Published in final edited form as:

Angew Chem Int Ed Engl. 2010 ; 49(1): 177–180. doi:10.1002/anie.200905173.

Expanded Genetic Alphabets in the Polymerase Chain Reaction**

Zunyi Yang, Fei Chen, Stephen G. Chamberlin, and Steven A. Benner*

Foundation for Applied Molecular Evolution, 720 SW 2nd Avenue, Suite 201, Gainesville, FL 32601 (USA)

Abstract

Cleaning up polymerase chain reactions: Artificially expanded genetic information systems (AEGIS) add extra nucleotide "letters" to DNA alphabets; oligonucleotides containing AEGIS nucleotides do not bind to natural DNA. This "orthogonality" is exploited here by placing two AEGIS nucleotides (P and Z) in external tags for primers targeting three cancer genes in a nested PCR architecture. AEGIS tags support multiplexed PCR with fewer primer dimers and off-target amplicons than multiplexed PCR without AEGIS components.

Keywords

PCR; DNA replication; genetic alphabets; polymerases; nucleobases

The information content of natural DNA molecules is limited through its construction from four nucleotide building blocks, which form just two pairs (A:T and G:C). Considerable effort has been directed towards this content by developing additional nucleotides that might form additional pairs.^[1,2] One way that this can be done is by rearranging hydrogen bond donor and acceptor groups in the nucleobases to allow them to form to pairs having the same geometry as A:T and G:C pairs that are joined by different hydrogen-bond patterns (Figure 1). This can yield an artificially expanded genetic information system (AEGIS) that has as many as 12 nucleotides and as many as six base pairs.^[3]

AEGIS nucleotides already are being applied as "orthogonal" binding elements in human diagnostic tools, including in FDA-approved assays for HIV, hepatitis B and hepatitis C viruses.^[4] Here, AEGIS oligonucleotides to bind to each other and not to natural DNA, lowering background noise in blood and other complex biological assay mixtures.

To date, however, no practical tools have exploited additional nucleotides as components of *dynamic* assays, where artificial nucleotides are copied by polymerases in, for example, a polymerase chain reaction (PCR). Successful academic PCRs using expanded genetic alphabets either require a mesophilic reverse transcriptase,^[5] lose the nucleobase due to unfavorable tautomeric equilibria,^[6] generate highly sulfurated amplicons,^[7] or disfavor incorporation of adjacent non-standard bases.^[8]

If it could be implemented in a fully flexible form, "six letter PCR" should offer new approaches to analyze DNA. One challenge that has long avoided solution is multiplexed

**This work was supported by the US National Institutes of General Medical Sciences (R01GM081527), the National Human Genome Research Institute (for the bioinformatics portion; R01HG004647), and the Defense Threat Reduction Agency (through its program in basic research, HDTRA1-08-1-0052).

*Fax: (+1) 352-271-7076, sbenner@ffame.org, Website: <http://www.ffame.org>.

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

PCR, which is often complicated by off-target priming by primers presented at high concentrations in standard PCR architectures and the resulting undesired amplicons.^[9] Clever approaches to minimize such complications include various kinds of nested PCR^[10] and MegaPlex PCR.^[11] These architectures use composite primers having 3'-segments specific for their target and 5'-tags common to all primers (Figure 2a), where PCR is initiated with composite primers in low concentration and continued using high concentrations of the common external primers that, if designed carefully, bind nothing in the assay environment. If built from A, T, G, and C, however, this design is often difficult to achieve. Essentially every 11mer is present in the human genome, for example, meaning that essentially any sequence built from standard nucleotides will primer somewhere eventually.

Placing AEGIS nucleotides in the external tags could solve this problem. Because AEGIS-containing tags do not pair with DNA in *any* natural genomic background, they cannot primer to create amplicons through off-target priming regardless of how complex the assay environment becomes. This requires, however, polymerase-AEGIS combinations that support fully flexible nested PCR architectures.

To develop a fully flexible nested PCR with an artificially expanded genetic alphabet, a pyrimidine analog (6-amino-5-nitro-3-(1'-*beta*-D-2'-deoxyribofuranosyl)-2(1H)-pyridone, **Z**) and its complementary purine analog (2-amino-8-(1'-*beta*-D-2'-deoxyribofuranosyl)-imidazo[1,2-*a*]-1,3,5-triazin-4(8H)-one, **P**) (Figure 1, available as both triphosphates and protected phosphoramidites through reagents@ffame.org), were examined as extra nucleotide genetic "letters". Incorporated into DNA strands, these form **Z:P** pairs that contribute to duplex thermostability more than C:G pairs.^[12] Further, **Z** and **P** discriminate against mismatches as well as standard nucleobases,^[13] and are accepted by many polymerases.^[14]

To develop flexible nested PCR, we first addressed a concern that arises because many polymerases hold primers and templates together in their active sites sufficiently to extend the primers even when the primer-template duplex melts below the temperature used for primer extension. We deliberately designed two primers that might form dimers via **Z:G** and **P:C** mismatches (5'-CAGGAAGGAGCGATCGC-3', 5'-CAGGAAGGAGCPATCPC-3', and 5'-CAGGAAGGAGZGATZGC-3'). We then incubated these with *Taq* polymerase to see if they generated primer dimer products (Supplementary Figure 1 and 2). They did not, laying this concern to rest.

We then screened for polymerases that might accept PCR primers that contained multiple **Z**'s and **P**'s, including **Z**'s and **P**'s at adjacent positions. Most polymerases tested amplified oligonucleotides containing multiple **Z**'s and **P**'s (Supplementary Figure 3). In incubation mixtures lacking dZTP and dPTP, many of these paused or stopped entirely rather than misincorporate standard triphosphates opposite **Z** or **P**; others failed to elongate efficiently following such misincorporation.

These experiments identified several polymerases that incorporated dZTP and dPTP opposite their non-standard complements especially well and were also good at rejecting mismatches. These included Deep Vent (both *exo*⁺ and *exo*⁻), Vent (both *exo*⁺ and *exo*⁻), and *Taq* DNA polymerases (Supplementary Figure 3). Other polymerases, such as Phusion (a high-fidelity DNA polymerase), 9 °N, Herculase and many *Taq* mutants also efficiently incorporate dZTP and dPTP opposite multiple template **P**'s and **Z**'s (data not shown).

We then asked whether oligonucleotides containing **P** or **Z** might serve as external tags in an unplexed nested PCR using a synthetic target and *Taq* polymerases (Figure 2a). To this end, two types of experiments were designed. The first (Type A) had standard nucleotides in the external tags of the composite primers (F-External-Std, R-External-Std, F-Composite-Std,

and R-Composite-Std, Table 1). The second (Type B) had four **P**'s in the external tags and composite primers replacing four G's (F-External-4P, R-External-4P, F-Composite-4P, and R-Composite-4P, Table 1). Initial PCR amplicons obtained with these composite primers in low concentration (2.5 nM) were then further amplified with external primers in high concentration (250 nM), where the external primers were either matched to the tags used in the composition primers (these are the A-Std and B-AEGIS experiments) or were mismatched to the tags (these are the A-AEGIS and B-Std experiments). Both the A-Std (standard nested PCR, lanes 1 to 4) and B-AEGIS (P-containing nested PCR, lanes 13 to 16) experiments gave desired PCR amplicons (Figure 2b), showing that **P**-containing external and composite primers support nested PCR. In contrast, nested PCR experiments that mismatched the external primers (standard or P-containing) with the incorrect composite primers (**P**-containing and standard, respectively) gave no PCR products (A-AEGIS, lanes 5 to 8, Figure 2b; B-Std, lanes 9 to 12, Figure 2b). This demonstrated the orthogonality of priming.

To further show that the **P**-containing external primers could not anneal and amplify standard mismatch semi-complementary sequences, we monitored the amplification in A-Std and A-AEGIS experiments using real-time PCR (Roche LightCycler® 480). The A-Std experiments showed significant enhancement of fluorescence signal (SYBR green) after 10 PCR cycles. In contrast, the A-AEGIS experiment gave no PCR amplicon under all annealing temperatures tested (50.4°C, 54.2°C, 58.9°C, and 63.9°C) even after 40 PCR cycles (data not shown). This clearly shows that four **P**:C mismatches in both forward and reverse external primers (F-External-4P and R-External-4P) prevent **P**-containing primers from annealing to **Z**-containing templates.

To show that nested PCR with **Z**- and **P**-containing primers can produce long amplicons, composite primers with and without **P** in their external tags were designed to amplify the 2.5 kb gene for *Taq* DNA polymerase itself. Standard nested PCR (Figure 3, lane 3) generated the desired amplicon (2.5 kb), but with significant amounts of an undesired amplicon (750 bp, lane 4). The undesired amplicon was unassigned, but may have arisen from off-target priming involving five or six 3'-terminal matches. In contrast, nested PCR exploiting **P**-containing primers (lane 6) produced clean products without undesired amplicons.

We then asked which polymerases amplified targets most efficiently in a nested PCR architecture with composite primers and **P** or **Z** in the external tags. As a metric for evaluating real-time PCR, the polymerase generating higher fluorescence (SYBR green) at an earlier cycle was awarded a higher efficiency score, the number of cycles required to reach 50% of the end point fluorescence. The efficiency of each polymerase was further confirmed using the intensity of PCR amplicon stained by ethidium bromide on agarose gel. *Taq*, Deep Vent (both exo^+ and exo^-), Vent (both exo^+ and exo^-), Phusion, 9 °N, and Herculanase were tested.

For both **P**- and **Z**-containing nested PCR, Deep Vent (exo^-) was the most efficient polymerase lacking exonuclease activity (data not shown). Phusion, however, had the highest efficiency among polymerases having proofreading activity. **P**-containing primers gave higher nested PCR yields than **Z**-containing tags and primers with these polymerases (data not shown). Therefore, we further developed the nested PCR system using **P**-containing primers and Phusion (Supporting Information, Supplementary Figure 4).

To demonstrate the possibility of downstream cloning of amplicons obtained with external primers containing **P**, an 81 nucleotide segment within the *Taq* gene was amplified with **P**-containing primers and Phusion. PCR products were cloned at the polylinker site of Vector pSC-B-amp/kan, and transformed (StrataClone Blunt PCR Cloning Kit) into *E. coli*. Blue-

white screening gave 50 colonies that were submitted for sequencing (BioBasic, Canada); 33 sequences were returned. No point mutation errors were found in the 1377 sites sequenced, placing an upper limit on the error in this PCR (< 0.1%).

PCR artifacts increase in multiplexed PCR as the number of amplicons and primers increases. We therefore tested whether nested multiplexed PCR with **P**-containing external primers might give cleaner products than with standard primers when applied to human genomic DNA (Promega, G147A). Here, composite primers having four **P**'s in their external tags were synthesized to target the three cancer-associated genes examined by Makrigiorgos *et al.* (TOP1, HBEGF, and MYC).^[15] Sequences of gene-specific primer segments of these composites were taken from this paper (Supplementary Table 1).

Three-plexed nested PCR with standard primers (Figure 4, left panel) gave the desired amplicons, but only with substantial amounts of undesired amplicons, again presumably from off-target priming with mismatches; a similarity search suggested candidate regions for this undesired priming. At lower annealing temperatures (from 53.4 to 58.8 °C), considerable primer dimer (about 40 bp) also appeared; at higher annealing temperatures (from 61.5 to 65.5 °C), primer dimers disappeared, while off-target PCR artifacts did not. In contrast, three-plexed nested PCR with **P**-containing primers produced the desired amplicons with negligible amounts of undesired products and no primer dimers were detected under any annealing temperatures tested (Figure 4, right panel).

Interestingly, in controls *lacking* genomic DNA, 45 PCR cycles with standard primers gave primer dimers and considerable amounts of long PCR artifacts, presumably arising from further priming from dimers (Figure 4, far left). However, the same 45 PCR cycles with **P** in the external and composite primers gave primer dimer (chimeric product) *without further artifacts* (Figure 4, far right). Further, increasing the number of PCR cycles led, with standard primers, to a "PCR mess"; this was not seen with **P**-containing external primers, even after 55 PCR cycles (Supplementary Figure 5).

While we expected AEGIS to improve PCR cleanliness, we did not expect improvement to this extent. Interestingly, external tags containing **P** still produced better results when the analyte-specific sequences designed by Makrigiorgos *et al.* were replaced by external primers carefully designed by genomic analysis to bind *nowhere* in the human genome (Supplemental Table 2). Even here, standard external primers gave off-target priming (Supplementary Figure 6).

We believe that nested PCR performs better with **P** in the external primers because (a) **P**-containing external primers cannot form primer dimers with **P**:C mismatches, even at high concentrations; (b) **P**-containing external primers cannot touch down elsewhere in genomic DNA to produce off-target PCR amplicons; and (c) all PCR amplicons end up having a **P**-containing sequence at their 5'-ends and a **Z**-containing sequence at 3'-ends; therefore, even when **P**-containing composite primers *do* produce unwanted species, they do not produce further PCR artifacts by binding to other components in the assay mixture.

The orthogonality of AEGIS components should improve the performance of other architectures that detect DNA in dynamic assays. Further, synthetic genetic alphabets other than **P** and **Z** are now becoming available in other laboratories.^[1,8] We expect that several of these alphabets will also eventually be used to clean up multiplexed PCR. By providing the first example of a fully flexible nested PCR, however, the work described here should encourage their application of all artificial genetic alphabets to this goal, especially now that it is clear that AEGIS orthogonality substantially improves the quality of products produced by multiplexed PCR. This sets the stage for the next challenges, which include

demonstrating that multiple AEGIS components can be placed within the amplicon as well as in PCR tags.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Reference

1. a) Krueger AT, Lu HG, Lee AHF, Kool ET. *Acc. Chem. Res.* 2007; 40:141–150. [PubMed: 17309194] b) Hirao I. *Curr. Opin. Chem. Biol.* 2006; 10:622–627. [PubMed: 17035074] c) Henry AA, Romesberg FE. *Curr. Opin. Chem. Biol.* 2003; 7:727–733. [PubMed: 14644182] d) Kim Y, Leconte AM, Hari Y, Romesberg FE. *Angew Chem Int Ed Engl.* 2006; 45:7809–7812. [PubMed: 17075934] e) Sintim HO, Kool ET. *Angew Chem Int Ed Engl.* 2006; 45:1974–1979. [PubMed: 16506248] f) Leconte AM, Matsuda S, Hwang GT, Romesberg FE. *Angew Chem Int Ed Engl.* 2006; 45:4326–4329. [PubMed: 16733840] g) Yu C, Henr AA, Romesberg FE, Schultz PG. *Angew Chem Int Ed Engl.* 2002; 41:3841–3844. [PubMed: 12386865]
2. Benner SA. *Accounts. Chem. Res.* 2004; 37:784–797.
3. a) Piccirilli JA, Krauch T, Moroney SE, Benner SA. *Nature.* 1990; 343:33–37. [PubMed: 1688644] b) Bain JD, Switzer C, Chamberlin AR, Benner SA. *Nature.* 1992; 356:537–539. [PubMed: 1560827]
4. a) Gleaves CA, Welle J, Campbell M, Elbeik T, Ng V, Taylor PE, Kuramoto K, Aceituno S, Lewalski E, Joppa B, Sawyer L, Schaper C, McNairn D, Quinn T. *J. Clin. Virol.* 2002; 25:205–216. [PubMed: 12367656] b) Elbeik T, Surtihadi J, Destree M, Gorlin J, Holodniy M, Jortani SA, Kuramoto K, Ng V, Valdes R, Valsamakis A, Terrault NA. *J. Clin. Microbiol.* 2004; 42:563–569. [PubMed: 14766817]
5. Sismour AM, Lutz, S. S, Park J-H, Lutz MJ, Boyer PL, Hughes SH, Benner, S. A. *SA. Nucl. Acids. Res.* 2004; 32:728–735. [PubMed: 14757837]
6. Johnson SC, Sherrill CB, Marshall DJ, Moser MJ, Prudent JR. *Nucleic Acids Res.* 2004; 32:1937–1941. [PubMed: 15051811]
7. Sismour AM, Benner SA. *Nucl. Acids Res.* 2005; 33:5640–5646. [PubMed: 16192575]
8. a) Hirao I, Mitsui T, Kimoto M, Yokoyama A. *J. Am. Chem. Soc.* 2007; 129:15549–15555. [PubMed: 18027940] b) Kimoto M, Kawai R, Mitsui T, Yokoyama S, Hirao I. *Nucleic Acids Res.* 2009; 37:e14. [PubMed: 19073696]
9. Fan JB, Chee MS, Gunderson KL. *Nat. Rev. Genet.* 2006; 7:632–644. [PubMed: 16847463]
10. a) Brownie J, Shawcross S, Theaker J, Whitcombe D, Ferrie R, Newton C, Little S. *Nucleic Acids Res.* 1997; 25:3235–3241. [PubMed: 9241236] b) Lin ZW, Cui XF, Li HH. *Proc. Natl. Acad. Sci. USA.* 1996; 93:2582–2587. [PubMed: 8637917]
11. Meuzelaar LS, Lancaster O, Pasche JP, Kopal G, Brookes AJ. *Nat. Methods.* 2007; 4:835–837. [PubMed: 17873887]
12. Yang ZY, Hutter D, Sheng PP, Sismour AM, Benner SA. *Nucleic Acids Res.* 2006; 34:6095–6101. [PubMed: 17074747]
13. Sheng PP, Yang ZY, Kim YM, Wu YR, Tan WH, Benner SA. *Chem. Commun.* 2008; 41:5128–5130.
14. Yang ZY, Sismour AM, Sheng PP, Puskar NL, Benner SA. *Nucleic Acids Res.* 2007; 35:4238–4249. [PubMed: 17576683]
15. Li J, Makrigiorgos GM. *Nat. Protocols.* 2007; 2:50–58.

Figure 1. Comparison of the standard **C:G** pair with the non-standard **Z:P** pair forming an extra two genetic letters in an artificially expanded genetic information system (AEGIS).

Figure 2.

Nested PCR with standard and AEGIS components using *Taq* DNA polymerase at four annealing temperatures (50.4, 54.2, 58.9, and 63.9 °C). The initial concentration of template DNA is 0.25 nM.

a) Nested PCR with different combinations of external and composite primers.

Type A nested PCR: Target first amplified by standard composite primers at low concentrations, and then standard external primers (A-Std) or AEGIS external primers (A-AEGIS) at high concentrations.

Type B nested PCR: Target first amplified by AEGIS composite primers at low concentrations and then standard (B-Std) or AEGIS external primers (B-AEGIS).

b) Autoradiograph of a PAGE gel resolving nested PCR products.

Lanes 1 to 4 (A-Std): Products from composite primers and external primers both without AEGIS.

Lanes 5 to 8 (A-AEGIS): Absence of products with standard composite primers and AEGIS external primers.

Lanes 9 to 12 (B-Std): Absence of products with AEGIS composite primers and standard external primers.

Lanes 13 to 16 (B-AEGIS). Products from composite and external primers both with AEGIS.

Figure 3.

Products from nested PCR amplification of *Taq* gene using standard or AEGIS-containing primers.

Lane 1: Hi-LO™ DNA marker 50bp – 10,000pb (Bionexus).

Lane 2: Control without target; composite and external primers without AEGIS.

Lane 3: Standard nested PCR with standard primers.

Lane 4: Control lacking composite primers, but with standard external primers and target.

Lane 5: Control without target; composite and external primers with AEGIS.

Lane 6: Standard nested PCR with AEGIS primers.

Lane 7: Control lacking composite primers, but with AEGIS external primers and target.

Figure 4.

Seven parallel nested PCR amplifications of three genes (TOP1, HBEGF, and MYC) from human genomic DNA (Phusion high fidelity DNA polymerase) under seven different annealing temperatures (53.4, 54.6, 56.5, 58.8, 61.5, 63.8 and 65.5 °C).

Left: Amplification with standard external and composite primers.

Right: Amplification using **P**-containing external and composite primers.

Controls A (for standard nested PCR) and B (for **P**-containing nested PCR) were performed at 58.8°C without genomic DNA.

Table 1

Oligonucleotides used in Figure 2 and 3.

F-External-Std	5`-CTAGGACGACGGACTGC-3`	
F-Composite-Std	5`-CTAGGACGACGGACTGCCCATGGGAGACCGCGGT-3`	
R-Template	3`-GGTACCCTCTGGCGCCACCCGGGCGGCCCATGGTAGCTATGCGCAA-5`	
F-Template	5`-CCATGGGAGACCGCGGTGGGCCCCGGCCGGGTACCATCGATACGCGTT-3`	
R-Composite-Std		3`-CCATGGTAGCTATGCGCAACGCTAGCGAGGGAAGGAC-5`
R-External-Std		3`-CGCTAGCGAGGGAAGGAC-5`
F-External-4P	5`-CTAPGACPACGPACTPC-3`	
F-Composite-4P	5`-CTAPGACPACGPACTPCCCATGGGAGACCGCGGT-3`	
R-Composite-4P		3`-CCATGGTAGCTATGCGCAACPCTAPCGAPGAAPGAC-5`
R-External-4P		3`-CPCTAPCGAPGAAPGAC-5`