Synthetic Biology

DOI: 10.1002/ange.201001977

Artificial Genetic Systems: Self-Avoiding DNA in PCR and Multiplexed PCR**

*Shuichi Hoshika, Fei Chen, Nicole A. Leal, and Steven A. Benner**

Many applications of DNA chemistry in biology and medicine would be enhanced if procedures for the efficient analysis of single DNA molecules also worked well for the analysis of many DNA molecules (multiplexing). Unfortunately, multiplexing often requires the addition of many DNA probes and primers to an assay at the same time, often in great excess with respect to the targeted DNA molecules. Multiple primers built from standard nucleotides can easily interact with each other, even when well-designed. These interactions can create artifacts and noise that defeat the analysis, especially when polymerases are involved in the analytic architecture, as in multiplexed PCR. With more than a dozen target amplicons, multiplexed PCR generally fails because of PCR artifacts.[1]

Recently, we reported that the efficiency and consistency of multiplexed PCR could be greatly improved by placing components of our artificially expanded genetic information system (AEGIS) in the external primers in a nested PCR architecture.[2] AEGIS increases the number of independently replicable nucleotides from the natural four (A, T, G, and C) to as many as 12. AEGIS is now in the clinic, where it personalizes the care of some 400000 patients annually infected with the HIV, hepatitis B, and hepatitis C viruses.[3] However, a nested PCR architecture still does not prevent the analyte-specific segments of the chimeric primers from interacting with each other, as these segments must be constructed from natural nucleotides.

In a different strategy, multiplexed PCR might be enabled if the analyte-specific portions of the primers were built from a "self-avoiding molecular-recognition system" (SAMRS). SAMRS DNA can be viewed as the opposite of AEGIS DNA in that it binds to natural DNA, but *not* to other members of the same SAMRS species. Schematically, an SAMRS replaces T, A, G, and C with the nucleotide analogues T*, A*, G* and C*, whereby T* pairs with A, A* pairs with T, G* pairs with C, and C* pairs with G, but neither the T*–A* pair nor the G*–C* pair contributes substantially to the stability of a duplex. In particular, if PCR primers were built from SAMRS components, they should enable multiplexed PCR without artifacts arising from primer–primer interactions.

Empirical studies have shown that pairs joined by two hydrogen bonds contribute to duplex stability, but not pairs joined by one hydrogen bond.[4] Accordingly, a candidate for G* in a "first-generation" SAMRS heterocycle might be hypoxanthine (found in inosine), which pairs with C by using the top two hydrogen-bonding units of C (Scheme 1, top left). The corresponding first-generation candidate for C* would be pyrimidin-2-one (found in zebularine),[5] which pairs with standard G by using the bottom two hydrogen-bonding units of G (Scheme 1, top left). As hypoxanthine and pyrimidin-2-one can form only one hydrogen bond in a standard Watson–Crick arrangement, the resulting pair should not contribute to duplex stability; the inosine–zebularine pair would be a G*–C* self-avoiding pair.

For the second self-avoiding pair, pyridone might be a first-generation T* candidate. It would pair with standard A by using the top two hydrogen-bonding units of A (Scheme 1, top right). As standard adenine lacks a "bottom" hydrogen-bonding unit, 2-aminopurine would be a candidate for A*: it would pair with standard T at the bottom two sites. 2-Aminopurine and pyridone would form only one hydrogen bond (Scheme 1, top right) and therefore would not contribute to duplex stability. The aminopurine–pyridone pair would then be an A*–T* self-avoiding pair. Some representative melting temperatures of duplexes incorporating these SAMRS components are shown in Tables 1 and 2 of the Supporting Information.

SAMRS should be effective for simple binding assays. For example, in 1996, Kutyavin et al.[6] reported that "pseudocomplementary" diaminopurine and 2-thiothymine[7] bound to thymine and adenine, respectively, but that diaminopurine did not bind to 2-thiothymine. The use of 2-thiothymine instead of pyridone as a T* candidate is consistent with a need for minor-groove solvation to stabilize double helices.[8] Indeed, 2-thiothymine pairs with A slightly better than T itself (see Table 3 in the Supporting Information).

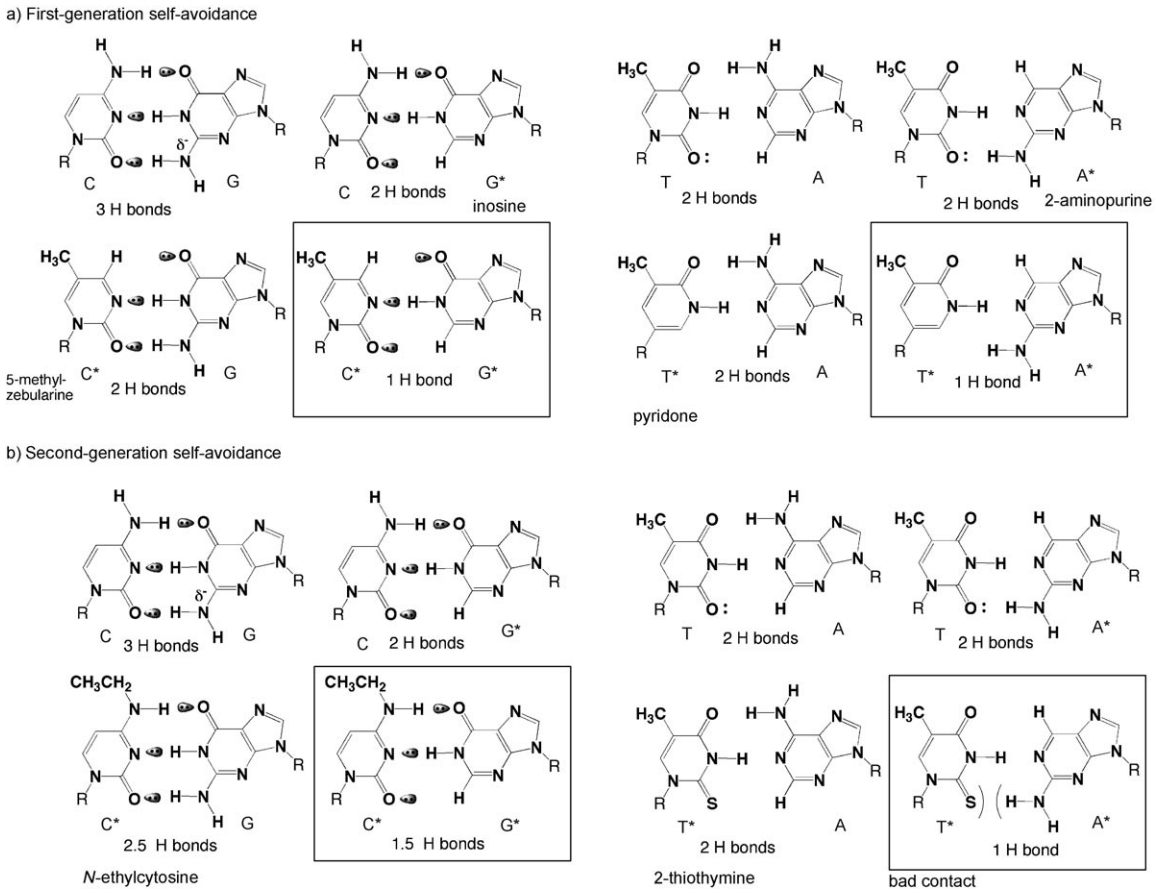
Pseudocomplementarity of this limited type has been used in peptide nucleic acids (PNAs) to invade duplex DNA.[9,10] Gamper and co-workers showed that similar species could be incorporated into DNA as triphosphates, and suggested that the products from this incorporation might not fold and might therefore be more uniformly captured on arrays.[11,12]

Accordingly, we attempted to extend the SAMRS concept to PCR by incorporating various SAMRS candidates into PCR primers on the basis of what we learned by analyzing duplexes built from a first-generation SAMRS alphabet (see Tables 1 and 2 in the Supporting Information). We encountered multiple difficulties. First, 2′-deoxy-5-methylzebularine

[*] Dr. S. Hoshika, Dr. F. Chen, Dr. N. A. Leal, Dr. S. A. Benner
Foundation for Applied Molecular Evolution
The Westheimer Institute for Science and Technology
720 SW 2nd Avenue, Suite 201, Gainesville, FL 32601 (USA)
Fax: (+1) 352-271-7076
E-mail: sbenner@ffame.org
Homepage: http://www.ffame.org

[**] This research was supported by a grant from Nucleic Acids Licensing LLC and by the National Human Genome Research Institute under 1R01 HG004831 and the Defense Threat Reduction Agency under its basic research program (HDTRA-08-1-0052).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201001977.



Scheme 1. Two generations of candidate nucleobases for self-avoiding molecular-recognition systems (SAMRS). a) First-generation SAMRS candidates are simple implementations in which the top two hydrogen-bonding units of the standard nucleobase are used for one pair and the bottom two are used for the other. b) Second-generation SAMRS exploits 2-thiothymine as T* to resolve issues arising from the weak bonding of the first-generation T* nucleobase to adenine and N^4-ethylcytosine as C* to resolve issues arising from the chemical instability of the first-generation C* nucleobase.

proved to be insufficiently stable in both acid and base to be useful in standard phosphoramidite DNA synthesis.[13] This problematic chemical reactivity was only partly mitigated by placing substituents on the heterocycle. Unfortunately, 5-phenyl- and 5-propynyl-substituted 2′-deoxyzebularines could not be made, and DNA containing 4,5-dimethylzebularine had a low T_m value.

Further problems were encountered with hypoxanthine as a G* candidate. A dozen thermophilic DNA polymerases were tested for their ability to support PCR with primers containing five or six inosine units as G* (data not shown). Most polymerases from extreme thermophiles rejected hypoxanthine, possibly because it is a deamination product of adenosine that occurs at very high temperatures, in the natural environment of extreme thermophiles.[14] In contrast, Taq DNA polymerase performed well in reading through SAMRS components in a template (see Figure 1 in the Supporting Information). We therefore focused on Taq to develop PCR with primers that incorporated SAMRS components (see Figure 2 in the Supporting Information).

During these studies, we encountered the surprising result that standard DNA duplexes held together entirely by pairs joined by just two hydrogen bonds were remarkably poor primers. The melting temperatures of such duplexes were also surprisingly low. To mitigate this problem, we first sought to replace zebularine derivatives as C* units. N^4-Methyl- and N^4-ethylcytosines[15] with adjacent 5-methyl groups proved not to form stable pairs; however, both N^4-methyl- and N^4-ethylcytosine performed well as C* (see Table 4 in the Supporting Information). The N^4-ethyl variant was chosen because it better distinguished various matches (see Table 5 in the Supporting Information).

We then developed a complete SAMRS based on 2-thiothymine, 2-aminopurine, hypoxanthine, and N^4-ethylcytosine as T*, A*, G*, and C*. When introduced individually into a reference DNA duplex, the corresponding SAMRS:standard pairs contributed to duplex stability to the same extent as an A:T pair (Table 1).[16] In every case, however, the SAMRS:SAMRS pair contributed less to the stability of the reference duplex than the corresponding SAMRS:standard pair.

We then turned to the development of polymerases that were compatible with this optimized chemistry, whereby we recognized that the properties of polymerases are rarely predictable.[17] Surprisingly, when we used the Klenow fragment of DNA polymerase 1, we found that 25-mer primers that formed duplexes joined uniformly by two hydrogen bonds performed unpredictably, even at low temperatures (see Figure 3 in the Supporting Information). The inefficiency

Table 1: Melting temperatures (T_m) for 5′-ACCAAGCXATCAAGT-3′ and 3′-TGGTTCGYTAGTTCA-5′.[a]

X\Y	T	T*	A	A*	C	C*	G	G*
A	55.5	*56.8*	43.7	46.5	45.1	43.5	46.7	49.8
A*	*54.5*	**52.0**	46.8	45.5	48.1	44.0	45.8	46.8
T	46.3	48.0	54.0	*52.5*	44.6	45.0	48.4	46.3
T*	47.0	50.0	*54.0*	**50.3**	40.9	41.3	44.6	45.1
G	49.5	47.0	47.0	45.1	58.8	*52.0*	47.0	46.0
G*	48.8	47.0	50.5	45.1	*54.1*	**49.3**	46.0	46.3
C	44.0	40.6	42.8	47.1	43.8	41.0	59.0	*52.6*
C*	44.0	42.0	42.0	43.0	41.1	39.5	*52.0*	**47.8**

[a] Outlined boxes contain T_m values for complementary pairs in two contexts. The italicized T_m values are for duplexes with matched SAMRS:SAMRS pairs; these values are lower than the corresponding italicized T_m values for duplexes with SAMRS:standard pairs. The T_m values in bold type are similar to those of duplexes with A:T pairs, which are also joined by two hydrogen bonds. The off-diagonal T_m values corresponding to formal mismatches are lower than those for the duplexes with N:N* pairs.

in priming correlated with the low T_m values of their SAMRS:standard duplexes: the T_m values of a set of SAMRS 25-mers paired with complementary standard DNA were all approximately 40 °C (see Table 6 in the Supporting Information), far below the melting temperature of 60–70 °C of a typical 25-mer duplex built from equal proportions of A, T, G, and C.

To mitigate this problem, we examined various backbones, including 2′-*O*-alkyl ribonucleosides, for their compatibility with SAMRS. Although 2′-*O*-alkyl ribonucleosides improved the stability of duplexes joined by SAMRS:standard pairs, they diminished the ability of the oligonucleotide to support PCR.

We therefore wondered whether chimeric primers containing SAMRS in their 3′ segments and standard nucleotides in their 5′ segments would still display useful self-avoidance. We prepared a pair of primers that targeted the *Taq* gene and were perfectly matched in their last nine nucleotides (Figure 1) with zero, four, or eight SAMRS components in their 3′ segments and a standard 3′-terminal nucleotide (to lower the cost of synthesis).

The PCR results were striking. Primer pairs built from standard nucleotides failed completely to yield the desired amplicon (1109 base pairs, Figure 1); only primer dimer was observed. When one of the primers was built from standard nucleotides and the other contained four SAMRS nucleotides, the amplicon was formed only inefficiently, and primer dimer resulted from SAMRS–standard mismatching between the primers. However, when both primers had four or eight SAMRS components, PCR amplification efficiently gave *only* the desired amplicon. This result was a surprising demonstration of the SAMRS effect in PCR, even for short SAMRS segments. It was confirmed in real-time PCR by using primers with eight SAMRS components near their 3′ ends in a chimeric {16 + 8* + 1} architecture (see Figure 4 in the Supporting Information).

We then tested multiplexed PCR with SAMRS primers. Ten pairs of chimeric {16 + 8* + 1} primers were prepared to target 14 cancer-relevant genes. The primers were chosen to give a ladder of amplicons of increasing length to facilitate analysis of the 10 PCR products by agarose gel electrophoresis. They were *not* designed by computer programs to explicitly avoid PCR artifacts. Control primers had analogous sequences built entirely from standard nucleotides.

Singleplex PCR was successful with all of the {16 + 8* + 1} primer pairs (Figure 2c). With standard primer pairs, singleplex PCR was also successful, except with the PTPN11 amplicon, which failed because of primer-dimer formation (Figure 2a). The grouping of these primer pairs in sets showed the advantage of SAMRS primer pairs over standard primer pairs in multiplexed PCR. For example, with standard primer pairs, fivefold multiplexing (FLT3, TSHR, EGFR, CTNNB1, APC; Figure 2b, right) gave only two of the five desired amplicons. In contrast, analogous multiplexing with {16 + 8* + 1} SAMRS primer pairs generated all desired amplicons (Figure 2d, right). PCR with all 10 {16 + 8* + 1} SAMRS primer pairs gave all 10 amplicons (Figure 2d). In contrast, PCR with standard primer pairs gave only five (or possibly six) of the 10 desired amplicons (Figure 2b).

This study reinforces the evolving view of DNA as a complex organic molecule rather than a simple linear string that pairs according to simple rules (the first-generation model for DNA of Watson and Crick).[18] Thus, although we expected that duplexes that contained more A:T-like-base pairs would have lower T_m values, we did not expect the

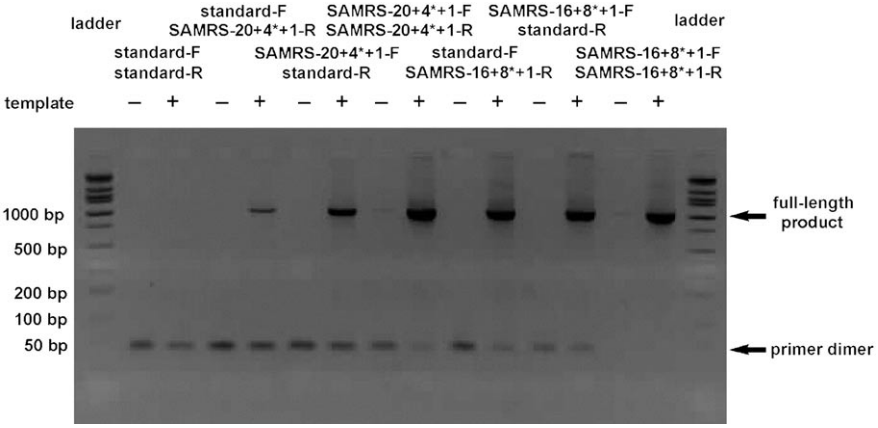


Figure 1. Amplification of the *Taq* gene and demonstration of the ability of SAMRS to manage PCR artifacts in a "worst-case-design" scenario, in which the forward and reverse primers are formally complementary in their last nine nucleotides. Standard primer pairs gave only primer dimers in these cases. Primer pairs that contained SAMRS components (even as few as four) in the 3′ segment gave the desired 1109 nucleotide amplicon. The symbols + and – indicate the presence and absence of the target gene. A* = 2-aminopurine, G* = hypoxanthine, T* = 2-thiothymine, C* = N^4-ethylcytosine. Standard-F: 5′-TATCTGCGTGCCCTGTCTCTGGAGG-3′, standard-R: 5′-CCAATGCCAACCTCTACCTCCAGAG-3′, SAMRS-20 + 4* + 1-F: 5′-TATCTGCGTGCCCTGTCTCTG*G*A*G*G-3′, SAMRS-20 + 4* + 1-R: 5′-CCAATGCCAACCTC-TACCTCC*A*G*A*G-3′, SAMRS-16 + 8* + 1-F: 5′-TATCTGCGTGCCCTGTC*T*C*T*G*G*A*G*G-3′, SAMRS-16 + 8* + 1-R: 5′-CCAATGCCAACCTCTAC*C*T*C*C*A*G*A*G-3′.

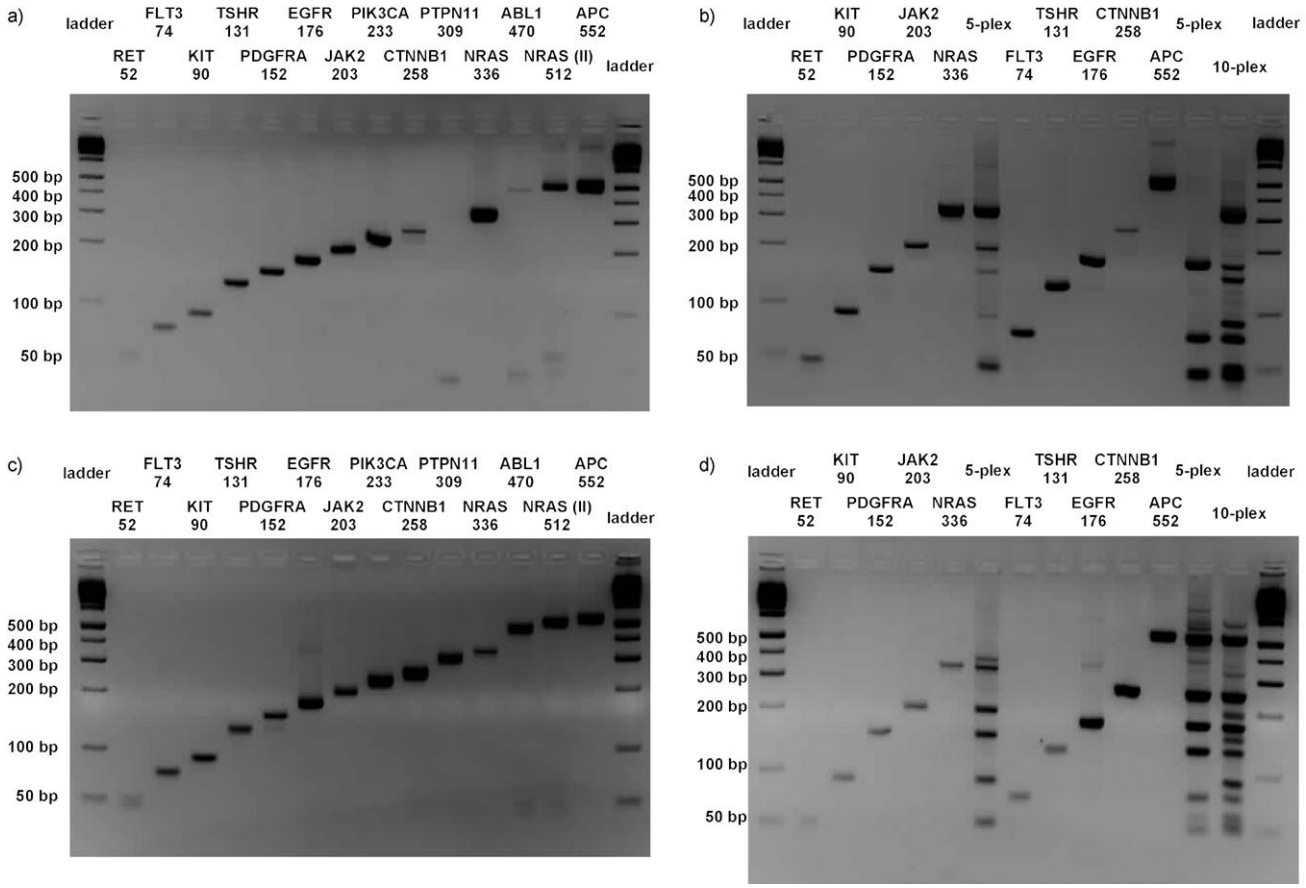


Figure 2. PCR amplification of 14 cancer genes with primer pairs chosen to enable the amplicons to be conveniently separated by size on a 3% agarose gel. The sequences of the primer pairs are given in Table 7 of the Supporting Information. a) Singleplex PCR with indicated primer pairs containing only standard nucleotides. b) Attempted multiplexed PCR with indicated pairs containing only standard nucleotides. c) Singleplex PCR with the indicated {16 + 8* + 1} SAMRS:standard primer pairs. d) Multiplexed PCR with the indicated {16 + 8* + 1} SAMRS:standard primer pairs.

T_m value to drop so severely as the fraction of such pairs approached unity. Considering the etiology of nucleic acids,[19] it is tempting to infer from this effect a need in DNA for at least one pair to be joined by three hydrogen bonds.

Received: April 2, 2010
Published online: June 28, 2010

Keywords: DNA analogues · DNA polymerases · polymerase chain reaction · pseudocomplementarity · synthetic biology

[1] S. Fredriksson, J. Banér, F. Dahl, A. Chu, H. Ji, K. Welch, R. W. Davis, *Nucleic Acids Res.* **2007**, *35*, e47.
[2] Z. Yang, F. Chen, S. G. Chamberlin, S. A. Benner, *Angew. Chem.* **2010**, *122*, 181 – 184; *Angew. Chem. Int. Ed.* **2010**, *49*, 177 – 180.
[3] S. A. Benner, *Acc. Chem. Res.* **2004**, *37*, 784 – 797.
[4] C. R. Geyer, T. R. Battersby, S. A. Benner, *Structure* **2003**, *11*, 1485 – 1498.
[5] D. Cech, A. Holy, *Collect. Czech. Chem. Commun.* **1977**, *42*, 2246 – 2260.
[6] I. V. Kutyavin, R. L. Rhinehart, E. A. Lukhtanov, V. V. Gorn, R. B. Meyer, Jr., H. B. Gamper, Jr., *Biochemistry* **1996**, *35*, 11170 – 11176.
[7] B. A. Connolly, P. C. Newman, *Nucleic Acids Res.* **1989**, *17*, 4957 – 4974.
[8] T. Lan, L. W. McLaughlin, *J. Am. Chem. Soc.* **2000**, *122*, 6512 – 6513.
[9] T. Ishizuka, J. Yoshida, Y. Yamamoto, J. Sumaoka, T. Tedeschi, R. Corradini, S. Sforza, M. Komiyama, *Nucleic Acids Res.* **2008**, *36*, 1464 – 1471.
[10] V. V. Demidov, E. Protozanova, K. I. Izvolsky, C. Price, P. E. Nielsen, M. D. Frank-Kamenetskii, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 5953 – 5958.
[11] H. B. Gamper, Jr., A. Gewirtz, J. Edwards, Y.-M. Hou, *Biochemistry* **2004**, *43*, 10224 – 10236.
[12] G. Lahoud, V. Timoshchuk, A. Lebedev, M. de Vega, M. Salas, K. Arar, Y.-M. Hou, H. Gamper, *Nucleic Acids Res.* **2008**, *36*, 3409 – 3419.
[13] M. Vives, R. Eritja, R. Tauler, V. E. Marquez, R. Gargallo, *Biopolymers* **2004**, *73*, 27 – 43.
[14] H. Kamiya, T. Sakaguchi, N. Murata, M. Fujimuro, H. Miura, H. Ishikawa, M. Shimizu, H. Inoue, S. Nishimura, A. Matsukage, C. Masutani, F. Hanaoka, E. Ohtsuka, *Chem. Pharm. Bull.* **1992**, *40*, 2792 – 2795.
[15] H. K. Nguyen, E. Bonfils, P. Auffray, P. Costaglioli, P. Schmitt, U. Asseline, M. Durand, J. C. Maurizot, D. Dupret, N. T. Thuong, *Nucleic Acids Res.* **1998**, *26*, 4249 – 4258.
[16] S. Bommarito, N. Peyret, J. SantaLucia, Jr. *Nucleic Acids Res.* **2000**, *28*, 1929 – 1934.
[17] J. Horlacher, M. Hottiger, V. N. Podust, U. Huebscher, S. A. Benner, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 6329 – 6333.
[18] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737 – 738.
[19] A. Eschenmoser, *Science* **1999**, *284*, 2118 – 2124.

