BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

# Directed Evolution of Polymerases To Accept Nucleotides with Nonstandard Hydrogen Bond Patterns

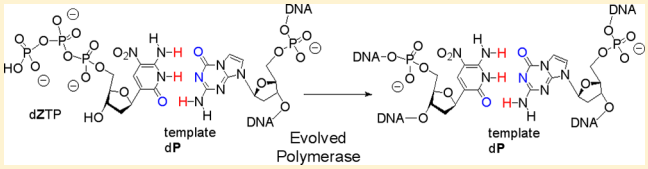
Roberto Laos,[†] Ryan Shaw,[†] Nicole A. Leal,[†] Eric Gaucher,[‡] and Steven Benner*[,†]

[†]Foundation for Applied Molecular Evolution and The Westheimer Institute of Science and Technology, Gainesville, Florida 32601, United States

[‡]School of Biology, School of Chemistry, and Parker H. Petit Institute for Bioengineering and Biosciences, Georgia Institute of Technology, Atlanta, Georgia 30332, United States

Supporting Information

**ABSTRACT:** Artificial genetic systems have been developed by synthetic biologists over the past two decades to include additional nucleotides that form additional nucleobase pairs independent of the standard T:A and C:G pairs. Their use in various tools to detect and analyze DNA and RNA requires polymerases that synthesize duplex DNA containing unnatural base pairs. This is especially true for nested polymerase chain reaction (PCR), which has been shown to dramatically lower noise in multiplexed nested PCR if nonstandard nucleotides are used in their external primers. We report here the results of a directed evolution experiment seeking variants of *Taq* DNA polymerase that can support the nested PCR amplification with external primers containing two particular nonstandard nucleotides, 2-amino-8-(1′-β-D-2′-deoxyribofuranosyl)imidazo[1,2-*a*]-1,3,5-triazin-4(8*H*)-one (trivially called **P**) that pairs with 6-amino-5-nitro-3-(1′-β-D-2′-deoxyribofuranosyl)-2(1*H*)-pyridone (trivially called **Z**). Variants emerging from the directed evolution experiments were shown to pause less when challenged *in vitro* to incorporate d**Z**TP opposite **P** in a template. Interestingly, several sites involved in the adaptation of *Taq* polymerases in the laboratory were also found to have displayed "heterotachy" (different rates of change) in their natural history, suggesting that these sites were involved in an adaptive change in natural polymerase evolution. Also remarkably, the polymerases evolved to be less able to incorporate d**P**TP opposite **Z** in the template, something that was not selected. In addition to being useful in certain assay architectures, this result underscores the general rule in directed evolution that "you get what you select for".




One of the many outcomes of synthetic biology over the past two decades has been the generation of alternative versions of nucleic acids having additional nucleotide "letters" that form additional nucleobase pairs.[1−11] Our laboratories have constructed an artificially expanded genetic information system (AEGIS).[10,12] The AEGIS is a species of DNA having additional nucleobases that present alternative hydrogen bonding patterns, allowing its members to form mutually exclusive pairing schemes. Some members of the AEGIS are used in the clinic to monitor the viral load of patients infected with human immunodeficiency virus and hepatitis C virus.[13] The AEGIS base pairs are formed "orthogonally", meaning that pairing between additional nucleotides and standard nucleotides contributes little (or negatively) to the stability of the duplex. The extra base pairs become even more important when DNA polymerases that synthesize them in duplex DNA products arising from template-directed polymerization or, if repeated, polymerase chain reaction (PCR) can be found.

One of the AEGIS base pairs is formed between 2-amino-8-(1′-β-D-2′-deoxyribofuranosyl)imidazo[1,2-*a*]-1,3,5-triazin-4(8*H*)-one (trivially named **P**) and 6-amino-5-nitro-3-(1′-β-D-2′-deoxyribofuranosyl)-2(1*H*)-pyridone (trivially named **Z**) (Figure 1). The **Z**:**P** pair has a standard Watson−Crick geometry joined by three hydrogen bonds, differing from the standard C:G pair in the arrangement of donor and acceptor groups that form the connecting hydrogen bonds. Further, both nucleobases place electron density into the minor groove, a density that can accept a hydrogen bond from a polymerase[10] (Figure 1). These features allow polymerases to accept d**Z**TP and d**P**TP as substrates to form duplexes containing **Z**:**P** pairs in primer extension reactions, PCR, and nested PCR (Figure 2) architectures.

We recently showed that **Z**:**P** pairs can be exploited in nested PCR.[17] Here, nested PCR (Figure 2) begins with "chimeric" primers, which are presented in the PCR mixture at low concentrations. The analyte-specific portions of the chimeric primers are built from only standard nucleotides, necessary to complement sequences within the natural PCR target. The primers are, however, tagged with DNA sequences that contain multiple **P**'s. The small amounts of chimeric primers are rapidly consumed in the initial rounds of PCR, requiring the external primers to later "carry" the amplification. Because the external primers must be present at high concentrations for amplification to occur, these can bind off-target to create

**Received:** May 2, 2013
**Revised:** June 20, 2013
**Published:** July 1, 2013

   dx.doi.org/10.1021/bi400558c | *Biochemistry* 2013, 52, 5288−5294

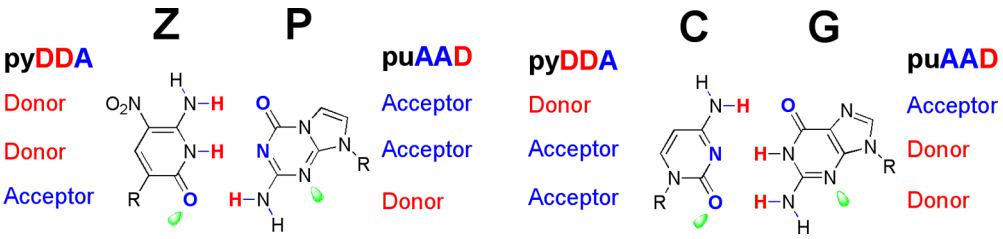



**Figure 1.** **Z**:**P** pair from an artificially expanded genetic information system (AEGIS, left); note the green lobes indicating electron density presented to the minor groove of DNA, electron density that polymerases seek. The natural **C**:**G** pair (right) also has this electron density.

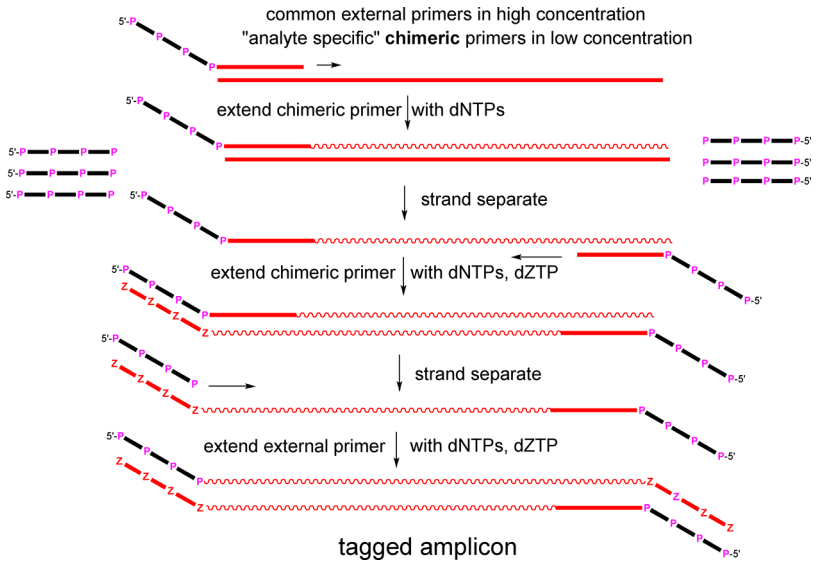



**Figure 2.** Nested PCR involving **Z**:**P** pairs begins with chimeric primers presented at low concentrations. The analyte-specific portion of the chimeric primer is built from only standard nucleotides, to complement the sequence of the natural target, which also contains only standard nucleotides. The tag, however, contains multiple **P**'s (here, four). After the chimeric primers are consumed in the initial rounds of PCR, the amplification is "carried" by external primers present at high concentrations that have four **P**'s in a primer 17 nucleotides in length. This external primer cannot bind anywhere to natural DNA, despite its high concentrations. This allows the nested PCR architecture using **Z**:**P** pairs to be extremely clean relative to that of nested PCR that uses external primers made from standard nucleotides only.

noise. However, if the external primers have multiple **P**'s, they cannot bind anywhere to natural DNA, despite their high concentrations. This allows nested PCR architectures using **Z**:**P** pairs to be extremely clean, when compared with that of nested PCR that uses external primers made from standard nucleotides only.[14]

Unfortunately, natural polymerases still see something slightly "unnatural" in **Z**:**P** pairs, despite their structural similarities to standard pairs. Speculatively, this might arise from the fact that **Z** is a "C-glycoside", having its heterocycle joined to the sugar via a carbon−carbon bond. Natural nucleotides have their heterocycles joined to their sugars via carbon−nitrogen bonds.

To improve polymerases that can support nested PCR, we began by noting that the nested PCR architecture is easily incorporated into directed evolution experiments using compartmentalized self-replication (CSR) procedures. As applied to polymerases by Holliger and co-workers,[15] CSR uses water droplets emulsified in oil as artificial cells.[16] Each droplet contains a single *Escherichia coli* bacterium expressing one member of a library of variant polymerase genes. PCR is done within the emulsion, where the first heat step releases to the droplet (which contains PCR primers) each variant polymerase as well as the plasmid that encodes it. If the variant polymerase is active under the condition of the PCR, it amplifies its own gene. This enriches the emulsion polymerases that are most "fit" to support PCR in the environment presented (Figure 3).

We report the application of CSR in obtaining *Taq* DNA polymerase variants that better incorporate d**Z**TP opposite **P** in a template. Here, the nested PCR[17] was reproduced inside of the emulsified droplets that contained chimeric primers and external primers containing **P** (Figure 4). Polymerase variants that are inactive do not deliver their genes in amplified form to the extract after the emulsion is broken. Variants that skip, mismatch, or inefficiently place **Z** opposite **P** are expected to deliver less of their encoding genes, as their products are less efficiently primed by the **P**-containing primers. However, genes for variants that synthesize **Z**:**P** pairs in duplex DNA more rapidly, with greater fidelity, or otherwise more efficiently should be most enriched at the end of the emulsion-based PCR.

## EXPERIMENTAL PROCEDURES

**Synthesis of Nonstandard Nucleotides and Oligonucleotides.** Nonstandard nucleotides 2-amino-8-(1′-*β*-D-2′-deoxyribofuranosyl)imidazo[1,2-*a*]-1,3,5-triazin-4(8*H*)-one (d**P**) that pairs with 6-amino-5-nitro-3-(1′-*β*-D-2′-deoxyribofuranosyl)-2(1*H*)-pyridone (d**Z**) were prepared in our laboratory as the phosphoramidite and triphosphate forms using previously published methods.[18−20] Oligonucleotides contain-


dx.doi.org/10.1021/bi400558c | *Biochemistry* 2013, 52, 5288−5294

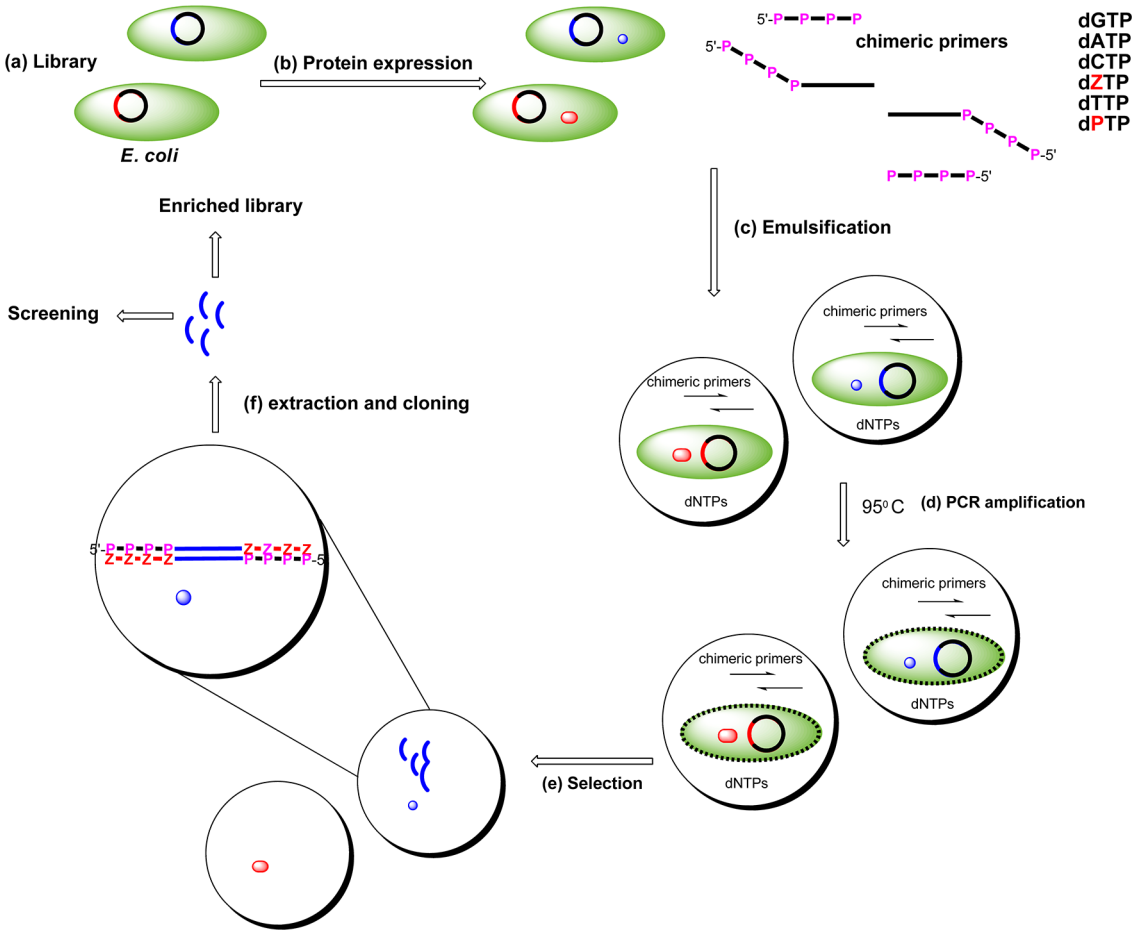



**Figure 3.** Compartmentalized self-replication (CSR) experiments start with the creation of a library of genes encoding variants of a polymerase (a). Here, we show the two variant genes (red and blue). The actual experiment contains ca. $50 \times 10^6$ *E. coli* cells each potentially carrying a different variant of polymerase. Each *E. coli* expresses the mutant polymerase gene (b). Then, the cells are emulsified along with the deoxynucleotide triphosphates (dNTPs) and chimeric primers. The resulting water-in-oil-emulsion contains ideally one cell per droplet (c). The first cycle of PCR breaks the cell wall of *E. coli*, exposing the expressed polymerase molecules and their gene to the contents of a water droplet containing all of the necessary components for a PCR amplification: (i) primers, (ii) dNTPs, (iii) a mutated gene of the polymerase, and (iv) the enzyme expressed by this gene (d). The droplets are heat resistant and maintain the original genotype−phenotype linkage present in *E. coli* cells. During PCR, active polymerases (blue) amplify their respective genes, enriching the pool of mutants having the desired properties; inactive polymerases (red) fail to do so (e). The emulsion is then broken, and the amplified genes enriched in those encoding polymerases having the desired behaviors are extracted and inserted into a plasmid vector to be expressed and screened or can enter in a new cycle of selection (a).

ing only the four natural nucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Oligonucleotides containing **Z** and **P** were prepared in our laboratories using an Applied Biosystems ABI394 DNA/RNA synthesizer.

**Compartmentalized Self-Replication (CSR).** An experimental procedure based on water-in-oil emulsions adapted from that developed in the Holliger laboratory[15] was followed. The aqueous phase of the emulsion {1× emulsion buffer [2 mM Tris-HCl, 2 mM $(NH_4)_2SO_4$, and 0.25 mM $MgCl_2$ (pH 8.8)]} contained two chimeric PCR primers (0.2 pmol each) whose 3′-ends were composed of standard nucleotides complementary to a region of the gene for *Taq* DNA polymerase and whose 5′-ends contained four nonconsecutive d**P** nucleotides. The forward chimeric primer was 5′-CTA**P**GACP**A**CG**P**ACT**P**CCAGGAAGCAGCCATCACACAAATGGCTAGCGGGACCATGGG-3′; the reverse chimeric primer was 5′-CAG**P**AAG**P**AGC**P**ATC**P**CCAGGAAGCAGCCATCACGTCCCGAATTCGAGCTCGGTACCT-3′. When presented with the *Taq* DNA polymerase gene, these chimeric primers initiated a nested PCR, which was supported by two d**P**-containing external primers (5′-CTA**P**GAC**P**ACG**P**ACT-**P**C-3′ and 5′-CAG**P**AAG**P**AGC**P**ATC**P**C-3′, 20 pmol each). Other components of the aqueous phase of the emulsion were dNTPs (0.2 mM each), d**Z**TP (0.02 mM), and betaine (0.3 M), in a final volume of 100 μL.

The gene for *Taq* DNA polymerase does not, of course, contain any **Z** or **P**. Therefore, we designed a selection system in which the primers contain AEGIS components and the polymerase needs extend the primers first to copy its own gene. The primers present a nested configuration with the longer primers being present at low concentrations so they are used up in the first cycles of the PCR. The later PCR cycles use the shorter primers, which would anneal only to copies that have been correctly extended by placing **Z** opposite **P**.

*E. coli* cells (~50 million), each harboring a member of a library of *Taq* genes, were induced by addition of anhydrotetracycline as described previously,[15] resuspended in the aqueous solution (100 μL), and added dropwise to 200 μL of a solution of 3% Arlacel P135 and 0.1% soy lecithin in light mineral oil with constant stirring at 1000 rpm (7 min, room


dx.doi.org/10.1021/bi400558c | *Biochemistry* 2013, 52, 5288−5294

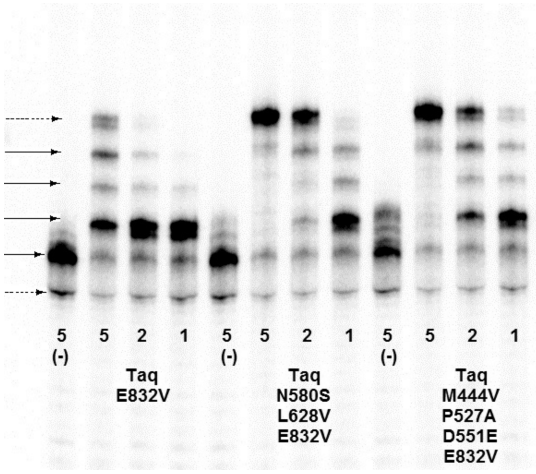



**Figure 4.** Acrylamide gel (20% with 7 M urea) showing the products of the extension by three *Taq* polymerase variants of a primer on a template with four **P**'s. Solid arrows indicate sites where d**Z**TP is incorporated opposite template d**P**. Dotted arrows show the labeled primer (bottom) and fully extended product (top, extended sequence **Z**GAT**Z**GCT**Z**CTT**Z**CTG). Numbers below lanes indicate picomoles of polymerase used. Experiments without d**Z**TP are indicated by dashes. Note the diminishing level of pausing with increasing numbers of mutations from compartmentalized self-replication. The reaction volume was 20 $\mu$L, and the reaction used 2.6 pmol of template, 1.7 pmol of primer, 0.33 pmol of $^{32}$P-labeled primer; 0.02 mM dNTPs, and 0.02 mM d**Z**TP where indicated. The primer/template was 5′-AGGTACCGAGCTCGAATTCGGGACGTGATGGCTGCTTCCT-GG-3′/3′-TCCATGGCTCGAGCTTAAGCCCTGCACTACCGAC-GAAGGACC**P**CTA**P**CGA**P**GAA**P**GAC-5′.

temperature). The emulsions were then placed in a PCR instrument (MJ thermal cycler) and subjected to thermocycling: 94 °C for 30 s; 30 cycles of 94 °C for 30 s, 63 °C for 30 s, and 72 °C for 1 min; and 72 °C for 10 min.

After thermocycling, the emulsions were broken with a phenol/chloroform mixture (200 $\mu$L). The DNA was then recovered using a Promega Wizard PCR cleanup system. The recovered PCR products were further amplified in a secondary PCR using the high-fidelity Pfu UltraII (Stratagene) polymerase and 5′-CAGGAAGCAGCCATCAC-3′ as a primer. The products of the secondary PCR were then digested and cloned back into the original vector (a derivative of pASK43IBAplus) using the NcoI and EcoRI restriction sites. The ligated plasmids were then used to transform electrocompetent *E. coli* BL21(DE3)gold cells, and transformants were grown on plates containing carbenicillin. Colonies were randomly selected from plates for subsequent screening (see below).

**Library Creation.** The library of *Taq* mutants was created by random mutagenesis of a codon-optimized *Taq* gene lacking the exonuclease domain. This is analogous to the Stoffel fragment, in which the first 289 amino acids are deleted from the *Taq* polymerase protein, removing the 5′−4′ exonuclease domain. The variants used here have the first 279 amino acids deleted. Additionally, the O-helix region was mutated by PCR using degenerate primers.

**Screening.** Transformants obtained following CSR were screened for polymerase activity by using real-time PCR in a multiwell plate directly on samples of individual colonies. Activities were tested with standard primers and nucleotides using SYBR green to detect PCR products (for details, see the Supporting Information). Transformants that showed activity

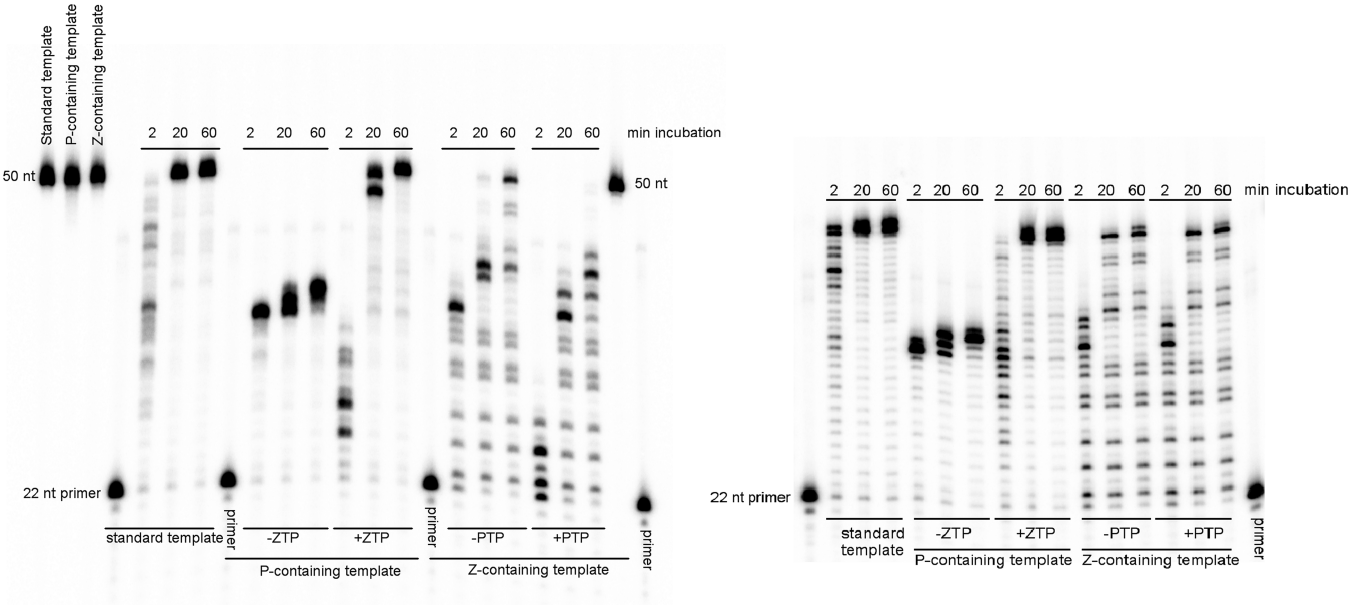



**Figure 5.** Acrylamide gel (10% with 7 M urea) showing the products of primer extension using variant N580S/L628V/E832V. The reaction volume was 20 $\mu$L with 4 pmol of primer and 6 pmol of template. When the polymerase is presented as a **P**-containing template in the absence of d**Z**TP, extension is stopped and no full-length product (50 nucleotides) is produced, even after incubation for 60 min. This shows that the variant does not mismatch any standard triphosphate opposite template d**P**. After d**Z**TP is added, the primer is extended through the d**P** to give the full-length product. In contrast, with a template containing d**Z**, primer extension does not stop at the d**Z** site in the absence of d**P**TP, indicating that the variant does mismatch one (or more) standard triphosphates opposite template d**Z**. Adding d**P**TP diminishes the level of primer extension involving mismatching, suggesting that d**P**TP occupies the triphosphate binding site when d**Z** is in the template, competing with mismatching, but is either not added or not extended. Parallel experiments with lower concentrations of d**P**TP confirm this. Here, the pattern of extended products involving misincorporation opposite d**Z** is not altered by the presence of small amounts of d**P**TP.


dx.doi.org/10.1021/bi400558c | *Biochemistry* 2013, 52, 5288−5294

were used to grow liquid cultures from which plasmids and purified enzymes were obtained. The *Taq* genes in the plasmids were sequenced, and the enzymes expressed by those genes were purified by heat shock as described by Grimm and Arbuthnot.[21]

**Primer Extension Experiments.** The primer/template used was 5′-AGGTACCGAGCTCGAATTCGGGACGTGAT-GGCTGCTTCCTGG-3′/3′-TCCATGGCTCGAGCTTAAG-CCCTGCACTACCGACGAAGGACC**P**CTA**P**CGA**P**GAA**P**-GAC-5′. The reaction mixture had a volume of 20 $\mu$L and contained 20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton X-100 (pH 8.8) 2.6 pmol of template, 1.7 pmol of primer, 0.33 pmol of $^{32}$P-labeled primer, 0.02 mM dNTPs, and 0.02 mM d**Z**TP (Figure 4). The mixture was heated to 95 °C (5 min), cooled to room temperature (0.1 °C/s), and placed in a cold block at 0 °C before *Taq* Pol was added at different enzyme concentrations. The tube was heated at 75 °C for 90 s and then the reaction quenched when the tube was placed on a cold block at 0 °C and 3 volumes of loading buffer (90% formamide, 20 mM EDTA, and 1 mg/mL bromophenol blue/xylene cyanol) was added. Finally, 5 $\mu$L of the quenched sample was loaded in a 7 M urea/acrylamide mixture.

In a parallel experiment with standard nucleotides (Figure 5), the following primer/template pairs were used: primer/standard template, 5′-GCGTAATACGACTCACTATAGG-3′/3′-CGCATTATGCTGAGTGATATCCGTCTCTCCTTC-TTCATGCTGTCCGTTCG-5′; primer/AEGIS (**P**) template, 5′-GCGTAATACGACTCACTATAGG-3′/3′-CGCATTATG-CTGAGTGATATCCGTCTCTCCTTCTT**PAPGP**TGTCC-GTTCG-5′; primer/AEGIS (**Z**) template, 5′-GCGTAATAC-GACTCACTATAGG-3′/3′-CGCATTATGCTGAGTGATA-TCCGTCTCTCCTTCTT**ZAZGZ**TGTCCGTTCG-5′.

The reaction mixtures (20 $\mu$L reaction volume) contained $\gamma$-$^{32}$P-labeled primer/cold primer (4 pmol) and standard template, **P**-containing template, or **Z**-containing template (6 pmol). These were annealed by incubation at 80 °C for 1 min and cooled to room temperature. Assays contained the primer/template complex, 20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton X-100 (pH 8.0), and 100 $\mu$M dNTP with and without dZTP or dPTP. Reactions were then initiated by the addition of 2.5 $\mu$mol of enzyme variant N580S/L628V/E832V (0.125 $\mu$M), and mixtures were incubated at 50 °C for 2, 20, and 60 min. Reactions were then quenched with 10 mM EDTA in formamide, and mixtures were resolved via 10% polyacrylamide gel electrophoresis.

## RESULTS

CSR exploits a feedback loop in which a polymerase copies the gene that encodes it. The gene for *Taq* DNA polymerase does not, of course, contain any **Z** or **P**. Therefore, to couple the selective pressure experienced by the polymerase to its ability to incorporate d**Z**TP opposite a d**P** in a template, we designed a selection system in which the primer binding sites containing d**Z** were added to the *Taq* gene early in the PCR with scarce chimeric primers and external primers at higher concentrations in a nested PCR[22] architecture that contained four non-consecutive d**P**'s, ca. 23.5% of their nucleotides.

Because they were present only at low concentrations, the chimeric primers were consumed in the first cycles of the PCR. Thus, a *Taq* DNA polymerase variant was further amplified only if it successfully incorporated four d**Z**TPs opposite four d**P**'s when it reads the d**P**-containing tag. Any *Taq* polymerase variant that omitted the d**Z**'s or mismatched some other nucleotide opposite the four template d**P**'s primed less efficiently in the next round of the PCR and therefore was depleted in the mixture collected after the emulsion is broken.

After 96 colonies harboring different polymerases had been screened, 13 were shown to have active polymerases. Four of the most active were sequenced: variant R334C/D578N/E832V, variant T447N/M673I/E832V, variant M444V/P527A/D551E/E832V, and variant N580S/L628V/E832V (see the Supporting Information for a comparison among these polymerases).

The E832V replacement had been previously identified as being beneficial for **Z**:**P** incorporation (unpublished). The variants recovered in these CSR experiments conserve this beneficial amino acid replacement and also have some additional replacements. Variants M444V/P527A/D551E/E832V and N580S/L628V/E832V showed more activity in the primer extension experiments reported here.

The emergence of these variants from the CSR experiment is, of course, evidence that their amino acid replacements allow the polymerase to survive better under the specific PCR conditions that they saw during CSR. It was interesting, however, to see whether the recovered polymerase variants would also display improvements in *in vitro* assays where they were challenged to incorporate **Z**TP opposite **P** in a template. To this end, primer extension experiments were performed in which variants M444V/P527A/D551E/E832V and N580S/L628V/E832V were challenged to incorporate d**Z**TP opposite **P** in four positions in the template.

Figure 4 shows that the polymerases containing the E832V replacement alone noticeably pause at each point in the primer extension reaction where **Z** must be incorporated. The "pausing bands" on the gel are less intense if more polymerase is added, of course. However, the intensity of those bands can be used as a metric to indicate the relative ability of polymerase variants to incorporate **Z** opposite **P** in a template.

Perhaps as expected as the products of a directed evolution experiment under this selection pressure, both variants M444V/P527A/D551E/E832V and N580S/L628V/E832V give less pausing in these experiments. Indeed, M444V/P527A/D551E/E832V gives only very light pausing bands when 5 pmol of polymerase is present. Thus, in addition to being active survivors of the CSR experiment, these variants show an enhanced facility to synthesize duplex DNA containing **Z**:**P** pairs.

## DISCUSSION

CSR is, of course, a process intended to solve a technological problem: the creation of polymerases that do useful things. This represents the first time that CSR has been applied to obtain polymerases able to accept components of an expanded genetic alphabet. It differs from recent results obtained by the Holliger laboratory,[23] where the information in the artificial genetic system remained the same, with only the backbone structures changing.

CSR results can also address scientific issues by providing a glimpse into the "fitness landscape" that relates protein function to protein sequence "space". For example, we might compare the number and distribution of amino acid replacements that appeared in this selection experiment with the number and distribution of replacements that emerged in other selection experiments with polymerases. For example, when selecting for *Taq* polymerase variants resistant to inhibition by heparin,

**Table 1. Summary of Mutations Found in Improved *Taq* Variants**

| mutation | nearest homologue with indicated mutation | crystal structure number | overall variability | heterotachy? |
|---|---|---|---|---|
| N580S | trypanosome/*Caenorhabditis elegans* | 580 | 6 amino acids represented | heterotachous;[a] conserved in eubacteria and their viruses; variable in eukaryotes |
| L628V | none | 628 | 10 amino acids represented | heterotachous;[a] conserved in Gram-negative bacteria; variable in metazoans; Gram-positive |
| E832V | none | absent | 7 amino acids represented | heterotachous;[a] conserved Arg in standard eukaryotes; variable in eubacteria |
| M444V | *Salmonella* | 444 | | no analysis |
| P527A | T3 and T7 virus, also mammals | 527 | 7 amino acids represented | no; frequent Pro-Glu homoplasy |
| D551E | Gram-positive bacteria | 551 | 11 amino acids represented | heterotachous; conserved Gln in bilateral and fungi; elsewhere variable |
| E832V | none | absent | 7 amino acids represented | heterotachous; conserved Arg in standard eukaryotes; variable in eubacteria |

[a]Further comments are provided in the Supporting Information.

Holliger and his group mentioned one survivor having six replacements (K225E/E388V/K540R/D578G/N583S/M747R).[15] When selecting for variants that accepted 2′-deoxycytidine derivatives carrying appended Cy3 and Cy5 fluorescent dyes, they recovered variants of *Pfu* DNA polymerase each having two to six amino acid replacements.[24] The four amino acid replacements seen in both the M444V/P527A/D551E/E832V and N580S/L628V/E832V *Taq* variants are comparable in quantity to these variants.

Alternatively, we can ask whether the specific sites undergoing replacement in our experiments recapitulate those seen previously in other directed evolution experiments. Interestingly, this is the case for several of the sites reported here (see the Supporting Information). One of the amino acid changes found by our selection of polymerases better able to synthesize duplexes containing **Z**:**P** pairs (D578N) occurred in a site that underwent substitution in the CSR experiment that yielded a *Taq* variant resistant to heparin inhibition (D578G).[15] This provides support for a general hypothesis that sites involved in adaptation to one environmental novelty might also be involved in adaptation to environmental novelties more generally. This is the rationale behind the REAP strategy for directed evolution,[25] in which mutations in a directed evolution experiment focus on sites where adaptive evolution appears to have already occurred in the natural history of the protein family.

This, of course, prompts a question that asks whether the individual sites and/or replacements recovered in the CSR experiments to obtain polymerase variants that accept the **Z**:**P** pair recapitulated events that actually occurred in natural history. For this analysis, the polymerases in Pfam PF00476 were examined. Several hundred bacterial and viral DNA polymerases homologous to the *Taq* polymerase were placed in a phylogenetic context to determine the distributions of amino acid replacements across the polymerase phylogeny. We then analyzed this collection to find sites that were conserved or highly variable or, most relevant to this work, "heterotachous".[26] Heterotachy is a sequence pattern such that the rate of evolution acting at an individual site can be slow in one portion of the phylogeny while the rate at the same site can be rapid in a different portion of the phylogeny. Such sequence/phylogenetic patterns can arise because of shifts in the selective constraints acting at individual sites throughout the evolutionary history of a gene family, and by extension, the precise biomolecular behaviors of the homologous proteins are not identical across the phylogeny.[25,27] A qualitative analysis is given in Table 1.

The partial recapitulation by these *in vitro* selection experiments of protein sequence variation observed in natural history does not necessarily mean, of course, that different polymerases or ancient polymerase ancestors physiologically incorporated d**Z**TP opposite d**P**. Rather, we believe that it reflects the fact that the sites can be changed (and indeed, in natural history, have been changed) to meet new challenges presented to polymerases without damaging the catalytic power or fidelity of the proteins. It will be interesting to see how frequently contemporary *in vitro* selection experiments recapitulate the natural history of their proteins in this way.

Most striking, however, was the observation that if **Z** and **P** were reversed in their roles as template and triphosphate, the polymerases that evolved *in vitro* to incorporate d**Z**TP opposite d**P** in a template did not perform as well. Strong termination bands were observed at sites where the polymerase was challenged to incorporate d**P**TP opposite d**Z** in the template (Figure 5), indicating that the polymerase neither accepted the dP match nor introduced a standard nucleotide as a mismatch.

This result confirms that incorporating d**P**TP opposite template d**Z** and incorporating d**Z**TP opposite template d**P** are not symmetrical processes. Thus, we have obtained by this selection a "unidirectional" polymerase, something largely unknown in Nature (but see ref 28). In these experiments, only one process (the second) must be successfully catalyzed for the polymerase catalyzing it to survive the selection. In contrast, an inability to incorporate d**P** opposite d**Z** on a primer has no impact on the ability of the polymerase to survive this selection. Therefore, our failure to recover polymerases capable of doing this might not be viewed as surprising. Instead, this result underscores the general rule in directed evolution that "you get what you select for".

## ■ ASSOCIATED CONTENT

### Supporting Information

Additional details of the evolutionary analysis of amino acid replacements of the *Taq* variants studied as well as experiments to determine activities of screened polymerases. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION

### Corresponding Author

*E-mail: sbenner@ffame.org. Telephone: (352) 271-7005. Fax: (352) 271-7076.

## Funding

This work was supported by the NASA Exobiology program (NNX08AO23G), by the Defense Threat Reduction Agency (HDTRA1-13-1-0004), and in part by the NASA Astrobiology Institute (via Jet Propulsion Laboratory, RSA1371457).


## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank Mr. Clement Vulin for help trouble-shooting library creation and emulsion PCR while he interned at the Foundation and Dr. Shuichi Hoshika and Dr. Zunyi Yang for preparing some of the AEGIS-containing oligonucleotides used.

## ■ REFERENCES


(1) Sismour, A. M., and Benner, S. A. (2005) Synthetic biology. *Expert Opin. Biol. Ther. 5*, 1409−1414.

(2) Switzer, C., Moroney, S. E., and Benner, S. A. (1989) Enzymatic incorporation of a new base pair into DNA and RNA. *J. Am. Chem. Soc. 111*, 8322−8323.

(3) Kool, E. T. (2002) Replacing the nucleobases in DNA with designer molecules. *Acc. Chem. Res. 35*, 936−943.

(4) Hirao, I., Harada, Y., Kimoto, M., Mitsui, T., Fujiwara, T., and Yokoyama, S. (2004) A two-unnatural-base-pair system toward the expansion of the genetic code. *J. Am. Chem. Soc. 126*, 13298−13305.

(5) Ishikawa, M., Hirao, I., and Yokoyama, S. (2000) Synthesis of 3-(2-deoxy-β-D-ribofuranosyl)pyridin-2-one and 2-amino-6-(N,N-dimethylamino)-9-(2-deoxy-β-D-ribofuranosyl)purine derivatives for an unnatural base pair. *Tetrahedron Lett. 41*, 3931−3934.

(6) Minakawa, N., Kojima, N., Hikishima, S., Sasaki, T., Kiyosue, A., Atsumi, N., Ueno, Y., and Matsuda, A. (2003) New base pairing motifs. The synthesis and thermal stability of oligodeoxynucleotides containing imidazopyridopyrimidine nucleosides with the ability to form four hydrogen bonds. *J. Am. Chem. Soc. 125*, 9970−9982.

(7) Rappaport, H. P. (1988) The 6-thioguanine 5-methyl-2-pyrimidinone base pair. *Nucleic Acids Res. 16*, 7253−7267.

(8) Tae, E. J. L., Wu, Y. Q., Xia, G., Schultz, P. G., and Romesberg, F. E. (2001) Efforts toward expansion of the genetic alphabet: Replication of DNA with three base pairs. *J. Am. Chem. Soc. 123*, 7439−7440.

(9) Benner, S. A. (2004) Understanding nucleic acids using synthetic chemistry. *Acc. Chem. Res. 37*, 784−797.

(10) Geyer, C. R., Battersby, T. R., and Benner, S. A. (2003) Nucleobase pairing in Watson-Crick-like genetic expanded information systems. *Structure 11*, 1485−1498.

(11) Henry, A. A., and Romesberg, F. E. (2003) Beyond A, C, G and T: Augmenting nature's alphabet. *Curr. Opin. Chem. Biol. 7*, 727−733.

(12) Piccirilli, J. A., Krauch, T., Moroney, S. E., and Benner, S. A. (1990) Enzymatic Incorporation of a New Base Pair into DNA and RNA Extends the Genetic Alphabet. *Nature 343*, 33−37.

(13) Collins, M. L., Irvine, B., Tyner, D., Fine, E., Zayati, C., Chang, C. A., Horn, T., Ahle, D., Detmer, J., Shen, L. P., Kolberg, J., Bushnell, S., Urdea, M. S., and Ho, D. D. (1997) A branched DNA signal amplification assay for quantification of nucleic acid targets below 100 molecules/ml. *Nucleic Acids Res. 25*, 2979−2984.

(14) Yang, Z., Chen, F., Chamberlin, S. G., and Benner, S. A. (2010) Expanded Genetic Alphabets in the Polymerase Chain Reaction. *Angew. Chem., Int. Ed. 49*, 177−180.

(15) Ghadessy, F. J., Ong, J. L., and Holliger, P. (2001) Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. U.S.A. 98*, 4552−4557.

(16) Tawfik, D. S., and Griffiths, A. D. (1998) Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol. 16*, 652−656.

(17) Yang, Z., Chen, F., Alvarado, J. B., and Benner, S. A. (2011) Amplification, Mutation, and Sequencing of a Six-Letter Synthetic Genetic System. *J. Am. Chem. Soc. 133*, 15105−15112.

(18) Yang, Z. Y., Hutter, D., Sheng, P. P., and Benner, S. A. (2005) Artificially Expanded Genetic Information System: New base pair with alternative hydrogen bonding pattern. *Abstracts of Papers of the American Chemical Society*, Vol. *229*, No. U561, American Chemical Society, Washington, DC.

(19) Yang, Z., Sismour, A. M., Sheng, P., Puskar, N. L., and Benner, S. A. (2007) Enzymatic incorporation of a third nucleobase pair. *Nucleic Acids Res. 35*, 4238−4249.

(20) Yang, Z., Hutter, D., Sheng, P., Sismour, A. M., and Benner, S. A. (2006) Artificially expanded genetic information system: A new base pair with an alternative hydrogen bonding pattern. *Nucleic Acids Res. 34*, 6095−6101.

(21) Grimm, E., and Arbuthnot, P. (1995) Rapid purification of recombinant taq DNA-polymerase by freezing and high-temperature thawing of bacterial expression cultures. *Nucleic Acids Res. 23*, 4518−4519.

(22) Brownie, J., Shawcross, S., Theaker, J., Whitcombe, D., Ferrie, R., Newton, C., and Little, S. (1997) The elimination of primer-dimer accumulation in PCR. *Nucleic Acids Res. 25*, 3235−3241.

(23) Pinheiro, V. B., Taylor, A. I., Cozens, C., Abramov, M., Renders, M., Zhang, S., Chaput, J. C., Wengel, J., Peak-Chew, S.-Y., McLaughlin, S. H., Herdewijn, P., and Holliger, P. (2012) Synthetic Genetic Polymers Capable of Heredity and Evolution. *Science 336*, 341−344.

(24) Ramsay, N., Jemth, A.-S., Brown, A., Crampton, N., Dear, P., and Holliger, P. (2010) CyDNA: Synthesis and Replication of Highly Cy-Dye Substituted DNA by an Evolved Polymerase. *J. Am. Chem. Soc. 132*, 5096−5104.

(25) Cole, M. F., and Gaucher, E. A. (2011) Exploiting Models of Molecular Evolution to Efficiently Direct Protein Engineering. *J. Mol. Evol. 72*, 193−203.

(26) Lopez, P., Casane, D., and Philippe, H. (2002) Heterotachy, an important process of protein evolution. *Mol. Biol. Evol. 19*, 1−7.

(27) Chen, F., Gaucher, E. A., Leal, N. A., Hutter, D., Havemann, S. A., Govindarajan, S., Ortlund, E. A., and Benner, S. A. (2010) Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and SNP detection. *Proc. Natl. Acad. Sci. U.S.A. 107*, 1948−1953.

(28) Foerstemann, K., Zaug, A. J., Cech, T. R., and Lingner, J. (2003) Yeast telomerase is specialized for C/A-rich RNA templates. *Nucleic Acids Res. 31*, 1646−1655.