(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 December 2004 (09.12.2004)

PCT

(10) International Publication Number
WO 2004/106554 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number:
PCT/US2004/016206

(22) International Filing Date: 22 May 2004 (22.05.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/472,967 23 May 2003 (23.05.2003) US
60/473,054 23 May 2003 (23.05.2003) US

(71) Applicant *(for all designated States except US)*: **ASTRAZENECA AB** [SE/SE]; Global Intellectual Property, S-151 85 Södertälje (SE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **CHEN, Guo, Qiang** [CN/US]; AstraZeneca R & D Boston, 35 Gatehouse Drive, Waltham, MA 02451 (US). **LIU, Ce, Feng** [CN/US]; AstraZeneca R & D Boston, 35 Gatehouse Drive, Waltham, MA 02451 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

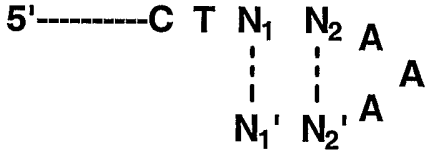
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2004/106554 A2

(54) Title: PRIMASE DNA TEMPLATES



(57) Abstract: Novel DNA primase recognition sites and templates, methods for identifying DNA primase recognition sites and templates, and methods for their use.

PRIMASE DNA TEMPLATES

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is related to U.S. Provisional Application 60/472,967, filed May 23, 2003, and U.S. Provisional Application 60/473,054, filed May 23, 2003, each of which is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

The present invention relates to novel DNA primase recognition sites and templates, methods for identifying DNA primase recognition sites and templates, and methods for their use.

BACKGROUND

DNA primases (primases) play essential roles in the DNA replication process (Frick *et al.*, 2001, Annu. Rev. Biochem., 70:39-80). At the DNA replication fork, DNA primases (DnaG) catalyze the synthesis of short RNA molecules (of about 10 to 12 bases long) from ribonucleoside triphosphates. These RNA molecules are used by DNA polymerases as primers to synthesize the complementary strand of the parent DNA.

Structural differences among primases from various species are significant (Augustin *et al.*, 2001, Nat. Struct. Biol., 8:57-61). Most bacterial and bacteriophage primases are single polypeptide proteins. These single polypeptide primases contain three distinct domains: a Zn-binding domain involved in DNA binding function, a central domain with catalytic activity, and a C-terminal domain. The C-terminal domain has helicase function for some bacteriophage primases such as T7 and P4, or interacts with a helicase in most bacteria. Eukaryotic primases are heterodimeric proteins that complex with a DNA polymerase. The DNA polymerase also complexes with another protein subunit that may be involved in DNA polymerase nuclear translocation. Different from prokaryotic and eukaryotic primases, herpes simplex virus (HSV) primase comprises three subunits, namely UL52, UL5 and UL8.

The three-polypeptide complex contains both primase and helicase activities (Frick *et al.*, 2001, *supra*).

Based on their fundamental functions in the cell and structural differences among various species, primases are potential drug targets for several different therapy areas. For example, compounds that inhibit HSV primase were selected for trial clinical treatment of HSV infections (Kleymann *et al.*, 2001, Nat. Med., 8:392-398). Bacterial primases are considered attractive antibiotic targets, and high-throughput screening (HTS) assays have been developed to find bacterial primase inhibitors (U.S. Patent No. 6,043,038). Genes and protein complexes of primases in some cancer cells have been identified as targets for cancer intervention.

Prior to RNA synthesis, primase binds to a DNA template, by recognizing a certain DNA sequence, termed the primase recognition site, to initiate RNA synthesis (Frick *et al.*, 2001, *supra*). The characteristic DNA sequence of primase recognition sites varies from one primase to another (Table 1).

Table 1. Primase recognition sites

E. coli primase	5'-CTG	
Bacteriophage T7 primase	5'-GTC	
SP6 primase	5'-GCA	
T4 primase	5'-GTT, 5'-GCT	
Herpes simplex virus primase	5'-ACCCTCCCGA	SEQ ID NO:1
Mouse primase	5'-CCA, 5'-CCC	

For example, 5'-CTG-3' is considered to be the recognition site for *Escherichia coli* (*E. coli*) primase (Khopde *et al.*, 2002, Biochemistry, 41:14820-14830). No information is available about the recognition site for primases from many pathogenic bacterial strains such as *Staphyloccus aureus*, *Streptococcus pneumoniae*, *Haemophilus influenzae*. In order to identify inhibitors of primases from these pathogenic strains, it is essential to develop reliable biochemical assays using appropriate templates based upon the recognition site of the relevant primase.

Investigations of primases began nearly 30 years ago; *E. coli* and bacteriophage T7 primases being the most thoroughly studied (Scherzinger *et al.*, 1975, Mol. Gen. Gent., 141:213-32; Swart *et al.*, 1995, Biochemistry, 34:16097-16106). The sequence of a 5'-CTG has generally been considered as the recognition site for *E. coli* primase. The concept of trinucleotide primase recognition site has been generally accepted and applied to studies of many other primases.

Primase recognition site information has been typically determined by identification of the RNA primer sequences that are synthesized by primase, with further experiments to test the sequences in the region where RNA synthesis is initiated. For example, using M13 ssDNA as a template, T7 primase synthesizes RNA primers 5'-pppACCC, 5'-pppACCA and 5'-pppACAC. The primers share the same sequence 5'-pppAC. On the DNA template, all the initiation sites share a 5'-GTC sequence that was considered the recognition site for T7 primase. It has been demonstrated that the C in the 5'-GTC site was essential for T7 primase activity (Frick *et al.*, 1999, J. Biol. Chem., 274:35889-35898).

The development of biochemical assays for primases is important to the identification of specific inhibitors of DNA primase through methods such as screening in the drug discovery process. Biochemical assays tailored to primases from different species are important to the elucidation of and characterization of compounds that inhibit particular isozymes or otherwise exhibit broad specificity. There is a need for the identificaton of DNA primase-specific DNA templates to support assay development and subsequent screening.

SUMMARY

The invention is based, in part, on methods that allow for the rapid identification of DNA primase recognition sites for specific DNA primases. Moreover, the methods of the invention have been used to identify novel DNA primase-recognition sites for specific DNA primases. The identified DNA primase-recognition sites can then serve as DNA templates in order to identify compounds which modulate DNA primase activity.

Accordingly, the invention includes a method for identifying a DNA primase-recognition site. The method includes providing a plurality of oligonucleotides, a DNA primase and ribonucleoside triphosphates; and determining DNA primase activity in order to identify an appropriate DNA primase-recognition site. The oligonucleotides used in the method can include the sequence:

$(X \text{ or } XX \text{ or } XXX)_n\ N_1N_2N_3\ (Y \text{ or } YY \text{ or } YYY)_m$, wherein

X is at each occurrence independently A, T, C, or G;

Y is at each occurrence independently A, T, C, or G;

n is at least 2;

m is at least 2; and

N_1, N_2, and N_3, are independently A, T, C, or G.

In one embodiment, $N_1N_2N_3$ is selected from the group consisting of TTT, TTC, TTA, TTG, TCT, TCC, TCA, TCG, TAT, TAC, TAA, TAG, CTT, CTC, CTA, CTG, CCT, CCC,

CCA, CCG, CAT, CAC, CAA, CAG, CGT, CGC, CGA, CGC, ATT, ATC, ATA, ATG, ACT, ACC, ACA, ACG, AAT, AAC, AAA, AAG, AGT, AGC, AGA, AGG, GTT, GTC, GTA, GTG, GCT, GCC, GCA, GCG, GAT, GAC, GAA, GAG, GGT, GGC, GGA, and GGG.

In one embodiment, the XX or YY is selected from the group consisting of AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, CC, CG, GA, GT, GC, and GG.

Alternatively, the oligonucleotides can include the sequence

N_1N_2 $(X)_p$ $N_2'N_1'$, wherein

X is at each occurrence independently A, T, C, or G;

N_1, N_2, N_1' and N_2' are independently A, T, C, or G;

p is at least 3;

N_1 and N_1' are complementary; and

N_2 and N_2' are complementary.

In another aspect, the invention includes a method for identifying a DNA primase-recognition site by (i) providing a first plurality of oligonucleotides, a DNA primase and a ribonucleoside triphosphate; (ii) providing a second plurality of oligonucleotides, a DNA primase and a ribonucleoside triphosphate; and determining DNA primase activity for (i) and (ii), wherein primase activity is indicative that the oligonucleotide has a DNA primase-recognition site. The first plurality of oligonucleotides can include the sequence

(X or XX or $XXX)_n$ $N_1N_2N_3$ (Y or YY or $YYY)_m$, wherein

X is at each occurrence independently A, T, C, or G;

Y is at each occurrence independently A, T, C, or G;

n is at least 2;

m is at least 2; and

N_1, N_2, and N_3, are independently A, T, C, or G.

In one embodiment, $N_1N_2N_3$ is selected from the group consisting of TTT, TTC, TTA, TTG, TCT, TCC, TCA, TCG, TAT, TAC, TAA, TAG, CTT, CTC, CTA, CTG, CCT, CCC, CCA, CCG, CAT, CAC, CAA, CAG, CGT, CGC, CGA, CGC, ATT, ATC, ATA, ATG, ACT, ACC, ACA, ACG, AAT, AAC, AAA, AAG, AGT, AGC, AGA, AGG, GTT, GTC, GTA, GTG, GCT, GCC, GCA, GCG, GAT, GAC, GAA, GAG, GGT, GGC, GGA, and GGG.

In one embodiment, the XX or YY is selected from the group consisting of AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, CC, CG, GA, GT, GC, and GG. In another embodiment, XXX or YYY is CTT or CCT.

The second plurality of oligonucleotides can include the sequence

N_1N_2 $(X)_p$ $N_2'N_1'$, wherein

X is at each occurrence independently A, T, C, or G;

N_1, N_2, N_1' and N_2' are independently A, T, C, or G;

p is at least 3;

N_1 and N_1' are complementary; and

N_2 and N_2' are complementary.

Any DNA primase can be used in the methods described above, for example, the DNA primase can be prokaryotic, eukaryotic or viral. The DNA primase can be from a bacteria such as *Escherichia coli*, *Staphyloccus aureus*, *Streptococcus pneumoniae*, or *Haemophilus influenzae*. In another example, the DNA primase is a bacteriophage DNA primase.

DNA primase activity can be determined by any method known in the art, for example, the DNA primase can be detected by detecting an RNA product, DNA-RNA heterohybrid regions, or pyrophosphate.

In another aspect, the invention includes an oligonucleotide including the sequence, as described above,

(X or XX or XXX)$_n$ $N_1N_2N_3$ (Y or YY or YYY)$_m$, wherein

X is at each occurrence independently A, T, C, or G;

Y is at each occurrence independently A, T, C, or G;

n is at least 2;

m is at least 2; and

N_1, N_2, and N_3, are independently A, T, C, or G.

In another aspect, the invention includes an oligonucleotide including the sequence of

N_1N_2 $(X)_p$ $N_2'N_1'$, wherein

X is at each occurrence independently A, T, C, or G;

N_1, N_2, N_1' and N_2' are independently A, T, C, or G;

p is at least 3;

N_1 and N_1' are complementary; and

N_2 and N_2' are complementary.

In another aspect, the invention includes an oligonucleotide including the sequence of

$(XX)_n$ NNN $(YY)_m$,

wherein

XX is selected from the group consisting of AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, CC, CG, GA, GT, GC, and GG;

YY is selected from the group consisting of AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, CC, CG, GA, GT, GC, and GG;

n is at least 2;

m is at least 2; and

NNN is selected from the group consisting of TTT, TTC, TTA, TTG, TCT, TCC, TCA, TCG, TAT, TAC, TAA, TAG, CTT, CTC, CTA, CTG, CCT, CCC, CCA, CCG, CAT, CAC, CAA, CAG, CGT, CGC, CGA, CGC, ATT, ATC, ATA, ATG, ACT, ACC, ACA, ACG, AAT, AAC, AAA, AAG, AGT, AGC, AGA, AGG, GTT, GTC, GTA, GTG, GCT, GCC, GCA, GCG, GAT, GAC, GAA, GAG, GGT, GGC, GGA, and GGG.

In one embodiment, the XX or YY is selected from the group consisting of CT, TC, GT, and TG. In another embodiment, NNN is selected from the group consisting of ACC and AGT.

In yet another aspect, the invention includes an oligonucleotide including the sequence of SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, or SEQ ID NO:43.

The invention further includes a method for identifying compounds that modulate DNA primase activity. The method includes providing an oligonucleotide comprising the sequence of

$(X \text{ or } XX \text{ or } XXX)_n\ N_1N_2N_3\ (Y \text{ or } YY \text{ or } YYY)_m$, wherein

X is at each occurrence independently A, T, C, or G;

Y is at each occurrence independently A, T, C, or G;

n is at least 2;

m is at least 2; and

N_1, N_2, and N_3, are independently A, T, C, or G;

contacting the oligonucleotide with DNA primase, a ribonucleoside triphosphate and a compound; and determining DNA primase activity, wherein an increase or decrease in primase activity in the presence of the compound compared to a control (no compound), indicates that the compound modulates DNA primase activity.

In one embodiment, $N_1N_2N_3$ is selected from the group consisting of TTT, TTC, TTA, TTG, TCT, TCC, TCA, TCG, TAT, TAC, TAA, TAG, CTT, CTC, CTA, CTG, CCT, CCC, CCA, CCG, CAT, CAC, CAA, CAG, CGT, CGC, CGA, CGC, ATT, ATC, ATA, ATG, ACT, ACC, ACA, ACG, AAT, AAC, AAA, AAG, AGT, AGC, AGA, AGG, GTT,

GTC, GTA, GTG, GCT, GCC, GCA, GCG, GAT, GAC, GAA, GAG, GGT, GGC, GGA, and GGG.

In one embodiment, the XX or YY is selected from the group consisting of AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, CC, CG, GA, GT, GC, and GG.

The invention further includes a method for identifying compounds that modulate DNA primase activity. The method includes providing an oligonucleotide comprising the sequence of

$N_1N_2\ (X)_p\ N_2'N_1'$, wherein

X is at each occurrence independently A, T, C, or G;

N_1, N_2, N_1' and N_2' are independently A, T, C, or G;

p is at least 3;

N_1 and N_1' are complementary; and

N_2 and N_2' are complementary;

contacting the oligonucleotide with DNA primase, a ribonucleoside triphosphate and a compound; and determining DNA primase activity, wherein an increase or decrease in primase activity in the presence of the compound compared to a control (no compound), indicates that the compound modulates DNA primase activity.

The invention further includes a method for identifying compounds that modulate *S. aureus* DNA primase activity. The method includes providing an oligonucleotide selected from the group consisting of SEQ ID NO:7 and SEQ ID NO:43, contacting the oligonucleotide with *S. aureus* DNA primase, a ribonucleoside triphosphate and a compound; and determining DNA primase activity, wherein an increase or decrease in primase activity in the presence of the compound compared to a control (no compound), indicates that the compound modulates DNA primase activity.

The invention further includes a method for identifying compounds that modulate *S. pneumoniae* DNA primase activity. The method includes providing an oligonucleotide of SEQ ID NO:7, contacting the oligonucleotide with *S. pneumoniae* DNA primase, a ribonucleoside triphosphate and a compound; and determining DNA primase activity, wherein an increase or decrease in primase activity in the presence of the compound compared to a control (no compound), indicates that the compound modulates DNA primase activity.

The invention further includes a method for identifying compounds that modulate *E. coli* DNA primase activity including providing an oligonucleotide selected from the group consisting of SEQ ID NO: 8, SEQ ID NO:36, and SEQ ID NO:40; contacting the

oligonucleotide with *E. coli* DNA primase, a ribonucleoside triphosphates and a compound; and determining DNA primase activity, wherein an increase or decrease in primase activity in the presence of the compound compared to a control (no compound), indicates that the compound modulates DNA primase activity.

The invention further includes a method for identifying compounds that modulate *H. influenzae* DNA primase activity. The invention includes providing an oligonucleotide selected from the group consisting of SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, and SEQ ID NO:41; contacting the oligonucleotide with DNA primase and ribonucleoside triphosphates; and determining DNA primase activity, wherein an increase or decrease in primase activity in the presence of the compound compared to a control (no compound), indicates that the compound modulates DNA primase activity.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows an example schematic of the dinucleotide base pairing pattern at the 3' end of a DNA primase DNA template oligonucleotide in which N_1 pairs with N_1' and N_2 pairs with N_2'.

Figure 2 is a bar graph depicting the relative activities (FU = fluorescent units based upon detection of fluorescently labeled RNA product) of *S. aureus* and *H. influenzae* DNA primases using DNA template oligonucleotides having the general sequence of $(CT)_{16}NNN(CT)_3$ (SEQ ID NO:2). The precise triplet represented by NNN in a particular oligonucleotide is provided along the X-axis.

Figure 3 is a bar graph depicting the activity (FU = fluorescent units based upon detection of a SYBR Green-bound RNA product) of *E. coli* DNA primase using the set of DNA template oligonucleotides listed in Table 3. For a particular oligonucleotide, the precise nucleotides at the N_1 and N_2 positions (in accordance with Figure 1) are indicated along the X-axis.

Figure 4 is a line graph depicting *E. coli* DNA primase activity (based upon absorbance readings of malachite green / phosphate complexes at A650) using three different DNA template oligonucleotides having the general sequence of $(CT)_{15}$CTGCAAANN (SEQ ID NO:3). Specifically, *E. coli* DNA primase activity is shown using DNA template oligonucleotides having fifteen CT repeats followed by CTGCAAAGC (SEQ ID NO:4) (filled squares), CTGCAAACC (SEQ ID NO:5) (open squares), or CTGCAAAGT (SEQ ID

NO:6) (open triangles) at the 3' end. Thus, the complete sequences for the oligonucleotides tested in Figure 4 are as follows: $(CT)_{16}$GCAAAGC (SEQ ID NO:40) (filled squares), $(CT)_{16}$GCAAACC (SEQ ID NO:43) (open squares), and $(CT)_{16}$GCAAAGT (SEQ ID NO:44) (open triangles).

Figure 5 is a bar graph depicting *S. aureus* DNA primase activity (FU = fluorescent units based upon detection of a SYBR Green-bound RNA product) using the DNA template oligonucleotides presented in Table 3 and the DNA template oligonucleotide having the sequence $(CT)_{16}ACC(CT)_3$ (SEQ ID NO:7) (referred to herein as "DNA30" and indicated as "30" along the X-axis in the graph. The DNA template oligonucleotides of Table 3 are indicated along the X-axis by the nucleotides in positions N_1 and N_2 at their 3' ends, in accordance with the schematic of Figure 1.

Figure 6 is a bar graph depicting *H. influenzae* DNA primase activity (FU = fluorescent units based upon detection of a SYBR Green-bound RNA product) using the DNA template oligonucleotides presented in Table 3. The DNA template oligonucleotides are indicated along the X-axis by the nucleotides in positions N_1 and N_2 at their 3' ends, in accordance with the schematic of Figure 1.

Figure 7 is a line graph depicting K_m values for *S. aureus* DNA primase activity using DNA template oligonucleotide "DNA7" ($(CT)_{17}$GCAAAGC (SEQ ID NO:8)) or DNA template oligonucleotide "DNA30".

Figure 8 is a line graph depicting the linear relationship between RNA formation, as monitored by fluorescent signal, and reaction time using *S. aureus* DNA primase and DNA template oligonucleotide "DNA30".

DETAILED DESCRIPTION

The present invention provides methods for identifying a DNA primase-recognition sites for the determination of appropriate DNA templates for particular DNA primases. These methods can identify DNA templates that can be used in DNA primase assays for the study of DNA primase function and for screening for compounds that modulate (either stimulate/increase/augment or inhibit/decrease/diminish) DNA primase activity.

The present invention is based upon our discovery that the 5'-CTG trinucleotide sequence is neither essential nor sufficient for *in vitro E. coli* DNA primase activity. We have further discovered that a dinucleotide pairing pattern is crucial to *E. coli* DNA primase activity.

We have identified specific oligonucleotide sequences that can serve as single stranded DNA templates (DNA template oligonucleotides) to support the activity of specific bacterial DNA primases, including templates for DNA primase from *S. aureus, S. pneumoniae, H. influenzae* and *E. coli.* We have discovered that the nucleotide sequence ACC in the 3'-end region of a DNA template oligonucleotide is important for the enzyme activity of DNA primases from *S. aureus* and *S. pneumoniae*, and that DNA template oligonucleotides with GC pairings in the 3'-end region are important for *E. coli* and *H. influenzae* DNA primase activities. Our discoveries can be harnessed to provide DNA template oligonucleotides appropriate for use with specific DNA primases, which can be used in assays of DNA primase activity and in assays to screen for DNA primase modulators. Our identification of these new DNA primase/DNA template oligonucleotide recognition reactions permits the design of assays of DNA primase activity using low concentrations of DNA template oligonucleotide (for example, about 100 to about 200 nM DNA) and low concentrations of DNA primase (for example, about 30 to about 50 nM primase).

A primase recognition site is a sequence on a nucleic acid template recognized by a primase to initiate RNA synthesis. We have established novel screening methods to determine DNA template oligonucleotides for *in vitro* DNA primase function assays. These rational screening methods are based on the concept that a primase recognition site is composed of a trinucleotide sequence or a dinucleotide pairing pattern.

In one aspect, the present invention provides methods for identifying a DNA primase-recognition site. The methods comprise screening a DNA primase with a plurality of oligonucleotides in order to identify an appropriate DNA primase recognition site.

Two types of libraries of oligonucleotides can be used in the DNA primase recognition site screening assays: trinucleotide screening libraries and dinucleotide pairing screening libraries.

In an example of a trinucleotide screening library, the oligonucleotides comprise the sequence

$(X \text{ or } XX \text{ or } XXX)_n\ N_1N_2N_3\ (Y \text{ or } YY \text{ or } YYY)_m$, wherein

X is at each occurrence independently A, T, C, or G;

Y is at each occurrence independently A, T, C, or G;

n is at least 2;

m is at least 2; and

N_1, N_2, and N_3, are independently A, T, C, or G.

This plurality of oligonucleotides is designed to screen a DNA primase against trinucleotide sequences to identify trinucleotide sequences that are most supportive of DNA primase activity. Trinucleotide screening of primase DNA templates is based on the concept that a recognition site is composed of three consecutive nucleotides. The core sequence NNN can present up to 64 possible combinations of a trinucleotide sequence. In particular embodiments, the screening library will contain oligonucleotides carrying each of the 64 possible trinucleotides. If a particular DNA primase has a recognition site that is composed of a trinucleotide sequence, trinucleotide screening will permit the identification of the recognition site.

Trinucleotide screening oligonucleotides can be of any length.

The adjacent nucleotide sequences (X or XX or $XXX)_n$ and (Y or YY or $YYY)_m$ that are located to the 5' and 3' sides, respectively, of a core triucleotide sequence NNN, can be of any length and sequence. For example, $(X)_n$ can represent a stretch of "n" nucleotides, where each nucleotide can be independently A, T, C, or G. Thus the ajacent sequences can be random, but they preferably have a defined sequence. The defined sequence makes it readily ascertainable which particular trinucleotide recognition site a particular DNA primase is interacting with. The defined set of nucleotides for (X or XX or $XXX)_n$ and (Y or YY or $YYY)_m$ can be a single nucleotide, a set of dinucleotide repeats or a set of trinucleotide repeats. In one example, XX and/or YY are selected from the group consisting of CT, TC, GT, and TG, yielding oligonucleotides with repeats of CT, TC, GT, or TG to the 5' and/or 3' sides of the core trinucleotide sequence.

Moreover, the adjacent nucleotide sequences (X or XX or $XXX)_n$ and (Y or YY or $YYY)_m$ can have different defined sets of nucleotides or the same sets of nucleotides. For example, adjacent nucleotide sequences (X or XX or $XXX)_n$ and (Y or YY or $YYY)_m$ can both be repeats of CT, or one adjacent nucleotide sequence can be repeats of CT, while the other is repeats of a single nucleotide or a different dinucleotide repeat. However, uniformity of sequence between (X or XX or $XXX)_n$ and (Y or YY or $YYY)_m$ is beneficial to identify the particular trinucleotide recognition sites.

Any number of repeats can be used. For example, there can be from 2 to 10, 5 to 15, 10 to 20, 15 to 30, or 25 to 50 repeats of either or both adjacent nucleotide sequences (X or XX or $XXX)_n$ and (Y or YY or $YYY)_m$. The number of repeats for (X or XX or $XXX)_n$ and (Y or YY or $YYY)_m$ need not be equivalent. For example, oligonucleotides can have 16 repeats of the (X or $XX)_n$ ($n = 16$), and 3 repeats of (Y or $YY)_m$ ($m = 3$).

Alternatively, instead of having to make all 64 combinations of $N_1N_2N_3$, an oligonucleotide set of the following sequence can be used:

$(X \text{ or } XX \text{ or } XXX)_n\ N_{4\text{-}9}\ (Y \text{ or } YY \text{ or } YYY)_m$, wherein

X is at each occurrence independently A, T, C, or G;

Y is at each occurrence independently is A, T, C, or G;

n is at least 2;

m is at least 2; and

$N_{4\text{-}9}$ are independently A, T, C, or G.

By generating oligonucleotides that have $N_{4\text{-}9}$ it is possible to use fewer oligonucleotides in the initial screening for the DNA primase recognition site. Table 2 shows such a set of oligonucleotides. Once it has been identified which oligonucleotide contains a trinucleotide recognition site, it is possible to generate a subgroup of oligonucleotides which have each of the trinucleotide sequences. For example, should Oligo 1 (from Table 2) be found to be the best at supporting DNA primase activity upon screening, then 5 further oligonucleotides would be made, each having as $N_1N_2N_3$ the following: CTA, TAA, AAA, AAC, ACT.

Another DNA primase recognition site which has been identified in the present invention is a dinucleotide pairing pattern.

In an example of a dinucleotide screening library, the oligonucleotides comprise the sequence $N_1N_2\ (X)_p\ N_2'N_1'$, wherein

X is at each occurrence independently A, T, C, or G;

N_1, N_2, N_1', and N_2' are independently A, T, C, or G;

p is at least 3;

N_1 and N_1' are complementary; and

N_2 and N_2' are complementary.

This plurality of oligonucleotides is designed to screen a DNA primase against dinucleotide pairing oligonucleotides to identify dinucleotide pairing patterns that are most supportive of DNA primase activity. In this oligonucleotide set, it is essential that $5'-N_2'N_1'-3'$ represents the complement of $5'-N_1N_2-3'$. The purpose of this is to form at the 3' end a dinucleotide repeat which can form a loop or hairpin, as exemplified in Figure 1. For example, the dinucleotide pairing pattern can be two consecutive nucleotides that can form G-C and/or A-T pairings with two other consecutive nucleotides in the same nucleic acid, thereby forming a loop or hairpin. There are 16 possible dinucleotide pairing patterns. An example of a dinucleotide pairing library of oligonucleotides is presented in Table 3.

Dinucleotide pairing oligonucleotides can be of any length.

Dinucleotide pairing oligonucleotides can include additional sequences adjacent to the 5' side of the pairing pattern core (N_1N_2 $(X)_p$ $N_2'N_1'$). These additional sequences preferably have a defined sequence. The defined sequence makes it readily ascertainable which particular dinucleotide pairing serves as a recognition site for a specific DNA primase. Preferably, the additional sequences can be represented by (X or XX or $XXX)_n$, for example, (X or XX or $XXX)_n$ (N_1N_2 $(X)_p$ $N_2'N_1'$), where X can be a single nucleotide, a set of dinucleotide repeats or a set of trinucleotide repeats. In one example, XX is selected from the group consisting of CT, TC, GT, and TG, yielding oligonucleotides with repeats of CT, TC, GT, or TG, on the 5' side of the dinucleotide pairing pattern core. For example, in Table 3, the oligonucleotides include a sequence of sixteen CT repeats to the 5' side of the dinucleotide pairing core. In another example, XXX is selected from the group consisting of CTT, TCT, CTC, TTC, GTT, TTG, and TGT, yielding oligonucleotides with repeats of CTT, TCT, CTC, TTC, GTT, TTG, or TGT to the 5' side of the dinucleotide pairing pattern core.

Additional sequences can also be located at the 3' side of (N_1N_2 $(X)_p$ $N_2'N_1'$), for example, (X or XX or $XXX)_n$ (N_1N_2 $(X)_p$ $N_2'N_1'$) (Y or YY or $YYY)_m$. However, such additional 3' sequences are not preferred, as they can decrease DNA primase activity. If such sequences are included, they will preferably have a defined sequence and must not interfere with the ability of the dinucleotides, N_1N_2 and $N_2'N_1'$ to pair.

The pair of dinucleotides within a dinucleotide pairing oligonucleotide is separated by a spacer sequence of nucleotides $(X)_p$. Each X nucleotide in the $(X)_p$ spacer sequence can independently be A, T, C, or G. For example, the spacer sequence can be AAA, ATC, GTA, etc. Any number of nucleotides can be used for this spacer so long as dinucleotide pairing is capable and a loop can be made (see, for example, Figure 1). Thus, "p" can be any number, so long as pairing to generate a loop is possible. For example, the spacer can range in length from 3 to 30 nucleotides in length, 3 to 25 nucleotides, 3 to 20 nucleotides, 3 to 15 nucleotides, 3 to 10, or 3 to 7.

The oligonucleotides of the present invention can be any length. In some embodiments, the oligonucleotides range from 7 nucleotides in length to 150 nucleotides in length. In some embodiments, the oligonucleotides range from 10 to 100 nucleotides in length. In some embodiments, the oligonucleotides range from 15 to 75 nucleotides in length. In some embodiments, the oligonucleotides range from 20 to 60 nucleotides in length. In some embodiements, the oligonucleotides range from 25 to 55 nucleotides in length. In further embodiments the oligonucleotides range from 30 to 50 nucleotides in

length. In further embodiments, the oligonucleotides range from 35 to 45 nucleotides in length.

In some embodiments, the screening of a DNA primase to identify a DNA primase-recognition site is carried out by running a series of screens against different sets or libraries of oligonucleotides. A DNA primase, in the presence of ribonucleoside triphosphates is tested for activity with a first plurality of oligonucleotides, and also tested for activity with a second plurality of oligonucleotides. This comparative testing can be done simultaneously or sequentially. The DNA primase can be tested against further pluralities of oligonucleotides until an oligonucleotide comprising an appropriate DNA primase-recognition site is identified based upon determining DNA primase activity.

The oligonucleotides of the invention can be synthesized using chemical synthesis known in the art (See, *e.g.*, Sambrook *et al.*, eds., Molecular Cloning: A Laboratory Manual (3^{rd} ed.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (2001), herein incorporated by reference).

In another aspect, the present invention provides novel oligonucleotides that serve as appropriate DNA primase templates.

In some embodiments the oligonucleotide includes the sequence

$(X \text{ or } XX \text{ or } XXX)_n\ N_1N_2N_3\ (Y \text{ or } YY \text{ or } YYY)_m$, wherein

X is at each occurrence independently A, T, C, or G;

Y is at each occurrence independently A, T, C, or G;

n is at least 2;

m is at least 2; and

N_1, N_2, and N_3, are independently A, T, C, or G.

In some embodiments the oligonucleotide includes the sequence

$N_1N_2\ (X)_p\ N_2'N_1'$, wherein

X is at each occurrence independently A, T, C, or G;

N_1, N_2, N_2' and N_1' are independently A, T, C, or G;

p is at least 3;

N_1 and N_1' are complementary; and

N_2 and N_2' are complementary.

In some embodiments the oligonucleotide includes the sequence

$(XX)_n\ ACC\ (YY)_n$, wherein,

X is independently A, T, C, or G;

Y is independently A, T, C, or G; and

n is at least 2.

In some embodiments the oligonucleotide includes the sequence

$(XX)_n$ AGT $(YY)_n$, wherein,

X is independently A, T, C, or G;

Y is independently A, T, C, or G; and

n is at least 2.

In some embodiments, the oligonucleotide includes a sequence selected from SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, and SEQ ID NO:43.

In another aspect, the invention provides methods for screening for and identifying compounds that modulate DNA primase activity. These methods use the oligonucleotides of the present invention for the screening. In the methods, an oligonucleotide is contacted with a DNA primase, ribonucleotide triphosphates and a test compound, and DNA primase activity is determined. Any change/alteration (increase or decrease) in DNA primase activity in the presence of the test compound as compared to DNA primase activity in the absence of the test compound (control assay without compound present) indicates that the test compound modulates DNA primase activity.

As used herein, the terms "modulate" or "modulates" in reference to DNA primase activity includes any measurable alteration, either an inhibition or enhancement, of DNA primase activity.

Any of the oligonucleotides of the present invention can be used in the assays for identifying compounds that modulate DNA primase.

In some embodiments, particular oligonucleotides are used for screening for compounds with particular DNA primases. For example, $(CT)_{16}ACC(CT)_3$ (SEQ ID NO:7) and $(CT)_{16}AGT(CT)_3$ (SEQ ID NO:43) can be used for *S. aureus*; $(CT)_{16}ACC(CT)_3$ (SEQ ID NO:7) for *S. pneumoniae*; $(CT)_{17}GCAAAGC$ (SEQ ID NO:8), $(CT)_{16}CCAAAGG$ (SEQ ID NO:36), and $(CT)_{16}GCAAAGC$ (SEQ ID NO:40) for *E. coli*; $(CT)_{16}GAAAATC$, $(CT)_{16}GTAAAAC$, $(CT)_{16}GCAAAGC$, and $(CT)_{16}GGAAACC$ (SEQ ID NOs:38, 39, 40, and 41, respectively) for *H. influenzae*.

In the screening methods, compounds such as peptides, peptidomimetics, small molecules, DNA, RNA, or other drugs can be used to determine if they modulate DNA primase activity.

Natural or synthetic nucleoside triphosphates are employed as primase substrates in the assays of the invention for testing DNA primase recognition sites and for identifying

compounds that modulate DNA primase. Nucleoside triphosphates used for the primase reaction can be labeled, for example with a radiolabel or fluorescent label, or they can be modified for activity or detection purposes in accordance with methods and techniques well known to the art.

Assays of the present invention are conducted under conditions such that the DNA primase is catalytically active. These are conditions under which the DNA primase is capable of polymerizing the substrate ribonucleoside triphosphates to form RNA. Such conditions are known to those of skill in the art. A wide variety of incubation conditions can be used, depending on the enzyme used. Reaction conditions in which DNA primase is active *in vitro* are exemplified below and/or are otherwise known in the art. U.S. Provisional Application 60/473,054 discloses detailed DNA primase assay conditions, including concentrations of DNA template, enzymes, NTP and buffer components, pH, temperature and time, incorporated herein by reference.

Many assays of DNA primase activity are known to those of skill in the art, including assays using antibody and enzyme coupling to detect the formation of DNA-RNA heterohybrid regions (U.S. Patent No. 6,043,038), assays using incorporated radioactivity (U.S. Patent No. 6,096,499; Johnson *et al.*, 2000, Biochemistry, 39:736-744), and scintillation proximity assays (Earnshaw & Pope, 2001, J. Biomol. Screen., 6:39-46; Zhang *et al.*, 2002, Anal. Biochem., 304:174-179. DNA primase activity can also be determined by measuring the production of the RNA product using a fluorescent marker that binds to RNA, such as SYBR Green II, as disclosed in the U.S. Provisional Application 60/473,054, incorporated herein by reference. DNA primase activity can also be measured by detection of pyrophosphate formed during the primase reaction. Pyrophosphate can be detected by many methods known to the art, including, but not limited to the Malachite Green assay of phosphate after hydrolysis of the pyrophosphate by pyrophosphatase (Shatton *et al.*, 1983, Anal. Biochem., 130:114-119). Any of these assays of DNA primase activity can be used with the methods of the present invention.

The methods of the present invention can be used for the determination of primase recognition sites and DNA templates for any primase from any source including all prokaryotic, eukaryotic or viral primases. Different primases are potential drug targets for a number of therapeutic areas including antibacterial, antiviral and anticancer areas.

Any DNA primase can be used in the assays of the present invention. Primases have been identified in prokaryotes, including bacteria, bacteriophage, eukaryotes, and eukaryotic viruses.

Many bacterial primase nucleotide and amino acid sequences are known, including *Bacillus subtilis* (X03897), *Clostridium acetobutylicum* (Z23080), *Escherichia coli* (J01687), *Haemophilus influenzae* (L11044), *Helicobacter pylori* (AE000523), *Lactococcus lactis* (D10168), *Legionella pneumophila* (U63641), *Listeria monocytogenes* (U13165), *Myxococcus xanthus* (U20669), *Mycoplasm genitalium* (U39703), *Mycoplasma pneumoniae* (1674174), *Mycobacterium tuberculosis* (Z83860), *Pseudomonas putida* (U85774), *Rickettsia prowazekii* (M95860), *Salmonella typhimurium* (M14427), *Staphylococcus aureus* (AB001896), *Synechococcus elongatus* (PCC7942), *Cyanobacteria chroococcales* (X94247), Synechocystis sp. (D90912), *Cyanobacteria chroococcales* (D90912). Additonal bacterial primases for which nucleotide and amino acid sequences are known include *Aquifex aeolicus* (AE000743), *Bacillus stearothermophilus* (AF106033), *Borrelia burgdorferi* (AE001171), *Buchnera aphidicola* (P32000), *Campylobacter jejuni* (CAB73626), *Chlamydia trachomatis* (3329259), *Chlamydophila pneumoniae* (AE001674), *Deinococcus radiodurans* (AAF10180), *Mycobacterium smegmatis* (AF027507), *Neisseria meningitidis* (CAB84964), *Streptomyces coelicolor A3(2)* (CAB51551), *Thermotoga maritima* (AAD36520), *Treponema pallidum* (3322781).

In some embodiments, the DNA primase is bacterial.

In particular embodiments, the DNA primase is from *Escherichia coli*, *Staphyloccus aureus*, *Streptococcus pneumoniae*, or *Haemophilus influenzae*.

Bacteriophage primases for which nucleotide and amino acid sequences are known include, for example, T7 (gene 4 protein) (P03692), T3 (gene 4 protein) (P20315), P4 (α protein) (PRBPP4), and T4 (gene 41 protein) (ADD42502).

Herpes simplex virus type 1 is the prototype for the herpesvirus family. UL5, UL8, and UL52 proteins compose the heterotrimeric primase/helicase enzyme from this virus. Herpesviral UL52 primases include Human type 1, alphaherpesvirus (Acc# P10236), Equine type 1, alphaherpesvirus (Acc# M86664:7064..10301), Equine type 4, alphaherpesvirus (Acc# AF030027:6658..9900), Bovine type 1, alphaherpesvirus (Acc# AJ004801:4013..7237), Human type 3 = varicella-zoster virus, alphaherpesvirus (Acc# P09270), Pseudorabies type 1 = Suid herpesvirus, alphaherpesvirus (Acc# X87246:1087..3963), Human type 7, betaherpesvirus (Acc# AF037218:68725..71310), Human type 6, betaherpesvirus (Acc# P52540), Equine type 2, betaherpesvirus (Acc# S55651), Saimiriine type 2, gammaherpesvirus (Acc# M86409:6378..8885), Human type 8 = Kaposi's herpesvirus, gammaherpesvirus (Acc# U93872:79735..82266), Wildebeest type 1 = Alcelaphine herpesvirus, gammaherpesvirus (Acc# AF005370:81823..84336), Human type 4

= Epstein-Barr virus, gammaherpesvirus (Acc# P03193), Human Cytomegalovirus, betaherpesvirus (Acc# P17149), Mouse Cytomegalovirus, betaherpesvirus (Acc# L07319:3153..6047).

Eukaryotic primase activity is typically associated with two types of multimeric complex. Primases involved in DNA replication are generally found in a complex with DNA polymerase α; these complexes typically include two primase subunits (Pri1 (49 kDa) and Pri2 (58 kDa)), as well as DNA polymerase α and a p70-90 subunit. Primase activity is also found as a heterodimer which consists of the Pri1 and Pri2 subunits. The active site for RNA polymer elongation is found in the Pri1 subunit, while initiation generally involves the Pri2 subunit. Amino acid and nucleotide sequences for several eukaryotic primase Pri1 subunits have been characterized, including human (Swissprot Accession No. P49642, Genbank X74330), *Mus musculus* (GenBank J04620), *Rattus norvegicus* PID:g1763025, *Drosophila melanogaster* p50 (PID:g666989), *Caenorhabditis elegans* p48 (SP:P34471), *Saccharomyces cerevisiae* p48 (SP:P10363), *Schizosaccharomyces pombe* p53 (Acc# Z98531:22398 . . . 23846), *Plasmodium falciparum* p53 (Acc# X99254:486 . . . 3750). Archaeobacteria having primases that are similar to the eukaryotic Pri1 include *Methanococcus jannaschii* (Acc# Q58249), *Archaeoglobus fulgidus* (Acc# AE001054:8352 . . . 9434), and *Methanobacterium thermoautotrophicum* (Acc# AE000840:7684 . . . 8655). Pri2 primases have been sequenced from human Acc# P49643, *Mus musculus* Acc# S45629, *Caenorhabditis elegans* (Acc# Z81137) *Saccharomyces cerevisiae* (Acc# P20457) and the plant *Arabidopsis thalians* (Acc# AC002130:39565 . . . 46348).

DNA primase can be obtained for use in the present invention according to procedures well known to the art. DNA primase can be obtained by isolation or purification from natural sources or can be expressed using recombinant technology. Primase can be expressed as a single target protein or co-expressed with other proteins. Primase can be expressed with or without peptide tags or fusion proteins. Primase can be isolated as a cell extract, prepared in substantially pure form as a single protein, or prepared as a protein complex. Numerous techniques for obtaining DNA primase proteins, including bacterial DNA primase proteins, have been described in the literature. Catalytically active portions or fragments of DNA primase can also be used in the assays of the present invention. For example, the catalytic domain of DNA primase from *E. coli* is found within amino acid residues 111H – 433K, the catalytic domain for *H. influenzae* DNA primase is found within amino acid residues 115T – 434 K, and the catalytic domain for *S. pneumoniae* DNA primase is found within amino acid residues 105S – 455T).

Table 2. DNA oligonucleotides used for screening of *S. aureus* primase activity.

Oligo	Nucleotide sequence	SEQ ID NO
1	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**CTAAACT**CTCTCTCT CTCT	9
2	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**AAGCTCAA**CTCTCTCT CTCT	10
3	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**GAGCT**CTCTCTCTCT CT	11
4	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**AATCTGGA**CTCTCTCT CTCT	12
5	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**GATCT**CTCTCTCTCT CT	13
6	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**TAGATGC**CTCTCTCT CTCT	14
7	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**CAGGCT**CTCTCTCTC TCT	15
8	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**TCGGGTGTA**CTCTCTC TCTCT	16
9	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**TACACG**CTCTCTCTC TCTCT	17
10	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**GCATCCG**CTCTCTC TCTCT	18
11	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**ACCCTGAA**CTCTCT CTCTCT	19
12	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**GCCAGT**CTCTCTCTC TCTCT	20
13	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**ATTCTGCGT**CTCTCTC TCTCT	21
14	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**GTTGTCGA**CTCTCT CTCTCT	22
15	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**TCTTTATA**CTCTCTC TCTCT	23

Table 2 shows one embodiment of an oligonucleotide library for trinucleotide recognition site screening. Each of the 15 oligonucleotides listed in Table 2 has a unique sequence of 5 to 9 nucleotides in length, indicated in bold with underlining, surrounded by CT repeats (5' and 3' adjacent CT repeats). These 5 to 9 nucleotide long sequences are a condensed form of 64 possibilities after a process of combination and optimization. The process began with a list of 64 sequences of the format: CTNNNCT. Since the tricucleotide NNN is bordered by a CT on the 5' and 3' side, a number of overlapping trinucleotides are apparent from such a format, such as CTN, TNN, NNC and NCT. These CTNNNCT sequences were removed from the list of 64, leaving 36 CTNNCT sequences, called "single appearance" sequences. These 36 "single appearance" sequences were combined, and "loop" structure type sequences eliminated, leaving 15 "core sequences." These "core sequences" are represented as the nucleotides indicated in bold with underlining in the 15 oligonucleotides listed in Table 2.

Screening using the oligonucleotides in Table 2 revealed that Oligo 11 was the most active one (the signal to background ratio was the highest in the series) for *S. aureus* DNA primase and that Oligo 12 was the most active one for *H. influenzae* DNA primase. CTC and TCT were deemed not to be recognition sites because they are in all the oligonucleotides. Each oligonucleotide yielded activity information for a set of trinucleotides (NNNs). For example, Oligo 11 can provide information for the following trinucleotides: CTA, TAC, ACC, CCC, CCT, CTG, TGA, GAA, AAC and ACT. The activity data corresponding to each NNN were pooled and ranked according to their signal-to-background ratio. Based on this analysis, at least one of the following NNN sequences: ACC, AGT, CCA, CCC, CCT, GAA, GCC, GTC, GCA, CAG and CTG was the most promising sequences for *H. influenzae* DNA primase. Similarly, AAA, ACC, CCC, CCT and GAA were the most promising trinucleotides for *S. aureus* DNA primase. These most promising NNNs are included in most of the oligonucleotides used to generate the data shown in Figure 2. In addition to these promising NNNs, the least promising, GGG, was chosen as a negative control.

The choice of CT repeats in this embodiment (a) provides a simple sequence to avoid confusing results, (b) yields balanced AT and GC content in the DNA-RNA duplex formed, (c) maximizes the fluorescence signal by producing RNA with GA bases, and (d) avoids using relatively unstable UTP in the DNA primase reaction. The length of CT repeats at both

the 5' and 3' ends of each oligonucleotide in this library set is empirical and similar among oligonucleotides to simplify comparison.

The oligonucleotides shown in Table 2 were used to screen the activity of primases from *S. aureus* and *H. influenzae*. After initial screens, the most active trinucleotide sequences were selected.

Screening using the fifteen oligonucleotides listed in Table 2 can exclude most of the 64 possibilities. To decipher which NNN provided the most active sequence (supports the highest activity for a DNA primase), a new set of oligonucleotides with sequences in the format of $(CT)_{16}NNN(CT)_3$ was designed to confirm activity (Figure 2). According to the results shown in Figure 2, oligonucleotide $(CT)_{16}ACC(CT)_3$ (also referred to herein as "DNA30") was the most active DNA template oligonucleotide for *S. aureus* DNA primase. $(CT)_{16}CAG(CT)_3$ (SEQ ID NO:24) and $(CT)_{16}CTG(CT)_3$ (SEQ ID NO:25) were the most active DNA template oligonucleotides for *H. influenzae* DNA primase. The screening results of *S. pneumoniae* DNA primase using the $(CT)_{16}NNN(CT)_3$ trinucleotide library suggested that the differences among trinucleotide sequences was not significant. DNA30 gave a higher signal-to-background ratio with *S. pneumoniae* DNA primase than other DNA template oligonucleotidess.

In another screening method, sixteen DNA oligonucleotides (Table 3) were designed to cover the 16 possibilities of dinucleotide pairing patterns.

Table 3. DNA oligonucleotides used for screening of *E. coli* primase activity.

Oligo	Sequence	SEQ ID NO
16	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAAAAATT	26
17	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTATAAAAT	27
18	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTACAAAGT	28
19	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAGAAACT	29
20	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTAAAATA	30
21	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTTAAAAA	31

22	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTCAAAGA	32
23	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTGAAACA	33
24	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAAAATG	34
25	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAAAAG	35
26	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCAAAGG	36
27	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCGAAACG	37
28	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGAAAATC	38
29	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGTAAAAC	39
30	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGCAAAGC	40
31	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGGAAACC	41

Table 3 shows one embodiment of an oligonucleotide library set for the screening of all 16 possibilities of a 2-base (dinucleotide) pairing pattern in the context of sixteen CT repeats at the 5' end and a three nucleotide hairpin region composed of AAA. The choice of CT repeats in this embodiment (a) provides a simple sequence to avoid confusing results, (b) yields balanced AT and GC content in the DNA-RNA duplex formed, (c) maximizes the fluorescence signal by producing RNA with GA bases, and (d) avoids using relatively unstable UTP in the DNA primase reaction. The length of CT repeats at the 5' end of each oligonucleotide in this library set is empirical and similar among oligonucleotides to simplify comparison. The AAA sequence in this embodiment was placed between each of the 2-base (dinucleotide) sequences in each of the 16 oligonucleotides because (a) a known template for *E. coli* DNA primase contains the sequence CTGCAAAGC (SEQ ID NO:42) at the 3'-end, and (b) we wanted uniformity among the oligonucleotides in this embodiment for ease of

comparison of results. Our experimental data showed that the AAA sequence is not critical for activity. The AAA insert between the two sets of pairing dinucleotides allows the 3'-end of the template to form a loop during pairing as shown in Figure 1.

When *E. coli* primase was used to test the sixteen DNA template oligonucleotides shown in Table 3, we discovered that the *E. coli* primase activity was not correlated to the 5'-CTG trinucleotide sequence (Figure 3). We further discovered that it was critical to have C at position N_2. Thus, the $N_2 - N_2$' pairing was C-G. C-G and G-C were favorable $N_1 - N_1$' pairings. Therefore, the best dinucleotide pairing pattern for *E. coli* DNA primase was 5'----CC---GG. The second best pairing pattern was 5'----GC---GC. Any change of the dinucleotide pairing pattern had significant impact on *E. coli* DNA primase activity. For example, when the 5'----GC---GC-3' was changed to 5'----GC---CC-3' or 5'----GC---GT-3', the primase activity became minimal (Figure 4).

The dinucleotide pairing library in Table 3 was employed to screen *S. aureus* and *H. influenzae* primases. For *S. aureus* primase, DNA30 was more active than any other oligonucleotide in the two-base pairing library (Figure 5). For *H. influenzae* primase, it is critical to have G at position N_1 (Figure 6); variety at position N_2 did not significantly affect activity. Nucleotide specificity at positions N_1 and N_2 for *E. coli* and *H. influenzae* primases are summarized in Table 4.

Table 4. Selectivity of *E. coli* and *H. influenzae* primases.

DNA Primase	Position	A	C	T	G
***E.coli* DNA**	N_1	12	100	18	87
primase	N_2	2	100	1	17
H Influenzae	N_1	12	0	0	100
DNA primase	N_2	100	86	74	34

The numbers in Table 4 represent relative activity of DNA primase scaled from 0 to 100; 100 being the most active and 0 being inactive. Higher numbers (closer to 100) indicate higher activity of DNA primase. Each of the four oligonucleotides having G at position N_1 comprise a 5'-CTG sequence suggesting that *H. influenzae* DNA primase, like *S. aureus* DNA primase, preferentially recognizes a trinucleotide sequence over a dinucleotide pairing pattern.

The invention is further illustrated by way of the following examples, which are intended to elaborate several embodiments of the invention. These examples are not intended to, nor are they to be construed to, limit the scope of the invention. It will be clear that the invention may be practiced otherwise than as particularly described herein. Numerous modifications and variations of the present invention are possible in view of the teachings herein and, therefore, are within the scope of the invention.

EXAMPLES

Example 1. Measurement of DNA primase activity.

DNA primase activity was determined by measuring the production of the RNA product using a fluorescent marker. The RNA products for each DNA primase were analyzed by use of the fluorescent marker SYBR Green II to interact with the RNA, followed by fluorescence measurement, in accordance with the protocols disclosed in the U.S. Provisional Application 60/473,054, incorporated herein by reference. In the screening reactions, the primase mixtures were composed of 50 mM Tris, pH 7.5, 25 mM potassium glutamate (KGlu), 10 mM Mg(OAc)2, 10 mM DTT, 200 nM template, 200 μM ATP, 200 μM GTP, 200 μM UTP, 200 μM CTP, 0.003% Brij-35. Thirty microliters of the reaction mixture were taken at time points and mixed with 30 μl of 30 x SYBR Green II solution in a 384 well plate. The excitation wavelength was 485 nm, and the emission wavelength was 535 nm. Alternatively, the primase activity was measured by detection of pyrophosphate formed in the primase reaction (data presented in Figure 4). The pyrophosphate was detected by the Malachite Green assay of phosphate after hydrolysis of the pyrophosphate by pyrophosphatase (Shatton *et al.*, 1983, Anal. Biochem., 130:114-119).

Example 2. HTS reaction conditions for some bacterial primases.

In accordance with the screening results, a DNA template with the ACC trinucleotide sequence in the context of CT repeats, "DNA30" $(CT)_{16}ACC(CT)_3$ (SEQ ID NO:7), was chosen for assay development of *S. aureus* and *S. pneumoniae* primases. An oligonucleotide with 5'-----GC---GC pairing, "DNA7" $(CT)_{17}GCAAAGC$ (SEQ ID NO:8), was chosen for *E. coli* and *H. influenzae* primases.

For *S. aureus* DNA primase, the K_m for DNA30 was about 25 times lower than the K_m for DNA7 (Figure 7). Furthermore, under the following reaction conditions (20 mM MES, pH 6.5, 10 mM $Mg(AcO)_2$, 1 mM DTT, 25 mM NH_4Cl, 10 mM $CaCl_2$, 0.003 % Brij-35, 100 nM DNA template, 20 nM DnaG, 200 μM ATP, 200 μM GTP), RNA synthesis by *S.*

aureus DNA primase using DNA7 was negligible. Under the same reaction conditions, DNA30 gave strong signals with *S. aureus* DNA primase (Figure 8).

The optimized HTS conditions for *S. aureus*, *H. influenzae*, and *S. pneumoniae* primases were as follows:

***S. aureus* primase**: The assay reaction contained 20 mM MES, pH 6.5, 10 mM $Mg(AcO)_2$, 1 mM DTT, 25 mM NH_4Cl, 10 mM CaCl2, 0.003 % Brij-35,100 nM DNA30, 30 nM *S. aureus* DnaG, 200 μM ATP, 200 μM GTP. The reaction (30 μl) was carried out on a 384-well assay plate at room temperature for 30 min., and then the RNA product was analyzed.

***H. influenzae* primase**: The assay reaction contained 20 mM Tris-HCl, pH 7.5, 10 mM $Mg(AcO)_2$, 1 mM DTT, 25 mM NH_4Cl, 5 mM $CaCl_2$, 0.003 % Brij-35,150 nM DNA7, 75 nM *H. influenzae* DnaG, 20 nM *H. influenzae* DnaB, 100 μM ATP, 100 μM GTP. The reaction (30 μl) was carried out on a 384-well assay plate at room temperature for 30 min., and then the RNA product was analyzed.

***S. pneumoniae* primase**: The assay reaction contained 20 mM MES, pH 6.5, 10 mM $Mg(AcO)_2$, 1 mM DTT, 25 mM NH_4Cl, 10 mM $CaCl_2$, 0.003 % Brij-35,100 nM DNA30, 40 nM *S. pneumoniae* DnaG, 75 μM ATP, 75 μM GTP. The reaction (30 μl) was carried out on a 384-well assay plate at room temperature for 40 min., and then the RNA product was analyzed.

The foregoing examples are meant to illustrate the invention and are not to be construed to limit the invention in any way. Those skilled in the art will recognize modifications that are within the spirit and scope of the invention.

We claim:

1. A method for identifying a DNA primase-recognition site, comprising:

providing a plurality of oligonucleotides, a DNA primase and ribonucleoside triphosphates; and

determining DNA primase activity in order to identify an appropriate DNA primase-recognition site.

2. The method of claim 1, wherein the oligonucleotides comprise the sequence

$(X \text{ or } XX \text{ or } XXX)_n\ N_1N_2N_3\ (Y \text{ or } YY \text{ or } YYY)_m$, wherein

X is at each occurrence independently A, T, C, or G;

Y is at each occurrence independently A, T, C, or G;

n is at least 2;

m is at least 2; and

N_1, N_2, and N_3, are independently A, T, C, or G.

3. The method of claim 2, wherein XX is selected from the group consisting of AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, CC, CG, GA, GT, GC, and GG.

4. The method of claim 2, wherein YY is selected from the group consisting of AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, CC, CG, GA, GT, GC, and GG.

5. The method of claim 1, wherein the oligonucleotides comprise the sequence

$N_1N_2\ (X)_p\ N_2'N_1'$,

wherein

X is at each occurrence independently A, T, C, or G;

N_1, N_2, N_1' and N_2' are independently A, T, C, or G;

p is at least 3;

N_1 and N_1' are complementary; and

N_2 and N_2' are complementary.

6. The method of claim 1, wherein the DNA primase is prokaryotic.

7. The method of claim 6, wherein the prokaryote is bacterial.

8. The method of claim 7, wherein the bacteria is *Escherichia coli*, *Staphyloccus aureus*, *Streptococcus pneumoniae*, or *Haemophilus influenzae*.

9. The method of claim 1, wherein the DNA primase is eukaryotic.

10. The method of claim 1, wherein the DNA primase is viral.

11. The method of claim 1, wherein the DNA primase a bacteriophage DNA primase.

12. The method of claim 1, wherein the DNA primase activity is determined by detecting an RNA product, DNA-RNA heterohybrid regions, or pyrophosphate.

13. A method for identifying a DNA primase-recognition site, comprising:

(i) providing a first plurality of oligonucleotides, a DNA primase and ribonucleoside triphosphates;

(ii) providing a second plurality of oligonucleotides, a DNA primase and ribonucleoside triphosphates; and

determining DNA primase activity for (i) and (ii), wherein primase activity is indicative of the oligonucleotide having a DNA primase-recognition site.

14. The method of claim 13, wherein the first plurality of oligonucleotides comprise the sequence of

$(X \text{ or } XX \text{ or } XXX)_n\ N_1N_2N_3\ (Y \text{ or } YY \text{ or } YYY)_m$, wherein

X is at each occurrence independently A, T, C, or G;

Y is at each occurrence independently A, T, C, or G;

n is at least 2;

m is at least 2; and

N_1, N_2, and N_3, are independently A, T, C, or G.

15. The method of claim 14, wherein XX is selected from the group consisting of AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, CC, CG, GA, GT, GC, and GG.

16. The method of claim 14, wherein YY is selected from the group consisting of AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, CC, CG, GA, GT, GC, and GG.

17. The method of claim 13, wherein the second plurality of oligonucleotides comprise the sequence of

$N_1N_2 (X)_p N_2'N_1'$,

wherein

X is at each occurrence independently A, T, C, or G;

N_1, N_2, N_1' and N_2' are independently A, T, C, or G;

p is at least 3;

N_1 and N_1' are complementary; and

N_2 and N_2' are complementary.

18. The method of claim 17, wherein the DNA primase is prokaryotic.

19. The method of claim 18, wherein the prokaryote is bacterial.

20. The method of claim 19, wherein the bacteria is *Escherichia coli*, *Staphyloccus aureus*, *Streptococcus pneumoniae*, or *Haemophilus influenzae*.

21. The method of claim 13, wherein the DNA primase is eukaryotic.

22. The method of claim 13, wherein the DNA primase is viral.

23. The method of claim 13, wherein the DNA primase a bacteriophage DNA primase.

24. The method of claim 13, wherein the DNA primase activity is determined by detecting an RNA product, DNA-RNA heterohybrid regions, or pyrophosphate.

25. An oligonucleotide comprising the sequence of

$(X \text{ or } XX \text{ or } XXX)_n N_1N_2N_3 (Y \text{ or } YY \text{ or } YYY)_m$, wherein

X is at each occurrence independently A, T, C, or G;

Y is at each occurrence independently A, T, C, or G;

n is at least 2;

m is at least 2; and

N_1, N_2, and N_3, are independently A, T, C, or G.

26. An oligonucleotide comprising the sequence of

$N_1N_2 (X)_p N_2'N_1'$,

wherein

X is at each occurrence independently A, T, C, or G;

N_1, N_2, N_1' and N_2' are independently A, T, C, or G;

p is at least 3;

N_1 and N_1' are complementary; and

N_2 and N_2' are complementary.

27. An oligonucleotide comprising the sequence of

$(XX)_n N_1N_2N_3 (YY)_m$;

wherein

XX is selected from the group consisting of AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, CC, CG, GA, GT, GC, and GG;

YY is selected from the group consisting of AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, CC, CG, GA, GT, GC, and GG;

n is at least 2;

m is at least 2; and

N_1 N_2, and N_3 are independently A, T, C, or G.

28. The method of claim 27, wherein XX is selected from the group consisting of CT, TC, GT, and TG.

29. The method of claim 27, wherein YY is selected from the group consisting of CT, TC, GT, and TG.

30. The oligonucleotide of claim 27, wherein $N_1N_2N_3$is selected from the group consisting of ACC and AGT.

31. An oligonucleotide comprising the sequence of SEQ ID NO:7.

32. An oligonucleotide comprising the sequence of SEQ ID NO:8.

33. An oligonucleotide comprising the sequence of SEQ ID NO:36.

34. An oligonucleotide comprising the sequence of SEQ ID NO:38.

35. An oligonucleotide comprising the sequence of SEQ ID NO:39.

36. An oligonucleotide comprising the sequence of SEQ ID NO:40.

37. An oligonucleotide comprising the sequence of SEQ ID NO:41.

38. An oligonucleotide comprising the sequence of SEQ ID NO:43.

39. A method for identifying compounds that modulate DNA primase activity comprising:

providing the oligonucleotide of any one of claims 25 – 38;

contacting the oligonucleotide with DNA primase, a ribonucleoside triphosphate and a compound; and

determining DNA primase activity, wherein an increase or decrease in primase activity in the presence of the compound compared to a control, indicates that the compound modulates DNA primase activity.

40. The method of claim 39, wherein the DNA primase is prokaryotic.

41. The method of claim 40, wherein the prokaryote is bacterial.

42. The method of claim 41, wherein the bacteria is *Escherichia coli*, *Staphyloccus aureus*, *Streptococcus pneumoniae*, or *Haemophilus influenzae*.

43. The method of claim 39, wherein the DNA primase is eukaryotic.

44. The method of claim 39, wherein the DNA primase is viral.

45. The method of claim 39, wherein the DNA primase a bacteriophage DNA primase.

46. The method of claim 39, wherein the DNA primase activity is determined by detecting an RNA product, DNA-RNA heterohybrid regions, or pyrophosphate.

47. A method for identifying compounds that modulate *S. aureus* DNA primase activity comprising:

providing an oligonucleotide selected from the group consisting of SEQ ID NO:7 or SEQ ID NO:43;

contacting the oligonucleotide with *S. aureus* DNA primase, a ribonucleoside triphosphate and a compound; and

determining DNA primase activity, wherein an increase or decrease in primase activity in the presence of the compound compared to a control, indicates that the compound modulates DNA primase activity.

48. A method for identifying compounds that modulate *S. pneumoniae* DNA primase activity comprising:

providing an oligonucleotide of SEQ ID NO:7;

contacting the oligonucleotide with *S. pneumoniae* DNA primase, a ribonucleoside triphosphate and a compound; and

determining DNA primase activity, wherein an increase or decrease in primase activity in the presence of the compound compared to a control, indicates that the compound modulates DNA primase activity.

49. A method for identifying compounds that modulate *E. coli* DNA primase activity comprising:

providing an oligonucleotide selected from the group consisting of SEQ ID NO: 8, SEQ ID NO:36, and SEQ ID NO:40;

contacting the oligonucleotide with *E. coli* DNA primase, a ribonucleoside triphosphate and a compound; and

determining DNA primase activity, wherein an increase or decrease in primase activity in the presence of the compound compared to a control, indicates that the compound modulates DNA primase activity.

50. A method for identifying compounds that modulate *H. influenzae* DNA primase activity comprising:

providing an oligonucleotide selected from the group consisting of SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, and SEQ ID NO:41;

contacting the oligonucleotide with *H. influenzae* DNA primase, a ribonucleoside triphosphate and a compound; and

determining DNA primase activity, wherein an increase or decrease in primase activity in the presence of the compound compared to a control, indicates that the compound modulates DNA primase activity.

Figure 1

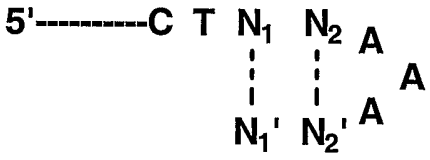


Figure 2

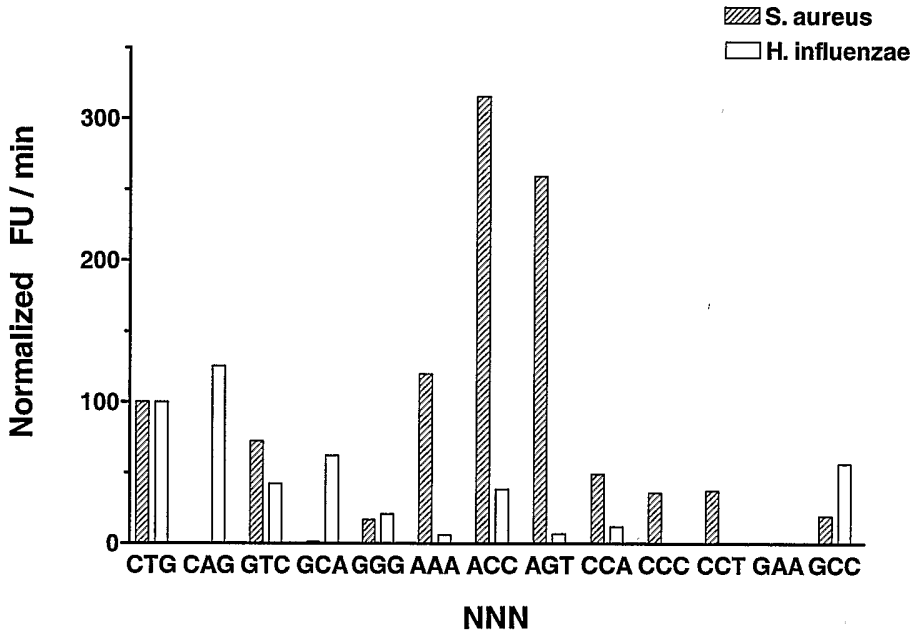

S. aureus
H. influenzae
Normalized FU / min
300
200
100
0
CTG CAG GTC GCA GGG AAA ACC AGT CCA CCC CCT GAA GCC
NNN

Figure 3

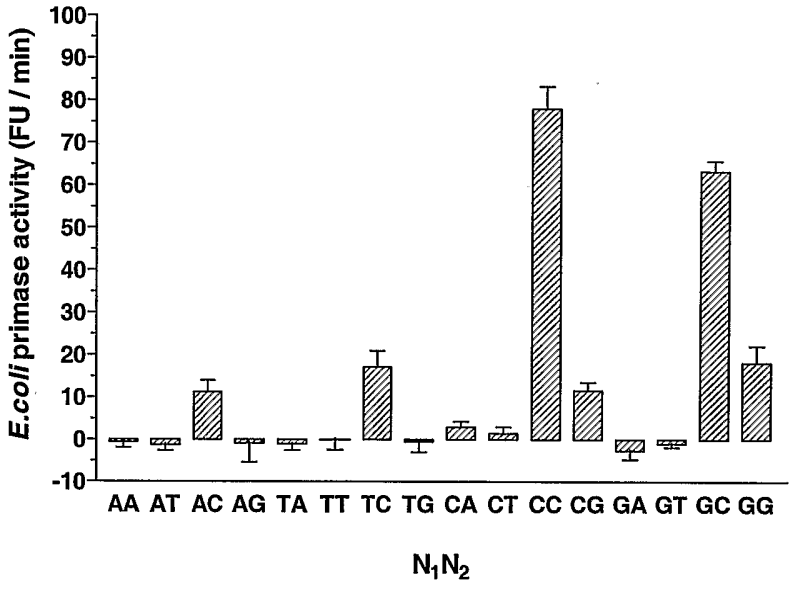

E.coli primase activity (FU / min)
100
90
80
70
60
50
40
30
20
10
0
-10
AA AT AC AG TA TT TC TG CA CT CC CG GA GT GC GG
N1N2

Figure 4

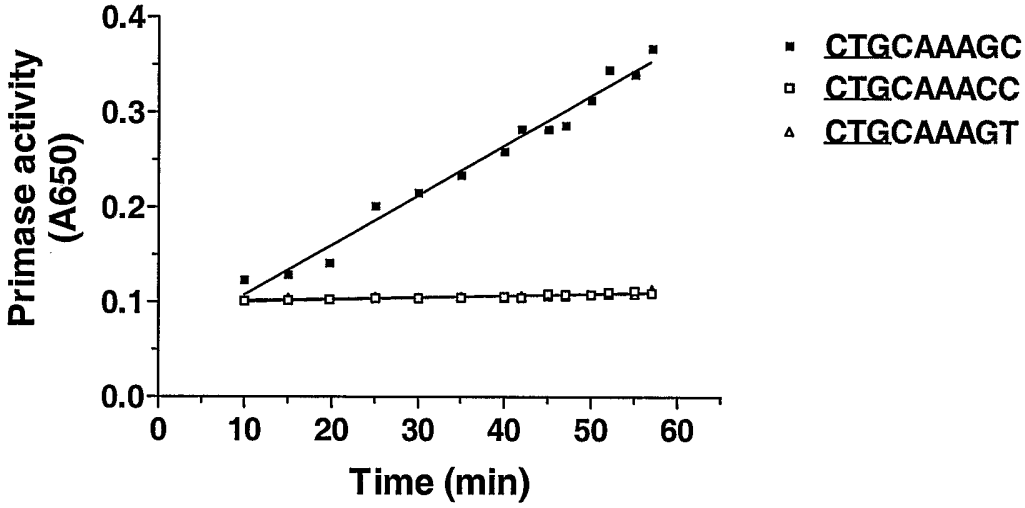

Primase activity (A650)
0.4
0.3
0.2
0.1
0.0
0
10
20
30
40
50
60
Time (min)
CTGCAAAGC
CTGCAAACC
CTGCAAAGT

Figure 5

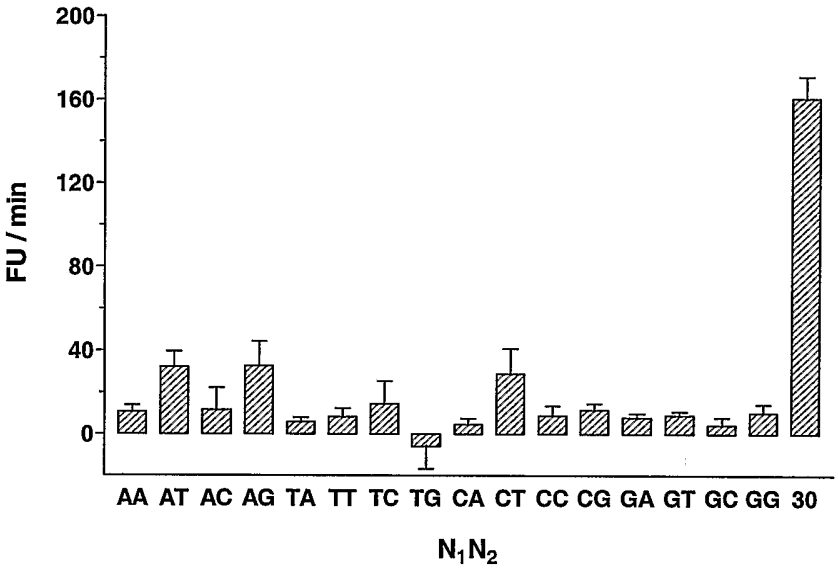

200
160
120
80
40
0
FU / min
AA AT AC AG TA TT TC TG CA CT CC CG GA GT GC GG 30
N1N2

Figure 6

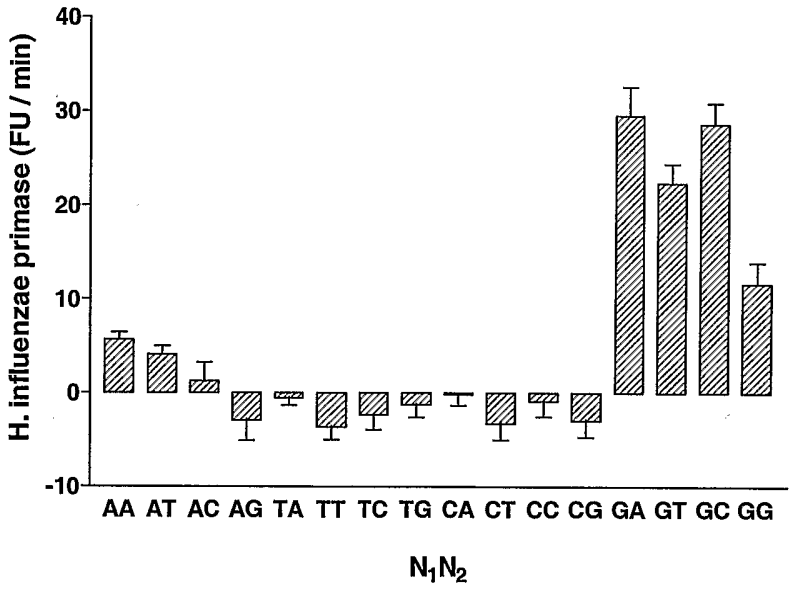

H. influenzae primase (FU / min)
40
30
20
10
0
-10
AA AT AC AG TA TT TC TG CA CT CC CG GA GT GC GG
N1N2

Figure 7

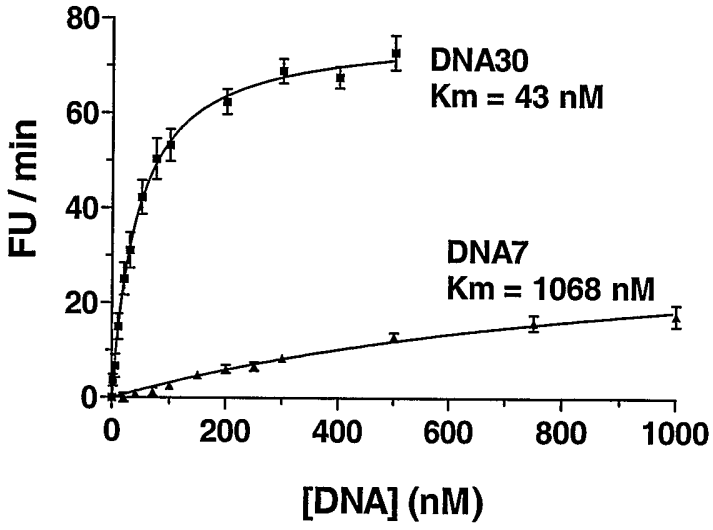
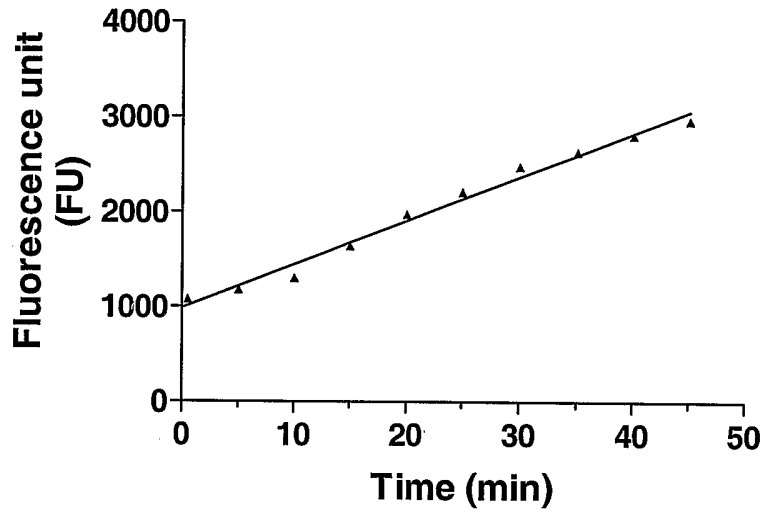

80
60
40
20
0
FU / min
0
200
400
600
800
1000
[DNA] (nM)
DNA30
Km = 43 nM
DNA7
Km = 1068 nM

Figure 8


Fluorescence unit (FU)
4000
3000
2000
1000
0
0
10
20
30
40
50
Time (min)