Published in final edited form as:

Angew Chem Int Ed Engl. 2010 July 26; 49(32): 5554–5557. doi:10.1002/anie.201001977.

Artificial Genetic Systems. Self-Avoiding DNA in PCR and Multiplexed PCR

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

Foundation for Applied Molecular Evolution, The Westheimer Institute for Science and Technology, 720 SW 2nd Avenue, Suite 201, Gainesville FL 32601

Graphical Abstract

A “Self-Avoiding Molecular Recognition System” (SAMRS) is a species of DNA that binds to natural DNA but not to other members of the same SAMRS species. SAMRS should be useful in clinical analyses where many DNA molecules must interact with target DNA but not with each other. We report here a SAMRS based on 2-aminopurine (A*), 2-thiothymine (T*), 2'-hypoxanthine (G*) and N4-ethylcytosine (C*) and its use in multiplexed polymerase chain reactions.

Keywords

PCR. Self-avoiding genetic systems; Pseudocomplementarity; DNA polymerases; Synthetic biology; DNA analogs

Many applications of DNA chemistry in biology and medicine would be enhanced if procedures that efficiently analyze *single* DNA molecules also worked well to analyze *many* DNA molecules (multiplexing). Unfortunately, multiplexing often requires adding many DNA probes and primers to an assay at the same time, often in great excess over 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 defeats 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.^[i]

Recently, we reported that the efficiency and consistency of multiplexed PCR could be greatly improved by placing components of our artificially expanded genetic information

*Fax (+1) 352 271 7076, sbenner{at}ffame.org, Homepage: <http://www.ffame.org>.

Supporting information for this article is available on the WWW under <http://www.angewandte.org> or from the author.

system (Aegis) in the external primers in a nested PCR architecture.^[iii] 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 today personalizes the care of some 400,000 patients annually infected with the HIV, hepatitis B, and hepatitis C viruses.^[iii] However, a nested PCR architecture still does not prevent the analyte-specific segments of the chimeric primers from interacting with each other, as these 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). The opposite of Aegis DNA, Samrs DNA binds to natural DNA, but *not* to other members of the same Samrs species. Schematically, a Samrs replaces T, A, G and C by the nucleotide analogs T*, A*, G* and C*, where 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 show that pairs joined by two hydrogen bonds contribute to duplex stability, but not pairs joined by one hydrogen bond.^[iv] Accordingly, a candidate for G* in a “first generation” Samrs heterocycle might be hypoxanthine (found in inosine), which pairs with C using the top two hydrogen bonding units (Figure 1, top left). The corresponding first generation candidate for C* would be pyrimidin-2-one (found in zebularine),^[v] which pairs with standard G using the bottom two hydrogen bonding units (Figure 1, top left). As hypoxanthine and pyrimidin-2-one can form only one hydrogen bond in a standard Watson-Crick geometry, their 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, pairing with standard A using its top two hydrogen bonding units (Figure 1, top right). As standard adenine lacks a “bottom” hydrogen bonding unit, 2-aminopurine would be a candidate for A*, pairing with standard T at the bottom two sites. 2-Aminopurine and pyridone would form only one hydrogen bond (Figure 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 Suppl. Tables 1 and 2.

Samrs should work for simple binding assays. For example, in 1996, Kutyavin *et al.*^[vi] reported that “pseudocomplementary” diaminopurine and 2-thiothymine^[vii] bound to thymine and adenine respectively, but that diaminopurine did not bind to 2-thiothymine. Using 2-thiothymine instead of pyridone as a T* candidate is consistent with a need for minor groove solvation to stabilize double helices.^[viii] Indeed, 2-thiothymine pairs with A slightly better than T itself (Suppl. Table 3).

Pseudocomplementarity of this limited type has been used in peptide nucleic acids (PNAs) to invade duplex DNA.^[ix,x] Gamper showed that similar species as triphosphates could be

incorporated into DNA, suggesting that the products from this incorporation might not fold and therefore be more uniformly captured on arrays.^[xi,xii]

Accordingly, we attempted to extend the Samrs concept to PCR by incorporating various Samrs candidates into PCR primers based on what we learned by analyzing duplexes built from a first generation Samrs alphabet (Suppl. Tables 1 and 2). We encountered multiple difficulties. First, 2'-deoxy-6-methylzebularine proved to be insufficiently stable in both acid and base to be useful in standard phosphoramidite DNA synthesis.^[xiii] 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, while DNA containing 4,5-dimethylzebularine had a low T_m .

Further problems were encountered with hypoxanthine as a candidate G*. A dozen thermophilic DNA polymerases were tested for their ability to support PCR using primers containing 5 or 6 inosines as G* (data not shown). Most polymerases from extreme thermophiles rejected hypoxanthine, possibly because it is a deamination product of adenosine occurring at very high temperatures where extreme thermophiles live.^[xiv] In contrast, *Taq* DNA polymerase performed well reading through Samrs components in a template (Suppl. Figure 1). We therefore focused on *Taq* to develop PCR with primers that incorporated Samrs components (Suppl. Figure 2).

Here, we encountered the surprising result that standard DNA duplexes held together entirely by pairs joined by just two hydrogen bonds were remarkably poor primers. Melting temperatures of such duplexes were also surprisingly low.

To mitigate this problem, we first sought to replace zebularine derivatives as C*'s. N⁴-Methyl and N⁴-ethyl cytosines^[xv] with adjacent 5-methyl groups proved not to form stable pairs. However, both N⁴-methyl and N⁴-ethyl cytosine performed well as C* (Suppl. Table 4); the N⁴-ethyl variant was chosen because it better distinguished various matches (Suppl. Table 5).

A complete Samrs using 2-thiothymine, 2-aminopurine, hypoxanthine, and N⁴-ethylcytosine as T*, A*, G* and C* was then developed. When introduced individually into a reference DNA duplex, the corresponding Samrs:standard pairs contributed to duplex stability as well as an A:T pair (Table 1).^[xvi] In every case, however, the Samrs:Samrs pair contributed to the stability of the reference duplex *less* than the corresponding Samrs:standard pair.

We then turned to develop polymerases that used this optimized chemistry, recognizing that the properties of polymerases are rarely predictable.^[xvii] Surprisingly, we found that 25-mer primers forming duplexes joined uniformly by two hydrogen bonds performed unpredictably, even at low temperatures using the Klenow fragment of DNA polymerase 1 (Suppl. Figure 3). The inefficiency in priming correlated with the low T_m s of their Samrs:standard duplexes. Thus, the T_m s of a set of Samrs 25-mers paired with complementary standard DNA were all approximately 40 °C (Suppl. Table 6), far below the 60–70 °C T_m of a typical 25-mer duplex build from equal proportions of A, T, G, and C.

To mitigate this problem, various backbones were examined to support Samrs, including 2'-O-alkyl ribonucleosides. While improving the stability of duplexes joined by Samrs:standard pairs, these diminished the ability of the oligonucleotide to support PCR.

We therefore asked whether chimeric primers containing Samrs at their 3'-segments and standard nucleotides in their 5'-segments would still display useful self-avoidance. A pair of primers targeting the *Taq* gene that perfectly matched in their last nine nucleotides (Figure 2) were prepared with zero, four or eight Samrs components in their 3'-segments, with the 3'-terminal nucleotide remaining standard (to lower the cost of synthesis).

PCR results were striking. Primer pairs built from standard nucleotides failed completely to yield the desired amplicon (1109 base pairs, Figure 2); only primer-dimer was observed. The combination where one of the primers was built from standard nucleotides while the other contained four Samrs nucleotides gave the amplicon only inefficiently, with primer dimer arising 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 was a surprising demonstration of the Samrs effect in PCR, even for short Samrs segments. This result was confirmed in real time PCR using primers with eight Samrs components near their 3'-ends in a chimeric {16+8*+1} architecture (Suppl. Fig. 4).

We then tested multiplexed PCR using 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, facilitating analysis of the ten 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.

Single-plexed PCR was successful with all of the {16+8*+1} primer pairs (Figure 3c). With standard primer pairs, singleplexed PCRs were also successful, except with the PTPN11 amplicon, which failed because of primer dimer formation. Grouping 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, 5-fold multiplexing (FLT3, TSHR, EGFR, CTNNB1, APC, Figure 3b, right) gave only two of the five desired amplicons. In contrast, the analogous multiplexing with {16+8*+1} Samrs primer pairs generated all desired amplicons (Figure 3d, right). PCR with all ten {16+8*+1} Samrs primer pairs gave all ten amplicons (Figure 3d). In contrast, PCR with standard primer pairs gave only five (or possibly six) of the ten desired amplicons (Figure 3b).

This work reinforces the evolving view of DNA as a complex organic molecule rather than a simple linear string that pairs following simple rules, the first generation model for DNA of Watson and Crick.^[xviii] Thus, while we expected that duplexes joined more A:T-like base would have lower T_m s, we did not expect the T_m to drop so severely as the fraction of such pairs approached unity. Considering the etiology of nucleic acids,^[xix] it is tempting to infer from this a need in DNA for at least one pair to be joined by three hydrogen bonds.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by a grant from Nucleic Acids Licensing LLC and by the National Human Genome Research Institute under 1R01 HG004831.

References

- i. Fredriksson S, Banér J, Dahl F, Chu A, Ji H, Welch K, Davis RW. *Nucleic Acids Res.* 2007; 35:e47, 6. [PubMed: 17317684]
- ii. Yang Z, Chen F, Chamberlin SG, Benner SA. *Angew Chem Int Ed.* 2010; 49:177–180.
- iii. Benner SA. *Accounts Chem Res.* 2004; 37:784–797.
- iv. Geyer CR, Battersby TR, Benner SA. *Structure.* 2003; 11:1485–1498. [PubMed: 14656433]
- v. Cech D, Holy A. *Coll Czech Chem Comm.* 1977; 42:2246–2260.
- vi. Kutayav IV, Rhinehart RL, Lukhtanov EA, Gorn VV, Meyer RB Jr, Gamper HB Jr. *Biochemistry.* 1996; 35:11170–11176. [PubMed: 8780521]
- vii. Connolly BA, Newman PC. *Nucleic Acids Res.* 1989; 17:4957–4974. [PubMed: 2762115]
- viii. Lan T, McLaughlin LW. *J Am Chem Soc.* 2000; 122:6512–6513.
- ix. Ishizuka T, Yoshida J, Yamamoto Y, Sumaoka J, Tedeschi T, Corradini R, Sforza S, Komiyama M. *Nucl Acids Res.* 2008; 36:1464–1471. [PubMed: 18203747]
- x. Demidov VV, Protozanova E, Izvol'sky KI, Price C, Nielsen PE, Frank-Kamenetskii MD. *Proc Natl Acad Sci USA.* 2002; 99:5953–5958. [PubMed: 11972051]
- xi. Gamper HB, Gewirtz A, Edwards J, Hou YM. *Biochemistry.* 2004; 43:10224–10236. [PubMed: 15287750]
- xii. Lahoud G, Timoshchuk V, Lebedev A, de Vega M, Salas M, Arar K, Hou YM, Gamper H. *Nucleic Acids Res.* 2008; 36:3409–3419. [PubMed: 18448471]
- xiii. Vives M, Eritja R, Tauler R, Marquez VE, Gargallo R. *Biopolymers.* 2004; 73:27–43. [PubMed: 14691937]
- xiv. Kamiya H, et al. *Chem Pharm Bull.* 1992; 40:2792–2795. [PubMed: 1464111]
- xv. Nguyen HK, Bonfils E, Auffray P, Costaglioli P, Schmitt P, Asseline U, Durand M, Maurizot JC, Dupret D, Thuong NT. *Nucleic Acids Res.* 1998; 26:4249–4258. [PubMed: 9722646]
- xvi. Bommarito S, Peyret N, SantaLucia J. *Nucleic Acids Res.* 2000; 28:1929–1934. [PubMed: 10756193]
- xvii. Horlacher J, Hottiger M, Podust VN, Huebscher U, Benner SA. *Proc Natl Acad Sci USA.* 1995; 92:6329–6333. [PubMed: 7541538]
- xviii. Watson JD, Crick FHC. *Nature.* 1953; 171:737–738. [PubMed: 13054692]
- xix. Eschenmoser A. *Science.* 1999; 284:2118–2124. [PubMed: 10381870]

Figure 1.

Two generations of candidate nucleobases for self-avoiding molecular recognition systems (Samrs). (a) First generation Samrs candidates are simple implementations having 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 the first generation T* to adenine, and N⁴-ethylcytosine as C* to resolve issues arising from the chemical instability of the first generation C*.

Figure 2.

Amplification of the *Taq* gene demonstrating the ability of Samrs to manage PCR artifacts in a “worst case design” scenario, where the forward and reverse primers are formally complementary in their last nine nucleotides. Standard primer pairs give only primer dimers in these cases. Primer pairs having Samrs components in the 3′-segment, even as few as four, give the desired 1109 nucleotide amplicon. + and – indicate with and without target gene. A* = 2-aminopurine. G* = hypoxanthine. T* = 2-thiothymine. C* = N⁴-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′-CCAATGCCAACCTCTACCTCC*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 3.

PCR amplification of 10 cancer genes using primer pairs chosen to allow the amplicons to be conveniently separated by size on a 3% agarose gel, using sequences in Suppl. Table 7. (a) Singleplexed PCR with indicated primer pairs containing only standard nucleotides. (b) Attempted multiplexed PCR with indicated pairs containing only standard nucleotides. (c) Singleplexed PCR with indicated pairs containing $\{16+8^*+1\}$ standard-Samrs-standard primer pairs. (d) Multiplexed PCR with indicated pairs containing $\{16+8^*+1\}$ standard-Samrs-standard primer pairs.

Table 1

Melting temperatures (T_m s) for 5'-ACCAAGCXATCAAGT-3' and 3'-TGGTTCGYTAGTTCA-5'. Boxes with bold outline hold T_m s for complementary pairs, in two contexts. Doubly underlined T_m s are from duplexes having matched Samrs:Samrs pairs; these are lower than the corresponding singly underlined T_m s, which have Samrs:standard or standard:Samrs pairs. The underlined T_m s are similar to those having A:T and T:A pairs, also joined by 2 hydrogen bonds. Note also how the off-diagonal terms representing formal mismatches have lower T_m s than the N:N* pairs.

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