J|A|C|S
ARTICLES
Published on Web 01/23/2004

Expanding the Substrate Repertoire of a DNA Polymerase by Directed Evolution

Ming Fa, Annalisa Radeghieri, Allison A. Henry, and Floyd E. Romesberg*

Contribution from the Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California, 92037-1000

Received September 15, 2003; E-mail: floyd@scripps.edu

Abstract: Nucleic acid polymerases are the most important reagents in biotechnology. Unfortunately, their high substrate specificity severely limits their applications. Polymerases with tailored substrate repertoires would significantly expand their potential and allow enzymatic synthesis of unnatural polymers for in vivo and in vitro applications. For example, the ability to synthesize 2′-*O*-methyl-modified polymers would provide access to materials possessing properties that make them attractive for biotechnology and therapeutic applications, but unfortunately, no known polymerases are capable of efficiently accepting these modified substrates. To evolve such enzymes, we have developed an activity-based selection method which isolates polymerase mutants with the desired property from libraries of the enzyme displayed on phage. In this report, mutants that could efficiently synthesize an unnatural polymer from 2′-*O*-methyl ribonucleoside triphosphates were immobilized and isolated by means of their activity-dependent modification of a DNA oligonucleotide primer attached to the same phage particle. In each case, directed evolution resulted in relocating a critical side chain to a different position in the polypeptide, thus re-engineering the overall active site while preserving critical protein−DNA interactions. Remarkably, one evolved polymerase is shown to incorporate the modified substrates with an efficiency and fidelity equivalent to that of the wild-type enzyme with natural substrates.

1. Introduction

Template-directed replication of nucleic acid polymers is perhaps the most important reaction in biology and biotechnology. In addition to PCR and its various applications, the ability to replicate and amplify rare DNA and RNA sequences has been used to evolve polymers with the properties desired for applications such as sensors,[1−9] therapeutics,[1,9−11] and catalysts.[12−14] However, the high substrate-selectivity of natural polymerases limits the application of these reactions to the natural biopolymers, DNA and RNA, or a small group of related derivatives.[15] Expansion of the substrate repertoire of these enzymes to include unnatural substrates would extend the scope of these reactions and allow for the template-directed synthesis of other polymers with interesting biological or physical properties.

Initial efforts to manipulate the substrate recognition of DNA polymerases were focused on the rational design of site-directed mutants[16,17] which required the mutants to be constructed and characterized individually. In addition, choosing the residues to mutate is seldom straightforward due to our limited structural and mechanistic understanding of these enzymes. In contrast, directed evolution strategies subject large libraries of polymerase mutants to activity-based selections. To evolve rare activities, directed evolution is more likely to be successful than rational design. For example, using an innovative water-in-oil emulsion technology,[18] Holliger and co-workers evolved DNA polymerases that were either more thermally stable or more resistant to an inhibitor.[19] Winters and co-workers showed that phage with pIII-displayed DNA polymerases could be isolated based on their activity-dependent modification of an attached oligonucleotide substrate.[20] In these studies, the oligonucleotide substrate was attached to the major phage coat protein, pVIII, raising the concern that intermolecular reactions would compromise the critical association of phenotype with genotype. In

(1) Osborne, S. E.; Matsumura, I.; Ellington, A. D. *Curr. Opin. Chem. Biol.* **1997**, *1*, 5−9.
(2) Jhaveri, S. D.; Kirby, R.; Conrad, R.; Maglott, E. J.; Bowser, M.; Kennedy, R. T.; Glick, R.; Ellington, A. D. *J. Am. Chem. Soc.* **2000**, *122*, 2469−2473.
(3) Hesselberth, J. R.; Robertson, M. P.; Knudsen, S. M.; Ellington, A. D. *Anal. Biochem.* **2002**, *312*, 106−112.
(4) Kleinjung, F.; Klussmann, S.; Erdmann, V. A.; Scheller, F. W.; Fürste, J. P.; Bier, F. F. *Anal. Chem.* **1998**, *70*, 328−331.
(5) Jaysena, S. D. *Clin. Chem.* **1999**, *45*, 1628−1650.
(6) Lee, M.; Walt, D. R. *Anal. Biochem.* **2000**, *282*, 142−146.
(7) O'Sullivan, C. K. *Anal. Bioanal. Chem.* **2002**, *372*, 44−48.
(8) Lu, Y. *Chem.−Eur. J.* **2002**, *8*, 4589−4596.
(9) Verma, S.; Jager, S.; Thum, O.; Famulok, M. *Chem. Rec.* **2003**, *3*, 51−60.
(10) Sullenger, B. A. *J. Clin. Invest.* **2000**, *106*, 921−922.
(11) White, R. R.; Sullenger, B. A.; Rusconi, C. P. *J. Clin. Invest.* **2000**, *106*, 929−934.
(12) Beaudry, A. A.; Joyce, G. F. *Science* **1992**, *257*, 635−641.
(13) Breaker, R. R. *Curr. Opin. Chem. Biol.* **1997**, *1*, 26−31.
(14) Pan, T. *Curr. Opin. Chem. Biol.* **1997**, *1*, 17−25.
(15) Bittker, J. A.; Phillips, K. J.; Liu, D. R. *Curr. Opin. Chem. Biol.* **2002**, *6*, 367−374.
(16) Astatke, M.; Ng, K.; Grindley, N. D. F.; Joyce, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3402−3407.
(17) Joyce, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 1619−1622.
(18) Griffiths, A. D.; Tawfik, D. S. *Curr. Opin. Biotechnol.* **2000**, *11*, 338−353.
(19) Ghadessy, F. J.; Ong, J. L.; Holliger, P. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4552−4557.
(20) Jestin, J.-L.; Kristensen, P.; Winter, G. *Angew. Chem., Int. Ed.* **1999**, *38*, 1124−1127.

10.1021/ja038525p CCC: $27.50

contrast, we have developed a system wherein the Stoffel fragment (SF) of *Thermus aquaticus* DNA polymerase I (*Taq*) and its oligonucleotide primer substrate are attached to the phage particle via its minor phage coat protein, pIII.[21] Phage produce about five pIII proteins per virus, and these proteins are localized to one end of the phage particle. This localization is expected to favor intramolecular reactivity and ensure the unique association of phenotype with genotype. The selection of active mutants depends on the enzyme-catalyzed biotinylation of the phage particle which in turn depends on the extension of the attached oligonucleotide primer, first by unnatural substrates, then by a biotinylated substrate. Thus, only those phage particles bearing polymerase mutants that accept the unnatural substrates are biotinylated and will be isolated on a solid streptavidin support.

We recently used this system to evolve SF into an RNA polymerase.[21] Phage particles displaying polymerases that were able to extend the attached oligonucleotide primer by incorporating rNTPs and biotinylated UTP were immobilized on streptavidin-coated magnetic beads and subsequently recovered. After four rounds of selection, a SF library yielded three mutants that could extend a DNA primer by incorporating up to eight rNTPs virtually as efficiently as the wild-type enzyme incorporated dNTPs.

The greatest potential of this method lies in tailoring polymerases appropriately for the synthesis of unnatural polymers. For example, polymers composed of 2′-*O*-methyl ribonucleosides are inexpensive and have properties that make them attractive for several biotechnology applications,[22−28] yet they have not been developed for these purposes because no known polymerase accepts them as substrates. We now report the evolution of SF into an enzyme that synthesizes these unnatural polymers from DNA templates using 2′-*O*-methyl ribonucleoside triphosphates ($_{OMe}$NTPs).

2. Experimental Section

2.1. Polymerase Library Construction. Two SF libraries were constructed. In each, six amino acid residues were subjected to random mutagenesis. The library **IEFYNQ** was randomly mutated at Ile614, Glu615, Phe667, Tyr671, Asn750, and Gln754, and the library **RMQVHE** was randomly mutated at Arg573, Met747, Gln754, Val783, His784, and Glu786. Randomization was accomplished using the 1.2 kb *Fse*I-*Not*I fragment of the SF gene, generated by overlapping extension PCR (Figure 1). In the first step, four fragments were generated from primers containing mutations. Second, the overlapping fragments were used as templates in conjunction with two external end primers to generate full-length polymerase gene products. Finally, the full-length PCR product was gel-purified and reamplified using two internal end primers, digested with *Fse*I/*Not*I, and ligated into gel-purified *Fse*I/*Not*I digested pFAB−SF.[21] To minimize background, pFAB−SF contained a nonfunctional SF gene with an amber stop codon at position 674.

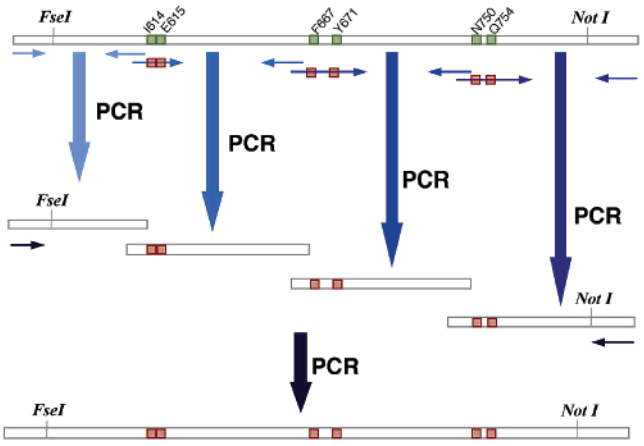

Figure 1. Construction of the **IEFYNQ** library by overlapping extension PCR. Mutated residues are indicated by red boxes. Primers are shown as thin arrows. The **RMQVHE** library was similarly constructed.

2.2. SF Phage Preparation and Substrate Attachment. SF library phage were prepared by electroporating 300 μL of XL1-blue MRF′ (Stratagene) competent cells with 10 μg of pFAB−SF library phagemid. The transformed cells were first transferred to 4 mL of SOC for 1 h at 37 °C, diluted in 150 mL of 2×YT containing 7.5 μg/mL tetracycline and 50 μg/mL spectinomycin, and grown at 30 °C to an OD600 of 0.6. The culture was then inoculated with 75 μL of 10^{13} cfu/mL X30 helper phage and incubated at 37 °C for 1 h without being shaked. In X30, gIII was modified by fusion with DNA encoding an "acidic peptide", as described previously.[21] The infected cells were pelleted and resuspended in 300 mL of fresh 2×YT containing 7.5 μg/mL tetracycline, 50 μg/mL spectinomycin, 50 μg/mL kanamycin, and 0.4 mM IPTG. The resuspended culture was grown at 30 °C for 16 h before the supernatant, containing the phage particles, was separated from the host cells by centrifugation. The phage particles were purified by two PEG precipitations. On average, the resulting phage particles co-display a single copy of one member of the polymerase library and four copies of the acidic peptide, all localized to one end of the phage particle.

The "basic peptide", *C*(GGS)$_4$AQLKKKLQALKKKNAQLKWKLQALKKKLAQGG**C**, containing a free and a 2-nitro-4,5-dimethoxybenzyl-protected cysteine (underlined and bold, respectively), and the 5′-thiol-modified P50 oligonucleotide, 5′-**S**-d(TTATG TATGT ATTTT CGACG TTTGC TAACA AGATA CGACT CACTA TAGGG), were synthesized, purified, and coupled via *N*,*N*′-bis(3-maleidopropiryl)-2-hydroxy-1,3-propanediamine, as described previously.[21] Briefly, 300 μg of basic peptide was dissolved in 150 μL of water and combined with 75 μL of 1 M sodium phosphate, pH 7.0, 30 μL of 5 M NaCl, and 22 μL of 3 mM P50 oligonucleotide. The mixture was incubated at 23 °C for 14 h under inert atmosphere to avoid oxidation, and the basic peptide-P50 DNA conjugate was purified by anion exchange FPLC.

The T28 template (1 nmol), 5′-d(CCTCC AAGGT CCCTA TAGTG AGTCG TAT)−NH_2, which is 3′ amino-modified to prevent extension and unintended biotinylation at that end, was annealed to the basic peptide−P50 conjugate (0.2 nmol) in 40 μL of TBS and then deprotected by irradiation at 365 nm for 45 min at room temperature. The deprotected basic peptide−P50/T28 assembly was diluted with 400 μL of attachment buffer (TBS containing 2.5 mM KCl, 1 mM EDTA, and 1 mM cystamine) and attached to the phage particles by adding 100 μL of 1.5 × 10^{13} cfu/mL phage and incubating for 1 h at 37 °C. The substrate-attached phage (pol/phage) were PEG-precipitated once to remove the free peptide−DNA conjugate.

(21) Xia, G.; Chen, L.; Sera, T.; Fa, M.; Schultz, P. G.; Romesberg, F. E. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6597−6602.
(22) Verma, S.; Eckstein, F. *Annu. Rev. Biochem.* **1998**, *67*, 99−134.
(23) Freier, S. M.; Altmann, K.-H. *Nucleic Acids Res.* **1997**, *25*, 4429−4443.
(24) Monia, B. P.; Lesnik, E. A.; Gonzalez, C.; Lima, W. F.; McGee, D.; Guinosso, C. J.; Kawasaki, A. M.; Cook, P. D.; Freier, S. M. *J. Biol. Chem.* **1993**, *268*, 14514−14522.
(25) Lesnik, E. A.; Guinosso, C. J.; Kawasaki, A. M.; Sasmor, H.; Zounes, M.; Cummins, L. L.; Ecker, D. J.; Cook, P. D.; Freier, S. M. *Biochemistry* **1993**, *32*, 7832−7838.
(26) Kawasaki, A. M.; Casper, M. D.; Freier, S. M.; Lesnik, E. A.; Zounes, M. C.; Cummins, L. L.; Gonzalez, C.; Cook, P. D. *J. Med. Chem.* **1993**, *36*, 831−841.
(27) Griffey, R. H.; Lesnik, E.; Freier, S.; Sanghvi, Y. S.; Teng, K.; Kwasaki, A.; Guinosso, C.; Wheeler, P.; Homan, V.; Cook, P. D.; Sanghvi, Y. S.; Cook, P. D., Eds.; *Carbohydrate Modifications in Antisense Research*; ACS Symposium Series 580; American Chemical Society: Washington, DC, 1994; pp 212−224.
(28) Tereshko, V.; Portmann, S.; Tay, E. C.; Martin, P.; Natt, F.; Altmann, K.-H.; Egli, M. *Biochemistry* **1998**, *37*, 10626−10634.

2.3. Phage Selection. In the first round of screening, 400 μL of pol/phage (3×10^{12} cfu) was combined with 1.1 mL of 20 μM Tris-HCl, pH 7.2, 10 mM $MgCl_2$, 40 μM $_{OMe}$CTP and $_{OMe}$ATP (TriLink), and 2 μM biotin−UTP (Enzo Diagnostic). The reaction was initiated by transferring the mixture from ice to a 50 °C water bath. After 5 min of incubation, 100 μL of 0.5 M EDTA was added to stop the reaction, and the phage particles were PEG-precipitated twice to remove unreacted biotin−UTP, followed by resuspension in 400 μL of TBS. Residual precipitate was removed by centrifugation (20 800 × g for 10 min), and the supernatant was added to 200 μL of streptavidin-coated magnetic beads (Dynal Corp., Norway). The beads were incubated at room temperature for 1 h with continuous inversion and then washed 10 times with TBS containing 0.5% (v/v) Tween-20. To release bound phage particles, the beads were resuspended in 50 μL of TBS containing 10 mM $MgCl_2$ and 1 mg/mL DNase I and incubated at 37 °C for 30 min. The amount of phage present in the supernatant was determined by titering, and the recovered phage were subjected to additional rounds of selection. In each subsequent round of selection, 100 μL of pol/phage was used in a final reaction volume of 400 μL.

2.4. Isolation and Small-Scale Purification of Polymerase Mutants. After four rounds of selection, SF genes from the recovered phagemids were cloned into a modified pET-23b(+) vector for overexpression of the polymerases in *Escherichia coli* BL21(DE3) pLysS (Novagen). Single colonies were used to inoculate 10 mL of 2×YT containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol and grown at 37 °C to an OD600 of 0.4. The cells were pelleted and resuspended in 10 mL of fresh 2×YT supplemented with 100 μg/mL ampicillin, 34 μg/mL chloramphenicol, and 0.4 mM IPTG. After an additional 4 h of growth, the cells were pelleted and resuspended in 1 mL of BugBuster reagent (Novagen). Lysis was complete after 20 min at room temperature, and cellular debris was removed by centrifugation at 20800g for 15 min. The clarified cell lysates were incubated at 75 °C for 30 min and centrifuged to remove denatured cellular proteins. Polymerase concentrations of 50−100 μg/mL were routinely obtained and were used directly in primer extension reactions to determine $_{OMe}$ATP incorporation activity.

2.5. Large-Scale Polymerase Purification. SF mutants exhibiting $_{OMe}$NTP incorporation activity were prepared on a larger scale. Freshly grown cells were diluted in 300 mL of 2×YT containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol and grown at 37 °C until the OD600 reached 0.4. IPTG was added to a final concentration of 0.4 mM, and the culture was incubated for an additional 4 h with shaking. The overexpressed SF proteins were partially purified by heat treatment and further purified using nickel affinity chromatography (Novagen), according to the manufacturer's instructions. Using NAP-10 columns (Pharmacia), the eluted proteins were exchanged into a buffer containing 50 mM Tris-HCl, pH 7.6, 0.5 mM EDTA, 1 mM DTT, and 50% (v/v) glycerol and concentrated to 0.5 mg/mL using 10 000 MWCO Ultrafree concentrators (Millipore). Yields of 100−300 μg of the SF mutant proteins (purity greater than 90%) were routinely obtained from 300 mL of cell culture.

2.6. Mutant Polymerase Activity Assay. SF mutant protein activities were assayed using a standard primer extension assay.[29] A typical 10 μL reaction mixture contained 40 nM 5′-^{33}P-radiolabeled primer annealed to the corresponding template (2-fold excess), 0.1−1.2 nM SF enzyme and varying concentrations of the triphosphate in a buffer containing 50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, and 50 μg/mL acetylated BSA. The reactions were incubated at 50 °C for 1−10 min and quenched with 20 μL of 95% formamide containing 20 mM EDTA and the loading dyes xylene cyanole and bromophenol blue. The products were separated on a 15% denaturing polyacrylamide gel and analyzed by phosphorimaging (Storm 860, Molecular Dynamics) and ImageQuant software. The apparent k_{cat} and K_M values were obtained by fitting the data to the Michaelis−Menten equation using KaleidaGraph (Synergy Software).

3. Results

3.1. Selection System. We have developed a selection method for DNA polymerase evolution that is based on the co-display (on a single phage particle) of a member of a mutant enzyme library and an "acidic" peptide that is used to attach a DNA primer-template substrate. The acidic peptide is designed to form a leucine zipper and a disulfide bond with a "basic" peptide that is covalently linked to the oligonucleotide primer strand. Association of the acidic and basic peptides results, after annealing the template strand to the primer, in the physical association of a primer-template substrate, a polymerase mutant, and the DNA encoding the mutant with a single phage particle.

Both the SF polymerase and the acidic peptide were displayed on phage as fusions with the phage minor coat protein, pIII. To prevent any deleterious effects on polymerase activity due to C-terminal fusion to the pIII protein, a synthetic DNA fragment encoding a flexible peptide linker (GGSGGS) was introduced between the SF gene and the truncated gIII gene in the pFAB−SF phagemid. Superinfection of bacteria harboring the pFAB−SF phagemid with X30 helper phage resulted in phage particles containing phagemid DNA, but whose coat consisted almost entirely of proteins encoded by the helper phage genome. During packaging of the phage, both polymerase-pIII and peptide-pIII fusions were included. The phage particles obtained from a typical preparation carried either one or zero copies of the polymerase-pIII fusion and three to five copies of the acidic peptide-pIII fusion. Western blots and phage ELISA indicated that the polymerase was successfully displayed on the phage particles, with a display level of 0.1−1% (percentage of total phage particles with polymerase displayed). The displayed polymerases retained significant DNA polymerase activity in a radiolabeled primer extension assay (data not shown).

The selection system is illustrated in Figure 2. Isolation of desired polymerase mutants is based on the catalysis of intramolecular primer extension with $_{OMe}$NTPs. Primer extension ultimately resulted in the incorporation of a biotinylated nucleotide tag, which was then used to immobilize and isolate the corresponding phage particles on streptavidin beads. Therefore, mutants with the desired activity were selected from the DNA polymerase library by incubating the phage particles first with the appropriate $_{OMe}$NTPs and biotin-labeled substrates and then with streptavidin-coated magnetic beads. After sufficient enrichment, through multiple rounds of reinfection, selection, and isolation, phage were isolated and their displayed polymerases were characterized.

3.2. Library Construction and Polymerase Evolution. Two SF mutant libraries were designed based on the *Thermus aquaticus* DNA polymerase I ternary complex crystal structure.[30] Each library contained random mutations of six amino acids. The libraries are referred to as **IEFYNQ** and **RMQVHE**, indicating the residues targeted for mutation. Forty-eight clones from each library were selected and sequenced to assess the size and quality of the library. More than 75% of the sequenced clones contained in-frame polymerase genes with no stop codons. On the basis of the initial library size ($\sim 6.4 \times 10^7$) and polymerase display efficiency ($\sim$0.1%), we used 3×10^{12}

(29) Tae, E. L.; Wu, Y. Q.; Xia, G.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2001**, *123*, 7439−7440.

(30) Li, Y.; Korolev, S.; Waksman, G. *EMBO J.* **1998**, *17*, 7514−7525.

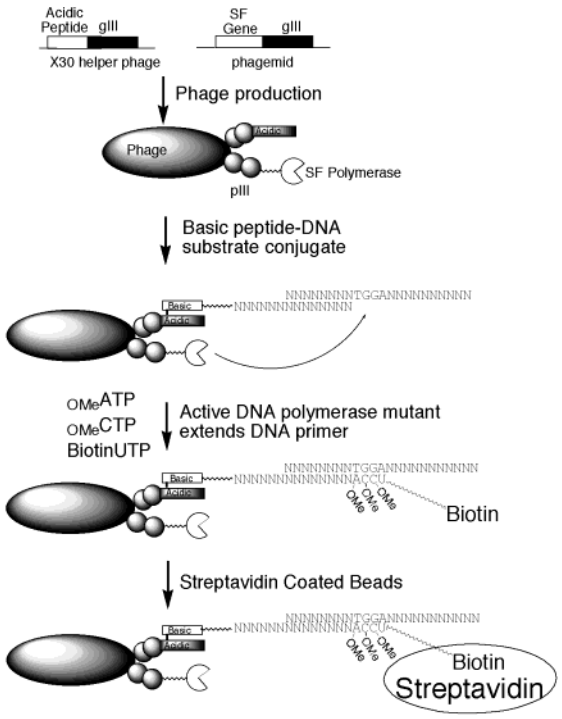

Figure 2. Activity-based phage selection system.

Table 1. Residues Selected at Positions 614 and 615

	614	615
SF wt	Ile	Glu
SFM01	Tyr	Gly[a]
SFM18	Tyr	Gly[b]
SFM19	Glu	Gly
SFM30	Glu	Ala

[a] GGG codon. [b] GGT codon.

phage particles in the first round of screening to ensure that the average mutant would be displayed on multiple phage particles.

After four rounds of selection, mutant SF genes were isolated and cloned into the pET-23b(+) vector for protein overexpression. In 80 randomly chosen colonies from each library, more than half expressed SF. After partial purification (see the Experimental Section), four of the proteins from the **IEFYNQ** library that could efficiently extend a DNA primer by inserting $_{OMe}$ATP opposite dT in the template were identified. The corresponding plasmids were isolated and sequenced (Table 1). Each sequence revealed two mutations at the adjacent residues 614 and 615, with no mutation of Phe667, Tyr671, Asn750, or Gln754. Ile614 was mutated to Tyr (SFM01 and SFM18) or Glu (SFM19 and SFM30), while Glu615 was mutated to Gly (SFM01, SFM18, SFM19) or Ala (SFM30). Clones SFM01 and SFM18 encoded identical amino acid sequences, but had different DNA sequences. Thus, this particular protein sequence was independently isolated twice, implying that a sufficient number of clones were assayed for activity.

His-tagged mutant proteins were purified by nickel affinity chromatography. To quantify the unnatural activity of each clone, we employed a steady-state single nucleotide incorporation assay using a 23-mer/45-mer DNA primer-template. The initial kinetics revealed a remarkable increase in $_{OMe}$NTP incorporation efficiency that was most pronounced for clone

Table 2. SFM19 Kinetic Data[a]

```
5'-ATACGACTCACTATAGGG
3'-TATGCTGAGTGATATCCCTGGAACCTCC
          Triphosphate
5'-ATACGACTCACTATAGGGX
3'-TATGCTGAGTGATATCCCTGGAACCTCC
```

triphosphate	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (min^{-1} M^{-1})
$_{OMe}$ATP	121.57 ± 27.2	54.17 ± 18.2	2.2×10^6
dATP	93.51 ± 12.4	55.69 ± 14.2	1.7×10^6
dCTP	0.53 ± 0.5	569.59 ± 243.5	9.2×10^2
dGTP	0.70 ± 1.0	410.74 ± 158.1	1.7×10^3
dTTP	1.55 ± 1.6	776.14 ± 71.4	2.0×10^3

[a] See text for details.

SFM19, which was thus chosen for more detailed characterization. To characterize SFM19, two distinct primer-templates were chosen, as depicted in Tables 2 and 3. One mimicked the conditions employed in the selection and one allowed evaluation of the kinetics in a different sequence context. As shown in Tables 2 and 3, the k_{cat}/K_M for correct $_{OMe}$NTP insertion by SFM19 ranged from 1×10^6 to 6×10^6 M^{-1} s^{-1}. Even the 2′-*O*-methyl UTP was inserted efficiently opposite dA; this reaction has proven especially difficult for DNA polymerase mutants to catalyze due to strong discrimination against this base.[31] SFM19 also retained wild-type-like recognition of 2′-deoxyribonucleoside triphosphates (Tables 2 and 3). In contrast to SFM19, the wild-type enzyme showed no detectable activity with any of the modified triphosphates ($k_{cat}/K_M < 10^2$ M^{-1} s^{-1}). Thus, the evolved activity of SFM19 corresponded to an increase in k_{cat}/K_M of at least 10 000-fold. To determine whether the $_{OMe}$NTP incorporation activity was obtained by compromising the overall fidelity of the polymerase, we also characterized the efficiency of mispair synthesis by this enzyme (Tables 2 and 3). No mispair between 2′-modified or unmodified nucleotides was synthesized at a detectable rate (greater than $\sim 10^3$ M^{-1} s^{-1}), thus the fidelity of SFM19 remained high.

To determine whether SFM19 was able to continue primer extension after incorporation of an $_{OMe}$NTP, we used an 18-mer/35-mer primer-template (Figure 3). Lane 1 shows 18-mer primer. Lane 2 shows primer extension by SFM19 with dNTPs. In this case, full-length synthesis remained efficient. Lanes 3 and 4 show primer extension in the presence of dCTP, dGTP, and dTTP without (lane 3) or including (lane 4) $_{OMe}$ATP. In the absence of $_{OMe}$ATP, synthesis stopped at or immediately prior to the first dT in the template, again demonstrating the high fidelity of SFM19. In the presence of $_{OMe}$ATP, SFM19 inserted the unnatural triphosphate and continued primer extension until five dNTPs had been incorporated. Termination of synthesis after the incorporation of approximately five natural triphosphates after the modified nucleoside was observed with several templates of different sequence. Apparently, as the newly synthesized duplex moved through the polymerase active site, certain unfavorable protein−duplex interactions were encountered that hindered continued primer extension. This result was not unexpected as the libraries used in the selection were focused exclusively to the primer terminus.

(31) Patel, P. H.; Loeb, L. A. *J. Biol. Chem.* **2000**, *275*, 40266−40272.

Table 3. SFM19 Kinetic Data[a]

```
5'-TAATACGACTCACTATAGGGAGA
3'-ATTATGCTGAGTGATATCCCTCTXGCTAGGTTACGGCAGGATCGC
                  |
                  | Triphosphate
                  ▼
5'-TAATACGACTCACTATAGGGAGAN
3'-ATTATGCTGAGTGATATCCCTCTXGCTAGGTTACGGCAGGATCGC
```

X	triphosphate	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($min^{-1}M^{-1}$)	X	triphosphate	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($min^{-1}M^{-1}$)
dT	$_{OMe}$ATP	76.4[b]	11.09[b]	6.9×10^6	dG	$_{OMe}$CTP	40.99 ± 12.7	42.52 ± 3.2	9.6×10^5
dT	dATP	93.5[b]	55.69[b]	1.7×10^6	dG	dCTP	38.81 ± 5.8	101.17 ± 4.8	3.9×10^5
dT	dCTP	nd[c]	nd[c]	9.2×10^2	dG	dGTP	nd[c]	nd[c]	$<10^3$
dT	dGTP	nd[c]	nd[c]	1.7×10^3	dG	dTTP	nd[c]	nd[c]	$<10^3$
dT	dTTP	nd[c]	nd[c]	2.0×10^3	dG	dATP	nd[c]	nd[c]	$<10^3$
dA	$_{OMe}$UTP	58.92 ± 39.4	24.40 ± 6.2	2.4×10^6	dC	$_{OMe}$GTP	61.95 ± 49.9	10.34 ± 3.3	6.0×10^6
dA	dTTP	128.75 ± 6.2	21.64 ± 3.5	6.0×10^6	dC	dGTP	74.80 ± 7.9	4.35 ± 1.6	1.7×10^7
dA	dCTP	nd[c]	nd[c]	$<10^3$	dC	dATP	nd[c]	nd[c]	$<10^3$
dA	dGTP	nd[c]	nd[c]	$<10^3$	dC	dCTP	nd[c]	nd[c]	$<10^3$
dA	dATP	nd[c]	nd[c]	$<10^3$	dC	dTTP	nd[c]	nd[c]	$<10^3$

[a] See text for experimental details. [b] Single determination; no error reported. [c] Rate insufficient to determine k_{cat} and K_M independently.

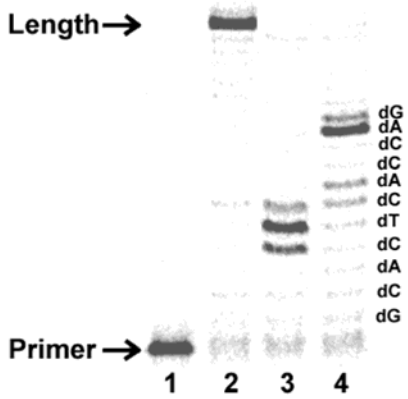

Figure 3. Primer extension by SFM19. Reactions (10 μL) contained 1 nM SFM19, 40 nM primer template, 10 mM Tris-HCl pH 7.5, 50 mM KCl, 0.1% TritonX-100, 5 mM $MgCl_2$, 1 mM $MnCl_2$, and no triphosphate (lane 1); 500 μM dNTPs (lane 2); 500 μM dCTP, dGTP, and dTTP (lane 3); 500 μM dCTP, dGTP, dTTP, and $_{OMe}$ATP (lane 4). Reactions were incubated at 50 °C for 45 min and quenched by addition of EDTA.

4. Discussion

The natural nucleic acid biopolymers DNA and RNA constitute a unique class of materials because they may be replicated and amplified many million-fold by nucleic acid polymerases. In addition to revolutionizing molecular biology and biotechnology, these amplification reactions are the basis for in vitro evolution techniques, such as SELEX, that allow the evolution of biopolymers with desired properties.[32−34] However, the power of such techniques is limited by the substrate specificity of natural polymerases; that is, many modified nucleic acid polymers cannot be produced in amplification reactions carried out by the known DNA and RNA polymerases. Thus, the potential properties and applications of the polymers are restricted to those that are compatible with DNA and RNA, which is a severe limitation compared to other synthetic polymers. For example, the instability of DNA and RNA in biological fluids precludes their use in many therapeutic and diagnostic applications. Thus, the ability to enzymatically replicate modified nucleic acid polymers that are more stable in such fluids would significantly expand the potential scope and application of these materials.

Biopolymers composed of 2′-*O*-methyl-modified ribonucleotides form stable heteroduplexes with both natural DNA and RNA,[23,35] and the nuclease resistance of these polymers enables their direct use in biological fluids. In general, nucleic acid polymers containing 2′-fluoro and 2′-*O*-alkyl subunits form chimeric duplexes that are more stable than the corresponding natural DNA duplexes. 2′-*O*-Methyl-modified ribonucleic acids, in particular, form the most stable chimeric duplexes, with increases in duplex melting temperature of up to 1.5 °C per modification.[24] These sugar rings appear to preferentially populate the C3′-endo sugar conformation associated with A-form duplexes.[25−28,36] 2′-*O*-Methyl-modified polymers are also attractive for practical applications due to the lower cost of the corresponding nucleotides relative to other 2′-modified nucleosides.

Developing the potential of these polymers requires their efficient enzymatic synthesis. In general, wild-type DNA polymerases discriminate strongly against all 2′-substitutents, including $_{OMe}$NTPs, most likely because these polymerases selectively synthesize DNA in the presence of a cellular excess of rNTPs. RNA polymerases from T7 bacteriophage and *E. coli* are more accommodating of modifications at the 2′ position, but neither recognizes $_{OMe}$NTPs with high efficiency.[37] 2′-Fluoro-ATP and 2′-azido-CTP[38] are also not recognized, whereas 2′-F- and 2′-NH_2-substituted pyrimidines are recognized, but incorporated with up to 1000-fold reduced efficiency.[39−44] Poor efficiency requires the use of excess

(32) Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611−647.
(33) Famulok, M.; Jenne, A. *Curr. Opin. Chem. Biol.* **1998**, *2*, 320−327.
(34) Famulok, M.; Mayer, G.; Blind, M. *Acc. Chem. Res.* **2000**, *33*, 591−599.
(35) Egli, M.; Gryaznov, S. M. *Cell. Mol. Life Sci.* **2000**, *57*, 1440−1456.
(36) Egli, M. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 1895−1910.
(37) Conrad, F.; Hanne, A.; Gaur, R. K.; Krupp, G. *Nucleic Acids Res.* **1995**, *23*, 1845−1853.
(38) Kay, B. K.; Winter, J.; McCafferty, J. *Phage Display of Peptides and Proteins: A Laboratory Manual*; Academic Press: San Diego, CA, 1996.
(39) Aurup, H.; Williams, D. M.; Eckstein, F. *Biochemistry* **1992**, *31*, 9636−9641.
(40) Cramer, P.; Bushnell, D. A.; Fu, J.; Gnatt, A. L.; Maier-Davis, B.; Thompson, N. E.; Burgess, R. R.; Edwards, A. M.; David, P. R.; Kornberg, R. D. *Science* **2000**, *288*.
(41) Lin, Y.; Nieuwlandt, D.; Magallanez, A.; Feistner, B.; Jayasena, S. D. *Nucleic Acids Res.* **1996**, *24*, 3407−3414.
(42) Kubik, M. F.; Bell, C.; Fitzwater, T.; Watson, S. R.; Tasset, D. M. *J. Immunol.* **1997**, *159*, 259−267.

enzyme (approaching millimolar concentrations) and error-prone conditions. These issues seriously complicate the synthesis of these unnatural polymers.

Efforts to synthesize these polymers are expected to be greatly facilitated by the use of polymerases that are specifically tailored to accept the corresponding modified substrates. We have developed an activity-based selection method in which both a DNA polymerase mutant and its primer-template substrate are attached to the minor phage coat protein, pIII. To evolve a nucleic acid polymerase that recognizes $_{OMe}$NTP substrates, we chose to display the Stoffel fragment of *Taq* DNA polymerase I based on our previous success in evolving this enzyme into an RNA polymerase and because it has been extensively characterized both structurally and functionally. We generated two separate SF libraries wherein six residues were completely randomized. In the **IEFYNQ** library, we mutated the residues Ile614, Glu615, Phe667, Tyr671, Asn750, and Gln754. On the basis of the crystal structure,[30] these residues were chosen to manipulate interactions between the polymerase and both the catalytic magnesium ions and the triphosphate substrate. In the second library, **RMQVHE**, we mutated the residues Arg573, Met747, Gln754, Val783, His784, and Glu786 to manipulate interactions between the polymerase and the primer terminus. Both libraries were constructed by overlapping extension PCR, as described above. The libraries were each estimated to contain $\sim 6.4 \times 10^7$ members and were displayed on 3×10^{12} phage particles, with a display efficiency of $\sim$0.1%, ensuring the display of each clone on multiple phage particles.

The selection strategy is based on the isolation of mutants that incorporate biotinylated UTP opposite dA in the template (Figure 2). The primer terminates just prior to a 3′-d(TGGA...) template sequence; thus, biotinylation requires the incorporation of a single $_{OMe}$ATP, two $_{OMe}$CTPs, and one biotin−UTP. Four rounds of selection were performed, each including phage preparation, substrate attachment, incubation with triphosphate substrates, isolation on streptavidin beads, and phage recovery after incubation with DNase. Both libraries showed a significant increase in phage recovery after each round of selection, relative to the previous round and the negative control (which included EDTA to prevent the reaction). After the final round, four mutants from the **IEFYNQ** library were identified that could extend a primer substrate by incorporating a single $_{OMe}$NTP. One, SFM19, was clearly the most efficient, and three others, SFM01, SFM18, and SFM30, showed significant activity compared to the wild type, although less than SFM19. The corresponding genes were sequenced and the mutations listed in Table 1. Interestingly, all four of the isolated polymerases have mutations only at positions 614 and 615. We did not isolate any active mutants from the **RMQVHE** library, and there was little enrichment of phage observed during the rounds of selection. Moreover, sequencing of 40 randomly selected clones from the **RMQVHE** library after the final round of selection did not reveal any sequence consensus.

Residues 614 and 615 are located in the C-terminus of the third strand of the four-stranded antiparallel β sheet of the "palm" region of the polymerase, at the junction with an α-helix.[30] The residues are part of the DYSQIELR motif A sequence, highly conserved in the DNA polymerase I enzyme from prokaryotes and eubacteria.[45] This strand is of obvious functional relevance as it contains Glu610, which chelates one of the two catalytic Mg^{2+} ions. The strand forms part of the floor and wall of the binding site proximal to the primer terminus and incoming triphosphate. Residues 614 and 615 pack on the sugar ring of the incoming triphosphate and are located near the N-terminus of the O-helix, which is also known to make functionally important contacts with the triphosphate substrate. In the *Klentaq* ternary complex crystal structure with bound ddNTP,[46] Glu615 coordinates a water molecule that makes sequence-independent H-bonds with the N3 or O2 atoms of the purines and pyrimidines, respectively. The analogous residues in several homologous polymerases, such as Glu710 in the Klenow fragment for *E. coli* DNA polymerase I, are thought to act as a "steric gate" to discriminate against the incorporation of triphosphates with 2′-substitutents.[16,17,47,48] To fulfill this function, the Glu side chain, possibly along with the 3′-OH of an incoming dNTP, may chelate a Mg^{2+} ion to selectively stabilize the transition state for dNTP incorporation.[49,50] As a result, the metal atom, or the Glu side chain itself, occupies a position in the active site that prevents productive binding of the modified substrate.[16] Consistent with this steric exclusion model, mutations of Glu615 in *Taq* or SF to the smaller Asp or Gly residue are found to facilitate insertion of rNTPs.[16,21,31] Mutation of this Glu residue to Gly in SFM01, SFM18, and SFM19 and to Ala in SFM30 is also consistent with the steric exclusion model and is likely to be important for creating space within the active site to accommodate the 2′-*O*-methyl substitutent.

The problem with simply creating space in the active site for the $_{OMe}$NTP by mutation of Glu615 to Gly or Ala is that the other functions of the Glu615 side chain will be lost. Thus, secondary mutations are required to restore these functions, possibly including binding a Mg^{2+} ion or a minor groove water molecule. This may explain why each of the isolated mutants also possesses mutation of Ile614 to Glu or Tyr, either of which should be capable of fulfilling these roles. In fact, two of the three unique protein sequences isolated (SFM19 and SFM30) contain the Ile614Glu mutation, including the most active SFM19 clone, possibly to most efficiently compensate for the loss of other Glu615 side-chain functions. Thus, it appears that the directed evolution process succeeded by first creating space for the 2′-substitutent and then retailoring the binding site as required for function. This model highlights the importance of multiple coupled mutations, as has been seen in the evolution of function in natural systems.[51] Because these secondary interactions are challenging to predict, these results also emphasize the critical role played by library size and design in evolving rare activities.

(43) Ruckman, J.; Green, L. S.; Beeson, J.; Waugh, S.; Gillette, W. L.; Henninger, D. D.; Claesson-Welsh, L.; Janjic, N. *J. Biol. Chem.* **1998**, *273*, 20556−20567.
(44) Lupold, S. E.; Hicke, B. J.; Lin, Y.; Coffey, D. S. *Cancer Res.* **2002**, *62*, 4029−4033.
(45) Wang, J.; Sattar, A. K.; Wang, C. C.; Karam, J. D.; Konigsberg, W. H.; Steitz, T. A. *Cell* **1997**, *89*, 1087−99.
(46) Li, Y.; Waksman, G. *Protein Sci.* **2001**, *10*, 1225−1233.
(47) Gao, G.; Orlova, M.; Georgiadis, M. M.; Hendrickson, W. A.; Goff, S. P. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 407−411.
(48) Gao, G.; Goff, S. P. *J. Virol.* **1998**, *72*, 5905−5911.
(49) Astatke, M.; Grindley, N. D. F.; Joyce, C. M. *J. Mol. Biol.* **1998**, *278*, 147−165.
(50) Gangurde, R.; Kaushik, N.; Singh, K.; Modak, M. J. *J. Biol. Chem.* **2000**, *275*, 19685−19692.
(51) Patten, P. A.; Gray, N. S.; Yang, P. L.; Marks, C. B.; Wedemayer, G. J.; Boniface, J. J.; Stevens, R. C.; Schultz, P. G. *Science* **1996**, *271*, 1086−1091.

5. Conclusions

From PCR to SELEX, enzymatic replication of DNA and RNA has revolutionized the biological sciences. The availability of polymerases with expanded substrate repertoires would expand the application of these techniques to allow the efficient production (or evolution) of unnatural polymers with interesting properties outside the scope of those compatible with the natural biopolymers. In particular, the ability to synthesize 2′-*O*-methyl-modified polymers would provide access to materials with both improved hybridization characteristics and stability toward nucleases and is expected to have important applications in therapeutic and biosensor development. Toward this goal, we have developed and used an activity-based selection system to evolve polymerases capable of efficiently synthesizing 2′-*O*-methyl-modified polymers with high efficiency and fidelity. The mutations appear to re-engineer the binding site at the primer terminus to accommodate the modified substrates while preserving critical intrapolymerase side chain or polymerase−substrate interactions. Given appropriate library design, it seems likely that selection schemes designed specifically to evolve other enzymatic activities will also be successful, including the recognition of other substrates as triphosphates or in the template or more processive synthesis. Such selections are currently in progress.

Acknowledgment. Funding was provided by USPHS NIH Grant No. R01 GM60005 (F.E.R.).

JA038525P