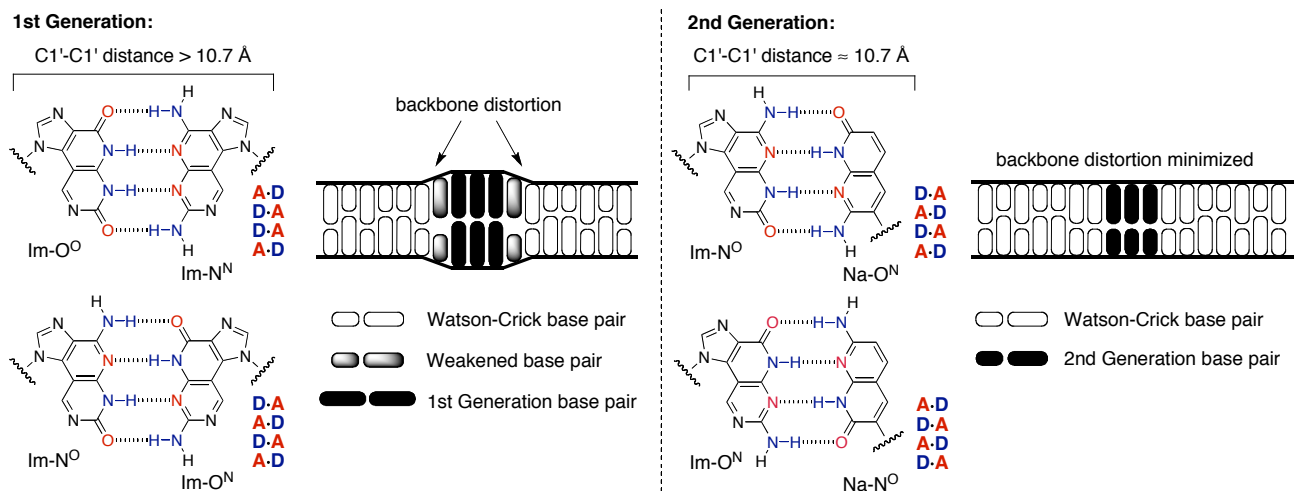


Figure 3. Non-natural quadruple hydrogen-bonding bases. Notation is as follows: Im = imidazopyridopyrimidine; Na = naphthyridine. (For example: Na-N^O is a naphthyridine nucleobase with an amine (N) group and a keto (O) group in the major groove and minor groove, respectively.)

Table 1. Melting temperature data of a DNA duplex containing an X•Y pair.^a

5' -GCACCGAAXAAACCACG-3'
3' -CGTGGCTTYTTTGGTGC-5'

X	Y	T_m (°C) ^b	ΔT_m (°C) ^c	X	Y	T_m (°C) ^d	ΔT_m (°C) ^c	
Im-N^N	Im-O^O	56.3	-4.2	Im-O^N	Na-N^O	57.2	+9.4	
	Im-N ^N	54.5	-6.0		Im-N^O	Na-O^N	56.4	+8.6
	Im-N ^O	54.5	-6.0		Im-O ^N	Im-N ^O	44.0	-3.8
	Im-O ^N	56.1	-4.4		Na-N ^O	Na-O ^N	50.1	+2.3
	Im-O^N	Im-N^O	58.4		-2.1	G	C	49.1
	Im-O ^O	53.3	-7.2	A	T	47.8	0	
	Im-N ^N	55.4	-5.1					
	Im-O ^N	56.9	-3.6					
G	C	61.9	+1.4					
A	T	60.5	0					

^a Matched base pairs are in bold. ^b 3 μ M duplex with 10 mM sodium cacodylate (pH 7.0) and 100 mM NaCl.

^c Determined by subtracting the T_m of the A•T pair from that of the X•Y pair. ^d Same as ^b except 1 mM NaCl.

Metal-Mediated Base Pairing

One of the bio-inspired approaches to elicit base pairing is the use of metal-ligand coordination interactions (Figure 4).²¹ In many biological systems, metal-ligand interactions play a significant role in terms of structural and functional effects such as the formation of G-quadruplexes in telomeres and the catalytic properties of RNA.²² It is anticipated that the stabilities of these metallo-base pairs are relatively higher than those of hydrogen-bonding base pairs because coordination bonds are usually much stronger than multiple hydrogen bonds. Similar to the hydrogen-bonding bases, the incorporation of these metallo-bases into oligonucleotides is feasible with standard protocols using an automated DNA synthesizer.²³ However, because of the much higher stability of the metal-ligand interactions, these

metallo-bases may not be ideal candidates for biological applications that require reversibility. Nevertheless, the interesting electronic and, in particular, magnetic properties of the metallo-DNAs may find applications in materials science, such as molecular wires and molecular magnets.²⁴

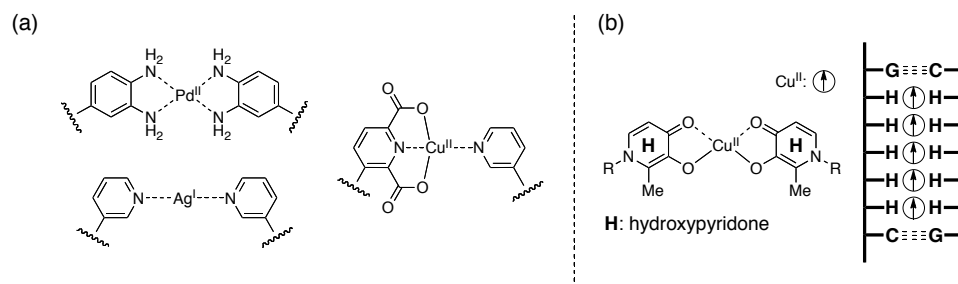


Figure 4. (a) Some metal-mediated base pairings. (b) An example of DNA-based molecular wire.^{24b}

Nonpolar Nucleoside Isosteres

Although scientists have been studying DNA for decades, some of the most important questions are still not answered.²⁵ For example, are hydrogen bonds important in duplex stabilization? Are hydrogen bonds necessary for selective base-pair formation? How important is the Watson–Crick hydrogen bonding in DNA replication? To this end, Kool and coworkers developed a set of nonpolar aromatic bases in the absence of hydrogen bonding donors and acceptors for Watson–Crick pairing. In addition, the shape of each natural base is mimicked as closely as possible by replacing the purine ring with indole or benzimidazole and the pyrimidine ring with substituted benzene (Figure 5). These nonpolar isosteres are therefore good candidates for studies to address the above questions, or at least provide insights into the roles of hydrogen bonding interactions in biological systems.²⁶ Unlike the previous three classes of non-natural bases, there is essentially no driving force for the base pairing and this allows one to probe the relative importance of hydrogen-bonding and steric effects. The structural similarities of the isosteres are confirmed by X-ray crystallography and NMR spectroscopy.²⁷ To address some of the mentioned questions, nonpolar isosteres were used to test the importance of Watson–Crick hydrogen bonding, base-pair geometry, and steric requirements in various biological systems. For example, they found that the Watson–Crick hydrogen bonding is essential for high

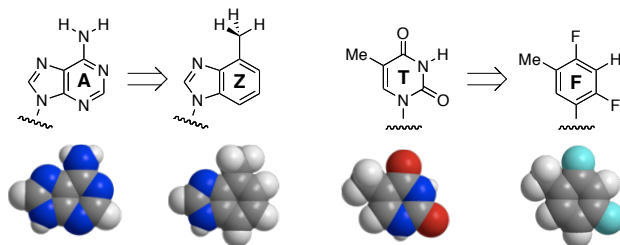


Figure 5. Nonpolar isosteres (Z and F) mimicking the size and shape of adenine (A) and thymine (T), respectively.

proofreading selectivity by the human mitochondrial DNA polymerase,²⁸ whereas in the DNA replication process in living bacterial cells, Watson–Crick hydrogen bonding is found to be unnecessary for high fidelity replication of a base pair *in vivo*, suggesting the geometric constraints in DNA polymerase is the most powerful determinant of fidelity of the process.²⁹

CONCLUSIONS AND OUTLOOK

Through the synthesis and application of non-natural nucleobases, the understanding of the properties and functions of the genetic material, DNA, and various biological processes involving DNA has been advanced. This progress is made possible only with proper molecular design and accessible synthetic methods. The preparations of non-natural bases also lead to the expansion of the genetic alphabet, which finds applications in biotechnology and clinical development. Enhancing DNA stability by using quadruple hydrogen bonding and robust metal-ligand interactions may also lead to potential applications in material science. Synthetic, non-natural nucleobases and DNA are indispensable tools for the emerging field of synthetic biology.³⁰ This field is not simple biomimetic chemistry, but also develops artificial chemical systems that can undergo replication, selection, and evolution that ultimately enables the relationship between chemistry and life to be better understood.

References:

-
- (1) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737–738.
 - (2) Mitsuya, H.; Yarchoan, R.; Broder, S. *Science* **1990**, *249*, 1533–1544.
 - (3) (a) Beaucage, S. L.; Caruthers, M. H. *Tetrahedron Lett.* **1981**, *22*, 1859–1862. (b) Letsinger, R. L. Lunsford, W. B. *J. Am. Chem. Soc.* **1976**, *98*, 3655–3661. (c) Letsinger, R. L.; Finnan, J. L.; Heavner, G. A.; Lunsford, W. B. *J. Am. Chem. Soc.* **1975**, *97*, 3278–3279.
 - (4) Kool, E. T. *Acc. Chem. Res.* **2002**, *35*, 936–943.
 - (5) (a) Berti, L.; Burley, G. A. *Nat. Nanotechnol.* **2008**, *3*, 81–87. (b) Tanaka, K.; Shionoya, M. *Chem. Lett.* **2006**, *35*, 694–699.
 - (6) Scopes, D. I. C.; Barrio, J. R.; Leonard, N. J. *Science* **1977**, *195*, 296–298.
 - (7) Dreyfuss, G.; Schwartz, K.; Blout, E. R.; Barrio, J. R.; Liu, F.-T.; Leonard, N. J. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 1199–1203.
 - (8) Leonard, N. J.; Keyser, G. E. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 4262–4264.
 - (9) Rosendahl, M. S.; Leonard, N. J.; DeLuca, M. *Photochem. Photobiol.* **1982**, *35*, 857–861.
 - (10) Rosendahl, M. S.; Omann, G. M.; Leonard, N. J. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 3480–3484.
 - (11) Leonard, N. J. *Acc. Chem. Res.* **1982**, *15*, 128–135.
 - (12) Leonard, N. J.; Sprecker, M. A.; Morrice, A. G. *J. Am. Chem. Soc.* **1976**, *98*, 3987–3994.

-
- (13) Krueger, A. T.; Lu, H.; Lee, A. H. F.; Kool, E. T. *Acc. Chem. Res.* **2007**, *40*, 141–150.
- (14) (a) Liu, H.; Gao, J.; Kool, E. T. *J. Org. Chem.* **2005**, *70*, 639–647. (b) Liu, H.; Gao, J.; Maynard, L.; Saito, D.; Kool, E. T. *J. Am. Chem. Soc.* **2004**, *124*, 1102–1109.
- (15) (a) Lu, H.; He, K.; Kool, E. T. *Angew. Chem. Int. Ed.* **2004**, *43*, 5834–5836. (b) Liu, H.; Gao, J.; Lynch, S. R.; Saito, D.; Maynard, L.; Kool, E. T. *Science* **2003**, *302*, 868–871.
- (16) Krueger, A. T.; Kool, E. T. *J. Am. Chem. Soc.* **2008**, *130*, 3989–3999.
- (17) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33–37.
- (18) Benner, S. A. *Acc. Chem. Res.* **2004**, *37*, 784–797.
- (19) (a) Hikishima, S.; Minakawa, N.; Kuramoto, K.; Fujisawa, Y.; Ogawa, M.; Matsuda, A. *Angew. Chem. Int. Ed.* **2005**, *44*, 596–598. (b) Minakawa, N.; Kojima, N.; Hikishima, S.; Sasaki, T.; Kiyosue, A.; Atsumi, N.; Ueno, Y.; Matsuda, A. *J. Am. Chem. Soc.* **2003**, *125*, 9970–9982.
- (20) Minakawa, N.; Kuramoto, K.; Hikishima, S.; Matsuda, A. *ARKIVOC* **2006**, *vii*, 326–337.
- (21) (a) Shionoya, M.; Tanaka, K. *Curr. Opin. Chem. Biol.* **2004**, *8*, 592–597. (b) Tanaka, K.; Yamada, Y.; Shionoya, M. *J. Am. Chem. Soc.* **2002**, *124*, 8802–8803. (c) Tanaka, K.; Shionoya, M. *J. Org. Chem.* **1999**, *64*, 5002–5003.
- (22) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*, 6th ed.; W. H. Freeman: New York, 2006.
- (23) Meggers, E.; Holland, P. L.; Tolman, W. B.; Romesberg, F. E.; Schultz, P. G. *J. Am. Chem. Soc.* **2000**, *122*, 10714–10715.
- (24) (a) Fischler, M.; Simon, U.; Nir, H.; Eichen, Y.; Burley, G. A.; Gierlich, J.; Gramlich, P. M. E.; Carell, T. *Small* **2007**, *3*, 1049–1055. (b) Tanaka, K.; Tengeiji, A.; Kato, T.; Toyama, N.; Shionoya, M. *Science* **2003**, *299*, 1212–1213.
- (25) (a) Kool, E. T. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 1–22. (b) Kool, E. T.; Morales, J. C.; Guckian, K. M. *Angew. Chem. Int. Ed.* **2000**, *39*, 990–1009.
- (26) (a) Moran, S.; Ren, R. X.-F.; Rumney, S., IV; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056–2057. (b) Schweitzer, B. A.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 1863–1872.
- (27) (a) Guckian, K. M.; Morales, J. C.; Kool, E. T. *J. Org. Chem.* **1998**, *63*, 9652–9656. (b) Guckian, K. M.; Kool, E. T. *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 2825–2828.
- (28) Lee, H. R.; Helquist, S. A.; Kool, E. T.; Johnson, K. A. *J. Biol. Chem.* **2008**, *283*, 14411–14416.
- (29) Delaney, J. C.; Henderson, P. T.; Helquist, S. A.; Morales, J. C.; Essigmann, J. M.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4469–4473.
- (30) Benner, S. A.; Sismour, A. M. *Nat. Rev. Genet.* **2005**, *6*, 533–543.