Hyperthermophilic *Aquifex aeolicus* initiates primer synthesis on a limited set of trinucleotides comprised of cytosines and guanines

Marilynn A. Larson[1,2], Rafael Bressani[1,2], Khalid Sayood[3], Jacob E. Corn[4], James M. Berger[4], Mark A. Griep[5,*] and Steven H. Hinrichs[1,2]

[1]Department of Microbiology and Pathology, University of Nebraska Medical Center, Omaha, NE 68198-6495, [2]University of Nebraska Center for Biosecurity, Omaha, NE 68198-4080, [3]Department of Electrical Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588-0511, [4]University of California, Berkeley, CA 94720 and [5]Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588-0304, USA

ABSTRACT

The placement of the extreme thermophile *Aquifex aeolicus* in the bacterial phylogenetic tree has evoked much controversy. We investigated whether adaptations for growth at high temperatures would alter a key functional component of the replication machinery, specifically DnaG primase. Although the structure of bacterial primases is conserved, the trinucleotide initiation specificity for *A. aeolicus* was hypothesized to differ from other microbes as an adaptation to a geothermal milieu. To determine the full range of *A. aeolicus* primase activity, two oligonucleotides were designed that comprised all potential trinucleotide initiation sequences. One of the screening templates supported primer synthesis and the lengths of the resulting primers were used to predict possible initiation trinucleotides. Use of trinucleotide-specific templates demonstrated that the preferred initiation trinucleotide sequence for *A. aeolicus* primase was 5′-d(CCC)-3′. Two other sequences, 5′-d(GCC)-3′ and d(CGC)-3′, were also capable of supporting initiation, but to a much lesser degree. None of these trinucleotides were known to be recognition sequences used by other microbial primases. These results suggest that the initiation specificity of *A. aeolicus* primase may represent an adaptation to a thermophilic environment.

INTRODUCTION

Molecular strategies in bacteria for adaptation to growth and survival at elevated temperatures have not been fully characterized although structural and functional adaptations have been described. The tRNAs from hyperthermophiles have extensive hydrogen bonding and numerous base modifications that restrict bending at crucial locations, thereby increasing biomolecular thermostability (1). Chaperone proteins in the thermophilic bacterium *Thermus thermophilus* have been shown to assist in folding or maintenance of structure (2). However, little effort has been directed toward investigating adaptations of the DNA replication machinery in extreme thermophiles.

Most hyperthermophiles belong to the domain Archaea, with the exception of a few bacteria (3). Therefore, the primase from the hyperthermophilic bacterium *Aquifex aeolicus* was chosen to study the potential mechanisms of thermoadaptation during DNA replication for comparison to other microbes. Structural analyses of the bacterial primases studied to date show that these enzymes are remarkably similar with three conserved functional domains that includes the zinc binding, RNA polymerase and C-terminal domains (4). Primase, a specialized DNA-dependent RNA polymerase, generates short oligoribonucleotide polymers *de novo* at the replication fork (4). All known DNA polymerases are unable to initiate polymerization of nucleotides and require a free 3′-hydroxyl group for elongation. During DNA replication, primase provides this essential 3′-hydroxyl to initiate leading strand synthesis at least once and lagging strand synthesis numerous times. Conditional lethal mutations in the primase gene from *Escherichia coli*, *Bacillus subtilis* and *Staphylococcus aureus* demonstrate the essential nature of this enzyme (5–7).

The *Aquifex* lineage is composed of nonsporeforming, hydrogen-oxidizing, obligate chemolithotrophic, microaerophilic, Gram-negative, rod-shaped bacteria. *Aquifex aeolicus* is one of the few bacteria known to be capable of survival at extremely high temperatures with a growth temperature maxima near 95°C and an optimum growth

temperature of 85°C (8). Interestingly, the genomes of these hyperthermophiles are only one-third the size of the Gram-negative *E. coli* genome and demonstrate few specific indications of heat-resistance (9). Despite the intuitive expectation that thermophilic adaptations would occur, such as a higher G+C genome content with a higher melting temperature (10−12), the *A. aeolicus* chromosome has a G+C content of only 43% (8). Therefore, there appears to be little relationship between adaptation of the organism to high growth temperatures and high G+C content in the genome (9,13).

Inferences made from the phylogenetic analysis of the 16S ribosomal RNA sequence from *A. aeolicus* indicate that this microbe is one of the earliest bacteria to have diverged from the evolutionary tree of the eubacteria (14), but other studies suggest that RNA sequence changes were an adaptive response to the elevated temperature of the organism's environment (9,15). Protein phylogenies have placed *A. aeolicus* much further from the root, but some studies could not identify a consistent node (8,16). More recent cladistic studies using multiple signature proteins indicate that the *Aquifex* lineage emerged from Gram-positive bacteria, prior to the split of Gram-positive and Gram-negative bacteria (17,18). The proximity of *Aquifex* to Gram-positive bacteria is consistent with the low GC content of its genome and the encoding of both PolC and DnaE, replicative DNA polymerase III holoenzymes. In contrast, *Aquifex* possesses a double membrane structure similar to Gram-negative bacteria. Therefore, questions remain as to whether heat tolerance in *A. aeolicus* is a legacy of primordial bacterial evolution or a later adaptation to fill a vacant niche in the geothermal environment. We hypothesized that a bacterial hyperthermophile might utilize primase initiation sequences that differ from mesophilic microbes.

Primases from several viruses, phages and bacteria studied to date have required a specific sequence to initiate synthesis of an RNA polymer *de novo* (4). The bacteriophages T4, T7 and SP6 recognize the trinucleotides 5′-d(GTT)-3′, 5′-d(GTC)-3′ and 5′-d(GCA)-3′, respectively (19−21). In *E. coli*, the primer is most often initiated by the sequence pppApG *in vivo* (22) and *in vitro* studies determined that *E. coli* primase primed single-stranded DNA (ssDNA) templates containing 5′-d(CTG)-3′ in the absence of the replicative helicase (23). It was unknown whether bacteria from other genera would have different recognition sequences until the primases from *Geobacillus stearothermophilus* and *S. aureus* were recently shown to recognize 5′-d(CTA)-3′ and 5′-d(TTA)-3′ (24,25). Based solely on these three known examples, bacterial primase initiation specificity appears to be inherited and shared between phylogenetic siblings, but different between Gram-negative and Gram-positive bacteria. Nevertheless, the initiation sequences for microbial primases have several common features: the recognition sequences are three bases in length; primer synthesis initiates complementary to the middle base of the trinucleotide, which is a pyrimidine and the third base in the trinucleotide is cryptic, meaning its presence is required to support efficient primer synthesis, but its complement is not incorporated into the RNA primer (4).

Although the contiguous structure of *A. aeolicus* primase zinc-binding and RNA polymerase domains have been resolved by crystallography (26), the activity and the initiation specificity of this protein had not been previously characterized. The lack of prior biochemical and biological studies on this enzyme and the potential difference in trinucleotide specificity presented a challenge for screening. Therefore, a new approach was developed with the capability to not only determine the activity of *A. aeolicus* primase, but also any other microbial primase that had not been characterized. To test as many trinucleotide sequences as possible, two oligonucleotides were created using a genetic algorithm program such that all 64 trinucleotides were present in one or the other sequence. After the resulting primer lengths were determined, the potential initiation sequences were predicted from the template and then confirmed using trinucleotide-specific templates. Using this approach, the zinc binding and RNA polymerase domains of *A. aeolicus* primase were shown to be sufficient to synthesize primers *de novo*. Further, this replicative enzyme initiated RNA polymer synthesis on trinucleotides not previously known to be recognition sequences for any microbial primase. *Aquifex aeolicus* primase preferentially initiated primer synthesis on 5′-d(CCC)-3′ and to a lesser degree on 5′-d(GCC)-3′ and 5′-d(CGC)-3′. Elucidation of the initiation trinucleotides for *A. aeolicus* primase from this hyperthermophile has provided the basis for further enzymatic studies.

MATERIALS AND METHODS

Proteins, reagents and substrates

Residues 1−405 of *A. aeolicus* primase were cloned, expressed and purified as previously described (26). The C-terminal domain had been deleted to enhance protein solubility. The full-length primases from other bacteria, including *S. aureus* and *E. coli*, were used to evaluate the functionality of the two oligonucleotides. *Escherichia coli* DnaG was isolated as previously described (27) and the other bacterial primases were cloned into pET30 and expressed in *E. coli* BL21(DE3) according to the manufacturer's suggestions (Novagen EMD Biosciences, La Jolla, CA, USA). Protein purification was performed using nickel affinity chromatography as recommended by the manufacturer (GE Healthcare, Piscataway, NJ, USA). Enzyme purity and subunit molecular mass were estimated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. The concentration of *A. aeolicus*, *S. aureus* and *E. coli* primase were determined using the monomer extinction coefficient of $36\,010\,M^{-1}\,cm^{-1}$, $53\,890^{-1}\,cm^{-1}$ and $46\,215\,M^{-1}\,cm^{-1}$, respectively, at 280 nm.

Ribonucleoside triphosphates (NTPs) were obtained from Promega (Madison, WI, USA). Magnesium acetate, potassium glutamate, HEPES and dithiothreitol (DTT) were from Sigma (St. Louis, MO, USA). The Sephadex G-25 microspin columns were purchased from Amersham (Piscataway, NJ, USA). The WAVE HPLC Nucleic Acid Fragment Analysis System, HPLC Buffer A (0.1 M triethylammonium acetate, pH 7.0), HPLC Buffer B

(0.1 M triethylammonium acetate, 25% acetonitrile v/v) and the DNA Sep HPLC column were obtained from Transgenomic (Omaha, NE, USA).

All oligonucleotide templates used in this study were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA, USA). Each template was purified using urea–PAGE, UV-shadowing and elution of the oligonucleotide into Tris–EDTA buffer. The templates were quantified by spectrophotometry at 260 nm using the extinction coefficients obtained for each oligonucleotide from the online OligoTools Calculator from Integrated DNA Technologies, Inc.

Trinucleotide evaluation oligonucleotide sequence design

Two oligonucleotide sequences that comprised all 64 potential trinucleotide sequences were designed to be between 20 and 40 nt in length since oligonucleotides with this length have been successfully used to study the yields of RNA polymers that are readily separable from the DNA templates by denaturing HPLC (25,28). The sequences incorporated features previously shown to be essential for activity. Since the shortest length of a nucleotide sequence containing two unique trinucleotides is 4, the shortest length of a sequence containing m unique trinucleotides is $m + 2$. In consideration of the overall length constriction, two oligonucleotides sequences were generated; the base element of each sequence containing 32 unique combinations of the 64 possible trinucleotides.

A genetic algorithm (29,30) was used to generate a set of solutions under selection from the following fitness criterion: $f_k = \sigma_k - \mu_k$, where f_k was the fitness criterion, σ_k was the number of unique trinucleotides from the set of desired nucleotides in the sequence and μ_k was the number of bases in the final oligonucleotide that can self-hybridize. The algorithm was initialized with a population of $M = 4096$ random solution vectors, each reproduced through crossovers and modified through mutation at a rate of 5×10^{-3}. Each derived population was culled using the fitness criterion. The solution vectors were oligonucleotides of 34 nt comprised of four values corresponding to the four bases A, G, C and T.

The set of desired nucleotides was selected to ensure that the trinucleotides represented in each of the two 34-nt sequences were distinct from each other, but that all 64 trinucleotides were present in the final two oligonucleotides. The number of iterations required for convergence to the desired solution was approximately 10 000. For some solutions, the length constraint was relaxed to be 36, which substantially reduced the number of iterations. In those cases, the final solution had a single repeated trinucleotide that was identified by inspection and removed.

The two selected oligonucleotides were modified by adding d(CA) repeats at the 5′- and 3′-termini. The 3′-terminal hexanucleotide CACACA was added to achieve optimal binding of primase. CAGACACACA was added to the 5′-terminus to allow for elongation (31). Both oligonucleotides were terminated with a 3′-propanediol to prevent primase from elongating from the end
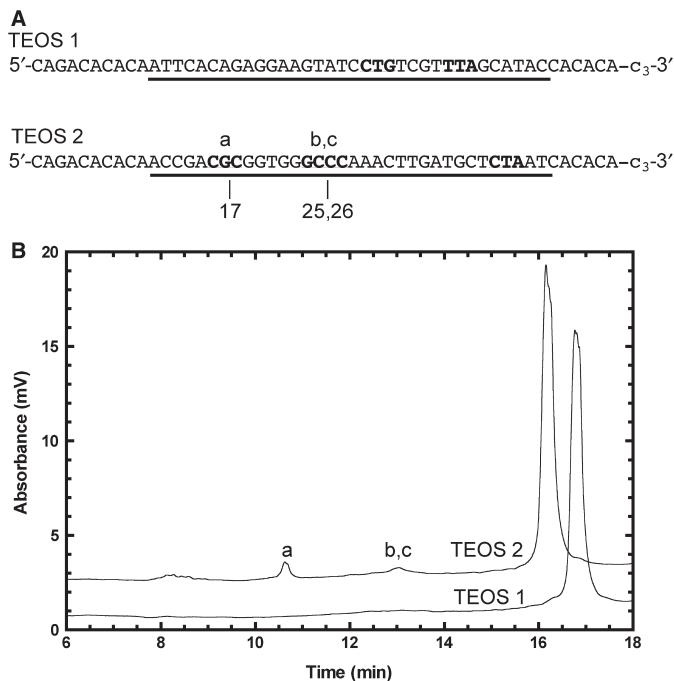

Figure 1. Assessment of *A. aeolicus* primase activity. (**A**) Sequence of TEOS 1 and TEOS 2 with the 64 trinucleotides underlined and bacterial trinucleotide initiation sites denoted in bold. Small letters above the selected trinucleotides correspond to peaks in the chromatograms in (B). Numbers below the sequences correspond to the lengths of primers that terminate at the 5′-end of TEOS 2. (**B**) Denaturing HPLC analysis of the resulting RNA polymers synthesized by *A. aeolicus* primase with TEOS 1 and TEOS 2. The peaks labeled a, b and c eluted at 10.6 and 13.1 min and corresponded to primers of ~17 and 26 nt in length, respectively.

of a stabilized 3′-hairpin (31,32). These two selected oligonucleotides are hereafter referred to as trinucleotide evaluation oligonucleotide sequence 1 (TEOS 1) and TEOS 2 (Figure 1A). TEOS 1 and TEOS 2 were each used in the primer synthesis assay to evaluate the activity of the primase of interest, as well as to screen for trinucleotide specificity.

Trinucleotide-specific templates and trinucleotide specificity confirmation

To delineate the trinucleotide recognition sequence of *A. aeolicus* primase, the most likely trinucleotide initiation sequence was predicted by evaluation of the length of the primers obtained in the activity screen using TEOS 1 or TEOS 2. *Aquifex aeolicus* primase was then used in more detailed analyses using trinucleotide-specific templates to test the predicted recognition trinucleotides. The trinucleotide-specific DNA templates consisted of the sequence 5′-CAGA(CA)₅ <u>XYZ</u> (CA)₃-1,3-propanediol-3′, where XYZ was CAC, CCA, CCC, CCG, CGA, CGC, CGG, CTA, CTC, CTG, CTT, GCC, GCG, GGC, GGG, GGT, TTA or TTG. These 23-mer templates contained the minimal elements needed to support bacterial primase activity, as described previously (33), and were used in the primer synthesis assay with a high

concentration of *A. aeolicus* primase to maximize the length of the primers produced. These experiments tested overall trinucleotide specificity and allowed for kinetic studies of *A. aeolicus* primase.

RNA primer synthesis assay and thermally denaturing HPLC analysis

All RNA primer synthesis reactions were performed in 100 µl nuclease-free water containing 50 mM HEPES, 100 mM potassium glutamate, pH 7.5, 10 mM DTT, 10 mM magnesium acetate and either 1 µM of TEOS 1 or TEOS 2 or 2 µM of the 23-mer trinucleotide-specific template. The concentration of each NTP was 400 µM unless otherwise noted and the primase concentration was 2 µM. The composition of these reaction mixtures were similar to those conditions previously determined to be optimal for *E. coli* and *S. aureus* primases (25,31). Since the optimum temperature for *A. aeolicus* activity was unknown, this primase was incubated with each of the TEOS templates sequentially for 30 min at 65, 75 and 85°C in the first set of experiments. After its template specificity was determined, the temperature for maximum enzyme activity was determined by incubating *A. aeolicus* primase in separate reactions, each incubated at a different temperature that ranged from 25°C to 85°C for 1 h. Since *A. aeolicus* primase exhibited the highest activity at 55 or 65°C, all subsequent primer synthesis reactions were incubated at 55°C for 1 h. After incubation, the samples were desalted in a Sephadex G-25 spin column and dried using a speed vacuum. The pellet was resuspended in water to 1/10th of the original volume of the sample and 8 µl of that sample was analyzed by HPLC under thermally denaturing conditions at 80°C as previously described (28).

The denaturing HPLC elution gradient was adjusted for optimal separation of primer products and ssDNA template peaks on a DNA Sep column. For the primase activity reactions, the primers synthesized in the reactions containing either TEOS 1 or TEOS 2 were eluted with a flow rate of 0.9 ml/min using a gradient of 0–10% acetonitrile over 24 min and for the reactions with the 23-mer trinucleotide-specific template, a gradient 0–8.8% acetonitrile over 16 min was used. Both the RNA and ssDNA templates were detected by UV absorbance at 260 nm. Retention times of the single-stranded oligoribonucleotide and oligodeoxyribonucleotide standards were correlated to the retention times of the appropriate oligonucleotide standard to confirm composition and length. The RNA standards utilized in this study consisted of the sequence 5′-AGUGUGUGUGUG-3′ (12-mer), 5′-AGUGUGUG UGUGUC-3′ (14-mer) and 5′-AGUGUGUGUGUGUC UG-3′ (16-mer).

The moles of primers were quantified by using the template as an internal standard (34). First, the area under each RNA peak and the template peak were background-corrected and each peak area summed such that it had units of mV × min. Next, all areas were then divided by their relative extinction coefficients. Finally, the moles of template in each reaction were used in conjunction with its mole-adjusted area to determine the moles of each primer

species. Total primer synthesis is the sum of moles of all primer lengths. This approach also normalized the variability introduced to the system during sample preparation and injection into the HPLC column.

RESULTS

Functional assessment of TEOS 1 and TEOS 2 oligonucleotide templates

To determine whether oligonucleotides that together comprise all 64 trinucleotides would support primase activity, TEOS 1 and TEOS 2 (Figure 1A) were tested in the primer synthesis assay using the primases from *E. coli* and *S. aureus*. Primers were produced by *E. coli* primase using TEOS 1 and by *S. aureus* primase with TEOS 2 (data not shown). This was expected since TEOS 1 contained the *E. coli* initiation trinucleotide d(CTG) and since d(CTA), the preferred recognition sequence for *S. aureus*, was present in TEOS 2. These results demonstrated that both TEOS 1 and TEOS 2 could be useful in evaluating the activity of an uncharacterized bacterial primase.

Analysis of *A. aeolicus* primase activity using TEOS 1 and TEOS 2

Since the optimum temperature for *A. aeolicus* primase activity was unknown, the samples were initially incubated at three sequential reaction temperatures, specifically 65, 75 and 85°C. TEOS 1 did not generate any primers and was not further utilized, whereas TEOS 2 generated two species of primers that eluted at 10.6 and 13.1 min (Figure 1B). These results demonstrated that the truncated *A. aeolicus* primase was active and contained the necessary components for *de novo* primer synthesis.

The predominant peak at 10.6 min using TEOS 2 comprised the most homogeneous and abundant population (peak a in Figure 1B), presumably due to *A. aeolicus* primase initiating primer synthesis from a single position in TEOS 2 and terminating elongation at the 5′ nucleotide of this template. A comparison of the elution profile of RNA polymer standards with the primer peak at 10.6 min indicated that this RNA polymer was ~17 nt in length. Similarly, a comparison of the peak that eluted at 13.1 min with the RNA oligonucleotide standards indicated that these primer products were about 26 nt in length. Examination of the TEOS 2 sequence suggested that initiation of primer synthesis by *A. aeolicus* primase for the 10.6 min peak corresponded to d(CGC), d(GCG), d(CGG) or d(GGT), whereas the 13.1 min peak correlated with d(GGG), d(GGC), d(GCC) or d(CCC). These predictions were then tested using trinucleotide-specific templates.

Preferred recognition trinucleotide sequence for *A. aeolicus* primase is d(CCC)

To identify the initiation trinucleotides recognized by *A. aeolicus* primase, a series of trinucleotide-specific templates were synthesized for testing as described in Materials and methods section. When templates 1 through 4 (Figure 2A) were analyzed, the full-length primer

A

Template 1: 5′-d(CAGACACACACACA $\overline{\text{C}}$**CC C**ACACA)-R-3′

Template 2: 5′-d(CAGACACACACACA $\overline{\text{G}}$**CC C**ACACA)-R-3′

Template 3: 5′-d(CAGACACACACACA $\overline{\text{C}}$**GC C**ACACA)-R-3′

Template 4: 5′-d(CAGACACACACACA $\overline{\text{C}}$**CG C**ACACA)-R-3′

Template 5: 5′-d(CAGACACACACACA CTG CACACA)-R-3′

Template 6: 5′-d(CAGACACACACACA CTA CACACA)-R-3′

Template 7: 5′-d(CAGACACACACACA TTA CACACA)-R-3′
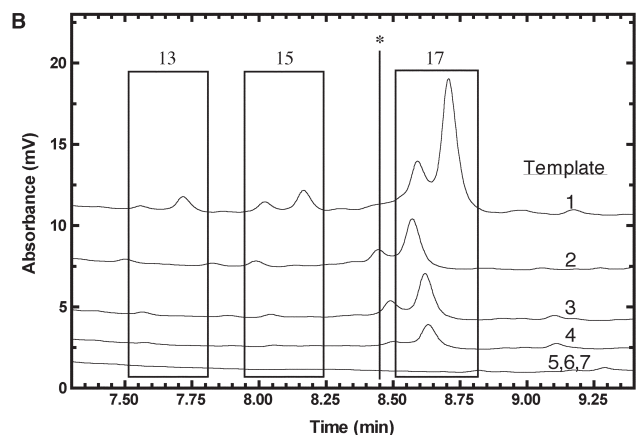
where R = 1,3-propanediol

B

Figure 2. Determination of *A. aeolicus* primase initiation specificity. (**A**) Selected sequences of the 23-mer templates containing the trinucleotide of interest that was tested in (**B**). (**B**) Denaturing HPLC detection and assessment of primer products obtained using selected trinucleotide-specific templates. The trinucleotide of interest in the template as originally designed is denoted in bold and the trinucleotide recognized by *A. aeolicus* primase is in bold. All reactions were performed in triplicate, but only representative chromatograms are shown. The elution times for the 15-mer and 17-mer primers are shown and were based on the elution profiles obtained for the RNA standards. The minor peak adjacent to the predominant full-length 17-mer is due to misincorporation and is discussed in the results. The asterisk indicates the elution time for a 16-mer primer synthesized by primases from either *S. aureus* or *G. stearothermophilus* on the d(CTA) template.

products obtained were longer than the predicted 16-mer when compared with the RNA standards (Figure 2B). The size of these full-length RNA polymers generated by *A. aeolicus* primase on the 23-mer templates were also compared with the more hydrophobic full-length 16-mer primers produced by the well-characterized primases from *S. aureus* and *G. stearothermophilus* on the d(CTA) 23-mer template (25,35). The 16-mer primers synthesized by *S. aureus* and *G. stearothermophilus* on their preferred trinucleotide contained the most hydrophobic ribonucleotide base, adenine, instead of a guanine. Together these comparisons indicated that the full-length RNA products produced by *A. aeolicus* primase on templates 1 through 4 were 17 nt in length. These findings prompted a review of the initiation sequence and it was noted that in each of the three templates, an additional nucleotide 3′ to the originally designed unique trinucleotide was potentially responsible for the longer product (Figure 2A). This notion was consistent with the model in which *A. aeolicus*

primase initiates primer synthesis on the middle nucleotide in the first recognition trinucleotide nearest to the 3′ end of the ssDNA template.

As shown in Figure 2B, template 1 yielded the largest number of primers (17.4 ± 0.5 pmol) in the presence of 2 μM primase (200 pmol) and 2 μM template (200 pmol). Template 2 generated the second highest amount of primers (7.6 ± 0.4 pmol), whereas templates 3 and 4 produced fewer primers (5.4 ± 0.4 and 2.9 ± 0.5 pmol, respectively). The minor peak adjacent to the predominant full-length 17-mer was due to misincorporation, as is discussed later. The predominant and full-length RNA polymers derived from template 1 eluted at 8.71 min, later than the full-length primers synthesized on templates 2 (8.57 min), 3 (8.62 min) and 4 (8.62 min). These primer elution profiles were consistent with the slightly more hydrophobic 17-mer primer derived from the overlined d(CCC) in template 1 (Figure 2A) eluting shortly after the full-length 17-mer RNAs derived from the initiation trinucleotides d(GCC) and d(CGC) in templates 3 and 4, respectively. This primer elution profile was also consistent with *A. aeolicus* primase initiating on d(CCC) in template 2 and the resulting 17-mer primer eluting slightly earlier than the 17-mer RNA polymer derived from d(CCC) in template 1, due to the presence of a hydrophilic cytosine instead of a more hydrophobic guanine. However, although *A. aeolicus* primase initiated primer synthesis on d(CCC) in both templates 1 and 2, the difference in the amount of RNA polymers produced on these ssDNA templates suggested that context, such as a single base adjacent to the initiation trinucleotide, influences enzyme activity.

The specificity of *A. aeolicus* primase activity was further evaluated by testing templates that contained the currently known recognition trinucleotides for the primases from *E. coli*, *S. aureus* and *G. stearothermophilus*, specifically d(CTG), d(CTA) and d(TTA) (Figure 2A). No observable primer products were synthesized by *A. aeolicus* primase on these templates (Figure 2B), confirming that this enzyme preferentially initiates *de novo* primer synthesis on trinucleotides that differ from those used by other bacterial primases.

Collectively, these results indicated that the trinucleotides d(CCC), d(GCC) and d(CGC) in templates 1 through 4, respectively, were generating a significant amount of primers (Figure 2). In addition, these results indicated that the majority of RNA polymers generated with TEOS 2 and eluting at 10.6 min were initiating from d(CGC) within that template, correlating with a RNA polymer that is also 17 nt in length (Figure 1). The broader peak that eluted at 13.1 min with the reactions containing TEOS 2 was consistent with primers initiating from the preferred trinucleotide d(CCC), as well as from d(GCC). The less than optimum positioning of these two later trinucleotides and the apparent limitations of *A. aeolicus* primase to produce primers that extend to the end of the TEOS 2 template most likely contributed to a reduction in the amount of full-length, template-derived, 25-mer and 26-mer RNA polymers. Moreover, the 9- to 13-mer primers observed at approximately 8.3 min in Figure 1B most likely initiated from d(CCC) or d(GCC) in TEOS 2 with
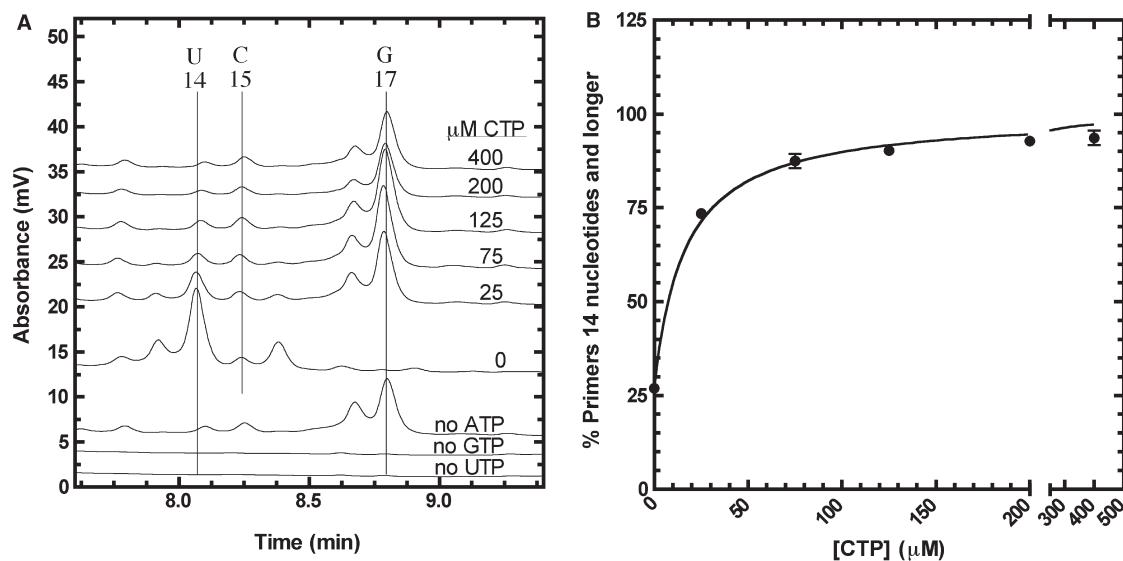
Figure 3. Initiating nucleotide in primers synthesize by *A. aeolicus* primase on the preferred d(CCC) template. (**A**) Chromatograms of primers produced as a function of CTP concentration and in the absence of ATP, GTP, or UTP, but with the presence of 400 μM CTP as denoted. (**B**) The fraction of synthesized RNA polymers 14 nt and longer in the presence of various CTP concentrations was quantified and then fit to a modified Michaelis–Menten equation that allowed for activity in the absence of CTP. The fit indicated that $Y(0\,\mu M\ CTP) = 27 \pm 1\%$, $K_{50\%} = 16 \pm 1\,\mu M$ and $R^2 = 0.994$. The error bars represent the standard deviation of triplicate experiments and in most cases the error was smaller than the symbol on the plot.

primer synthesis terminating prior to reaching the end of the template.

Identification of initial ribonucleotide in primer synthesized by *A. aeolicus* primase

The impact on *de novo* primer synthesis by *A. aeolicus* primase in the absence of GTP, UTP, ATP and CTP was evaluated to further verify trinucleotide specificity and to identify the initial ribonucleotide copied into the primer. No primers were synthesized when GTP or UTP were omitted from the reaction with the preferred d(CCC) template (Figure 3A), consistent with the require- ment for two GTPs during dinucleotide formation or initiation and the requirement for GTP and UTP during elongation. The omission of ATP from reactions with either the d(CCC) template (Figure 3A) or d(GCC) tem- plate (data not shown) increased *A. aeolicus* primase activ- ity by 10%. This was expected since both templates lacked thymine and the absence of ATP most likely enhanced the activity of *A. aeolicus* primase by reducing the time needed to sample the NTP pool at each template position during primer synthesis (36). In the absence of CTP with the d(GCC) template, no primers were generated, indicating that *A. aeolicus* primase required this ribonu- cleotide for dinucleotide polymerization (data not shown). However, increases in the concentration of CTP resulted in a proportional increase in the amount of full- length 17-mer primers, suggesting that *A. aeolicus* primase initiated primer synthesis complementary to the middle cytosine in the recognition trinucleotide d(GCC).

To verify that the 17-mer RNA primers synthesized by *A. aeolicus* primase initiated complementary to the middle nucleotide in its recognition sequences, CTP titration

experiments were performed with the preferred d(CCC) template. The template was designed so that the omission of CTP generated a primer that was 3 nt shorter than full-length RNA primers. When the preferred d(CCC) template with the unique guanine at the antepenultimate position was used in the CTP titration experiments, the absence of CTP in the primer synthesis reactions resulted in predominately a 14-mer RNA polymer (Figure 3A). These data confirmed that the initial ribonucleotide in the primers produced by *A. aeolicus* primase was com- plementary to the middle nucleotide in the recognition trinucleotide.

Interestingly, in the presence of CTP and the other NTPs, a minor peak adjacent to the predominant full- length 17-mer was apparent (Figures 2 and 3). Since the elution time for each RNA polymer length is spaced by about 0.3 min, the minor peak that eluted just prior to each 17-mer was due to a difference in sequence composi- tion and not length, indicating misincorporation. This observation was further supported when CTP was omitted from the primer synthesis reactions with the preferred d(CCC) template. In the absence of CTP with the pre- ferred d(CCC) template containing a unique guanine at the antepenultimate position, some primers were longer than the expected 14-mer, indicating that *A. aeolicus* pri- mase was misinserting nucleotides (Figure 3). In fact, the fit to the data indicated that $27 \pm 1\%$ of the primers were longer than expected, giving a misinsertion rate similar to the primases from *S. aureus* and *E. coli* (31,34).

As the CTP concentration was increased, the percent of primers that were longer than the 14-mer increased in a saturable manner to give a $K_{50\%}$ of $16 \pm 1\,\mu M$ CTP during the elongation part of the reaction (Figure 3B). This K_{M}- like kinetic affinity during the elongation step was similar to the same parameter for primases from *S. aureus* and
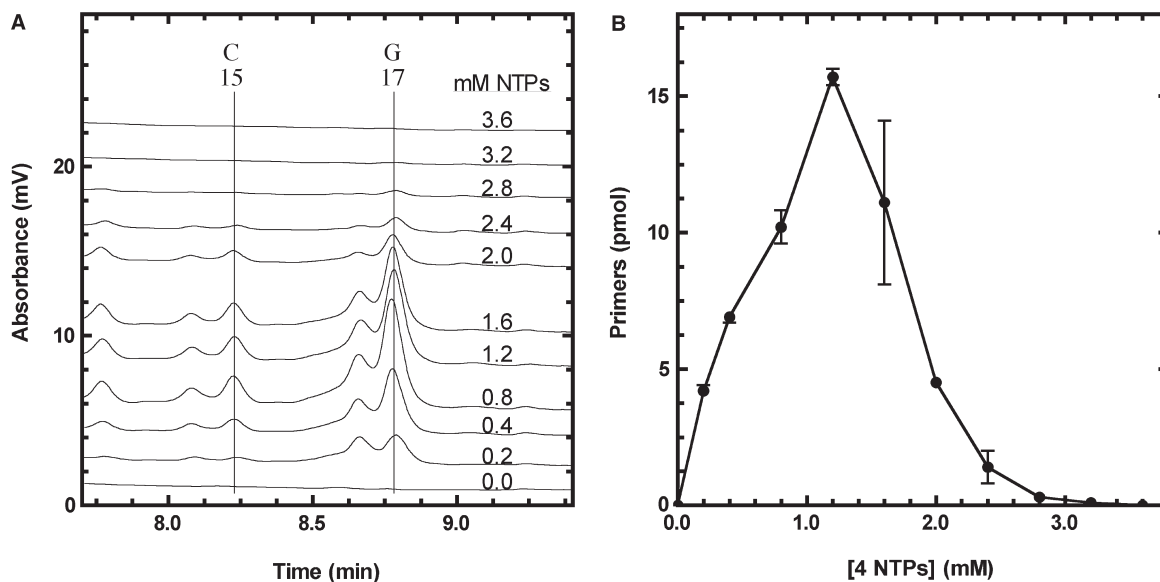
Figure 4. Ribonucleoside triphosphate dependence of *A. aeolicus* primase activity on the preferred d(CCC) template. (**A**) Chromatograms of primers synthesized in which the concentration of all four NTPs were varied simultaneously as indicated and detected by denaturing HPLC. (**B**) The picomole of primers produced were quantified and plotted versus NTP concentration. The error bars represent the SD of triplicate experiments.

E. coli, indicating that all primases share a common elongation mechanism. This elongation kinetic affinity appears to be orders of magnitude weaker than the overall NTP kinetic affinity (Figure 4), demonstrating the vast difference between the initiation and elongation phases of primer synthesis for the *A. aeolicus* primase. However, caution is warranted in these analyses since the kinetics of the incorporation of a single nucleotide during elongation is compared with overall primer synthesis in the presence of all four NTPs.

Aquifex aeolicus primase activity dependence for ribonucleotides

The concentration of the NTP mixture was varied to optimize *A. aeolicus* primase activity using the preferred d(CCC) template. The striking results showed that primer synthesis activity rose almost linearly to a maximum with 1.2 mM of each NTP and then fell even more rapidly until there was virtually no activity with 2.8 mM of each NTP (Figure 4). At the optimum NTP concentration, the enzyme synthesized 15.7 ± 0.3 pmol primers even though the reaction contained 200 pmol of primase and 200 pmol of template. These data indicated that only 8% of the templates were primed at the optimum temperature and NTP concentration for maximum activity. Under similar conditions with the primases from *S. aureus* and *E. coli* and their recognition trinucleotides, 49 and 25% of the templates were primed, respectively (28,34). *Aquifex aeolicus* primase appeared to have a very high K_M for the NTP mixture, but was even more strongly inhibited by it. Since the NTP inhibitory effect was certainly exerting itself as the activity neared its peak, 400 µM of each NTP was used in the primer synthesis reactions to minimize inhibition.

Temperature optimum for *A. aeolicus* primase activity

To delineate the optimum temperature for maximum activity of *A. aeolicus* primase, the primer synthesis assay was performed at separate temperatures from 25°C to 85°C using the d(GCC) template. Based on the amount of primers synthesized, *A. aeolicus* primase activity was observed to be the highest at 55°C (3.4 ± 0.4 pmol primers) and 65°C (3.2 ± 0.4 pmol primers) with about half the activity at 75°C (2.4 ± 0.2 pmol primers) (Figure 5). Little or no activity was observed at 42°C and below (Figure 5). At 55°C, the predominant primer synthesized was a full-length 17-mer, whereas at 65 and 75°C the abundance of the 17-mer decreased. The minor peak that eluted just prior to the full-length RNA polymer for the reactions incubated at 65 and 75°C indicated an increase in the amount of misinsertion, especially at the higher temperature. At 85°C, *A. aeolicus* primase was only slightly active, although this is the optimal temperature for culturing this microbe. A small amount of template hydrolysis was also seen at this elevated temperature. Collectively, these results suggested that auxiliary proteins and/or factors are most likely required to stabilize the interaction between the genomic template and primase at the higher physiological temperatures for this extreme thermophile.

DISCUSSION

Although neither the initiation specificity nor the optimum functional temperature of a hyperthermophilic bacterial primase had previously been determined, the two-step empirical method used in this study showed that *A. aeolicus* primase was active and that the preferred initiation trinucleotide for *A. aeolicus* primase was d(CCC) and to a lesser degree d(GCC) and d(CGC). These recognition sequences are not used by any other
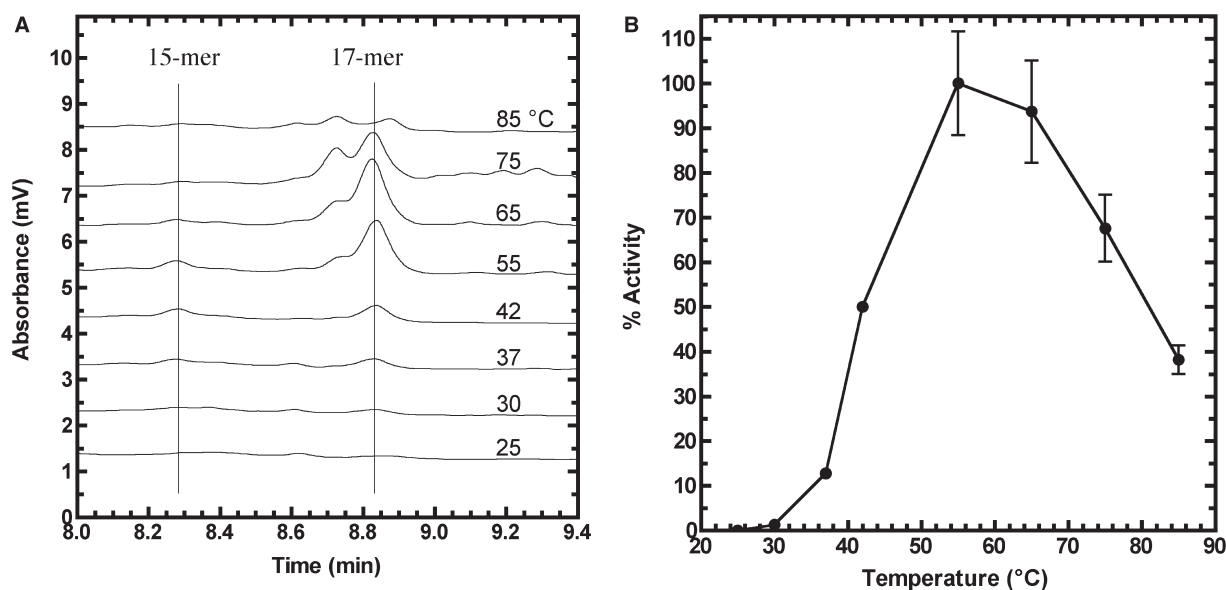
Figure 5. Temperature dependence of *A. aeolicus* primase activity using the d(GCC) template. (**A**) Chromatograms of primers synthesized by *A. aeolicus* primase and detected by denaturing HPLC, after incubation at the temperatures specified. (**B**) Plot of percentage activity versus temperature in which the total area under the RNA primer peaks was first normalized to the areas under the template peaks and was then normalized to the highest activity. The error bars represent the standard deviation of triplicate experiments.

bacterial or phage primase examined to date. Previous studies have demonstrated that the RNA primers generated by primase from both *S. aureus* and *E. coli* initiate complementary to the middle T in their respective initiation trinucleotide sequence (25,31). This study demonstrated that *A. aeolicus* primase also initiates *de novo* primer synthesis complementary to the middle nucleotide in the recognition trinucleotide, indicating that this is a conserved property for all microbial primases examined to date.

Knowledge of the preferred recognition sequence d(CCC) and temperature optimum for *A. aeolicus* primase has provided important information for further biochemical characterization of this enzyme. Although the *A. aeolicus* primase used in these studies did not have the C-terminal domain which had been deleted to obtain a soluble recombinant protein (26), the N-terminal zinc binding domain and RNA polymerase domain were sufficient for *de novo* primer synthesis. This is similar to findings in the *E. coli* and bacteriophage T7 systems in which only the domains containing the Cys_4 zinc finger and active site for RNA polymerization were necessary to retain initiation specificity and catalyze template-directed oligoribonucleotide synthesis (26,37).

In general, enzymes from thermophiles tend to be more stable, less flexible and less active than related proteins from mesophilic organisms (38). Although the relationship of molecular flexibility and activity is poorly defined, intermolecular and intracellular molecular associations most likely play a vital role in thermostability. For instance, crystal structure analyses indicated that the linker between the zinc binding domain and the RNA polymerase domain in *A. aeolicus* primase is shorter than the corresponding region in *E. coli* primase (26). Although the shorter linker in *A. aeolicus* primase is less flexible than the

corresponding linker in *E. coli* primase, the proposed two dominant transition states of this enzyme do not appear to have been compromised in terms of productive and nonproductive conformations (26). In the extended productive form, the zinc-binding domain is uncoupled from the associated RNA polymerase domain and presumably interacts with the template DNA that is bound to another primase's active site in the RNA polymerase domain in *trans*, resulting in primer synthesis. Therefore, the low activity of *A. aeolicus* primase may be due in part to reduced flexibility and the intermolecular protein–protein interactions required for primer synthesis, compared with the intramolecular interactions observed between the zinc binding and RNA polymerase domains of a primase from a mesophilic bacterium (34). Moreover, the C-terminal domain of *A. aeolicus* primase most likely plays a critical role in stabilizing intermolecular interactions *in vivo*.

Hyperthermophiles do not appear to replicate or repair their DNA by error-prone mechanisms (3). Despite the misincorporation propensities observed for *A. aeolicus* primase in this study and for other bacterial primases (31,34), other replications proteins may modulate primase activity as is implicated for archaeal primases (39). These misinsertions are most certainly corrected by primer-removing enzymes such as DNA polymerase I. In addition, although the deleterious effects of high temperatures on nucleic acids have been well documented (10,40), there is some evidence that biomolecular stability at high temperatures is achieved by elevated salt concentrations, as well as by the presence of polyamines, cationic proteins and/or supercoiling (38). Therefore, further study is required to provide additional insight on the mechanisms by which hyperthermophiles such as *A. aeolicus* accurately replicate their genome.

Our study demonstrated that the initiation sequences of *A. aeolicus* primase are comprised entirely of cytosines and/or guanines even though its chromosome composition is only 43% GC and there does not appear to be any obvious DNA strand bias for the recognition trinucleotides. There is also a disparity in the initiation trinucleotide sequences for primases from Firmicutes [d(CTA) and d(TTA) for *S. aureus* and *G. stearothermophilus*] and Proteobacteria [d(CTG) for *E. coli*] despite the presumed proximity of these phyla. In fact, the congruence in primase specificities for mesophilic *S. aureus* and thermophilic *G. stearothermophilus* becomes even more striking in this light. Therefore, our findings suggest that the hyperthermophilic environment may require a larger number of adaptations than does the thermophilic habitat.

Comparisons of the completed genomes from hyperthermophiles, including *A. aeolicus* and archaeon *Methanococcus jannaschii*, with bacterial genomes of mesophiles revealed that stable RNAs have a high $G + C$ content even if the genome has a comparatively low $G + C$ content (8). These stable RNAs include 16S RNAs and tRNAs where maintenance of structural stability is particularly important (38). The extensive hydrogen bonding and modified bases in these stable RNAs protects them against the deleterious effects of heat by increasing their melting temperature (1). Similarly, highly expressed genes have been determined to derive from relatively GC-rich codons more often than genes expressed at a low level in AT-rich microbial genomes (41). These findings suggest that GC base pairing facilitates microbial fitness to thermophily by stabilizing various factors of the translation machinery, particularly during cell growth. Our results show that primase initiation specificity represents another example of thermophilic adaptation, this time for the replication machinery.

Primases from bacteriophage T7, *S. aureus* and *E. coli* have been shown to have a higher affinity for ssDNA templates that contain the respective initiation sequence than for similar ssDNA not containing the preferred initiation trinucleotide (25,34,42,43). Accordingly, our findings suggest that the *A. aeolicus* initiation sequence that is comprised of cytosines and guanines is necessary for stabilizing the interaction among the primase–NTPs–ssDNA complex during initiation, especially due to the geothermal context of this microbe. The additional bonding energy between G–C versus A–T base pairs would provide greater thermostability in this complex during the initial steps of primer synthesis. Our results suggest that *A. aeolicus* primase appears to utilize the thermodynamic strength of GC basepair stacking to help assemble the complex between primase, CTP, GTP and ssDNA. In fact, the stacking free energy of the *A. aeolicus* primase initiating ribonucleotides with their complements in the template may be nearly twice as favorable (-2.17 kcal/mol) as the basepair stacking energy that takes place in mesophilic microbial DNA–RNA interactions (-1.00 kcal/mol) (44). Since dinucleotide polymerization is the rate-limiting step during *de novo* primer synthesis (31), our findings accentuate the importance of a stable interaction during the initial priming step for subsequent replication of the *A. aeolicus* genome.

We anticipate that primases from other hyperthermophilic bacteria will also utilize initiation trinucleotides that are comprised entirely of cytosines and/or guanines. Our future studies will focus on determining the relationship of template recognition and primase structure.

REFERENCES

1. Kowalak,J.A., Dalluge,J.J., McCloskey,J.A. and Stetter,K.O. (1994) The role of posttranscriptional modification in stabilization of transfer RNA from hyperthermophiles. *Biochemistry*, **33**, 7869–7876.
2. Schlee,S. and Reinstein,J. (2002) The DnaK/ClpB chaperone system from Thermus thermophilus. *Cell Mol. Life Sci.*, **59**, 1598–1606.
3. Grogan,D.W. (1998) Hyperthermophiles and the problem of DNA instability. *Mol. Microbiol.*, **28**, 1043–1049.
4. Frick,D.N. and Richardson,C.C. (2001) DNA primases. *Annu. Rev. Biochem.*, **70**, 39–80.
5. Grompe,M., Versalovic,J., Koeuth,T. and Lupski,J.R. (1991) Mutations in the *Escherichia coli* dnaG gene suggest coupling between DNA replication and chromosome partitioning. *J. Bacteriol.*, **173**, 1268–1278.
6. Liu,J., Dehbi,M., Moeck,G., Arhin,F., Bauda,P., Bergeron,D., Callejo,M., Ferretti,V., Ha,N., Kwan,T. *et al.* (2004) Antimicrobial drug discovery through bacteriophage genomics. *Nat Biotechnol.*, **22**, 185–191.
7. Kobayashi,K., Ehrlich,S.D., Albertini,A., Amati,G., Andersen,K.K., Arnaud,M., Asai,K., Ashikaga,S., Aymerich,S., Bessieres,P. *et al.* (2003) Essential *Bacillus subtilis* genes. *Proc. Natl Acad. Sci. USA*, **100**, 4678–4683.
8. Deckert,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L., Graham,D.E., Overbeek,R., Snead,M.A., Keller,M., Aujay,M. *et al.* (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature*, **392**, 353–358.
9. Galtier,N. and Lobry,J.R. (1997) Relationships between genomic $G + C$ content, RNA secondary structures, and optimal growth temperature in prokaryotes. *J. Mol. Evol.*, **44**, 632–636.
10. Marmur,J. and Doty,P. (1962) Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *J. Mol. Biol.*, **5**, 109–118.
11. SantaLucia,J. Jr. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA*, **95**, 1460–1465.
12. De Luchi,D., Gouyette,C. and Subirana,J.A. (2003) The influence of size on the thermal stability of oligonucleotides: the case of AT sequences. *Anal. Biochem.*, **322**, 279–282.
13. Hurst,L.D. and Merchant,A.R. (2001) High guanine-cytosine content is not an adaptation to high temperature: a comparative analysis amongst prokaryotes. *Proc. Biol. Sci.*, **268**, 493–497.
14. Burggraf,S., Olsen,G.J., Stetter,K.O. and Woese,C.R. (1992) A phylogenetic analysis of Aquifex pyrophilus. *Syst. Appl. Microbiol.*, **15**, 352–356.
15. Wang,H.C. and Hickey,D.A. (2002) Evidence for strong selective constraint acting on the nucleotide composition of 16S ribosomal RNA genes. *Nucleic Acids Res.*, **30**, 2501–2507.

16. Coenye,T. and Vandamme,P. (2004) A genomic perspective on the relationship between the Aquificales and the epsilon-Proteobacteria. *Syst. Appl. Microbiol.*, **27**, 313–322.
17. Griffiths,E. and Gupta,R.S. (2004) Signature sequences in diverse proteins provide evidence for the late divergence of the Order Aquificales. *Int Microbiol.*, **7**, 41–52.
18. Skophammer,R.G., Servin,J.A., Herbold,C.W. and Lake,J.A. (2007) Evidence for a gram-positive, eubacterial root of the tree of life. *Mol. Biol. Evol.*, **24**, 1761–1768.
19. Cha,T.A. and Alberts,B.M. (1986) Studies of the DNA helicase-RNA primase unit from bacteriophage T4. A trinucleotide sequence on the DNA template starts RNA primer synthesis. *J. Biol. Chem.*, **261**, 7001–7010.
20. Tabor,S. and Richardson,C.C. (1981) Template recognition sequence for RNA primer synthesis by gene 4 protein of bacteriophage T7. *Proc. Natl Acad. Sci. USA*, **78**, 205–209.
21. Tseng,T.Y., Frick,D.N. and Richardson,C.C. (2000) Characterization of a novel DNA primase from the Salmonella typhimurium bacteriophage SP6. *Biochemistry*, **39**, 1643–1654.
22. Kitani,T., Yoda,K., Ogawa,T. and Okazaki,T. (1985) Evidence that discontinuous DNA replication in *Escherichia coli* is primed by approximately 10 to 12 residues of RNA starting with a purine. *J. Mol. Biol.*, **184**, 45–52.
23. Bhattacharyya,S. and Griep,M.A. (2000) DnaB helicase affects the initiation specificity of *Escherichia coli* primase on single-stranded DNA templates. *Biochemistry*, **39**, 745–752.
24. Thirlway,J. and Soultanas,P. (2006) In the *Bacillus stearothermophilus* DnaB-DnaG complex, the activities of the two proteins are modulated by distinct but overlapping networks of residues. *J. Bacteriol.*, **188**, 1534–1539.
25. Koepsell,S.A., Larson,M.A., Griep,M.A. and Hinrichs,S.H. (2006) *Staphylococcus aureus* helicase but not *Escherichia coli* helicase stimulates *S. aureus* primase activity and maintains initiation specificity. *J. Bacteriol.*, **188**, 4673–4680.
26. Corn,J.E., Pease,P.J., Hura,G.L. and Berger,J.M. (2005) Crosstalk between primase subunits can act to regulate primer synthesis in trans. *Mol. Cell.*, **20**, 391–401.
27. Griep,M.A. and Lokey,E.R. (1996) The role of zinc and the reactivity of cysteines in *Escherichia coli* primase. *Biochemistry*, **35**, 8260–8267.
28. Koepsell,S., Bastola,D., Hinrichs,S.H. and Griep,M.A. (2004) Thermally denaturing high-performance liquid chromatography analysis of primase activity. *Anal. Biochem.*, **332**, 330–336.
29. Holland,J.H. (1992) *Adaptation in Natural and Artificial Systems*, 2nd edn. MIT Press, Cambridge.
30. Goldberg,D.E. (1989) *Genetic Algorithms in Search, Optimization, and Machine Learning*, 1st edn. Addison-Wesley Longman Publishing Co., Inc., Boston.
31. Swart,J.R. and Griep,M.A. (1995) Primer synthesis kinetics by *Escherichia coli* primase on single-stranded DNA templates. *Biochemistry*, **34**, 16097–16106.
32. Johnson,S.K., Bhattacharyya,S. and Griep,M.A. (2000) DnaB helicase stimulates primer synthesis activity on short oligonucleotide templates. *Biochemistry*, **39**, 736–744.
33. Swart,J.R. and Griep,M.A. (1993) Primase from Escherichia coli primes single-stranded templates in the absence of single-stranded DNA-binding protein or other auxiliary proteins. Template sequence requirements based on the bacteriophage G4 complementary strand origin and Okazaki fragment initiation sites. *J. Biol. Chem.*, **268**, 12970–12976.
34. Koepsell,S.A., Larson,M.A., Frey,C.A., Hinrichs,S.H. and Griep,M.A. (2008) *Staphylococcus aureus* primase has higher initiation specificity, binds DNA stronger, but is less stimulated by its helicase than *Escherichia coli* primase. *Mol. Microbiol.*, **68**, 1570–1582.
35. Chintakayala,K., Larson,M.A., Griep,M.A., Hinrichs,S.H. and Soultanas,P. (2008) Conserved residues of the C-terminal p16 domain of primase are involved in modulating the activity of the bacterial primosome. *Mol. Microbiol.*, **68**, 360–371.
36. Viljoen,S., Griep,M.A., Nelson,M. and Viljoen,H. (2005) A macroscopic kinetic model for DNA polymerase elongation and high-fidelity nucleotide selection. *Comput. Biol. Chem.*, **29**, 101–110.
37. Frick,D.N., Baradaran,K. and Richardson,C.C. (1998) An N-terminal fragment of the gene 4 helicase/primase of bacteriophage T7 retains primase activity in the absence of helicase activity. *Proc. Natl Acad. Sci. USA*, **95**, 7957–7962.
38. Daniel,R.M. and Cowan,D.A. (2000) Biomolecular stability and life at high temperatures. *Cell Mol. Life Sci.*, **57**, 250–264.
39. Lao-Sirieix,S.H., Pellegrini,L. and Bell,S.D. (2005) The promiscuous primase. *Trends Genet.*, **21**, 568–572.
40. Lindahl,T. (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709–715.
41. Herbeck,J.T., Wall,D.P. and Wernegreen,J.J. (2003) Gene expression level influences amino acid usage, but not codon usage, in the tsetse fly endosymbiont Wigglesworthia. *Microbiology*, **149**, 2585–2596.
42. Frick,D.N. and Richardson,C.C. (1999) Interaction of bacteriophage T7 gene 4 primase with its template recognition site. *J. Biol. Chem.*, **274**, 35889–35898.
43. Khopde,S., Biswas,E.E. and Biswas,S.B. (2002) Affinity and sequence specificity of DNA binding and site selection for primer synthesis by *Escherichia coli* primase. *Biochemistry*, **41**, 14820–14830.
44. Ornstein,R.L., Rein,R., Breen,D.L. and MacElroy,R.D. (1978) An optimized potential function for the calculation of nucleic acid interaction energies. I. base stacking. *Biopolymers*, **17**, 2341–2360.