

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 2011319755 B2**

(54) Title
Compositions of toehold primer duplexes and methods of use

(51) International Patent Classification(s)
C12Q 1/68 (2006.01)

(21) Application No: **2011319755** (22) Date of Filing: **2011.10.27**

(87) WIPO No: **WO12/058488**

(30) Priority Data

(31) Number	(32) Date	(33) Country
61/407,291	2010.10.27	US

(43) Publication Date: **2012.05.03**

(44) Accepted Journal Date: **2017.02.23**

(71) Applicant(s)
President and Fellows of Harvard College

(72) Inventor(s)
Zhang, David Yu; Yin, Peng

(74) Agent / Attorney
Pizzeys Patent and Trade Mark Attorneys Pty Ltd, GPO Box 1374, BRISBANE, QLD, 4001, AU

(56) Related Art
US 2009/0191546 A1
Zhang, D. Y. et al., 'Control of DNA Strand Displacement Kinetics Using Toehold Exchange', Journal of American Chemical Society, 2009, Vol. 131, Pages 17303-17314
Hall, B. et al., 'Kinetic Optimization of a Protein-Responsive Aptamer Beacon', Biotechnology and Bioengineering, 2009, Vol. 103, No. 6, Pages 1049-1059

CORRECTED VERSION

(19) World Intellectual Property Organization International Bureau



(10) International Publication Number WO 2012/058488 A9

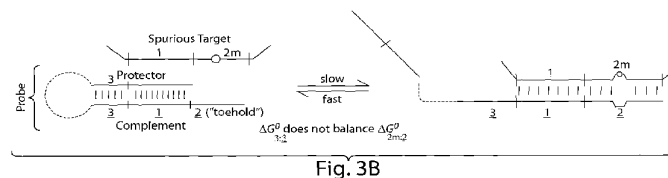
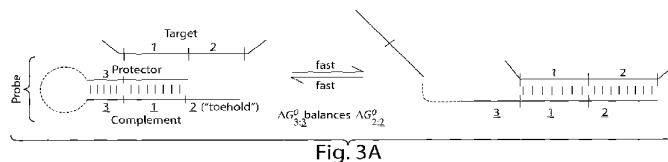
(43) International Publication Date 3 May 2012 (03.05.2012)

- (51) International Patent Classification: C12Q 1/68 (2006.01)
(21) International Application Number: PCT/US2011/058178
(22) International Filing Date: 27 October 2011 (27.10.2011)
(25) Filing Language: English
(26) Publication Language: English
(30) Priority Data: 61/407,291 27 October 2010 (27.10.2010) US
(71) Applicant: PRESIDENT AND FELLOWS OF HARVARD COLLEGE
(72) Inventors: and
(75) Inventors/Applicants: ZHANG, David, Yu; YIN, Peng
(74) Agent: TREVISAN, Maria, A.
(81) Designated States

CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

- Published: with international search report (Art. 21(3)) with sequence listing part of description (Rule 5.2(a))
(48) Date of publication of this corrected version: 5 July 2012
(15) Information about Correction: see Notice of 5 July 2012

(54) Title: COMPOSITIONS OF TOEHOLD PRIMER DUPLEXES AND METHODS OF USE



(57) Abstract: Provided herein are primers and primer systems having improved specificity and kinetics over existing primers, and methods of use thereof.

WO 2012/058488 A9

COMPOSITIONS OF TOEHOLD PRIMER DUPLEXES AND METHODS OF USE

RELATED APPLICATIONS

This application claims priority under 35 U.S.C. § 119(e) from U.S. provisional application serial number 61/407,291, filed October 27, 2010, the entire contents of which are incorporated by reference herein.

FIELD OF INVENTION

The embodiments described herein relate to partially double-stranded nucleic acid primers and their use in, for example, nucleic acid synthesis methods.

BACKGROUND OF INVENTION

Nucleic acids are vital information carriers of biology, and the detection, amplification, and identification of nucleic acids has formed the basis for a vast sector of biotechnology. In particular, methods such as the polymerase chain reaction (PCR) (Saiki *et al. Science* 239, 487-491 (1988)) have been used all over the world as a reliable means of amplifying DNA, while reverse transcriptase methods have been used to probe the transcriptome. The operation of DNA polymerase, RNA polymerase, and reverse transcriptase typically uses a short oligonucleotide fragment known as a primer to direct the portion of a long target to be replicated or transcribed.

Although the specificity of nucleic acid hybridization is frequently sufficient to direct enzymatic activity for most target sequences, targets with repetitive sequence, secondary structure, and high G/C content are difficult to amplify with high yield. Furthermore, high backgrounds of other nucleic acids can frequently lead to incorrect amplification, such as in the case of single copy human genome amplification. Finally, multiplexed amplification, such as from a DNA chip pool, can be difficult to achieve due to the large number of orthogonal amplification reactions that must occur simultaneously. Similar problems exist for transcription and for reverse transcription.

SUMMARY OF INVENTION

The hybridization of nucleic acids is specific at the single nucleotide level. For example, cytosine preferentially binds to guanine, and adenine preferentially binds to thymine or uracil.

However, for nucleic acid molecules composed of many nucleotides, the specificity of hybridization is reduced, and nucleic acids with near complementary sequences will bind almost as strongly as perfect complementary sequences. Given a heterogeneous mixture of target nucleic acid of interest (“targets”) and nucleic acids with sequences that differ from the target by, for example, one nucleotide (“spurious targets”), a significant portion of primers complementary to the target will hybridize instead to spurious targets.

Because the correct targets bind with a slightly higher affinity to a primer having a complementary sequence, given enough time, correct targets will eventually displace spurious targets in binding to a complementary primer. Though, this process is very slow, and would take months at the nanomolar concentrations typical of many experimental systems.

In order to mitigate the propensity of complementary primers binding to spurious targets it is often necessary to operate nucleic acid primer-based experimental systems near the melting temperature of the primer/target complex. Because this melting temperature is generally much higher than the temperature at which most biological systems naturally operate, this high temperature requirement precludes the experimental system from operating under normal biological conditions. Additionally, because the melting temperature will vary from target to target, the requisite narrow temperature range for such experimental systems restricts the simultaneous use of multiple primers to detect a plurality of targets.

Provided herein are primers (*e.g.*, primer duplexes and hairpin primer duplexes) that, in embodiments, are able to rapidly bind to nucleic acid targets with high specificity at a broad range of temperatures. These primers may be used, for example, in nucleic acid synthesis reactions (*e.g.*, PCR), microarray analyses, imaging methods, and single nucleotide polymorphism (SNP) analyses. The primers may also be used in nucleic acid detection assays where they function primarily as “probes”. Accordingly, regardless of the application, the primers of the invention may be referred herein to interchangeably as “probes”. Regardless of the application (or method of use), the primers of the invention overcome problems commonly experienced when specific hybridization is required in the presence of spurious targets, and more particularly when such spurious targets are present in excess.

The primers provided herein possess several unique properties that facilitate their use in combination with enzymes that act upon nucleic acids. First, the primers are thermodynamically designed to bind with high specificity to only their intended targets, and they show high

discrimination against even single-nucleotide changes. Second, the specificity of the primers enables PCR, transcription, and reverse transcription of traditionally difficult targets, such as those having significant sequence repetition, secondary structure, and/or high G/C content. The high degree of specificity that can be achieved with these primers further enables accurate processing even in high nucleic acid backgrounds such as single-copy human genome amplification. Third, the partially double-stranded nature of the primers means that they are unlikely to interact with each other, and consequently they are amenable to highly multiplexed replication, transcription, and/or reverse transcription reactions. Finally, the hybridization of these primers to targets is relatively robust to temperature and salinity, and therefore the primers may be of significantly greater length than standard primers, which in turn provides further enhanced specificity and primer design flexibility.

In some embodiments, the nucleic acid primers discussed here are rationally designed so that the standard free energy for hybridization (*e.g.*, theoretical standard free energy) between the specific target nucleic acid molecule and the primer is close to zero, while the standard free energy for hybridization between a spurious target (even one differing from the specific (actual) target by as little as a single nucleotide) and the primer is high enough to make their binding unfavorable by comparison. The inventors accomplished this by designing a primer having (a) a “toehold” single-stranded target specific region, (b) a “branch migration” double-stranded target specific region, and (c) a “balance” double-stranded target non-specific region.

In some embodiments, the primer may be comprised of a single strand that self hybridizes to form double-stranded regions. In some embodiments, the primer may be comprised of two strands. As an example of the latter embodiment, the primers is comprised of a first or complement strand and a second or protector strand. The complement strand, as its name implies, is partially complementary to the target of interest and will hybridize to the target. The protector strand, on the other hand, is designed to not hybridize to the target and rather to compete with the target (or spurious target) for binding of the complement.

The “toehold” region is present in the complement strand, is complementary to a target sequence and not complementary to a protector region. The “balance” region in the complement strand (*i.e.*, the complement balance region) is complementary to part of the protector (*i.e.*, to the protector balance region) and not complementary to target sequence. The hybridization energy of toehold to target is matched or nearly matched to the hybridization energy of complement

balance region to protector balance region (adjusting for various other thermodynamic considerations). The sequence of the balance region is rationally designed to achieve this matching under desired conditions of temperature and primer concentration. As a result, the equilibrium for the actual target and primer rapidly approaches 50% target:primer::protector:primer (or whatever ratio is desired), while equilibrium for the spurious target and primer greatly favors protector:primer. The abundant free primer in the presence of specific target facilitates its highly sensitive and specific detection.

In some embodiments, the nucleic acid primers discussed here are designed so that the concentration-adjusted free energy for hybridization between the specific target nucleic acid molecule and the primer is close to zero, while the concentration-adjusted standard free energy for hybridization between a spurious target and the primer is high enough to make their binding unfavorable by comparison. "Concentration-adjusted free energy," as used herein, refers to $\Delta G^\circ + (\Delta n)RT \ln(c)$, where R is the universal gas constant, T is temperature in Kelvins, c is concentration of the primer, and Δn is the change in the number of molecules through the course of the reaction ($\Delta n = -1$ for standard hybridization, $\Delta n = 0$ for two-stranded primer hybridization).

Aspects of the invention therefore provide the primer compositions comprising the primers, compositions comprising the complement and protector strands (for example in kits), methods of making the primers, and methods of using the primers in assays or reactions including without limitation nucleic acid synthesis and /or detection assays or reactions.

Thus, in one aspect, the invention provides a partially double-stranded primer comprised of (a) first nucleic acid strand (also referred to herein as a complement strand) and second nucleic acid strand (also referred to herein as a protector strand), wherein the first and second strands when hybridized to each other are arranged into (1) a double-stranded target-non-specific (balance) region, (2) a double-stranded target-specific (branch migration) region, and (3) a single-stranded target-specific (toehold) region contributed to by the first nucleic acid strand, wherein the double-stranded target-non-specific region has a standard free energy approximately equal to the standard free energy for the single-stranded target-specific region bound to a target nucleic acid. The partially double-stranded primer may comprise one or more double-stranded target-non-specific regions, one or more double-stranded target-specific regions, and/or one or more single-stranded target-specific regions. In some embodiments, the partially double-stranded primer may comprise one or two double-stranded target-non-specific (balance) regions,

one or more double-stranded target-specific (branch migration) regions, and/or one or more single-stranded target-specific (toehold) regions.

In some embodiments, the second nucleic acid strand comprises a non-extendable nucleotide at its 3' end and/or the first nucleic acid strand comprises a non-natural nucleotide at or near the 3' end of its target-non-specific region. In some embodiments, the non-extendable nucleotide is a non-natural nucleotide or a dideoxy nucleotide. In some embodiments, the non-natural nucleotide is iso-C, iso-G or deoxyuridine. These examples are intended as non-limiting.

In some embodiments, the double-stranded target non-specific region is about 4-20 nucleotides in length. The double-stranded target non-specific region may be longer than 20 nucleotides, such as for example 4-21 nucleotides in length. In some embodiments, it may be about 12-192 nucleotides in length.

In some embodiments, the single stranded target specific region is about 4-20 nucleotides in length. The single stranded target specific region may be longer than 20 nucleotides, such as for example 4-21 nucleotides in length. In some embodiments, it may be about 12-192 nucleotides in length.

In some embodiments, the double-stranded target non-specific region and the single stranded target specific region have similar or identical proportions of A/T nucleotides (and typically similar or identical proportions of G/C nucleotides). In some embodiments, the first and second nucleic acid strands are comprised of DNA or RNA or a combination thereof.

In another aspect, the invention provides a single-stranded primer that partially self-hybridizes to form (1) a double-stranded target-non-specific region, (2) a double-stranded target-specific region, (3) single-stranded target-specific region, and (4) a hairpin loop region, wherein the one or more double-stranded target-non-specific region has a concentration-adjusted standard free energy approximately equal to the concentration-adjusted standard free energy for the one or more single-stranded target-specific region bound to a target nucleic acid.

In another aspect, the invention provides a composition comprising the any of the aforementioned primers. The composition may further comprise a carrier such as a buffer, optionally comprising a preservative, one or more salts, etc. The composition may also comprise an excess of single-stranded protector strands, wherein each protector strand comprises a protector balance region and a protector branch migration region. The single stranded protector strands may each comprise a non-extendable and/or non-naturally occurring nucleotide, preferably at its 3' end.

In another aspect, the invention provides a system comprising a nucleic acid target, a polymerase, and any of the foregoing primers. In some embodiments, the primer is a partially double-stranded primer comprising a first and a second nucleic acid strand arranged into (1) a double-stranded target-non-specific region, (2) a double-stranded target-specific region, and (3) a single-stranded target-specific region contributed to by the first nucleic acid strand.

In some embodiments, the nucleic acid target is a single-stranded. In some embodiments, the nucleic acid target is DNA or RNA. In some embodiments, the nucleic acid target comprises repetitive sequence, secondary structure and/or high GC content. In some embodiments, the nucleic acid target is present in a plurality of different nucleic acids. In some embodiments, the nucleic acid target is present as a single copy or in low copy (e.g., less than 0.001%, less than 0.01%, less than 0.1%, or less than 1%) in a plurality of different nucleic acids.

In some embodiments, the system comprises a plurality of any of the foregoing primers such as a plurality of different partially double-stranded primers. In some embodiments, the system comprises at least two of the foregoing primers, such as at least two partially double-stranded primers, which together can be used to amplify a region of the nucleic acid target.

In another aspect, the invention provides a composition comprising the any of the aforementioned systems. The composition may further comprise a carrier such as a buffer, optionally comprising a preservative, one or more salts, one or more enzymes such as a polymerase, nucleotides suitable for nucleic acid synthesis, etc. The composition may also comprise an excess of single-stranded protector strands, wherein each protector strand comprises a protector balance region and a protector branch migration region. The single stranded protector strands may each comprise a non-extendable and/or non-naturally occurring nucleotide, preferably at its 3' end.

In another aspect, the invention provides a method comprising contacting any of the foregoing primers, including any of the foregoing partially double-stranded primers to a sample, and detecting hybridization of the primer to a target in the sample.

In some embodiments, the primer such as the partially double-stranded primer is labeled with a detectable moiety. In some embodiments, the detectable moiety comprises a fluorophore or a radioisotope.

The target will typically be a nucleic acid. In some embodiments, the target is a single-stranded nucleic acid. In some embodiments, the target is DNA or RNA. In some embodiments,

the target is a nucleic acid that comprises repetitive sequence, secondary structure and/or high GC content. In some embodiments, the target is present in a plurality of different nucleic acids. In some embodiments, the target is present as a single copy or in low copy (e.g., less than 0.001%, less than 0.01%, less than 0.1%, or less than 1%) in a plurality of different nucleic acids.

In some embodiments, this and other methods described herein are performed at a temperature below the melting temperature of the complement strand-target complex. In some embodiments, this and other methods described herein are performed at a temperature between and including room temperature up to and including 50°C, or up to and including 40°C, or up to and including 30°C. In some embodiments, this and other methods described herein are performed at about 37 °C. In some embodiments, this and other methods described herein are performed in an excess of protector strand that comprises a protector balance region and a protector branch migration region and that is identical to the protector strand in the partially double-stranded primer. In this and other methods described herein, the primer may be any of the foregoing primers including the partially double-stranded primers.

In another aspect, the invention provides a method comprising hybridizing a single-stranded target-specific (toehold) region of a first (complement) strand of any of the foregoing partially double-stranded primers to a nucleic acid target, thereby dissociating the first strand of the primer from the second (protector) strand of the primer, and extending the first strand at its 3' end, in a target-complementary manner, in the presence of a polymerase.

In another aspect, the invention provides a method comprising performing a nucleic acid synthesis reaction in the presence of a nucleic acid target, a polymerase, and one or more of the foregoing partially double-stranded primers.

In some embodiments, the nucleic acid synthesis reaction is a nucleic acid amplification reaction. In some embodiments, the nucleic acid amplification reaction is polymerase chain reaction (PCR). In some embodiments, the nucleic acid synthesis reaction is a transcription reaction. In some embodiments, the transcription reaction is a reverse transcription reaction.

In some embodiments, two partially double-stranded primers are used.

In another aspect, the invention provides a method of performing a multiplexed nucleic acid amplification reaction comprising amplifying multiple unique nucleic acid molecules using any of the foregoing primers including the partially double-stranded primer.

In another aspect, the invention provides a kit comprising one or more (including a plurality) of any of the foregoing partially double-stranded primers, and one or more nucleic acid synthesis reagents such as enzymes, nucleotides, salts, EDTA, a buffer, etc.

In some embodiments, the one or more nucleic acid synthesis reagents is selected from the group consisting of a buffer, nucleotides, and a polymerase.

In some embodiments, the kit further comprises an excess of protector strand that is identical to the protector strand comprised in the primer.

In some embodiments, the kit further comprises instructions for use.

In another aspect, the invention provides a kit comprising a first single-stranded (complement) nucleic acid in a first container, and a second single-stranded (protector) nucleic acid that is complementary to a region of the first single-stranded nucleic acid, in a second container, wherein, when the first and second single-stranded nucleic acids are hybridized to each other, a partially double-stranded nucleic acid is formed that comprises (1) a double-stranded target-non-specific region, (2) a double-stranded target-specific region, and (3) a single-stranded target-specific region contributed to by the first nucleic acid, wherein the first single-stranded nucleic acid comprises a non-natural nucleotide and/or the second single-stranded nucleic acid comprises a non-extendable nucleotide at its 3' end.

In some embodiments, the kit further comprises instructions for use. In some embodiments, the kit further comprises one or more nucleic acid synthesis reagents such as those recited above. In some embodiments, the one or more nucleic acid synthesis reagents is selected from the group consisting of a buffer, nucleotides, and a polymerase.

In some embodiments, the protector strand is provided in the kit in an amount (e.g., a molar amount) that is greater than the amount (e.g., a molar amount) of complement strand in the kit.

In some embodiments of the foregoing aspects and inventions, particularly those relating to two strand primers, the nucleotide sequence of the primer is selected such that:

$|\Delta G_1^\circ - \Delta G_2^\circ - \Delta G_3^\circ| \leq \Delta G_R^\circ$, wherein: ΔG_1° is the standard free energy of hybridization of the protector balance region to the complement balance region; ΔG_2° is the standard free energy of hybridization of the protector balance region to the sequence immediately adjacent in the first direction to the target nucleic acid sequence, if any; ΔG_3° is the standard free energy of

hybridization of the toehold region to the second target nucleic acid sequence; and ΔG_R° is 3.5 kcal/mol.

In one aspect, provided herein is a primer duplex system comprising a complement strand and a protector strand, wherein the protector strand comprises a nucleic acid having: a protector branch migration region having a first end, a second end, and a sequence that corresponds to a first target nucleic acid sequence having a first end and a second end, wherein the first end of the protector branch migration region and the first end of the first target nucleic acid sequence are either both 5' or else both 3'; and a protector balance region immediately adjacent to the first end of the protector branch migration region having a sequence that does not correspond to sequence immediately adjacent to the first end of the first target nucleic acid sequence, if any; and the complement primer comprises a nucleic acid having: a complement branch migration region having a first end and a second end, and a sequence that is complementary to the protector branch migration region, wherein the first end of the complement branch migration region and the first end of the first target nucleic acid sequence are either both 5' or else both 3'; a toehold region that is: immediately adjacent to the first end of the complement branch migration region; and complementary to a second target nucleic acid sequence that is immediately adjacent to the second end of the first target nucleic acid sequence; and a complement balance region that: is immediately adjacent to the second end of the complement branch migration region; is complementary to the protector balance region; and has a sequence such that:

$|\Delta G_1^\circ - \Delta G_2^\circ - \Delta G_3^\circ| \leq \Delta G_R^\circ$, wherein: ΔG_1° is the standard free energy of hybridization of the protector balance region to the complement balance region; ΔG_2° is the standard free energy of hybridization of the protector balance region to the sequence immediately adjacent in the first direction to the target nucleic acid sequence, if any; ΔG_3° is the standard free energy of hybridization of the toehold region to the second target nucleic acid sequence; and ΔG_R° is 3.5 kcal/mol.

In another aspect, provided herein is a primer duplex system comprising a nucleic acid having a protector strand, a hairpin region and a complement strand, wherein: the protector strand comprises a protector branch migration region and a protector balance region, wherein: the protector branch migration region has: a first end; a second end; and a sequence that corresponds to a first target nucleic acid sequence having a first end and a second end, wherein

the first end of the protector branch migration region and the first end of the first target nucleic acid sequence are either both 5' or else both 3'; and the protector balance region has: a first end; a second end immediately adjacent to the first end of the protector branch migration region; and a sequence that does not correspond to sequence immediately adjacent to the first end of the first target nucleic acid sequence, if any; the hairpin region comprises: a first end; and a second end immediately adjacent to the first end of the protector balance region; and the complement strand comprises a complement balance region, a complement branch migration region, and a toehold region, wherein: the complement balance region has: a first end; a second end immediately adjacent to the first end of the hairpin region; and a sequence that is complementary to the protector balance region; the complement branch migration region has: a first end; a second end immediately adjacent to the first end of the complement balance region; and a sequence that is complementary to the protector branch migration region, wherein the first end of the complement branch migration region and the first end of the first target nucleic acid sequence are either both 5' or else both 3'; the toehold region is: immediately adjacent to the first end of the complement branch migration region; and complementary to a second target nucleic acid sequence that is immediately adjacent to the second end of the first target nucleic acid sequence; and the complement balance region has a sequence such

that: $\left| \Delta G_1^\circ - \Delta G_2^\circ - \Delta G_3^\circ + \Delta G_4^\circ + RT \ln(c) \right| \leq \Delta G_R^\circ$, wherein: ΔG_1° is the standard free energy of hybridization of the protector balance region to the complement balance region; ΔG_2° is the standard free energy of hybridization of the protector balance region to the sequence immediately adjacent in the first direction to the target nucleic acid sequence, if any; and ΔG_3° is the standard free energy of hybridization of the toehold region to the second target nucleic acid sequence; ΔG_4° is the standard free energy of confinement of the hairpin region; R is the ideal gas constant; T is the temperature at which the primer duplex system is to be used; c is the concentration at which the primer duplex system is to be used; and ΔG_R° is 3.5 kcal/mol.

In yet another aspect, provided herein system having, in 3' to 5' order, a first protector strand, a first hairpin region, a complement strand, a second hairpin region and a second protector strand, wherein: the first protector strand comprises: a first protector branch migration region having a sequence that corresponds to a first target nucleic acid sequence; and a first protector balance region that: is immediately 5' to the first protector branch migration region; and

has a sequence that does not correspond to sequence immediately 5' to the first target nucleic acid sequence, if any; the first hairpin region is immediately 5' to the first protector balance region; the complement strand comprises: a first complement balance region that: is immediately 5' to the first hairpin region; and has a sequence complementary to the sequence of the first protector balance region; a first complement branch migration region that: is immediately 5' to the first complement balance region; and has a sequence complementary to a first protector branch migration region; a toehold region that: is immediately 5' to the first complement branch migration region; and has a sequence that is complementary to a second target nucleic acid sequence that is immediately 3' to the first target nucleic acid sequence; a second complement branch migration region that: is immediately 5' to the toehold region; and has a sequence complementary to a third target nucleic acid sequence that is immediately 3' to the second target nucleic acid sequence; a second complement balance region that: is immediately 5' to the second complement branch migration region; has a sequence that is not complementary to sequence immediately 3' to the third target nucleic acid sequence, if any; the second hairpin region is immediately 5' to the second complement balance region; and the second protector strand comprises: a second protector balance region that: is immediately 5' to the second hairpin region; and has a sequence complementary to the second complement balance region; and a second protector branch migration region that: is immediately 5' to the second protector balance region; and has a sequence complementary to the second complement branch migration region; wherein the first complement balance region and the second complement balance region have sequences such that: $\left| \Delta G_1^\circ - \Delta G_2^\circ + \Delta G_3^\circ - \Delta G_4^\circ - \Delta G_5^\circ + \Delta G_6^\circ + \Delta G_7^\circ + RT \ln(c) \right| \leq \Delta G_R^\circ$,

wherein: ΔG_1° is the standard free energy of hybridization of the first protector balance region to the first complement balance region; ΔG_2° is the standard free energy of hybridization of the first complement balance region to the sequence immediately 5' to the first target nucleic acid sequence, if any; ΔG_3° is the standard free energy of hybridization of the second protector balance region to the second complement balance region; ΔG_4° is the standard free energy of hybridization of the second complement balance region to the sequence immediately 3' to the third target nucleic acid sequence, if any; ΔG_5° is the standard free energy of hybridization of the toehold region to the second target nucleic acid sequence; ΔG_6° is the standard free energy of

confinement of the first hairpin region; ΔG_7° is the standard free energy of confinement of the second hairpin region; R is the ideal gas constant; T is the temperature at which the primer duplex system is to be used; and c is the concentration at which the primer duplex system is to be used; and ΔG_R° is 3.5 kcal/mol.

In still another aspect, provided herein is a primer duplex system comprising a hairpin primer and a protector strand, wherein: the hairpin primer comprises a nucleic acid having: a first protector strand having: a first protector branch migration region having a sequence that corresponds to a first target nucleic acid sequence; and a first protector balance region that: is immediately 5' to the first protector branch migration region; and has a sequence that does not correspond to sequence immediately 5' to the first target nucleic acid sequence, if any; a hairpin region immediately 5' to the first protector balance region; a complement strand having: a first complement balance region that: is immediately 5' to the first hairpin region; and has a sequence complementary to the sequence of the first protector balance region; a first complement branch migration region that: is immediately 5' to the first complement balance region; and has a sequence complementary to a first protector branch migration region; a toehold region that: is immediately 5' to the first complement branch migration region; and has a sequence that is complementary to a second target nucleic acid sequence that is immediately 3' to the first target nucleic acid sequence; a second complement branch migration region that: is immediately 5' to the toehold region; and has a sequence complementary to a third target nucleic acid sequence that is immediately 3' to the second target nucleic acid sequence; a second complement balance region that: is immediately 5' to the second complement branch migration region; has a sequence that is not complementary to sequence immediately 3' to the third target nucleic acid sequence, if any; and the protector comprises a nucleic acid having: a second protector strand having: a second protector balance region that has a sequence complementary to the second complement balance region; and a second protector branch migration region that: is immediately 5' to the second protector balance region; and has a sequence complementary to the second complement branch migration region; wherein the first complement balance region and the second complement balance region have sequences such that:

$|\Delta G_1^\circ - \Delta G_2^\circ + \Delta G_3^\circ - \Delta G_4^\circ - \Delta G_5^\circ + \Delta G_6^\circ| \leq \Delta G_R^\circ$, wherein: ΔG_1° is the standard free energy of hybridization of the first protector balance region to the first complement balance region; ΔG_2° is

the standard free energy of hybridization of the first complement balance region to the sequence immediately 5' to the first target nucleic acid sequence, if any; ΔG_3° is the standard free energy of hybridization of the second protector balance region to the second complement balance region; ΔG_4° is the standard free energy of hybridization of the second complement balance region to the sequence immediately 3' to the third target nucleic acid sequence, if any; ΔG_5° is the standard free energy of hybridization of the toehold region to the second target nucleic acid sequence; ΔG_6° is the standard free energy of confinement of the hairpin region; and ΔG_R° is 3.5 kcal/mol.

In a further aspect, provided herein is a primer duplex system comprising a protector strand and a hairpin primer, wherein: the protector strand comprises a nucleic acid having: a first protector strand having: a first protector branch migration region having a sequence that corresponds to a first target nucleic acid sequence; and a first protector balance region that: is immediately 5' to the first protector branch migration region; and has a sequence that does not correspond to sequence immediately 5' to the first target nucleic acid sequence, if any; the hairpin primer comprises a nucleic acid having: a complement strand having: a first complement balance region that has a sequence complementary to the sequence of the first protector balance region; a first complement branch migration region that: is immediately 5' to the first complement balance region; and has a sequence complementary to a first protector branch migration region; a toehold region that: is immediately 5' to the first complement branch migration region; and has a sequence that is complementary to a second target nucleic acid sequence that is immediately 3' to the first target nucleic acid sequence; a second complement branch migration region that: is immediately 5' to the toehold region; and has a sequence complementary to a third target nucleic acid sequence that is immediately 3' to the second target nucleic acid sequence; a second complement balance region that: is immediately 5' to the second complement branch migration region; has a sequence that is not complementary to sequence immediately 3' to the third target nucleic acid sequence, if any; a hairpin region immediately 5' to the second complement balance region; and a second protector strand having: a second protector balance region that: is immediately 5' to the second hairpin region; and has a sequence complementary to the second complement balance region; and a second protector branch migration region that: is immediately 5' to the second protector balance region; and has a

sequence complementary to the second complement branch migration region; wherein the first complement balance region and the second complement balance region have sequences such that: $|\Delta G_1^\circ - \Delta G_2^\circ + \Delta G_3^\circ - \Delta G_4^\circ - \Delta G_5^\circ + \Delta G_6^\circ| \leq \Delta G_R^\circ$, wherein: ΔG_1° is the standard free energy of hybridization of the first protector balance region to the first complement balance region; ΔG_2° is the standard free energy of hybridization of the first complement balance region to the sequence immediately 5' to the first target nucleic acid sequence, if any; ΔG_3° is the standard free energy of hybridization of the second protector balance region to the second complement balance region; ΔG_4° is the standard free energy of hybridization of the second complement balance region to the sequence immediately 3' to the third target nucleic acid sequence, if any; and ΔG_5° is the standard free energy of hybridization of the toehold region to the second target nucleic acid sequence; ΔG_6° is the standard free energy of confinement of the hairpin region; and ΔG_R° is 3.5 kcal/mol.

In another aspect, provided herein is a primer duplex system comprising a first protector strand, a complement strand and a second protector strand, wherein: the first protector strand comprises a nucleic acid having: a first protector branch migration region having a sequence that corresponds to a first target nucleic acid sequence; and a first protector balance region that: is immediately 5' to the first protector branch migration region; and has a sequence that does not correspond to sequence immediately 5' to the first target nucleic acid sequence, if any; the complement strand comprises a nucleic acid having: a first complement balance region that: is immediately 5' to the first hairpin region; and has a sequence complementary to the sequence of the first protector balance region; a first complement branch migration region that: is immediately 5' to the first complement balance region; and has a sequence complementary to a first protector branch migration region; a toehold region that: is immediately 5' to the first complement branch migration region; and has a sequence that is complementary to a second target nucleic acid sequence that is immediately 3' to the first target nucleic acid sequence; a second complement branch migration region that: is immediately 5' to the toehold region; and has a sequence complementary to a third target nucleic acid sequence that is immediately 3' to the second target nucleic acid sequence; a second complement balance region that: is immediately 5' to the second complement branch migration region; has a sequence that is not

complementary to sequence immediately 3' to the third target nucleic acid sequence, if any; and the second protector strand comprises: a second protector balance region that has a sequence complementary to the second complement balance region; and a second protector branch migration region that: is immediately 5' to the second protector balance region; and has a sequence complementary to the second complement branch migration region; wherein the first complement balance region and the second complement balance region have sequences such that: $|\Delta G_1^\circ - \Delta G_2^\circ + \Delta G_3^\circ - \Delta G_4^\circ - \Delta G_5^\circ - RT \ln(c)| \leq \Delta G_R^\circ$, wherein: ΔG_1° is the standard free energy of hybridization of the first protector balance region to the first complement balance region; ΔG_2° is the standard free energy of hybridization of the first complement balance region to the sequence immediately 5' to the first target nucleic acid sequence, if any; ΔG_3° is the standard free energy of hybridization of the second protector balance region to the second complement balance region; ΔG_4° is the standard free energy of hybridization of the second complement balance region to the sequence immediately 3' to the third target nucleic acid sequence, if any; and ΔG_5° is the standard free energy of hybridization of the toehold region to the second target nucleic acid sequence; R is the ideal gas constant; T is the temperature at which the primer duplex system is to be used; c is the concentration at which the primer duplex system is to be used; and ΔG_R° is 3.5 kcal/mol.

In yet another aspect, provided herein is a primer duplex system comprising, in 3' to 5' order, a first protector strand, a first hairpin region, a complement strand, a second hairpin region and a second protector strand, wherein: the first protector strand has a sequence that corresponds to a first target nucleic acid sequence; the first hairpin region is immediately 5' of the first protector strand; the complement strand comprises: a first complement branch migration region that: is immediately 5' of the first hairpin region; and has a sequence complementary to the sequence of the first protector strand; a toehold region that: is immediately 5' of the first complement branch migration region; and has a sequence complementary to a second target nucleic acid sequence that is immediately 3' of the first target nucleic acid sequence; and a second complement branch migration region that: is immediately 5' of the toehold region; and has a sequence complementary to a third target nucleic acid sequence that is immediately 3' of the second nucleic acid sequence; the second hairpin region is immediately 5' of the second

complement branch migration region; and the second protector strand has a sequence that is complementary to the sequence of the second complement branch migration region.

In any one of the foregoing aspects, ΔG_R° may be 2.0 kcal/mol, 1.0 kcal/mol, or 0.5 kcal/mol; and/or c may be about 10 nM; and/or T may be about 293 K or about 338 K; and/or the toehold region may be between 4 and 20 nucleotides in length, between about 4 and 15 nucleotides in length, or between about 4 and 10 nucleotides in length; and/or the first end of the protector branch migration region may be 5' or 3'; and/or the primer duplex system may further comprise a functionalized fluorescent group or dye; and/or the primer duplex system may be immobilized on a solid support; and/or the hairpin region may be no greater than 20 nucleotides in length or no greater than 10 nucleotides in length; and/or the sequence of the hairpin region may be selected from the group consisting of a poly-adenosine sequence, poly-deoxyadenosine sequence, a poly-5'-methyluridine sequence, a poly-thymidine sequence, a poly-guanosine sequence, a poly-deoxyguanosine sequence, a poly-cytidine sequence, a poly-deoxycytidine sequence, a poly-uridine sequence, and a poly-deoxyuridine sequence; and/or the first target nucleic acid sequence and/or the second target nucleic acid sequence may be sequences that naturally occur in an organism or a virus; and/or the first target nucleic acid sequence and/or the second target nucleic acid sequence may be sequences that naturally occur in a micro-RNA.

In one aspect, provided herein is a method of detecting a target nucleic acid in a sample comprising: contacting a target nucleic acid with a primer duplex system of any one of the embodiments described herein; and detecting the formation of a complex between the target nucleic acid and at least a part of the primer duplex system. In some embodiments, the primer duplex system further comprises a functionalized fluorescent group or dye. In some embodiments, the primer duplex system is immobilized on a solid support. In some embodiments, the contacting occurs in a cell. In some embodiments, the target nucleic acid is a nucleic acid that naturally occurs in an organism or a virus. In some embodiments, the target nucleic acid is a micro-RNA.

In another aspect, provided herein is a method of amplifying a sequence contained within a target nucleic acid comprising: forming a solution comprising: a target nucleic acid; a primer duplex system of any one of the embodiments described herein; and reagents for performing an amplification reaction; and incubating the solution under conditions such that a sequence

contained within the target nucleic acid is amplified. In some embodiments, the target nucleic acid is a nucleic acid that naturally occurs in an organism or a virus.

These and other aspects and embodiments of the invention will be explained in greater detail herein.

BRIEF DESCRIPTION OF DRAWINGS

FIGS. 1A, 1B, 2A, 2B, 3A, 3B, 4-8, 9A, and 9B depict exemplary nucleic acid probe systems.

FIGS. 10, 11, 12A, 12B, 13, and 14 depict exemplary methods of using nucleic acid probe systems.

FIGS. 15A and 15B depict highly specific polymerase chain reaction (PCR) using the primer duplexes provided herein.

FIGS. 16A-16D show experimental demonstrations of primer hybridization with single nucleotide discrimination.

FIGS. 17A-17D show additional experimental results and statistics on the single-base discrimination abilities of primer duplexes.

FIGS. 18A-18B show experimental results using duplex primers to improve the PCR yield of a quasi-repetitive target.

DETAILED DESCRIPTION OF INVENTION

A significant challenge in probe-based nucleic acid assays is that nucleic acids having sequences similar to that of a target will hybridize to the target's complement with strong thermodynamics and fast kinetics. However, as described herein, the kinetics and thermodynamics of strand displacement reactions can be partially decoupled, so that reactions that are only slightly thermodynamically favorable or even unfavorable can nonetheless have kinetics as fast as the hybridization of two complementary strands. The compositions and methods described herein take advantage of this decoupling mechanism to provide nucleic acid probe systems with improved specificity and kinetics.

Provided herein are highly specific nucleic acid probe systems and methods of using such probe systems. In certain embodiments, the nucleic acid probe systems described herein comprise complement probes having regions complementary to a target sequence that are

protected from hybridization to spurious targets by protector regions complementary to a portion of the complement probes. The free energy of the binding reaction between the target and the protected probe is finely controlled via the rationally designed bases of one or more balancing regions, which have sequences that do not correspond to the target nucleic acid sequence or its complement. In certain embodiments, a protector and a complement probe form regions on a single nucleic acid molecule and are separated from one another by one or more nucleic acid hairpins.

The methods and compositions described herein possess several unique properties that facilitate their use in hybridization assays. First, the nucleic acid probe systems described herein reliably convert small sequence differences between targets and spurious targets into large differences in binding affinity and reaction rates between hybridization of the target vs. spurious target with the probe. Second, the nucleic acid probe systems described herein can be designed to operate at any of a wide range of temperatures and salt concentrations, and can therefore function reliably under many different experimental conditions. Third, use of the nucleic acid probe systems described herein can result in hybridization reactions that are kinetically fast even at room temperature, which facilitates rapid and high-throughput analysis of nucleic acids. Fourth, the nucleic acid probe systems described herein are rationally designed, and therefore are unlikely to interact unfavorably or in unexpected ways with other biomolecules.

Accordingly, provided herein are primer compositions, methods of making such compositions, and methods of their use. The embodiments described herein are premised in part on the discovery that primer (*e.g.*, a pair of partially hybridized primers, or a single self-hybridizing primer) that are partially double-stranded and partially single-stranded, when used in a nucleic acid synthesis reaction for example, are able to discriminate between fully complementary targets and those having one or more mismatches (*i.e.*, spurious targets). As demonstrated herein, the primer duplexes described herein are superior to standard primers in, for example, PCR reactions using spurious targets such as those having quasi-repetitive sequences.

The primer duplexes herein comprise a single-stranded region referred to herein as a “toehold” from which the primer duplex initiates binding to a target, a double-stranded “balance region” which spontaneously dissociates so that a single primer strand does not complete hybridization (along the full length of the primer) to the target, and a double-stranded branch

migration region, in between the toehold and balance regions, which is fully complementary to a target nucleic acid sequence. Mechanistically, it is thought that hybridization to a target begins at the toehold and continues along the length of the complement strand until the primer is no longer “double-stranded”. This assumes complementarity between the target and the branch region as well. As used herein, a nucleic acid “region” or “domain” is a consecutive stretch of nucleotides of any length. When nucleotide mismatches exist between the “target” and the complement strand, displacement of the second strand (*i.e.*, the protector strand) is thermodynamically unfavorable and the association between the complement strand and the “target” is reversed. It is to be understood that in this latter description, the “target” is actually a spurious target since it comprises nucleotide differences or mismatches from the complement strand.

Because the standard free energy favors a complete match (fully complementary) between the target sequence of the nucleic acid and branch migration plus toehold regions of the primer rather than a mismatch (*e.g.*, single nucleotide change), the first (complement) strand of the primer will bind stably to a target in the absence of a mismatch but not in the presence of a mismatch. If a mismatch exists between the first (complement) strand of the primer and the target, the primer duplex prefers to reform via newly exposed single-stranded balance regions. In this way, the frequency of beginning a nucleic acid synthesis reaction at an incorrect position in a target (or in a sample, for that matter) is reduced. This type of discrimination is typically not possible using the standard single-stranded primers of the prior art because in those reactions there is no competing nucleic acid strand (such as the protector strand) to which a mismatched primer strand would prefer to bind. In some embodiments, the primers described herein may be significantly longer than conventional primers (*e.g.*, those used for polymerase chain reaction (PCR) amplification) because the instant primers rely on the presence of a competing, protector strand for specificity rather than on melting temperature to discriminate between complementary and mismatched sequences. Accordingly, the instant primers may be selected and used in a manner that is temperature independent.

The primer duplexes described herein therefore improve specificity of for example nucleic acid synthesis reactions and, in some embodiments, allow for a greater degree of multiplexing of primers. Preliminary experiments, the results of which are provided herein, show that the PCR yield of quasi-repetitive targets can be significantly improved using the

primer duplexes provided herein as compared to standard primers (*e.g.*, 75% vs. 30%). The primer duplexes described herein also provide for specific nucleic acid detection and amplification from a heterogeneous population of nucleic acids, such as for example, detecting and amplifying a bacterial DNA from a sample comprised of human DNA, which has broad applicability in detection of rare organisms such as biowarfare agents.

Primer Duplexes

As used herein, the primers of the invention may be referred to as “primer duplexes” to convey that they may be provided and/or exist in a conformation in which they comprise double-stranded regions. Accordingly, the terms “primer” and “primer duplex” may be used interchangeably.

The primer duplexes provide improved specificity and kinetics over existing primers. A “primer duplex” herein refers to a primer comprising a first strand (referred to herein as a “complement strand”) and a second strand (referred to herein as a “protector strand”) partially complementary to the first strand. In some embodiments, the complement strand and the protector strand are separate single-stranded nucleic acid molecules (FIGS. 1A and 1B). In other embodiments, the complement strand and the protector strand are connected to each other and separated by a hairpin region to form contiguous regions of a single nucleic acid molecule (FIGS. 2A and 2B). As used herein, a “hairpin region” is a single-stranded loop of nucleotides connecting two double-stranded regions of a nucleic acid. The general structure of exemplary primer duplexes is illustrated in the Figures and described herein. It is to be understood that, in most instances, when reference is made to a complement region or a protector region (or vice versa), each region is typically within a single “primer duplex” (or a single primer system). For example, a complement balance region in a primer of the invention is complementary to a protector balance region in the same primer such that a complement balance region of one primer of the invention does not hybridize to a protector balance region of different physically separate primer.

In embodiments in which the primer of the invention consists of only a single strand, the complement “strand” may be referred to as the complement region, and the protector “strand” may be referred to as the protector region.

In certain embodiments, the complement strand (or region) comprises a toehold region, a complement branch migration region, and a complement balance region, while the protector strand (or region) comprises a protector branch migration region and a protector balance region. As used herein, a nucleic acid “region” is a consecutive stretch of nucleotides of any length. Toehold and branch migration regions are each designed to be complementary to, and thus “base-pair” with (*e.g.*, hybridize to), adjacent regions in a target nucleic acid. A region of a complement strand that base-pairs with a region in a target nucleic acid is referred to as a “target-specific” region. Balance regions are designed to be not complementary to, and thus to not base-pair with, a target nucleic acid. Balance regions therefore are referred to as “target-non-specific” regions. In certain aspects, when the complement strand (or region) and the protector strand (or region) are hybridized to each other (are double-stranded), a primer duplex is formed. Thus, in some aspects, a primer duplex comprises a target-specific single-stranded toehold region, a target-specific double-stranded branch migration region, and a target-non-specific double-stranded balance region (FIGS. 1A and 1B). In some instances, the primer duplex may also comprise a hairpin loop, as described in greater detail below.

The primer duplexes described herein may be designed to hybridize specifically with a target nucleic acid. The efficacy of a primer, for example, in a nucleic acid amplification reaction, depends on the specificity, efficiency, and fidelity of the primer. Typical nucleic acid primers often bind to spurious targets with a thermodynamic and kinetic profile comparable to that of the same primer binding to its intended, specific target nucleic acid, except between the melting temperatures of the mismatched duplex and the perfectly hybridized duplex. Accordingly, mismatched and perfect duplexes can be distinguished by their melting temperatures. The primers of the invention, in contrast, distinguish between spurious and true target in a relatively temperature-independent manner.

A “spurious target” herein refers to a nucleic acid molecule that differs from a target nucleic acid molecule by at least one nucleotide within the region hybridizing to the complement strand. For example, TCGACGGGG is a spurious target, if the target is TCGAAGGG. In certain embodiments, a spurious target comprises at least 2, at least 3, at least 4, or more nucleotide changes relative to the target. Primer binding to spurious targets reduces the fidelity (accuracy) of, *e.g.*, nucleic acid amplification. The primer duplexes presented herein are designed to alter the standard free energy of strand displacement with spurious targets,

permitting discrimination between correct targets and spurious targets, including spurious targets that differ from a correct target by only one nucleotide. As described herein, the protector strand is responsible for altering the standard free energy to allow the complement strand to discriminate between correct and spurious targets.

The primers described herein are rationally designed to facilitate strand displacement reactions with finely tuned kinetics and thermodynamics such that kinetics and thermodynamics of strand displacement reactions are partially decoupled. As a result of this decoupling, reactions only slightly thermodynamically favorable or even unfavorable can nonetheless have kinetics as fast as the hybridization of two complementary strands.

For example, at 37° C and 1 M Na⁺, the concentration-adjusted standard free energy for hybridization of a primer to a perfectly complementary (correct or specific) target (*i.e.*, 100% nucleotide match) is between 1.9 kcal/mol and 6.6 kcal/mol more favorable than the concentration-adjusted standard free energy for hybridization of the same primer to a spurious target for every nucleotide that the spurious target differs from the intended target. In certain embodiments, the present primer duplexes use toehold exchange strand displacement reactions to translate this 1.9 to 6.6 kcal/mol difference in concentration-adjusted standard free energy to an optimal discrimination between the target and spurious targets. An example of the thermodynamics/kinetics of primer duplex binding to a target nucleic acid is described as follows in reference to FIGS 3A and 3B.

For purposes of this example, the target nucleic acid has at least two regions, (1) and (2). In certain embodiments, region (1) may be about 10 to about 200 (including 14-200 or 20-200) nucleotides long, while region 2 may be smaller, for example, about 4 to about 20 nucleotides long. As used herein, the terms “nucleotide” and “bases” are used interchangeably. The protector strand includes a protector branch migration region adjacent to a protector balance region (3). The protector branch migration region corresponds to target region 1, while the protector balance region (3) does not correspond to region (1) or region (2) or any region immediately 5' of the target regions. A nucleic acid sequence, domain or region is “immediately adjacent to”, “immediately 5'” or “immediately 3'” to another sequence if the two sequences are part of the same nucleic acid molecule and if no bases separate the two sequences. The complement strand includes a complement balance region (3), a complement branch migration region (1), and a toehold region (2). The complement balance region (3) is complementary to the

protector balance region (3), the complement branch migration region (1) is complementary to the protector branch migration region and target region (1) (i.e., the protector branch migration region and the target region (1) are identical in sequence and this both bind to the complement branch migration region (1), and the toehold region (2) is complementary to target region (2).

In certain embodiments, the balance region is designed so that its concentration-adjusted standard free energy ($\Delta G_{3,3}^{\circ}$) is the same or about the same as the concentration-adjusted standard free energy for the toehold region bound to target region (2) ($\Delta G_{2,2}^{\circ}$). In some instances, for a 10 nanomolar (nM) primer used in a reaction at 37 °C, $|\Delta G_{2,2}^{\circ}|$ and $|\Delta G_{3,3}^{\circ}|$ (the vertical bars denoting absolute value) should each be less than about 11.3 kcal/mol to ensure dissociation of the full protector strand from the target.

In some embodiments, when the primer duplex interacts with a specific (correct) target nucleic acid molecule (FIG. 3A), the dissociation of (3):(3) and the association of (2):(2) balance one another, and the (1):(1) hybridization thermodynamics are identical for the target nucleic acid and for the protector strand interacting with the complement strand. The total free energy change between the two states is relatively small (e.g., about 1 kcal/mol), and the reaction quickly (e.g., less than a minute) reaches an equilibrium of about 50:50. In certain embodiments, the balance region may be designed to have standard free energy very close to that of the toehold region binding to target region 2, so that the equilibrium balance is, for example, 60:40 or 70:30. In some embodiments, the design of the balance region may also take into account other contributors to free energy change during the reaction, such as hybridization between the protector balance region and upstream target sequences (which in some instances is negligible), confinement of a hairpin (if present), intended temperature of use, and intended primer concentration.

In some embodiments, when the primer duplex instead interacts with a spurious target nucleic acid molecule (FIG. 3B), the dissociation of (3):(3) and the association of (2m):(2) are not balanced because spurious target region (2m) is not fully complementary to the toehold region. The equilibrium is consequently shifted to the state in which the primer duplex does not bind the spurious target.

Explained another way, in some instances, the free energy of the complement strand bound to the protector strand is $\Delta G_{3,3}^{\circ} + \Delta G_{1,1}^{\circ}$ (ignoring contribution from the optional hairpin

region and other considerations), which balances the free energy of the complement strand bound to specific target, $\Delta G_{2;2}^{\circ} + \Delta G_{1;1}^{\circ}$. In this example, this is because the balance region (3) of the primer duplex has been designed to have a concentration-adjusted standard free energy equal to (or approximately equal to) that of target region (2). When the primer duplex interacts with a spurious target having a single-nucleotide (base) change in target region (2m), the system's free energy $\Delta G_{2m;2}^{\circ} + \Delta G_{1;1}^{\circ}$ is less negative than that of the primer duplex, and therefore disfavored in equilibrium.

As used herein, the term "approximately equal to" in reference to standard free energy means that the first referenced free energy is within 10% of the second referenced free energy. In some embodiments, a first free energy that is approximately equal to a second free energy is within about +3 kcal/mol to about -3 kcal/mol of the second free energy. It is to be understood that the differences between the first and second true energies may be less than or about 1 kcal/mol, less than or about 2 kcal/mol, less than or about 3 kcal/mol, less than or about 3.5 kcal/mol, or more, in some embodiments.

Although FIG. 3B illustrates a single nucleotide change corresponding to region (2)/(2m) of a target nucleic acid molecule, the present primer duplexes can also discriminate between a specific target and a spurious target having a nucleotide change in region (1). When the spurious target has a single-base change in target region 1, then the primer duplex's standard free energy after binding becomes $\Delta G_{2;2}^{\circ} + \Delta G_{1m;1}^{\circ}$, where $\Delta G_{1m;1}^{\circ}$ is the standard free energy of the mismatched target region (1) binding to the primer duplex's complement region (1). Because of the single-base change, the primer duplex's free energy is less negative than $\Delta G_{3;3}^{\circ} + \Delta G_{1;1}^{\circ}$ (free energy of complement primer bound to protector), so equilibrium is shifted to the state in which the primer duplex does not bind the spurious target region. Standard free energies can be calculated theoretically based on the knowledge in the art and the teachings provided herein.

Complement and Protector Strands, Regions or Domains

The complement domains of the nucleic acid probe systems described herein each include a plurality of regions, including a toehold region and one or more complement target regions. Both the toehold region and the one or more complement target regions have nucleic acid sequences that are complementary to nucleic acid sequences of the target nucleic acid. The

toehold region and the complement target region are therefore able to base-pair with and therefore form a complex with a sequence of a target nucleic acid when the nucleic acid probe system is contacted with a target nucleic acid under appropriate hybridization conditions. The complement domains may also include one or more complement balance regions. The one or more complement balance regions are rationally designed. Thus, the sequences of the one or more complement balance regions are not designed to be complementary to a target nucleic acid sequence.

A toehold region is complementary to (and thus hybridizes to) a sequence in the target nucleic acid molecule; however, a toehold region does not hybridize to a protector strand. Thus, when the complement strand is hybridized to the target nucleic acid molecule, the toehold region is also hybridized to the target nucleic acid molecule, but when the complement strand is hybridized to the protector strand, the toehold region remains single-stranded. A toehold region may be positioned at the 3' end or the 5' end of the complement strand (*e.g.*, is an extension of the 3' end or 5' end of the complement strand).

In certain embodiments, a toehold region is about 4 nucleotides to about 20 nucleotides in length, about 4 nucleotides to about 15 nucleotides in length, or about 4 nucleotides to about 10 nucleotides in length. In some embodiments, a toehold region is 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides in length. In some embodiments, the toehold region is greater than 20 nucleotides in length, including for example less than or about 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or more nucleotides.

The complement branch migration region is complementary to a sequence in the target nucleic acid molecule and to the protector branch migration region. Thus, when the complement strand hybridizes to a target nucleic acid molecule, the complement branch migration region hybridizes to the target nucleic acid. When the complement strand hybridizes to its protector strand, the complement branch migration region hybridizes to the protector branch migration region.

In certain embodiments, a branch migration region is no more than 200, 100, 75, 50, 40, 30, 25 or 20 nucleotides in length. In some embodiments, a branch migration region is about 10 nucleotides to about 200 nucleotides in length. In certain embodiments, a branch migration region is about 10 nucleotides to about 150 nucleotides, about 10 nucleotides to about 100 nucleotides, or about 10 nucleotides to about 50 nucleotides in length. In particular

embodiments, a branch migration region is 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, or 200 nucleotides in length. In particular embodiments, a branch migration region may be more than 200 nucleotides in length, depending on the target nucleic acid molecule of interest.

The balance regions of a complement strand and a protector strand are complementary to each other (*i.e.*, form a double-stranded nucleic acid) but are non-complementary to the target of interest (*i.e.*, neither forms a double-stranded nucleic acid with the target). Thus, when a complement strand hybridizes to a target nucleic acid molecule, the complement balance region does not hybridize to the target nucleic acid molecule. When the complement strand hybridizes to its protector strand, the complement balance region hybridizes to the protector balance region.

The design of the balance region is dependent on the design of the toehold region. In some embodiments, the balance region is designed such that the thermodynamic profile of the balance region is comparable to that of the toehold region. In some embodiments, the thermodynamic profile is based on a theoretic model, using for example, Mfold software available at the bioinfo website of RPI. The number and/or nature of nucleotides within a balance region is comparable to that of the toehold region. For example, if a toehold region is comprised of 40% A and T nucleotides and 60% G and C nucleotides, then the balance region should also be comprised of 40% A and T nucleotides and 60 % G and C nucleotides. In embodiments, the balance region is designed such that no more than three consecutive nucleotides are complementary to a sequence on the target nucleic acid to avoid binding of the balance region to the target nucleic acid.

In some embodiments, the length of a balance region is short enough so that the complement and protector spontaneously dissociate from each other. In some embodiments, a

balance region is about 4 nucleotides to about 20 nucleotides in length, about 4 nucleotides to about 15 nucleotides in length, or about 4 nucleotides to about 10 nucleotides in length. In some embodiments, a balance region is 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides in length. In some embodiments, a balance region is greater than 20 nucleotides, including for example less than about 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or more nucleotides. In some embodiments, the number of consecutive nucleotides that are complementary to a nucleotide sequence within the target nucleic acid may be greater than three provided that the balance region does not bind to the target nucleic acid.

In some embodiments, for example those where the primer duplex contains two separate nucleic acid strands, the design of a balance region does not depend on the concentration of the primer duplex or the temperature at which the primer duplex is formed/used. In some embodiments, a balance region is designed such that the standard free energy for the reaction in which the protector strand is displaced from the complement strand by the target nucleic acid molecule is close to zero kcal/mol. As used herein, "close to zero" means the standard free energy for the reaction is within 3.5 kcal/mol from 0 kcal/mol. In certain embodiments, the standard free energy of this displacement reaction is within 3.5, 3.0, 2.5, 2.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, or 0.1 kcal/mol of zero kcal/mol.

In other embodiments, for example those where the primer duplexes is formed by a single nucleic acid molecule (*e.g.*, a hairpin region separating the complement strand (or region or domain) and the protector strand (or region or domain)), the design of a balance region will be dependent on the primer duplex concentration as well as reaction temperature. In such embodiments, a balance region is designed so that the standard free energy for the reaction in which the protector strand is displaced from the complement strand by the target nucleic acid plus $RT\ln(c)$ is close to zero kcal/mol, where R is the universal gas constant (0.0019858775(34) kcal/mol•K), T is the temperature at which the primer duplex is used, and c is the concentration at which primer duplex is used. In some embodiments, the temperature at which the primer duplexes are used is about 273 K (0°C), 277 K, 283 K, 288 K, 293K, 298 K, 303 K, 308 K, 313 K, 318 K, 323 K, 328 K, 333 K, 338 K, 343 K, 348 K, 353 K, 358 K or 363 K (90°C). In some embodiments the concentration (c) at which the primer duplexes are used is about 1 nM, 2 nM, 3 nM, 4 nM, 5 nM, 10 nM, 15 nM, 20 nM, 25 nM, 30 nM, 35 nM, 40 nM, 45 nM, 50 nM, 55 nM, 60 nM, 65 nM, 70 nM, 75 nM, 80 nM, 85 nM, 90 nM, 95 nM, 100 nM, 125 nM, 150 nM, 175

nM, 200 nM, 225 nM, 250 nM, 300 nM, 350 nM, 400 nM, 450 nM, 500 nM, 600 nM, 700 nM, 800 nM, 900 nM or 1 μ M. In certain embodiments, the standard free energy of this displacement reaction plus $RT\ln(c)$ is within 3.5, 3.0, 2.5, 2.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, or 0.1 kcal/mol of zero kcal/mol.

In some embodiments, a primer duplex may include one or more hairpin regions that connect the complement strand to the protector strand. In certain embodiments, the hairpin region of a primer duplex can be of any length. In some embodiments, the hairpin region is more than 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4 or 3 nucleotides in length. In some embodiments, the sequence of the hairpin is not complementary to a sequence of the target nucleic acid molecule.

In certain embodiments, the hairpin region has a poly-mononucleotide sequence, such as a poly-adenosine sequence, poly-deoxyadenosine sequence, a poly-5'-methyluridine sequence, a poly-thymidine sequence, a poly-guanosine sequence, a poly-deoxyguanosine sequence, a poly-cytidine sequence, a poly-deoxycytidine sequence, a poly-uridine sequence or a poly-deoxyuridine sequence.

The primer duplex described herein may be one of at least two orientations. For example, in one orientation, the toehold region is located at the 5' end, immediately adjacent to the complement branch migration region (*i.e.*, no intervening nucleotides between the two regions), and the complement balance region is located at the 3' end, immediately adjacent to the complement branch migration region. In this orientation, the protector balance region is at the 5' end of the protector strand, immediately adjacent to the protector branch migration region (FIG. 1A). In another orientation, the toehold region is located at the 3' end, immediately adjacent to the complement branch migration region, and the complement balance region is located at the 5' end, immediately adjacent to the complement branch migration region. In this orientation, the protector balance region is at the 3' end of the protector strand, immediately adjacent to the protector branch migration region (FIG. 1B).

Regardless of orientation, the sequence of the complement balance region is such that such that:

$$\left| \Delta G_1^\circ - \Delta G_2^\circ - \Delta G_3^\circ \right| \leq \Delta G_R^\circ,$$

where:

ΔG_1° is the standard free energy of hybridization of the protector balance region to the complement balance region;

ΔG_2° is the standard free energy of hybridization of the protector balance region to the sequence immediately adjacent in the first direction to the target nucleic acid sequence, if any;

ΔG_3° is the standard free energy of hybridization of the toehold region to the second target nucleic acid sequence; and

ΔG_R° is 3.5 kcal/mol.

In some embodiments, a primer duplex comprises a complement strand longer than the protector, the difference in length being dependent on the length of the toehold region of the complement strand. The lengths of the primers are designed such that hybridization of the complement to the target of interest has a standard free energy (ΔG°) close to zero. Release of the protector strand (from the primer duplex) ensures that this hybridization reaction is entropically neutral and robust to concentration. As a result, in some embodiments, this reaction at room temperature (*e.g.*, about 25 °C or about 298 K) parallels the specificity of hybridization achieved at near melting temperature across many conditions .

As intended herein, a ΔG° (change in standard free energy) “close to zero” refers to an absolute value (amount) less than or about 1 kcal/mol, less than or about 2 kcal/mol, less than or about 3 kcal/mol, or less than or about 3.5 kcal/mol. In some embodiments, the standard free energy of a balance region or toehold region is >-1 kcal/mol to < 1 kcal/mol >-3 kcal/mol to <3 kcal/mol or >-3.5 kcal/mol to <3.5 kcal/mol.

The primer duplexes may be prepared at a ratio of protector strand to complement strand of about 2:1 to about 5:1. In some embodiments, the ratio of protector strand to complement strand is about 2:1, about 3:1, about 4:1, or about 5:1. In some embodiments, the ratio of protector strand to complement strand is 2:1, 2.1:1, 2.2:1, 2.3:1, 2.4:1, 2.5:1, 2.6:1, 2.7:1, 2.8:1, 2.9:1, 3:1, 3.1:1, 3.2:1, 3.3:1, 3.4:1, 3.5:1, 3.6:1, 3.7:1, 3.8:1, 3.9:1, 4:1, 4.2:1, 4.3:1, 4.4:1, 4.5:1, 4.6:1, 4.7:1, 4.8:1, 4.9:1, or 5:1. The primer duplexes may also be used together with excess protector strand in any of the assays or reactions described herein. The protector strand may be

in about equal to or more than 2-, 5-, 10-, 20-, 50-, 100-, or 500-fold molar excess relative to the primer.

Hairpin Primer Duplex Systems (e.g., Single-Stranded Systems)

In certain embodiments, a primer duplex includes a single nucleic acid that comprises a complement region or domain, a hairpin region, and a protector region or domain. In some embodiments, the complement domain hybridizes to the protector domain, forming a primer duplex having an intervening hairpin-loop region. Like other primer systems disclosed here, such hairpin primer systems are designed to specifically hybridize to a target nucleic acid molecule. Herein, a “hairpin primer duplex system” or a “hairpin system” includes a complement balance region, a complement branch migration region, a toehold region, a protector balance region, and a protector branch migration region. As described above, a complement balance region is complementary to a protector balance region; a complement branch migration region is complementary to a protector branch migration region and a target nucleic acid region; and a toehold region is complementary to a target nucleic acid region. A protector branch migration region corresponds to a target nucleic acid region and is complementary to a complement branch migration region. Because the hairpin primer duplex systems described herein are formed by a single nucleic acid molecule, the design of the complement balance region will be dependent on the temperature and concentration at which the primer system is to be used, as described herein. It is to be understood that though the sequence of the target nucleic acid molecule may be used to describe the characteristics of the primer systems, in some embodiments, the target nucleic acid itself may or may not be a component of the primer system (e.g., two-stranded or single-stranded systems).

For primer duplexes having a hairpin region, the standard free energy of the confinement of the hairpin region may be considered when determining the standard free energy for the reaction in which the protector strand is displaced from the complement strand by the target nucleic acid. Approximate values for the standard free energy of hairpin confinement for hairpins of various sizes are provided in Table 1 (from SantaLucia and Hicks, *Annu. Rev. Biophys. Biomol. Struct.*, 33:414-440, (2004)).

Hairpin Size	ΔG° of Hairpin Confinement
3 nt	3.5 kcal/mol
4 nt	3.5 kcal/mol
5 nt	3.3 kcal/mol
6 nt	4.0 kcal/mol
7 nt	4.2 kcal/mol
8 nt	4.3 kcal/mol
9 nt	4.5 kcal/mol
10 nt	4.6 kcal/mol
12 nt	5.0 kcal/mol
14 nt	5.1 kcal/mol
16 nt	5.3 kcal/mol
18 nt	5.5 kcal/mol
20 nt	5.7 kcal/mol
25 nt	6.1 kcal/mol
30 nt	6.3 kcal/mol

The standard free energy of the confinement of the hairpin regions having lengths not provided in Table 1 (*e.g.*, a length of n) can be estimated using the following equation:

$$\Delta G^\circ(\text{loop} - n) = \Delta G^\circ(\text{loop} - x) + 2.44RT \ln(n/x)$$

where $\Delta G^\circ(\text{loop} - n)$ is the unknown standard free energy of the confinement of a hairpin region of n nucleotides in length, $\Delta G^\circ(\text{loop} - x)$ is the known standard free energy of the confinement of a hairpin region of n nucleotides in length (*e.g.*, as provided in Table 1), R is the ideal gas constant, and T is the temperature at which the primer duplex is to be used. Additional information on the calculation of standard free energies of hairpin region confinement is provided in SantaLucia and Hicks, *id.*, which is hereby incorporated by reference in its entirety.

The hairpin primer duplex systems described herein may be one of at least two orientations. For example, in one orientation shown in FIG. 2A, the toehold region is located at the 5' end of the nucleic acid molecule. The 5' end of the complement branch migration region is immediately adjacent to the 3' end of the toehold region; the 5' end of the complement balance region is immediately adjacent to the 3' end of the complement branch migration region; the 5' end of the hairpin region is immediately adjacent to the 3' end of the complement balance region; the 5' end of the protector balance region is immediately adjacent to the 3' end of the

hairpin region; and the 5' end of the protector branch migration region is immediately adjacent to the 3' end of the protector balance region. In this orientation, when the nucleic acid molecule is subjected to conditions that permit annealing, the hairpin region forms a loop that extends from the complement balance region and the protector balance region. In another orientation shown in FIG. 2B, the toehold region is located at the 3' end of the nucleic acid molecule. The 3' end of the complement branch migration region is immediately adjacent to the 5' end of the toehold region; the 3' end of the complement balance region is immediately adjacent to the 5' end of the complement branch migration region; the 3' end of the hairpin region is immediately adjacent to the 5' end of the complement balance region; the 3' end of the protector balance region is immediately adjacent to the 5' end of the hairpin region; and the 3' end of the protector branch migration region is immediately adjacent to the 5' end of the protector balance region.

Regardless of orientation, the complement balance region of the hairpin primer duplex system has a sequence such that:

$$\left| \Delta G_1^\circ - \Delta G_2^\circ - \Delta G_3^\circ + \Delta G_4^\circ + RT \ln(c) \right| \leq \Delta G_R^\circ,$$

where:

ΔG_1° is the standard free energy of hybridization of the protector balance region to the complement balance region;

ΔG_2° is the standard free energy of hybridization of the protector balance region to the sequence immediately adjacent in the first direction to the target nucleic acid sequence, if any; and

ΔG_3° is the standard free energy of hybridization of the toehold region to the second target nucleic acid sequence;

ΔG_4° is the standard free energy of confinement of the hairpin region;

R is the ideal gas constant;

T is the temperature at which the primer system is to be used;

c is the concentration at which the primer system is to be used; and

ΔG_R° is 3.5 kcal/mol.

Other Primer Systems

Additional primer systems are depicted in FIGS. 4-7. Each of these primer systems include a complement domain and two protector domains. The complement domain has a single toehold region that is flanked by two complement branch migration regions and two complement balance regions. Each of the protector domains include a protector branch migration region that has a sequence complementary to the sequence of one of the complement branch migration regions, and a protector balance region that has a sequence that is complementary to the sequence of one of the complement balance regions. The difference between the primer systems of FIGS. 4-7 is in the number and location of the hairpin regions, which in turn affects the design of the complement balance regions. Like the other primer systems disclosed here, these primer systems are designed to specifically hybridize to a target nucleic acid. Though the sequence of the target nucleic acid is used to describe the characteristics of a primer system, in some embodiments, the target nucleic acid is not a component of a primer system.

As depicted in FIG. 4, a primer system may have, in 3' to 5' order, a first protector domain, a first hairpin region, a complement domain, a second hairpin region and a second protector domain.

In such embodiments, the first protector domain includes a first protector branch migration region and a first protector balance region. The first protector branch migration region has a sequence that corresponds to a first target nucleic acid sequence. The first protector balance region is immediately 5' to the first protector branch migration region and has a sequence that does not correspond to sequence immediately 5' to the first target nucleic acid sequence, if any, on the target nucleic acid.

In this embodiment, the first hairpin region is immediately 5' to the first protector balance region.

The complement domain of such target nucleic acids comprises a first complement balance region, a first complement branch migration region, a toehold region, a second complement branch migration region and a second complement balance region. The first complement balance region is immediately 5' to the first hairpin region and has a sequence complementary to the sequence of the first protector balance region. The first complement branch migration region is immediately 5' to the first complement balance region and has a sequence complementary to a first protector branch migration region. The toehold region is

immediately 5' to the first complement branch migration region and has a sequence that is complementary to a second target nucleic acid sequence that is immediately 3' to the first target nucleic acid sequence on the target nucleic acid. The second complement branch migration region is immediately 5' to the toehold region and has a sequence complementary to a third target nucleic acid sequence that is immediately 3' to the second target nucleic acid sequence on the target nucleic acid. The second complement balance region is immediately 5' to the second complement branch migration region and has a sequence that is not complementary to sequence immediately 3' to the third target nucleic acid sequence, if any, on the target nucleic acid.

In such embodiments, the second hairpin region is immediately 5' to the second complement balance region.

In this embodiment the second protector domain includes a second protector balance region and a second protector branch migration region. The second protector balance region is immediately 5' to the second hairpin region and has a sequence complementary to the second complement balance region. The second protector branch migration region is immediately 5' to the second protector balance region and has a sequence complementary to the second complement branch migration region.

According to this embodiment, the first complement balance region and the second complement balance region have sequences such that:

$$\left| \Delta G_1^\circ - \Delta G_2^\circ + \Delta G_3^\circ - \Delta G_4^\circ - \Delta G_5^\circ + \Delta G_6^\circ + \Delta G_7^\circ + RT \ln(c) \right| \leq \Delta G_R^\circ,$$

where:

ΔG_1° is the standard free energy of hybridization of the first protector balance region to the first complement balance region;

ΔG_2° is the standard free energy of hybridization of the first complement balance region to the sequence immediately 5' to the first target nucleic acid sequence, if any;

ΔG_3° is the standard free energy of hybridization of the second protector balance region to the second complement balance region;

ΔG_4° is the standard free energy of hybridization of the second complement balance region to the sequence immediately 3' to the third target nucleic acid sequence, if any;

ΔG_5° is the standard free energy of hybridization of the toehold region to the second target nucleic acid sequence;

ΔG_6° is the standard free energy of confinement of the first hairpin region;

ΔG_7° is the standard free energy of confinement of the second hairpin region;

R is the ideal gas constant;

T is the temperature at which the primer system is to be used; and

c is the concentration at which the primer system is to be used; and

ΔG_R° is 3.5 kcal/mol.

As depicted in FIG. 5, in certain embodiments, a primer system may have a hairpin primer and a protector, where the hairpin primer is a nucleic acid that includes a first protector domain, a first hairpin region, a complement domain and the protector is a nucleic acid that includes a second protector domain.

In such embodiments, the first protector domain includes a first protector branch migration region and a first protector balance region. The first protector branch migration region has a sequence that corresponds to a first target nucleic acid sequence. The first protector balance region is immediately 5' to the first protector branch migration region and has a sequence that does not correspond to sequence immediately 5' to the first target nucleic acid sequence, if any, on the target nucleic acid.

In this embodiment, the hairpin region is immediately 5' to the first protector balance region.

The complement domain of such target nucleic acids comprises a first complement balance region, a first complement branch migration region, a toehold region, a second complement branch migration region and a second complement balance region. The first complement balance region is immediately 5' to the hairpin region and has a sequence complementary to the sequence of the first protector balance region. The first complement branch migration region is immediately 5' to the first complement balance region and has a sequence complementary to a first protector branch migration region. The toehold region is immediately 5' to the first complement branch migration region and has a sequence that is complementary to a second target nucleic acid sequence that is immediately 3' to the first target nucleic acid sequence on the target nucleic acid. The second complement branch migration

region is immediately 5' to the toehold region and has a sequence complementary to a third target nucleic acid sequence that is immediately 3' to the second target nucleic acid sequence on the target nucleic acid. The second complement balance region is immediately 5' to the second complement branch migration region and has a sequence that is not complementary to sequence immediately 3' to the third target nucleic acid sequence, if any, on the target nucleic acid.

In this embodiment, the second protector domain includes a second protector balance region and a second protector branch migration region. The second protector balance region has a sequence complementary to the second complement balance region. The second protector branch migration region is immediately 5' to the second protector balance region and has a sequence complementary to the second complement branch migration region.

According to this embodiment, the first complement balance region and the second complement balance region have sequences such that:

$$\left| \Delta G_1^\circ - \Delta G_2^\circ + \Delta G_3^\circ - \Delta G_4^\circ - \Delta G_5^\circ + \Delta G_6^\circ \right| \leq \Delta G_R^\circ,$$

where:

ΔG_1° is the standard free energy of hybridization of the first protector balance region to the first complement balance region;

ΔG_2° is the standard free energy of hybridization of the first complement balance region to the sequence immediately 5' to the first target nucleic acid sequence, if any;

ΔG_3° is the standard free energy of hybridization of the second protector balance region to the second complement balance region;

ΔG_4° is the standard free energy of hybridization of the second complement balance region to the sequence immediately 3' to the third target nucleic acid sequence, if any;

ΔG_5° is the standard free energy of hybridization of the toehold region to the second target nucleic acid sequence;

ΔG_6° is the standard free energy of confinement of the hairpin region; and

ΔG_R° is 3.5 kcal/mol.

As depicted in FIG. 6, in certain embodiments a primer system may have a protector and a hairpin primer, where the protector is a nucleic acid that includes a first protector domain and

the hairpin primer is a nucleic acid that includes a complement domain, hairpin region and a second protector domain.

In such embodiments, the first protector domain includes a first protector branch migration region and a first protector balance region. The first protector branch migration region has a sequence that corresponds to a first target nucleic acid sequence. The first protector balance region is immediately 5' to the first protector branch migration region and has a sequence that does not correspond to sequence immediately 5' to the first target nucleic acid sequence, if any, on the target nucleic acid.

The complement domain of such target nucleic acids comprises a first complement balance region, a first complement branch migration region, a toehold region, a second complement branch migration region and a second complement balance region. The first complement balance region has a sequence complementary to the sequence of the first protector balance region. The first complement branch migration region is immediately 5' to the first complement balance region and has a sequence complementary to a first protector branch migration region. The toehold region is immediately 5' to the first complement branch migration region and has a sequence that is complementary to a second target nucleic acid sequence that is immediately 3' to the first target nucleic acid sequence on the target nucleic acid. The second complement branch migration region is immediately 5' to the toehold region and has a sequence complementary to a third target nucleic acid sequence that is immediately 3' to the second target nucleic acid sequence on the target nucleic acid. The second complement balance region is immediately 5' to the second complement branch migration region and has a sequence that is not complementary to sequence immediately 3' to the third target nucleic acid sequence, if any, on the target nucleic acid.

According to such embodiments, the hairpin region is immediately 5' to the second complement balance region.

In this embodiment the second protector domain includes a second protector balance region and a second protector branch migration region. The second protector balance region is immediately 5' to the hairpin region and has a sequence complementary to the second complement balance region. The second protector branch migration region is immediately 5' to the second protector balance region and has a sequence complementary to the second complement branch migration region.

According to this embodiment, the first complement balance region and the second complement balance region have sequences such that:

$$\left| \Delta G_1^{\circ} - \Delta G_2^{\circ} + \Delta G_3^{\circ} - \Delta G_4^{\circ} - \Delta G_5^{\circ} + \Delta G_6^{\circ} \right| \leq \Delta G_R^{\circ},$$

where:

ΔG_1° is the standard free energy of hybridization of the first protector balance region to the first complement balance region;

ΔG_2° is the standard free energy of hybridization of the first complement balance region to the sequence immediately 5' to the first target nucleic acid sequence, if any;

ΔG_3° is the standard free energy of hybridization of the second protector balance region to the second complement balance region;

ΔG_4° is the standard free energy of hybridization of the second complement balance region to the sequence immediately 3' to the third target nucleic acid sequence, if any; and

ΔG_5° is the standard free energy of hybridization of the toehold region to the second target nucleic acid sequence;

ΔG_6° is the standard free energy of confinement of the hairpin region; and

ΔG_R° is 3.5 kcal/mol.

As depicted in FIG. 7, in certain embodiments a primer system may have a first protector, a complement primer and a second protector, where the first protector is a nucleic acid that includes a first protector domain, the complement primer is a nucleic acid that includes a complement domain, and the second protector is a nucleic acid that includes a second protector domain.

In such embodiments, the first protector domain includes a first protector branch migration region and a first protector balance region. The first protector branch migration region has a sequence that corresponds to a first target nucleic acid sequence. The first protector balance region is immediately 5' to the first protector branch migration region and has a sequence that does not correspond to sequence immediately 5' to the first target nucleic acid sequence, if any, on the target nucleic acid.

The complement domain of such target nucleic acids comprises a first complement balance region, a first complement branch migration region, a toehold region, a second complement branch migration region and a second complement balance region. The first complement balance region has a sequence complementary to the sequence of the first protector balance region. The first complement branch migration region is immediately 5' to the first complement balance region and has a sequence complementary to a first protector branch migration region. The toehold region is immediately 5' to the first complement branch migration region and has a sequence that is complementary to a second target nucleic acid sequence that is immediately 3' to the first target nucleic acid sequence on the target nucleic acid. The second complement branch migration region is immediately 5' to the toehold region and has a sequence complementary to a third target nucleic acid sequence that is immediately 3' to the second target nucleic acid sequence on the target nucleic acid. The second complement balance region is immediately 5' to the second complement branch migration region and has a sequence that is not complementary to sequence immediately 3' to the third target nucleic acid sequence, if any, on the target nucleic acid.

In this embodiment the second protector domain includes a second protector balance region and a second protector branch migration region. The second protector balance region has a sequence complementary to the second complement balance region. The second protector branch migration region is immediately 5' to the second protector balance region and has a sequence complementary to the second complement branch migration region.

According to this embodiment, the first complement balance region and the second complement balance region have sequences such that:

$$\left| \Delta G_1^\circ - \Delta G_2^\circ + \Delta G_3^\circ - \Delta G_4^\circ - \Delta G_5^\circ - RT \ln(c) \right| \leq \Delta G_R^\circ,$$

where:

ΔG_1° is the standard free energy of hybridization of the first protector balance region to the first complement balance region;

ΔG_2° is the standard free energy of hybridization of the first complement balance region to the sequence immediately 5' to the first target nucleic acid sequence, if any;

ΔG_3° is the standard free energy of hybridization of the second protector balance region to the second complement balance region;

ΔG_4° is the standard free energy of hybridization of the second complement balance region to the sequence immediately 3' to the third target nucleic acid sequence, if any; and

ΔG_5° is the standard free energy of hybridization of the toehold region to the second target nucleic acid sequence;

R is the ideal gas constant;

T is the temperature at which the primer system is to be used;

c is the concentration at which the primer system is to be used; and

ΔG_R° is 3.5 kcal/mol.

Primer Duplex Systems Lacking Balance Domains

In some embodiments, primer systems may lack balance domains. Such nucleic acids will hybridize with a target nucleic acid with fast kinetics if the target nucleic acid has a sequence complementary to the sequence of the toehold region of the primer system, but with slow kinetics if the target nucleic acid is mutated so that it does not contain a sequence complementary to the toehold region of the primer system. Such primer systems are therefore useful, for example, for locating difference and/or mutations in nucleic acid targets using kinetic discrimination.

As depicted in FIG. 8, in certain embodiments a primer system may include a nucleic acid having, in 3' to 5' order, a first protector domain, a first hairpin region, a complement domain, a second hairpin region and a second protector domain. The first protector domain of such primer systems has a sequence that corresponds to a first target nucleic acid sequence. The first hairpin region is immediately 5' of the first protector domain. The complement domain has a first complement branch migration region, a toehold region and a second complement branch migration region. The first complement branch migration region is immediately 5' of the first hairpin region and has a sequence complementary to the branch migration sequence of the first protector domain. The toehold region is immediately 5' of the first complement branch migration region and has a sequence complementary to a second target nucleic acid sequence that is immediately 3' of the first target nucleic acid sequence on the target nucleic acid molecule. The second complement branch migration region is immediately 5' of the toehold region and has a sequence complementary to a third target nucleic acid sequence that is

immediately 3' of the second nucleic acid sequence. The second hairpin region is immediately 5' of the second complement branch migration region. The second protector domain has a sequence that is complementary to the sequence of the second complement branch migration region.

Primer Modifications Generally

Each primer described herein may be comprised of DNA, RNA, or analogs thereof, and/or combinations thereof. In certain embodiments, a primer comprises one or more non-natural nucleotides. The incorporation of non-natural nucleotides in the primers can further augment the performance of the primer duplexes. In particular, the protector strand, while not intended to serve to initiate transcription, may happen to be complementary to other regions of the target or other background molecules, and may spuriously initiate replication/transcription. To prevent this, the use of a non-natural nucleotide or a dideoxy nucleotide at the 3' end of the second protector strand may serve to prevent unintended priming by that strand. Examples of non-natural nucleotides include, but are not limited to, iso-C, iso-G, deoxyuridine (see also Krueger *et al. Chem Biol.* 16:242-48 (2009), the teachings which relating to non-natural nucleotides are incorporated by reference herein).

In some embodiments, for example, in a polymerase chain reaction (PCR) where a repeated primed enzymatic function is used, the extended complement strand can become a target for subsequent primer hybridization. To preserve the specificity of primer hybridization for subsequent rounds of amplification, a balance region of a primer cannot be replicated. Introducing a non-natural nucleotide at the interface between the branch migration and balance regions of the complement strand, for example, may prevent the balance region from being replicated.

In certain embodiments, the primers described herein serve as starting points for polymerase extensions. To facilitate analysis of amplified (nucleic acid) fragments, labeled primers can also be used in PCR reactions. Labeled primers are those that are coupled (or conjugated) to a detectable moiety. Examples include fluorescent dyes, radioactive labels, and identifiable metals, nucleic acid sequences, and proteins. When a reaction is carried out with fluorescently labeled primers, amplicons (nucleic acid products) with a fluorescent label may be generated.

The primers described herein can be synthesized by any method known in the art (see, *e.g.*, Ogilvie *et al. J. Amer. Chem. Soc.* 99 (23): 7741–7743; Reese, C. B. *Tetrahedron* 34(21): 3143 (1978); Efimov *et al. Nucleic Acids Res.* 11(23): 8369–8387 (1983); Garegg *et al. Tetrahedron Lett.* 27(34): 4051 (1986); Beaucage *et al. Tetrahedron* 48(12): 2223 (1992); Efimov *et al. Nucleosides, Nucleotides & Nucleic Acids* 26 (8–9): 1087–93 (2007), incorporated herein by reference).

Target Nucleic Acid Molecules

A “target” can be a single-stranded (ss) or double-stranded (ds) nucleic acid. Target nucleic acids can be, for example, DNA, RNA, or the DNA product of RNA subjected to reverse transcription. In some embodiments, a target may be a mixture (chimera) of DNA and RNA. In other embodiments, a target comprises artificial nucleic acid analogs, for example, peptide nucleic acids (Nielsen *et al. Science* 254(5037): 1497-500 (1991)) or locked nucleic acids (Alexei *et al. Tetrahedron* 54(14): 3607-30 (1998)). In some embodiments, a target may be naturally occurring (*e.g.*, genomic DNA) or it may be synthetic (*e.g.*, from a genomic library). As used herein, a “naturally occurring” nucleic acid sequence is a sequence that is present in nucleic acid molecules of organisms or viruses that exist in nature in the absence of human intervention. In some embodiments, a target is genomic DNA, messenger RNA, ribosomal RNA, micro-RNA, pre-micro-RNA, pro-micro-RNA, viral DNA, viral RNA or piwi-RNA. In certain embodiments, a target nucleic acid is a nucleic acid that naturally occurs in an organism or virus. In some embodiments, a target nucleic acid is the nucleic acid of a pathogenic organism or virus. In certain embodiments the presence or absence of a target nucleic acid in a subject is indicative that the subject has a disease or disorder or is predisposed to acquire a disease or disorder. In certain embodiments the presence or absence of a target nucleic acid in a subject is indicative that the subject will respond well or poorly to a treatment, such as a drug, to treat a disease or disorder.

The terms “polynucleotide,” “nucleic acid” and “nucleic acid molecule” are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three-dimensional structure, and may perform any function. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci

(locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. A polynucleotide may be further modified, such as by conjugation with a labeling component. The term “recombinant” polynucleotide means a polynucleotide of genomic, cDNA, semi-synthetic, or synthetic origin which either does not occur in nature or is linked to another polynucleotide in a non-natural arrangement. The term “isolated nucleic acid” refers to a polynucleotide of natural or synthetic origin or some combination thereof, which (1) is not associated with the cell in which the “isolated nucleic acid” is found in nature, and/or (2) is operably linked to a polynucleotide to which it is not linked in nature.

A nucleic acid may also encompass single- and double-stranded DNA and RNA, as well as any and all forms of alternative nucleic acid containing modified bases, sugars, and backbones. The term “nucleic acid” thus will be understood to include, but not be limited to, single- or double-stranded DNA or RNA (and forms thereof that can be partially single-stranded or partially double-stranded), cDNA, aptamers, peptide nucleic acids (“PNA”), 2'-5' DNA (a synthetic material with a shortened backbone that has a base-spacing that matches the A conformation of DNA; 2'-5' DNA will not normally hybridize with DNA in the B form, but it will hybridize readily with RNA), and locked nucleic acids (“LNA”). Nucleic acid analogues include known analogues of natural nucleotides that have similar or improved binding, hybridization of base-pairing properties. “Analogous” forms of purines and pyrimidines are well known in the art, and include, but are not limited to aziridinylcytosine, 4-acetylcytosine, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxymethylaminomethyluracil, inosine, N⁶-isopentenyladenine, 1-methyladenine, 1-methylpseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N^{sup}.6-methyladenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5-methoxyuracil, 2-methylthio- N⁶-isopentenyladenine, uracil-5-oxyacetic acid methylester, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-

thiouracil, 5-methyluracil, uracil-5-oxyacetic acid, and 2,6-diaminopurine. DNA backbone analogues provided herein include phosphodiester, phosphorothioate, phosphorodithioate, methylphosphonate, phosphoramidate, alkyl phosphotriester, sulfamate, 3'-thioacetal, methylene(methylimino), 3'-N-carbamate, morpholino carbamate, and peptide nucleic acids (PNAs), methylphosphonate linkages or alternating methylphosphonate and phosphodiester linkages (Strauss-Soukup, 1997, *Biochemistry* 36:8692-8698), and benzylphosphonate linkages, as discussed in U.S. Pat. No. 6,664,057; see also OLIGONUCLEOTIDES AND ANALOGUES, A PRACTICAL APPROACH, edited by F. Eckstein, IRL Press at Oxford University Press (1991); Antisense Strategies, *Annals of the New York Academy of Sciences*, Volume 600, Eds. Baserga and Denhardt (NYAS 1992); Milligan, 1993, *J. Med. Chem.* 36:1923-1937; Antisense Research and Applications (1993, CRC Press). The nucleic acids herein can be extracted from cells or synthetically prepared according to any means known to those skilled in the art; for example, the nucleic acids can be chemically synthesized or transcribed or reverse transcribed from cDNA or mRNA, among other sources.

As used herein, two nucleic acids or nucleic acid regions "correspond" to one another if they are both complementary to the same nucleic acid sequence. Two nucleic acids or nucleic acid regions are "complementary" to one another if they base-pair with each other to form a double-stranded nucleic acid molecule.

A target nucleic acids utilized herein can be any nucleic acid, for example, human nucleic acids, bacterial nucleic acids, or viral nucleic acids. A target nucleic acid sample can be, for example, a nucleic acid sample from one or more cells, tissues, or bodily fluids. Target samples can be derived from any source including, but not limited to, eukaryotes, plants, animals, vertebrates, fish, mammals, humans, non-humans, bacteria, microbes, viruses, biological sources, serum, plasma, blood, urine, semen, lymphatic fluid, cerebrospinal fluid, amniotic fluid, biopsies, needle aspiration biopsies, cancers, tumors, tissues, cells, cell lysates, crude cell lysates, tissue lysates, tissue culture cells, buccal swabs, mouthwashes, stool, mummified tissue, forensic sources, autopsies, archeological sources, infections, nosocomial infections, production sources, drug preparations, biological molecule productions, protein preparations, lipid preparations, carbohydrate preparations, inanimate objects, air, soil, sap, metal, fossils, excavated materials, and/or other terrestrial or extra-terrestrial materials and sources. The sample may also contain mixtures of material from one source or different sources. For example, nucleic acids of an

infecting bacterium or virus can be amplified along with human nucleic acids when nucleic acids from such infected cells or tissues are amplified using the disclosed methods. Types of useful target samples include eukaryotic samples, plant samples, animal samples, vertebrate samples, fish samples, mammalian samples, human samples, non-human samples, bacterial samples, microbial samples, viral samples, biological samples, serum samples, plasma samples, blood samples, urine samples, semen samples, lymphatic fluid samples, cerebrospinal fluid samples, amniotic fluid samples, biopsy samples, needle aspiration biopsy samples, cancer samples, tumor samples, tissue samples, cell samples, cell lysate samples, crude cell lysate samples, tissue lysate samples, tissue culture cell samples, buccal swab samples, mouthwash samples, stool samples, mummified tissue samples, autopsy samples, archeological samples, infection samples, nosocomial infection samples, production samples, drug preparation samples, biological molecule production samples, protein preparation samples, lipid preparation samples, carbohydrate preparation samples, inanimate object samples, air samples, soil samples, sap samples, metal samples, fossil samples, excavated material samples, and/or other terrestrial or extra-terrestrial samples.

In some embodiments, a target nucleic acids utilized herein comprise repetitive sequence, secondary structure, and/or a high G/C content.

In certain embodiments, a target nucleic acid molecule of interest is about 100 to about 1,000,000 nucleotides (nt) in length. In some embodiments, the target is about 100 to about 1000, about 1000 to about 10,000, about 10,000 to about 100,000, or about 100,000 to about 1,000,000 nucleotides in length. In some embodiments, the target is about 100, about 200, about 300, about 400, about 500, about 600, about 700, about 800, about 900, about 1,000, about 2,000, about 3,000, about 4,000, about 5,000, about 6,000, about 7,000, about 8,000, about 9000, about 10,000, about 20,000, about 30,000, about 40,000, about 50,000, about 60,000, about 70,000, about 80,000, about 90,000, about 100,000, about 200,000, about 300,000, about 400,000, about 500,000, about 600,000, about 700,000, about 800,000, about 900,000, or about 1,000,000 nucleotides in length. It is to be understood that the target nucleic acid may be provided in the context of a longer nucleic acid (e.g., such as a coding sequence or gene within a chromosome or a chromosome fragment).

In certain embodiments, a target of interest is linear, while in other embodiments, a target is circular (e.g., plasmid DNA, mitochondrial DNA, or plastid DNA).

Combined Primer-Target Systems

In some embodiments, provided herein are primer-target systems. A primer-target system comprises one or more nucleic acid targets, a polymerase, and one or more primers (*e.g.*, primer duplex and/or hairpin primer duplex). The term “primer” encompasses any one of the primers or primer systems described herein (*e.g.*, single-stranded primers, double-stranded primer duplexes, and hairpin primer duplexes). In certain embodiments, the primer-target systems described herein comprise a plurality of different primers. In some embodiments, a primer-target system can comprise at least two primers, which can be used to identify and, for example amplify, a target nucleic acid molecule. A target nucleic acid molecule may be present amongst a plurality of non-target nucleic acid molecules, for example, as a single copy or in low copy number. Any one of the primer-target systems described herein may comprises conditions similar to those used in nucleic acid amplification or sequencing reactions (*e.g.*, similar reagents, reaction temperature, *etc.*).

Methods of Use

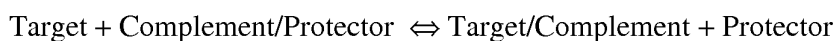
The primer systems described herein are able to discriminate a specific target from spurious targets either through a thermodynamic mechanism or through a kinetic mechanism. In distinguishing the target from spurious targets using a thermodynamic mechanism (described below), the strand displacement reaction is run to completion and the target is distinguished from the spurious targets based on differences in equilibrium binding affinity. To distinguish the target from spurious targets using a kinetic mechanism (described below), the strand displacement reaction is stopped before reaching equilibrium, and the differential rate in reaction completion is used to distinguish the targets from spurious targets.

Thermodynamic separation

The general strategy for both the thermodynamic and the kinetic mechanism is to use toehold exchange strand displacement reactions. In general, toehold exchange involves extending a target's complement with an additional region that is not complementary to the target, and pre-hybridizing a protector strand to this extended complement strand and a large number of bases adjacent to the extended region, but not to a single-stranded toehold region.

FIG. 9 depicts one implementation of toehold exchange. In this implementation, a two-stranded primer duplex system (as described above and depicted in FIG. 1A) is used to distinguish between a target nucleic acid and a spurious target. The complement branch migration region and the toehold region of the complement strand have sequences that correspond to a first target sequence and a second target sequence, respectively. A complement balance region was designed to be the same length and nucleotide base distribution as the toehold region. In this way, the standard free energy of the strand displacement reaction shown in FIG. 9A between the correct target and the protected complement is roughly $\Delta G^\circ = 0$ kcal/mol.

The strand displacement reaction can be written as:



ΔG° relates to the equilibrium constant K_{eq} by the following relation:

$$\Delta G^\circ = RT \ln(K_{\text{eq}}).$$

For a reaction with $\Delta G^\circ = 0$, the equilibrium constant (K_{eq}) is 1. The equilibrium constant also relates to the equilibrium; for this reaction, $K_{\text{eq}} = [\text{TC}] [\text{P}] / [\text{T}] [\text{PC}] = 1$. For an assay where $[\text{PC}]$ and $[\text{P}]$ are in excess of $[\text{T}]$, $[\text{TC}] / [\text{T}] = 1$, meaning that exactly half of all target molecules are hybridized at equilibrium. In the example shown in FIG. 9A, $\Delta G^\circ = +0.1$ kcal/mol, corresponding to a $K_{\text{eq}} = 0.85$, which means that 46% of the target molecules are hybridized at equilibrium.

The protector strand correspondingly changes the standard free energy of the strand displacement reaction with spurious targets. In the example shown in FIG. 9B, the spurious target differs from the correct target by a single base, which results in the strand displacement reaction with the same two-stranded nucleic acid primer system having a ΔG° of +3.7 kcal/mol, which corresponds to $K_{\text{eq}} = 1.9 \times 10^{-3}$. At equilibrium, only 0.19% of the spurious target will be hybridized to the complement. Thus, the exemplary nucleic acid primer system depicted in FIG. 8 preferentially binds to its target versus a spurious target having only a single nucleotide mismatch by more than 200-fold.

FIG. 10 is a plot of the equilibrium binding affinity as a function of the standard free energy of the reaction. When the standard free energy of the reaction is very negative (as in the case in a pure hybridization reaction), both the target and the spurious targets bind very strongly,

and it is difficult to distinguish between the two, leading to false positives. On the other hand, when the standard free energy of the reaction is very positive, the primer binds to the target very weakly, leading to false negatives. Designing a primer duplex system to have a standard free energy of near zero results in an optimal discrimination between targets and spurious targets, thereby minimizing false positives and false negatives.

Kinetic Separation

Kinetic separation relies on the differential kinetics of toehold exchange. The kinetics of the toehold exchange reaction depend on the binding strengths of the toehold regions. Each kcal/mol of difference in toehold binding energy can affect kinetics by a factor of 5.4 (Figure 11), so the +3.6 kcal/mol mismatch shown in FIG. 9B would yield a kinetic slowdown of 434.

Unlike in thermodynamic discrimination, kinetic discrimination occurs only when the mismatch is in the toehold region. Spurious targets differing from the correct target at a position complementary to the complement branch migration region are unlikely to yield significantly different reaction kinetics. As a consequence, methods that use the kinetic mechanism of distinguishing target from spurious target are useful in conjunction with thermodynamic separation as a means of pinpointing the locations of target/primer mismatches.

Significantly, primer duplex systems lacking complement balance regions, such as the primer systems depicted in FIG. 8, can be used in methods that exploit the kinetic mechanism to pinpoint the location of target/primer mismatches.

Microarrays

Nucleic acid microarrays are often used for high-throughput nucleic acid detection, but often are unable to distinguish between closely related nucleic acid sequences. In some embodiments, the primer duplex systems described herein can be used in nucleic acid microarrays in order to, for example, improve the specificity of microarray analysis. In some instance, microarrays assays can be performed using methods well known in the art, with the exception that the primer duplex systems described herein can be used in place of conventional nucleic acid primers.

For example, as depicted in FIG. 12A, in certain embodiments, a hairpin primer duplex system from can be directly synthesized or immobilized on a microarray chip using standard

techniques. In other embodiments, a two-stranded primer duplex system can be used in a nucleic acid microarray. In some embodiment, hairpin structures including two photocleavable bases at predefined positions can be synthesized as in FIG. 12A. Subsequent exposure to light cleaves the hairpin, yielding the two-stranded complexes functionalized to the array surface (FIG. 12B). Other methods, such as use of nicking or restriction enzymes, can also be used to prepare two-stranded complexes.

Nucleic Acid Synthesis Reactions, Including Amplification Reactions

Primer duplexes and systems disclosed here can be used in some embodiments to improve the specificity of a primer-based amplification reaction, including polymerase chain reaction (PCR), strand displacement amplification, or transcription mediated amplification, by substituting a primer duplex system described herein for the nucleic acid primers in a primer based amplification reaction known in the art.

For example, as depicted in FIG. 13, by using as PCR primers the hairpin primer duplex systems of the type depicted in FIG. 4, it is possible to improve the specificity of PCR for a variety of (*e.g.*, biotechnological) applications. In this example, a target nucleic acid sequence is amplified by forming a solution comprising a primer duplex system with the target nucleic acid and standard reagents for performing an amplification reaction and incubating the solution under conditions such that an amplification reaction occurs. In certain embodiments, non-natural bases are incorporated into a hairpin primer duplex primer systems in order to prevent replication of the hairpin itself.

In some embodiments, the primer duplexes described herein can be adapted for use in amplifying target nucleic acids that typically require amplification by any one or more of the following PCR methods: allele-specific PCR, assembly PCR, asymmetric PCR, helicase-dependent amplification, intersequence-specific PCR (ISSR), inverse PCR, ligation-mediated PCR, methylation-specific PCR (MSP), miniprimer PCR, multiplex PCR, nested PCR, overlap-extension PCR, quantitative PCR (Q-PCR), reverse transcription PCR (RT-PCR), solid phase PCR, thermal asymmetric interlaced PCR (TAIL-PCR), or touchdown PCR. In some instances, the primer duplexes and methods described herein may be used or adapted for use in any one of the foregoing PCR methods or may substitute (used instead of) any one of the foregoing PCR methods. A brief description of each of the foregoing PCR methods is presented below.

Allele-specific PCR is a diagnostic or cloning technique based on single-nucleotide polymorphisms (SNPs) (single-base differences in DNA). It typically requires prior knowledge of a DNA sequence, including differences between alleles.

Assembly PCR or polymerase cycling assembly (PCA) is an artificial synthesis of long DNA sequences by performing PCR on a pool of long oligonucleotides with short overlapping segments. The oligonucleotides alternate between sense and antisense directions, and the overlapping segments determine the order of the PCR fragments, thereby selectively producing the final long DNA product (Stemmer *et al. Gene* 164(1): 49-53 (1995)).

Asymmetric PCR preferentially amplifies one DNA strand in a double-stranded DNA target. It can be used in sequencing and hybridization probing where amplification of only one of the two complementary strands is required (Innis *et al. Proc. Natl. Acad. Sci. USA* 85(24): 9436-40 (1988)).

Helicase-dependent amplification is similar to traditional PCR, but typically uses a constant temperature rather than cycling through denaturation and annealing/extension cycles. DNA helicase, an enzyme that unwinds DNA, is used in place of thermal denaturation (Vincent *et al. EMBO Reports* 5(8): 795-800 (2004)).

Intersequence-specific PCR (ISSR) is a PCR method for DNA fingerprinting that amplifies regions between simple sequence repeats to produce a unique fingerprint of amplified fragment lengths (Zietkiewicz *et al. Genomics* 20(2): 176-83 (1994)).

Inverse PCR is commonly used to identify the flanking sequences around genomic inserts. It involves a series of DNA digestions and self-ligation, resulting in known sequences at either end of the unknown sequence (Ochman *et al. Genetics* 120 (3): 621-623 (1988)).

Ligation-mediated PCR uses small DNA linkers ligated to the DNA of interest and multiple primers annealing to the DNA linkers; it has been used for DNA sequencing, genome walking, and DNA footprinting (Mueller *et al. Science* 246(4931): 780-786 (1988)).

Methylation-specific PCR (MSP) is used to detect methylation of CpG islands in genomic DNA. DNA is first treated with sodium bisulfite, which converts unmethylated cytosine bases to uracil, which is recognized by primers as thymine.

Miniprimer PCR uses a thermostable polymerase (S-Tbr) and is used to amplify conserved DNA sequences, such as the 16S (or eukaryotic 18S) rRNA gene (Isenbarger *et al. Applied and Environmental Microbiology* 74(3): 840-9. (2008)).

Multiplex-PCR targets multiple genes at once, gaining additional information from a single test-run that otherwise would require several times the reagents and more time to perform.

Nested PCR increases the specificity of DNA amplification, by reducing background due to non-specific amplification of DNA. Two sets of primers are used in two successive PCRs. In the first reaction, one pair of primers is used to generate DNA products, which besides the intended target, may still consist of non-specifically amplified DNA fragments. The product(s) are then used in a second PCR with a set of primers whose binding sites are completely or partially different from and located 3' of each of the primers used in the first reaction.

Overlap-extension PCR or splicing by overlap extension (SOE) is a genetic engineering technique that is used to splice together two or more DNA fragments that contain complementary sequences. It is used to join DNA pieces containing genes, regulatory sequences, or mutations; the technique enables creation of specific and long DNA constructs.

Quantitative PCR (Q-PCR) is used to measure the quantity of a PCR product (commonly in real-time). It quantitatively measures starting amounts of DNA, cDNA, or RNA. Q-PCR is commonly used to determine whether a DNA sequence is present in a sample and the number of its copies in the sample.

Reverse Transcription PCR (RT-PCR) is used for amplifying DNA from RNA. Reverse transcriptase reverse transcribes RNA into cDNA, which is then amplified by PCR. RT-PCR is widely used in expression profiling, to determine the expression of a gene or to identify the sequence of an RNA transcript, including transcription start and termination sites. If the genomic DNA sequence of a gene is known, RT-PCR can be used to map the location of exons and introns in the gene. The 5' end of a gene (corresponding to the transcription start site) is typically identified by RACE-PCR (Rapid Amplification of cDNA Ends).

Solid Phase PCR encompasses multiple meanings, including polony amplification (where PCR colonies are derived in a gel matrix, for example), bridge PCR (primers are covalently linked to a solid-support surface), conventional solid phase PCR (where asymmetric PCR is applied in the presence of solid support bearing primer with sequence matching one of the aqueous primers) and enhanced solid phase PCR (where conventional solid phase PCR can be improved by employing high melting temperature (T_m) and nested solid support primer with optional application of a thermal 'step' to favor solid support priming).

Thermal asymmetric interlaced PCR (TAIL-PCR) is used for isolation of an unknown sequence flanking a known sequence. Within the known sequence, TAIL-PCR uses a nested pair of primers with differing annealing temperatures; a degenerate primer is used to amplify in the other direction from the unknown sequence (Liu *et al. Genomics* 25 (3): 674–81. (1995)).

Touchdown PCR (step-down PCR) is a variant of PCR that aims to reduce nonspecific background by gradually lowering the annealing temperature as PCR cycling progresses. The annealing temperature at the initial cycles is usually a few degrees (3-5°C) above the T_m of the primers used, while at the later cycles, it is a few degrees (3-5°C) below the primer T_m . The higher temperatures give greater specificity for primer binding, and the lower temperatures permit more efficient amplification from the specific products formed during the initial cycles.

The temperature of the reaction solutions may be sequentially cycled between a denaturing state, an annealing state, and an extension state for a predetermined number of cycles. The actual times and temperatures can be enzyme, primer, and target dependent.

For any given reaction, denaturing states can range in certain embodiments from about 75 °C to about 100 °C. The annealing temperature and time can influence the specificity and efficiency of primer binding to a particular locus within a target nucleic acid and may be important for particular PCR reactions.

For any given reaction, annealing states can range in certain embodiments from about 20 °C to about 75 °C. In some embodiments, the annealing state can be performed at about 20 °C to about 25 °C, about 25 °C to about 30 °C, about 30 °C to about 35 °C, or about 35 °C to about 40 °C, about 40 °C to about 45 °C, about 45 °C to about 50 °C. In certain embodiments, the annealing state can be performed at room temperature (*e.g.*, 20 °C or 25 °C). In some embodiments, the annealing state can be performed at a temperature of 20 °C, 21 °C, 22 °C, 23 °C, 24 °C, 25 °C, 26 °C, 27 °C, 28 °C, 29 °C, 30 °C, 31 °C, 32 °C, 33 °C, 34 °C, 35 °C, 36 °C, 37 °C, 38 °C, 39 °C, 40 °C, 41 °C, 42 °C, 43 °C, 44 °C, 45 °C, 46 °C, 47 °C, 48 °C, 49 °C, or 50 °C.

Extension temperature and time may impact the allele product yield and are understood to be an inherent property of the enzyme under study. For a given enzyme, extension states can range in certain embodiments from about 60 °C to about 75 °C.

In any of the foregoing embodiments, any DNA or RNA polymerase (enzyme that catalyzes polymerization of nucleotides into a nucleic acid strand) may be utilized, including

thermostable polymerases and reverse transcriptases (RTases). Examples include *Bacillus stearothermophilus* pol I, *Thermus aquaticus* (Taq) pol I, *Pyrococcus furiosus* (Pfu), *Pyrococcus woesei* (Pwo), *Thermus flavus* (Tfl), *Thermus thermophilus* (Tth), *Thermus litoris* (Tli) and *Thermotoga maritime* (Tma). These enzymes, modified versions of these enzymes, and combination of enzymes, are commercially available from vendors including Roche, Invitrogen, Qiagen, Stratagene, and Applied Biosystems. Representative enzymes include PHUSION® (New England Biolabs, Ipswich, MA), Hot MasterTaq™ (Eppendorf), PHUSION® Mpx (Finnzymes), PyroStart® (Fermentas), KOD (EMD Biosciences), Z-Taq (TAKARA), and CS3AC/LA (KlenTaq, University City, MO).

Salts and buffers include those familiar to those skilled in the art, including those comprising MgCl₂, and Tris-HCl and KCl, respectively. Buffers may contain additives such as surfactants, dimethyl sulfoxide (DMSO), glycerol, bovine serum albumin (BSA) and polyethylene glycol (PEG), as well as others familiar to those skilled in the art. Nucleotides are generally deoxyribonucleoside triphosphates, such as deoxyadenosine triphosphate (dATP), deoxycytidine triphosphate (dCTP), deoxyguanosine triphosphate (dGTP), and deoxythymidine triphosphate (dTTP), and are also added to a reaction adequate amount for amplification of the target nucleic acid.

Also provided herein are methods comprising (1) hybridizing a complement strand of a primer duplex to a target nucleic acid, thereby dissociating the complement strand from its protector strand, and (2) extending the complement strand at its 3' end, in a target-complementary manner, in the presence of a polymerase.

Also provided herein are methods comprising performing a nucleic acid synthesis reaction in the presence of a target nucleic acid, a polymerase, and one or more of the primer duplexes of any one of the embodiments described herein.

A "nucleic acid synthesis reaction" refers to any reaction in which a nucleic acid is synthesized. Examples include nucleic acid amplification reactions such as polymerase chain reaction (PCR) or a variation thereof (described elsewhere herein), a transcription reaction, a reverse transcription reaction, sequencing-by-synthesis, or other primer extension reactions (see also, Lizardi *et al. Nat. Genet.* 19: 225-32 (1998), incorporated by reference).

In some instances, a method is provided that comprises (1) synthesizing a complement strand having a target-non-specific balance region, a target-specific branch migration region, and

a target-specific toehold region; (2) synthesizing a protector strand having a balance region complementary to the complement strand and a branch migration region complementary to the complement strand; and (3) hybridizing the complement strand to the protector strand to form a primer duplex.

In some instances, a method is provided that comprises (1) providing a complement strand having a target-non-specific balance region, a target-specific branch migration region, and a target-specific toehold region; (2) providing a protector strand having a balance region complementary to the complement strand and a branch migration region complementary to the complement strand; and (3) combining the complement strand to the protector strand to form a primer duplex.

In some instances, a method is provided that comprises (1) providing a plurality of nucleic acid molecules comprising a target nucleic acid; (2) providing at least one primer duplex having (i) a balance region, (ii) a branch migration region complementary to the target nucleic acid, and (iii) a toehold region; and (3) combining in a single reaction the plurality of target nucleic acids, at least one primer duplex, and a polymerase under conditions suitable for nucleic acid hybridization.

Also provided herein are methods of amplifying at least one target nucleic acid of interest, comprising (1) providing a plurality of nucleic acid molecules comprising at least one target nucleic acid, (2) providing at least one primer duplex having (i) a balance region, (ii) a branch migration region, and (iii) a toehold region; and (3) combining in a single reaction the plurality of target nucleic acid molecules, at least one primer duplex, and a polymerase under conditions suitable for amplification of the at least one target nucleic acid. In certain embodiments multiple unique target nucleic acids are amplified in a single reaction or in multiple reactions, for example, in one or more multiplexed PCR amplification reaction. In some embodiments, about 10 to 100, about 100 to about 1000, about 1000 to about 10,000, or about 10,000 to about 100,000 nucleic acid targets are amplified. The number of different primer duplexes in a reaction will depend on the number of desired targets.

In some embodiments, provided herein are methods of discriminating against spurious nucleic acid molecules having one or more nucleotide changes relative to a target nucleic acid molecule, comprising (1) providing a plurality of nucleic acid molecules comprising at least one target nucleic acid, (2) providing at least one primer duplex having (i) a balance region, (ii) a

branch migration region, and (iii) a toehold region; and (3) combining in a single reaction the plurality of target nucleic acid molecules, at least one primer duplex, and a polymerase under conditions suitable for amplification of the at least one target nucleic acid molecule.

Any one of the methods described herein may further comprise providing or combining in a single reaction one or more of the following reagents: buffer (*e.g.*, KCl, MgCl₂, Tris-HCl), dNTPs (*e.g.*, dATP, dCTP, dGTP, dTTP at concentrations of, *e.g.*, about 50 to about 100 μM), polymerase (*e.g.*, at concentrations of about 0.5-2.0 units per 50 μl reaction), and/or water. The concentration of each strand of a primer duplex in a single reaction varies depending on, for example, the concentration of target nucleic acid. In some embodiments, about 5 to about 50 pg of plasmid or viral target may be used, or about 50 ng to about 500 ng of genomic target may be used. In such instances, the concentration each primer (the first strand and the second strand) may be, for example, about 0.05 μM to about 1 μM. In particular embodiments, the concentration of each primer is about 1 nM to about 1 μM.

In any one of the embodiments described herein, a single reaction may be subject to cyclic temperature changes such that a dsDNA structure undergoes multiple rounds of denaturation, subsequent primer annealing, and polymerase-based extension, for example, similar to those conditions used for standard PCR. In some embodiments, the temperature range for a denaturation step is about 90 to about 95 °C. In certain embodiments, an initial denaturation step of about 1 to about 5 minutes is required prior to cycling; the exact amount of time may depend on GC content of the nucleic acid target of interest. In certain embodiments, the denaturation step during a cycling reaction is about 15 to about 30 seconds. In some embodiments, the temperature range for an annealing step is about 50 °C to about 60 °C. In some embodiments, the annealing step is about 20 °C to about 40 °C. In particular embodiments, the annealing step is at room temperature (about 20°C or about 25°C). In certain embodiments, the annealing step during a cycling reaction is about 15 to about 30 seconds. In some embodiments, the temperature range for an extension step is about 70 °C to about 75 °C. In certain embodiments, the extension step during a cycling reaction is about 45 to about 60 seconds. The temperature, time of each step, and number of cycles of a cycling reaction may depend on the length of the nucleic acid target(s) of interest as well as the polymerase being used. Longer target may require, for example, longer extension times. One example of cycling conditions for a 500 nucleotide target is set forth in Table 2.

Table 2

<u>1 cycle</u>	<u>98 °C</u>	<u>2 minutes</u>
<u>25 cycles</u>	<u>98 °C</u>	<u>15 seconds</u>
	<u>30 °C, 35 °C, 40 °C, 45 °C, 50 °C, 55 °C, or 60 °C</u>	<u>15 seconds</u>
	<u>72 °C</u>	<u>45 seconds</u>
<u>1 cycle</u>	<u>72 °C</u>	<u>5 minutes</u>
<u>1 cycle</u>	<u>4 °C</u>	<u>indefinite</u>

In any one of the embodiments described herein, a single reaction (*e.g.*, nucleic acid amplification) may proceed at room temperature (*e.g.*, about 20°C or about 25°C). In certain embodiments, a single reaction proceeds at room temperature for about 1 hour.

In any one of the methods described herein, the second protector strand of a primer duplex may be provided in excess of the first complementary strand or in excess of the annealed primer duplexes. For example, in some embodiments, the second strand is provided at a concentration about 1x to about 10x (*e.g.*, 1x, 2x, 3x, 4x, 5x, 6x, 7x, 8x, 9x, or 10x) the concentration the first strand, or about 1x to about 10x (*e.g.*, 1x, 2x, 3x, 4x, 5x, 6x, 7x, 8x, 9x, or 10x) the concentration of the annealed primer duplex. In some embodiments, the first strand is provided at a concentration of about 0.05 μM to about 1 μM, while the second strand is provided at a concentration of about 0.10 μM to about 2 μM, or about 0.15 μM to about 3 μM, about 0.2 μM to about 4 μM, or about 0.25 μM to about 5 μM.

Any one of the methods described herein may comprise a method selected from: allele-specific PCR, assembly PCR, asymmetric PCR, helicase-dependent amplification, intersequence-specific PCR (ISSR), inverse PCR, ligation-mediated PCR, methylation-specific PCR (MSP), miniprimer PCR, multiplex PCR, nested PCR, overlap-extension PCR, quantitative PCR (Q-PCR), reverse transcription PCR (RT-PCR), solid phase PCR, thermal asymmetric interlaced PCR (TAIL-PCR), and touchdown PCR.

In any one of the methods described herein, the yield of amplified nucleic acid target may be about 30% to about 100%. In some embodiments, the yield is at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or at least 100%.

In any one of the methods described herein, the amplified nucleic acid product may be purified. Nucleic acid purification methods are well-known to those of skill in the art and include, phenol extraction, guanidinium isothiocyanate, alcohol precipitation, DEAE (ion exchange), size exclusion chromatography (SEC), cesium chloride, extraction from agarose, silica, and other column-based purification methods.

In any one of the methods described herein, a purified amplified target nucleic acid may be about 30% to about 100% pure. In some embodiments, the purity is at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% pure.

Imaging

The primer duplexes and systems described herein can also be used to improve the specificity of *in situ* imaging assays. Nonspecific interactions between biological RNAs and fluorophore-labeled primers are frequently a source of background noise. Thus, as depicted in FIG. 14, the use fluorophore-labeled nucleic acid primer systems described herein in the place of conventional primers, in some embodiments, greatly improves the performance of existing *in situ* imaging techniques. Notably, by labeling the complement strand or domain with a fluorophore and the protector strand or domain with a quencher, the primer duplex system will only produce a detectable signal when it is bound to the target.

Single Nucleotide Polymorphism (SNP) Detection

The accurate detection of the location and identity of single nucleotide polymorphisms (SNPs) is of great interest for both research and therapeutic purposes. The kinetic discrimination methods described herein are therefore useful for the convenient identification SNPs.

Kits

Provided herein are kits comprising (1) at least one complement strand having a balance region, a branch migration region, and a toehold region, and (2) at least one protector strand having a balance region and a branch migration region.

Provided herein are kits comprising at least one primer duplex comprising (1) at least one complement strand or region having a balance region, a branch migration region, and a toehold region, and (2) at least one protector strand or region having a balance region and a branch migration region.

Any one of the kits described herein may further comprise a polymerase. Any one of the kits provided herein may further comprise one or more agent selected from buffer (*e.g.*, KCl, MgCl₂, Tris-HCl), dNTPs (*e.g.*, dATP, dCTP, dGTP, dTTP), and water. Any one of the kits provided herein may comprise protector strand is molar excess of the primer. Any one of the kits provided herein may further comprise instructions or directions for obtaining instructions (*e.g.*, from a website) for using the components of the kits. Any one of the kits provided herein may further comprise at least one reaction tube, well, chamber, or the like.

Any one of the primers or primer systems described herein may be provided in the form of a kit or comprised within a kit.

EXAMPLES

In accordance with the invention, the above limitations of PCR, transcription, and reverse transcription can be overcome through the use of highly specific primer duplexes. The experiments described herein demonstrate that primer duplexes can reliably discriminate against targets with single-base changes (FIG. 16) for both DNA and RNA targets and primers (FIG. 17). The correct target hybridizes to the 7/5 primers with roughly 50% yield, but even a large excess (200x) of targets with a single-base change is insufficient to significantly hybridize. Primer duplexes were designed and tested for multiple different targets, and each primer duplex achieved high discrimination factors versus single-nucleotide changes (FIG. 17). Quantitatively, the median discrimination in hybridization yield to a spurious target with a single-nucleotide change is 26.

The primer duplexes were used for PCR in a proof-of-principle demonstration (FIGS. 18A and 18B). A semi-repetitive target nucleic acid was designed, which is difficult to amplify by traditional PCR (PCR without the use of the instant primer duplexes). The yield of standard 21 nucleotide primers and the primer duplexes were calculated. Many different thermal cycling schedules were determined in order to investigate the range of function. Based on the length and nucleotide content of the primer duplexes, standard PCR condition would predict that the

annealing temperature of the primers would 55 °C. Surprisingly, as an example, even under conditions most unfavorable for primer duplex annealing (35 °C and 40 °C), the fraction (50.2%) of correct-length product amplified using the primer duplexes was higher than the fraction (31.0%) of correct-length product amplified using standard primers under their most favorable PCR conditions (45 °C). Furthermore, in this particular experiment, the primer duplexes were arbitrarily designed (7 nucleotide toehold region and 5 nucleotide balance region), and were not optimized for PCR yield performance. Thus, it is likely that even higher PCR specificity can be achieved through optimization of the instant primer duplexes.

FIG. 15 shows highly specific PCR using the primer duplexes provided herein. In Figure 5A, the primer “PC” is comprised of a complement strand “C” and a protector strand “P”. When PC binds to the intended target at the correct position “X”, the single-stranded protector oligonucleotide “P” is released as an inert waste product, and the primed target is elongated by the DNA polymerase. In FIG. 5B, when the primer PC binds to an unintended target or to the correct target at an incorrect position (in either case, denoted “Y”), the displacement of the protector from the complementary strand “C” is thermodynamically unfavorable, and kinetically quick to reverse. Consequently, off-target amplification (*e.g.*, amplification of Y rather than X) is expected to be significantly reduced.

FIG. 16. shows an experimental demonstration of primer hybridization with single nucleotide discrimination. In FIG. 16A, short synthetic DNA target “X” or spurious target “Y” is reacted with the primer. (The poly-T tail on the protector strand “P” serves to distinguish products from reactants on a gel.) Shown in red boxes are the positions of single-base changes for spurious target Y. FIG 16B shows native polyacrylamide gel results. The primer “PC” was prepared at a 2:1 ratio of protector P to complement C, and annealed at 1 μM concentration of PC. Either the correct or spurious targets were added to achieve final concentrations of 200 nM target (X or Y), 100 nM PC, and 100 nM P. In some embodiments, a reaction may have an excess of the protector (P) primer. For example, in some embodiments, the protector strand is provided at a concentration of about 1x to about 10x (*e.g.*, 1x, 2x, 3x, 4x, 5x, 6x, 7x, 8x, 9x, or 10x) of the complement strand. All reactions proceeded at room temperature (25°C) for 1 hour. As an example, the designation “7/4” denotes a primer that possesses 7 nucleotides of single-stranded nucleotides (as a 3’ overhang) to initiate hybridization to the target, and the protector spontaneously dissociates 4 nucleotides to be released. FIG. 16C is a plot of hybridization yields

inferred from data shown in FIG. 16B. Shown as plot “X” is the hybridization of the primer to the correct target X, while the remaining “dotted” plots show the hybridization to the spurious targets Y. The 7/4, 7/5 and 7/6 primers all discriminate in their hybridization yields (χ) between the correct and the spurious targets. The 7/0 target does not. In FIG. 16D, the discrimination factor (Q) is a quantitative measurement of the specificity of the primer, and is calculated as the hybridization yield (χ) of the correct target divided by the hybridization yield (χ) of the spurious target. In 16E, there is little hybridization of the 7/5 primer to a spurious target Y even when such target is present in large excess (*i.e.*, 200-fold).

FIG. 17. shows additional experimental results and statistics on the single-base discrimination abilities of primer duplexes. FIG 17A shows that four additional targets and sets of primers were constructed and tested: two based on naturally occurring microRNA sequences, and two designed to intentionally possess significant secondary structure. FIG. 17B shows a histogram of the discrimination factors (Q) achieved by the 7/5 primers for each target. Due to limitations of the gel scanner, it was not possible to reliably measure discrimination factors above 100, and these were all grouped as “100+.” FIG. 17C show RNA target and primer. The target sequence is a synthetic RNA oligonucleotide with sequence identical to the human let7g microRNA. FIG. 17D shows native PAGE results. The PC primer was prepared at a 2:1 ratio of protector P to complement C, and annealed at 3 μ M concentration. Either the correct or spurious targets were added to achieve final concentrations of 2 μ M X or Y, 1 μ M PC, and 1 μ M P. The correct target successfully binds to the primer; the hybridization yield of targets with single-nucleotide mismatches is low.

FIG. 18 shows experimental results using duplex primers to improve the PCR yield of a quasi-repetitive target. FIG. 18A shows a quasi-repetitive PCR target (168 nt) that traditional PCR primers struggle to amplify with high yield. Here, a* is the correct target for X1. The remaining sites labeled a*m1 (which is X1-m17G), a*m2 (which is X1-m9T), and a*m3 (which is X1-m11G) are not the correct targets. Similarly, b* is the correct target for X2, and b*m1 (which is X2-m3T), b*m2 (which is X2-m11C), and b*m3 (which is X2-m18T) are not the correct targets. Thus, the outer-most binding sites are the perfect binding sites for the primers, but there are also 3 additional single-base mismatch primer binding sites between the perfect sites. The primer duplexes bind by 7 nucleotides to the target, and the protector must spontaneously dissociate 5 nucleotides to be released. The primer duplex was designed so that

its 3' end cannot be extended by the polymerase. The toehold region of the complement strand was designed at the 3' end, instead of the 5' end as in previous designs. In FIG. 18B, primer duplexes show significantly higher yield of correct length product, as compared to standard primers. Each lane is labeled with the primers used as well as the temperature cycling schedule (*e.g.*, "98-40-72" indicates denaturation at 98 °C, annealing at 40 °C, and elongation at 72 °C). The left-most lane shows the synthetic oligonucleotide reference. The lower numbers labeled "% Correct" indicate the relative intensity of band corresponding to the correct length product compared to the integrated intensity of all bands in the lane. The primer duplex PCR product appears as 10 nucleotides longer than the reference and the standard PCR product because of the 5 nucleotide of overhangs (toehold region) on each primer.

EQUIVALENTS

Those skilled in the art will recognize or be able to ascertain using no more than routine experimentation many equivalents to the specific embodiments described herein.

The articles "a" and "an" are used herein to refer to one or to more than one (*i.e.*, to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

Claims or descriptions that include "or" between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. The invention includes embodiments in which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. The invention includes embodiments in which more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process. Furthermore, it is to be understood that the invention encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, descriptive terms, etc., from one or more of the listed claims is introduced into another claim. For example, any claim that is dependent on another claim can be modified to include one or more limitations found in any other claim that is dependent on the same base claim.

Where elements are presented as lists, *e.g.*, in Markush group format, it is to be understood that each subgroup of the elements is also disclosed, and any element(s) can be

removed from the group. It should be understood that, in general, where the invention, or aspects of the invention, is/are referred to as comprising particular elements, features, certain embodiments of the invention or aspects of the invention consist, or consist essentially of, such elements and/or features. For purposes of simplicity those embodiments have not been specifically set forth in haec verba herein. It is also noted that the term “comprising” is intended to be open and permits the inclusion of additional elements or steps.

Where ranges are given, endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value or sub-range within the stated ranges in different embodiments of the invention, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise.

As used herein, the term “about” generally may refer to any value within a range of 10% of the recited value. In some instances, however, “about” may encompass a range of 20% of the recited value.

In addition, it is to be understood that any particular embodiment of the present invention that falls within the prior art may be explicitly excluded from any one or more of the claims. Because such embodiments are deemed to be known to one of ordinary skill in the art, they may be excluded even if the exclusion is not set forth explicitly herein. Any particular embodiment of the methods of the invention can be excluded from any one or more claims, for any reason, whether or not related to the existence of prior art. This invention is not limited in its application to the details of construction and the arrangement of components set forth in the following description or illustrated in the drawings. The invention is capable of other embodiments and of being practiced or of being carried out in various ways. Also, the phraseology and terminology used herein is for the purpose of description and should not be regarded as limiting. The use of “including,” “comprising,” or “having,” “containing,” “involving,” and variations thereof herein, is meant to encompass the items listed thereafter and equivalents thereof as well as additional items.

Each of the foregoing patents, patent applications and references is hereby incorporated by reference, particularly for the teachings referenced herein.

REFERENCES

- [1] Petersen, M. & Wengel, J. LNA: a versatile tool for therapeutics and genomics. *Trends Biotechnol.* 21, 74-81, (2003).
- [2] Krueger, A.T. & Kool, E.T. Redesigning the Architecture of the Base Pair: Toward Biochemical and Biological Function of New Genetic Sets. *Chem Biol.* 16, 242-248 (2009).
- [3] Lizardi, P.M. et al. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* 19, 225-232 (1998).
- [4] Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487-491 (1988).
- [5] Zhang, D. Y., Chen, X. & Yin, P. Optimizing Nucleic Acid Hybridization Specificity. submitted (2011).

CLAIMS

1. A system comprising
a nucleic acid target,
a polymerase, and
a partially double-stranded primer comprising a first and a second nucleic acid strand
arranged into
 - (1) one double-stranded target-non-specific region,
 - (2) one double-stranded target-specific region, and
 - (3) one single-stranded target-specific region contributed to by the first
nucleic acid strand,wherein the double-stranded target-non-specific region has a standard free energy that
is within 10% of the standard free energy for the single-stranded target-specific region bound
to a perfectly complementary region of a target nucleic acid, wherein the sequence of the first
nucleic acid strand that contributes to region (1) is not perfectly complementary to and does
not bind to the target nucleic acid, and wherein the sequences of regions (2) and (3) are
complementary to and bind to the target nucleic acid.
2. The system of claim 1, wherein the nucleic acid target is a single-stranded.
3. The system of any one of claims 1 or 2, wherein the nucleic acid target is DNA or
RNA.
4. The system of any one of claims 1-3, wherein the system comprises a plurality of
different partially double-stranded primers.
5. The system of one of claims 1-4, wherein the system comprises at least two partially
double-stranded primers which together can be used to amplify a region of the nucleic acid
target.
6. The system of any one of claims 1-5, wherein the nucleic acid target comprises
repetitive sequence, secondary structure and/or high GC content.
7. The system of any one of claims 1-6, wherein the nucleic acid target is present in a
plurality of different nucleic acid targets.

8. The system of claim 7, wherein the nucleic acid target is present as a single copy or in low copy in a plurality of different nucleic acid targets.

9. A method comprising
contacting a partially double-stranded primer to a sample, and detecting hybridization of the primer to a target in the sample, wherein the partially double-stranded primer is comprised of first and second nucleic acid strands arranged into

- (1) one double-stranded target-non-specific region,
- (2) one double-stranded target-specific region, and
- (3) one single-stranded target-specific region contributed to by the first nucleic acid strand,

wherein the double-stranded target-non-specific region has a standard free energy that is within 10% of the standard free energy for the single-stranded target-specific region bound to a perfectly complementary region of a target nucleic acid, wherein the sequence of the first nucleic acid strand that contributes to region (1) is not perfectly complementary to and does not bind to the target nucleic acid, and wherein the sequences of regions (2) and (3) are complementary to and bind to the target nucleic acid.

10. The method of claim 9, wherein the partially double-stranded primer is labeled with a detectable moiety.

11. The method of claim 10, wherein the detectable moiety comprises a fluorophore or a radioisotope.

12. The method of any one of claims 9-11, wherein the target is present as a single copy in the sample.

13. A method comprising
hybridizing a single-stranded target-specific region of a first strand of a partially double-stranded primer to a nucleic acid target, thereby dissociating the first strand of the primer from a second strand of the primer, and

extending the first strand at its 3' end, in a target-complementary manner, in the presence of a polymerase, wherein the partially double-stranded primer is comprised of first and second nucleic acid strands arranged into

- (1) one double-stranded target-non-specific region,
- (2) one double-stranded target-specific region, and
- (3) one single-stranded target-specific region contributed to by the first nucleic acid strand,

wherein the double-stranded target-non-specific region has a standard free energy that is within 10% of the standard free energy for the single-stranded target-specific region bound to a perfectly complementary region of a target nucleic acid, wherein the sequence of the first nucleic acid strand that contributes to region (1) is not perfectly complementary to and does not bind to the target nucleic acid, and wherein the sequences of regions (2) and (3) are complementary to and bind to the target nucleic acid.

14. A method comprising

performing a nucleic acid synthesis reaction in the presence of a nucleic acid target, a polymerase, and one or more partially double-stranded primers, wherein the partially double-stranded primer is comprised of first and second nucleic acid strands arranged into

- (1) one double-stranded target-non-specific region,
- (2) one double-stranded target-specific region, and
- (3) one single-stranded target-specific region contributed to by the first nucleic acid strand,

wherein the double-stranded target-non-specific region has a standard free energy that is within 10% of the standard free energy for the single-stranded target-specific region bound to a perfectly complementary region of a target nucleic acid, wherein the sequence of the first nucleic acid strand that contributes to region (1) is not perfectly complementary to and does not bind to the target nucleic acid, and wherein the sequences of regions (2) and (3) are complementary to and bind to the target nucleic acid.

15. The method of claim 14, wherein the nucleic acid synthesis reaction is a nucleic acid amplification reaction.

16. The method of claim 15, wherein the nucleic acid amplification reaction is polymerase chain reaction (PCR).

17. The method of claim 14, wherein the nucleic acid synthesis reaction is a transcription reaction.
18. The method of claim 17, wherein the transcription reaction is a reverse transcription reaction.
19. The method of any one of claims 13-18, wherein two partially double-stranded primers are used.

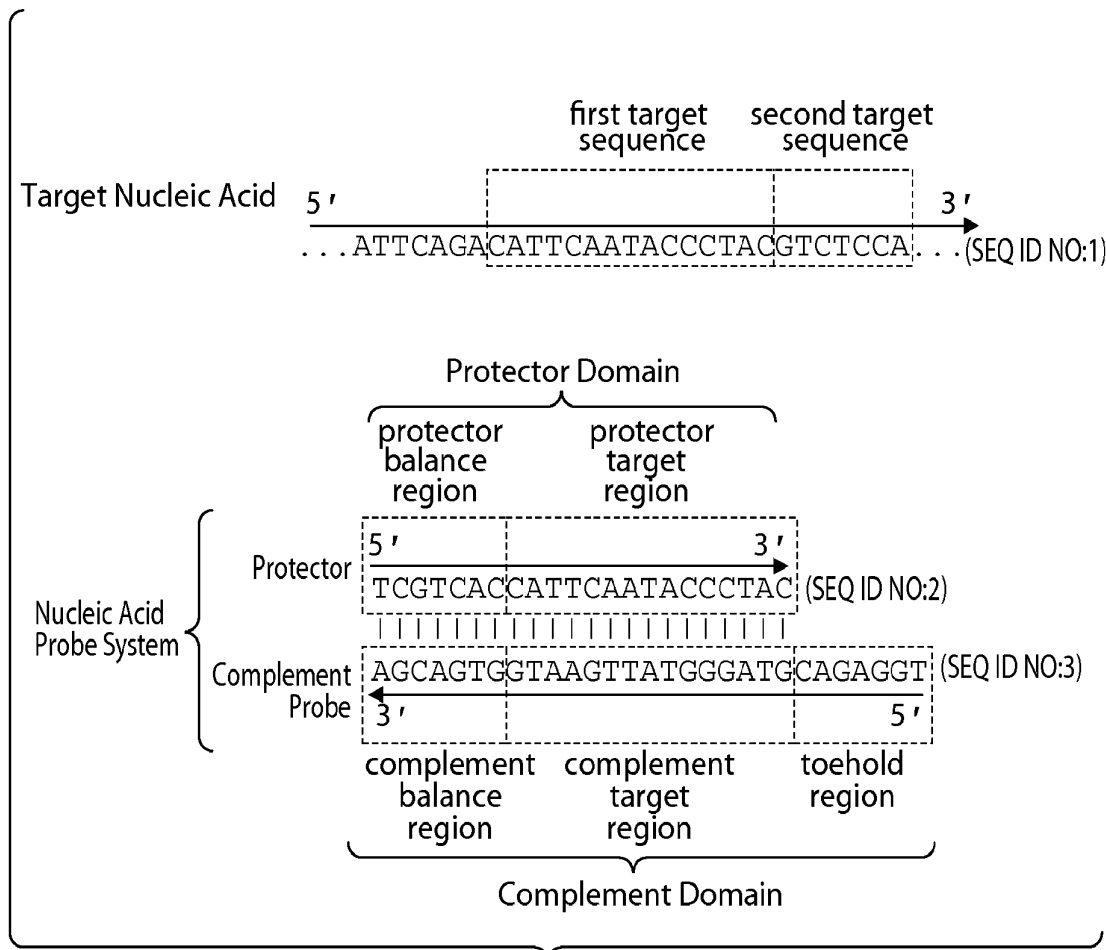


Fig. 1A

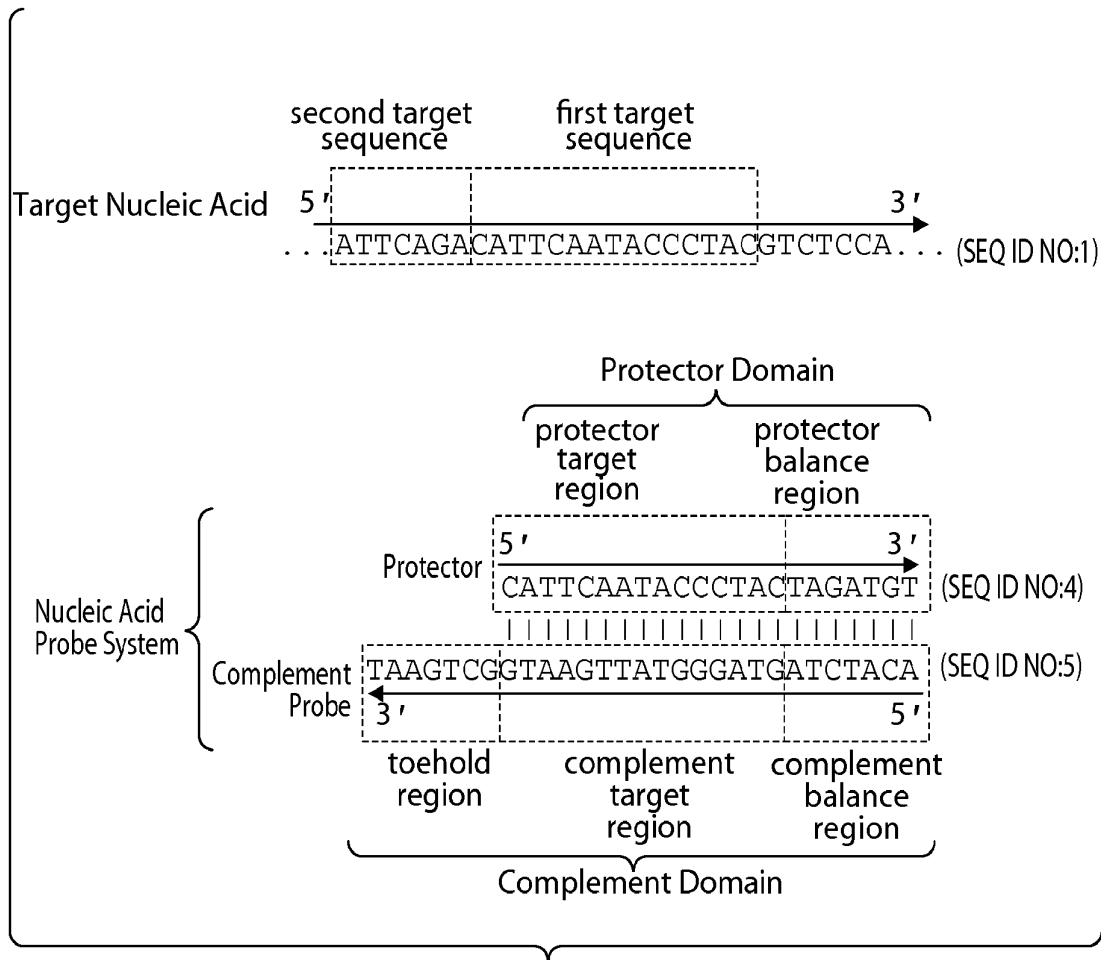


Fig. 1B

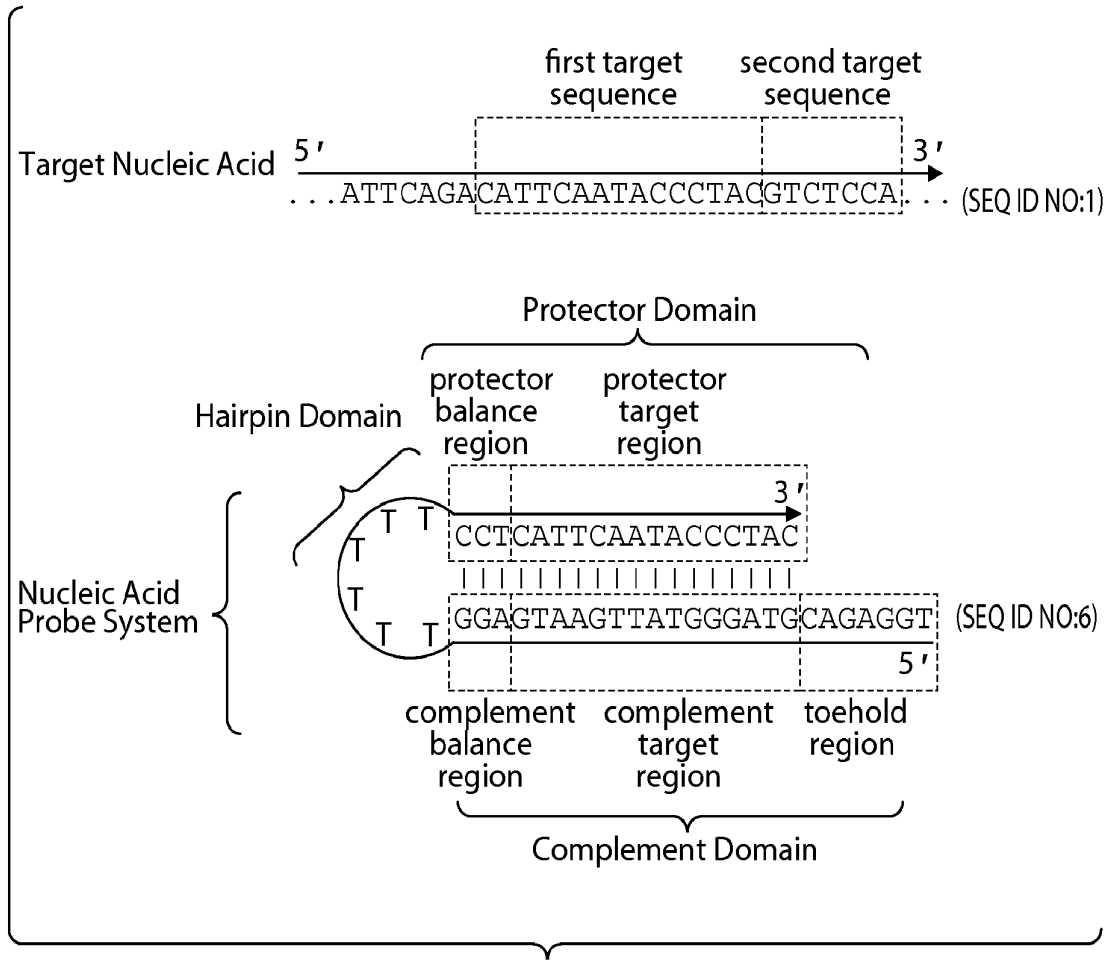


Fig. 2A

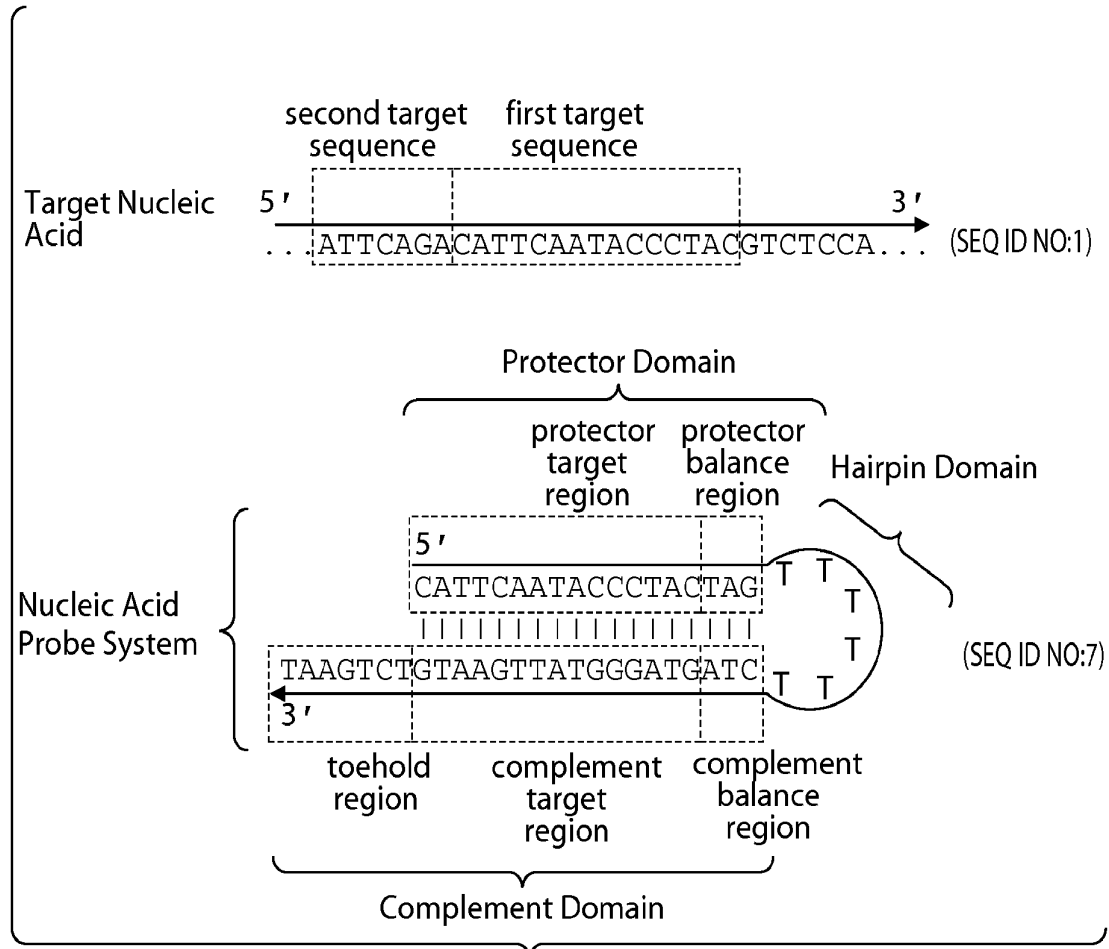


Fig. 2B

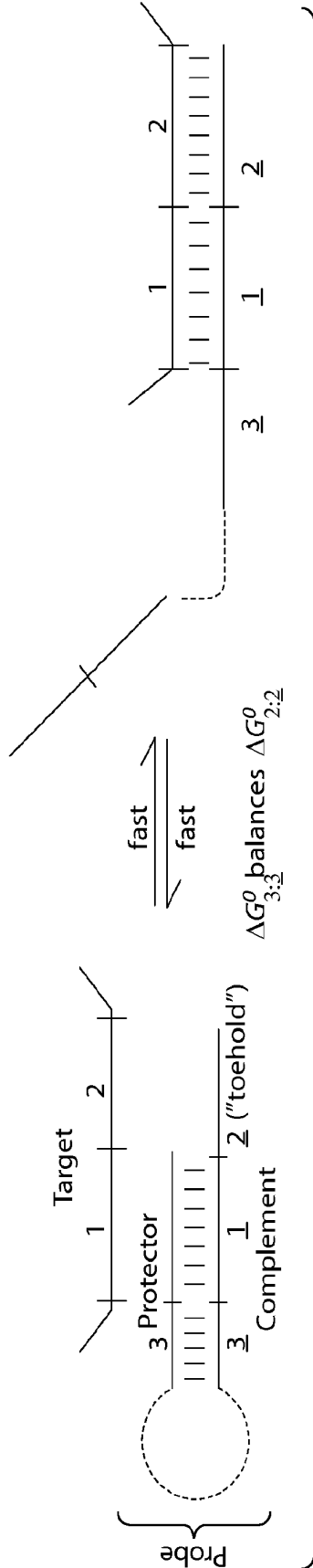


Fig. 3A

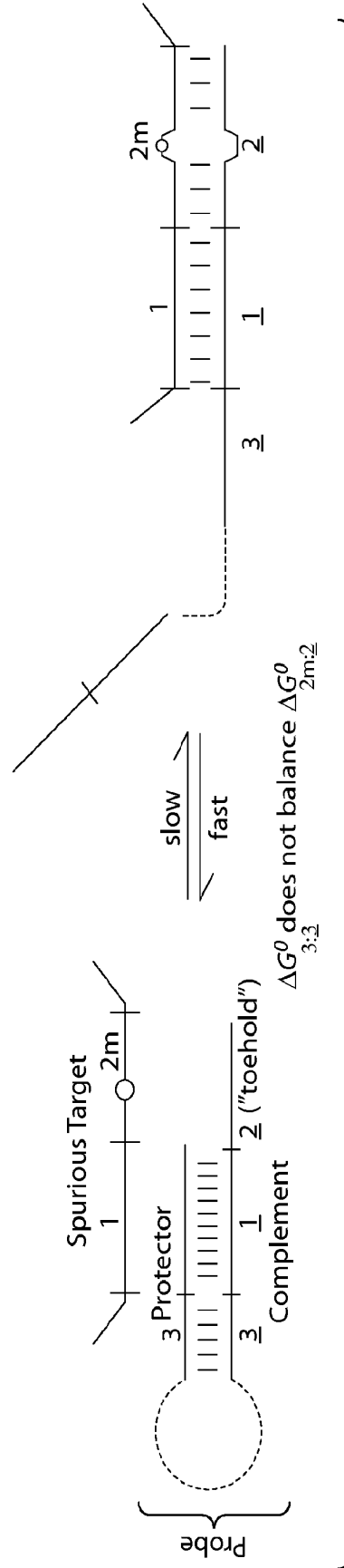


Fig. 3B

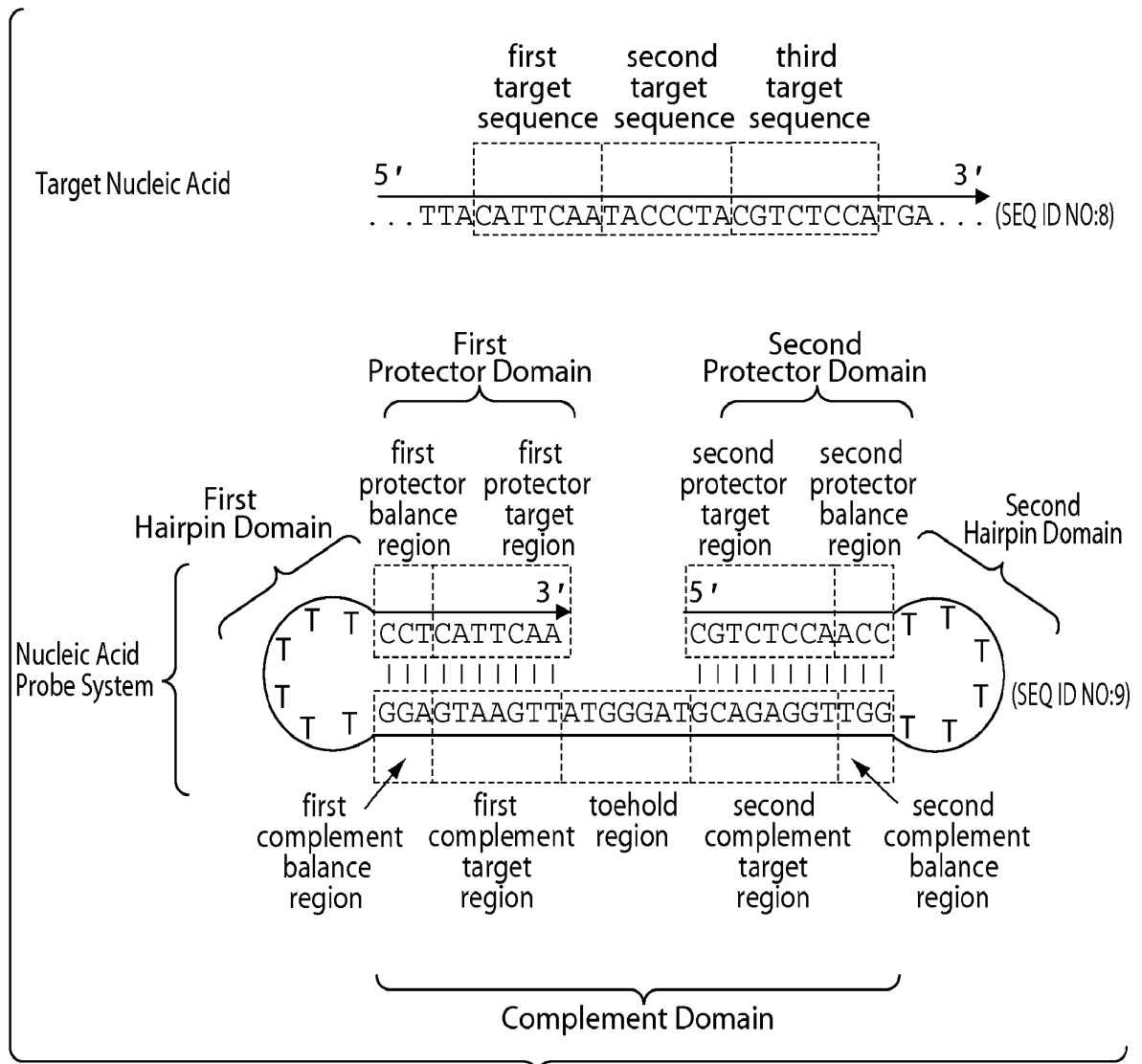


Fig. 4

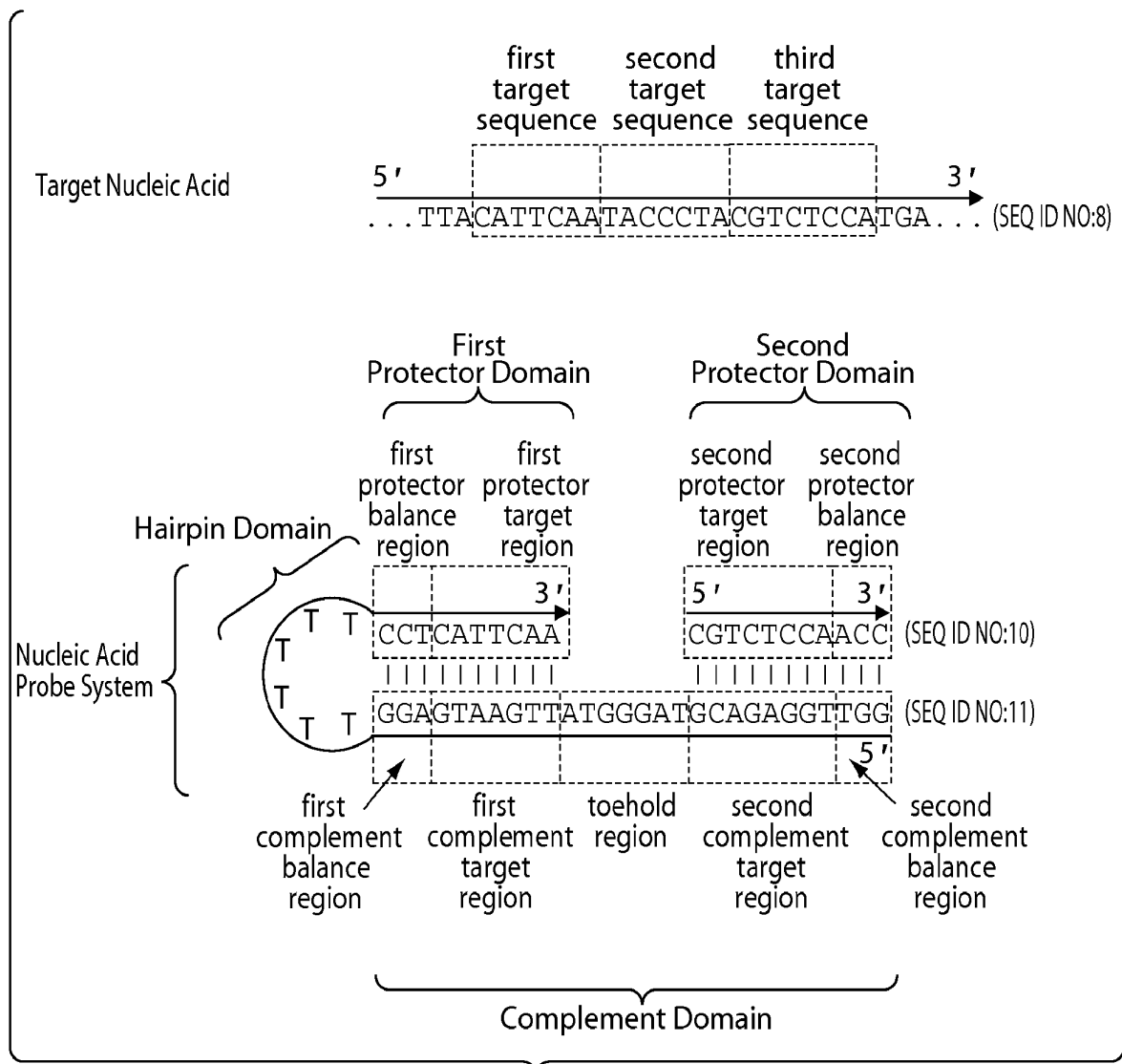


Fig. 5

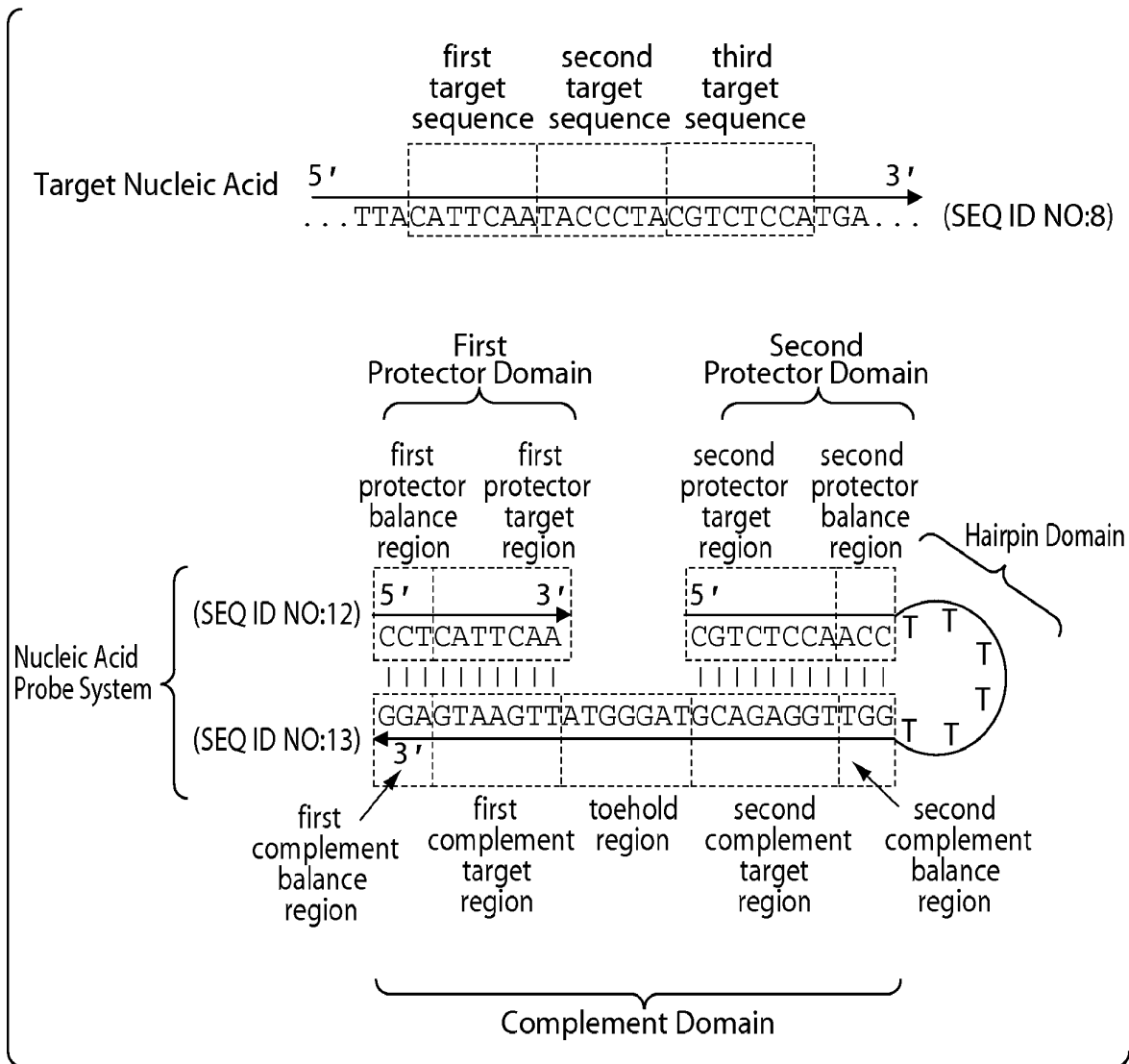


Fig. 6

9/25

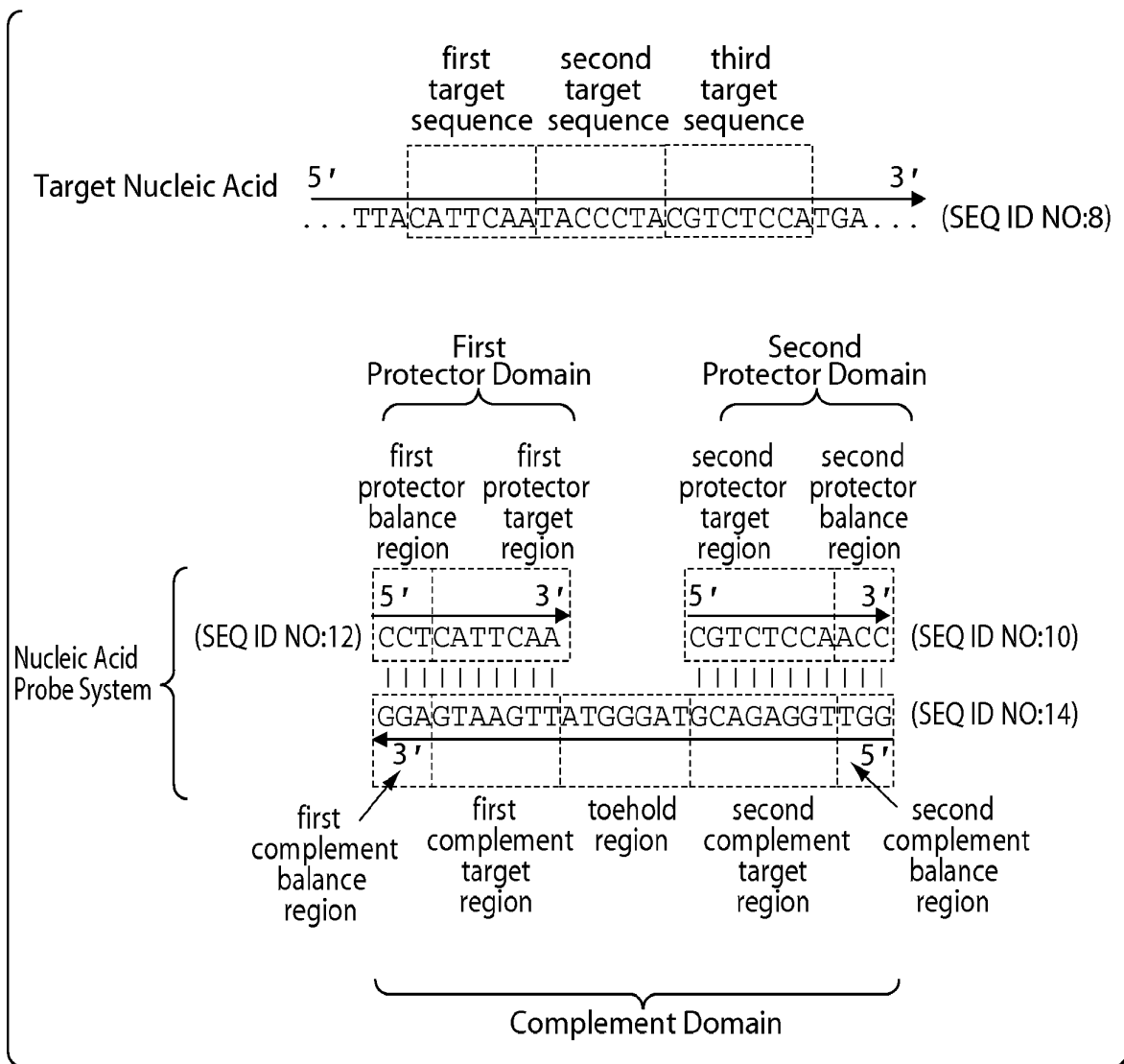


Fig. 7

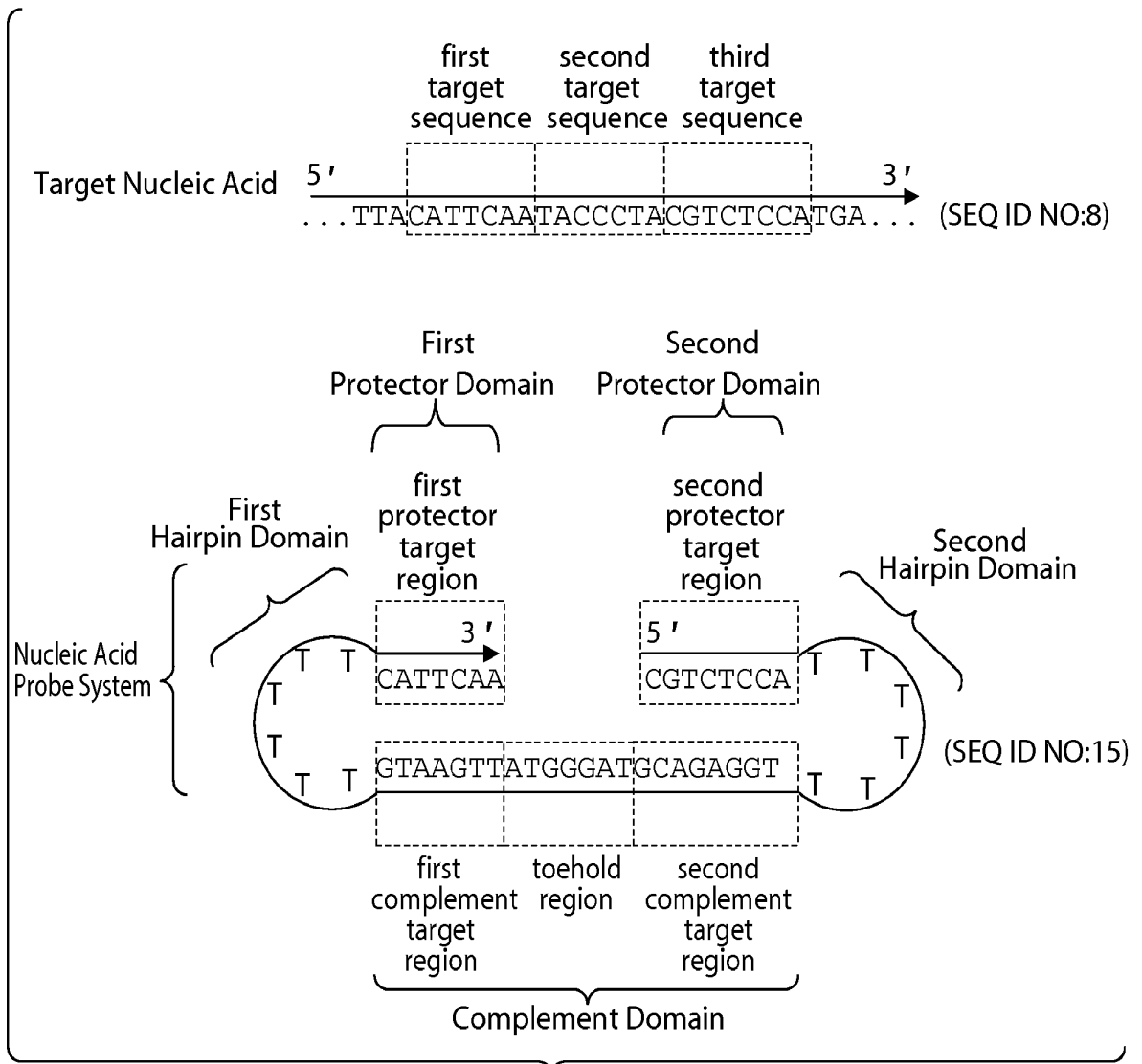


Fig. 8

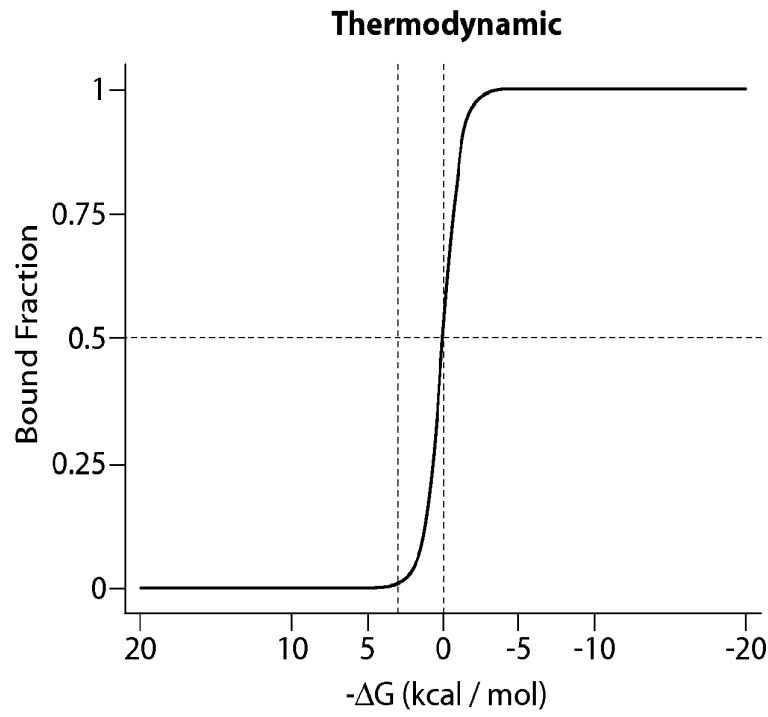


Fig. 10

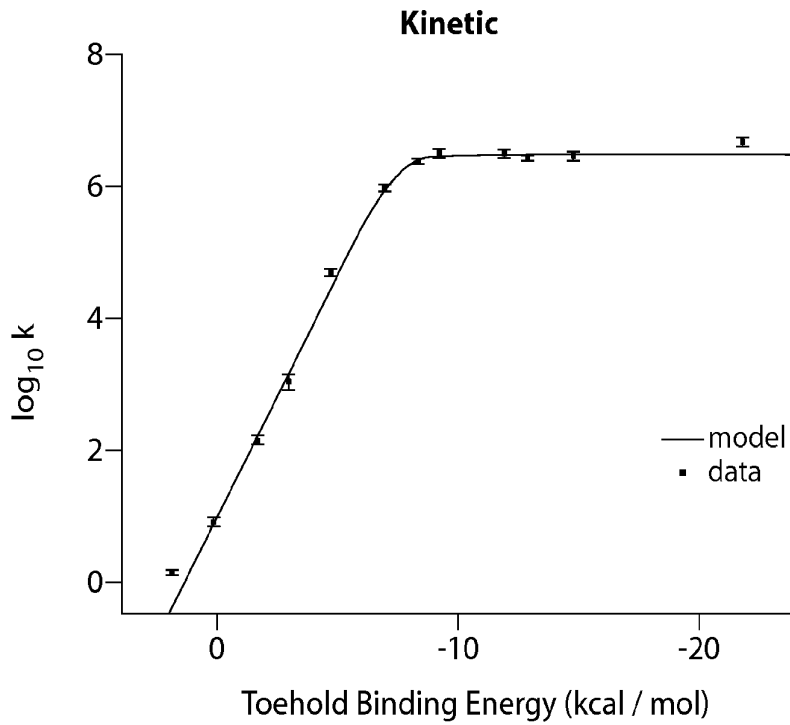


Fig. 11

14/25

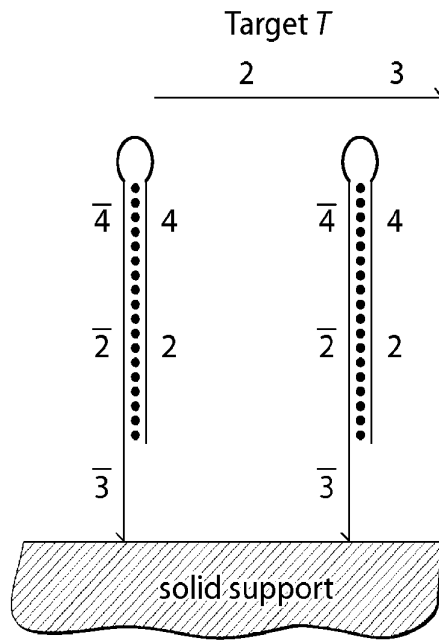


Fig. 12A

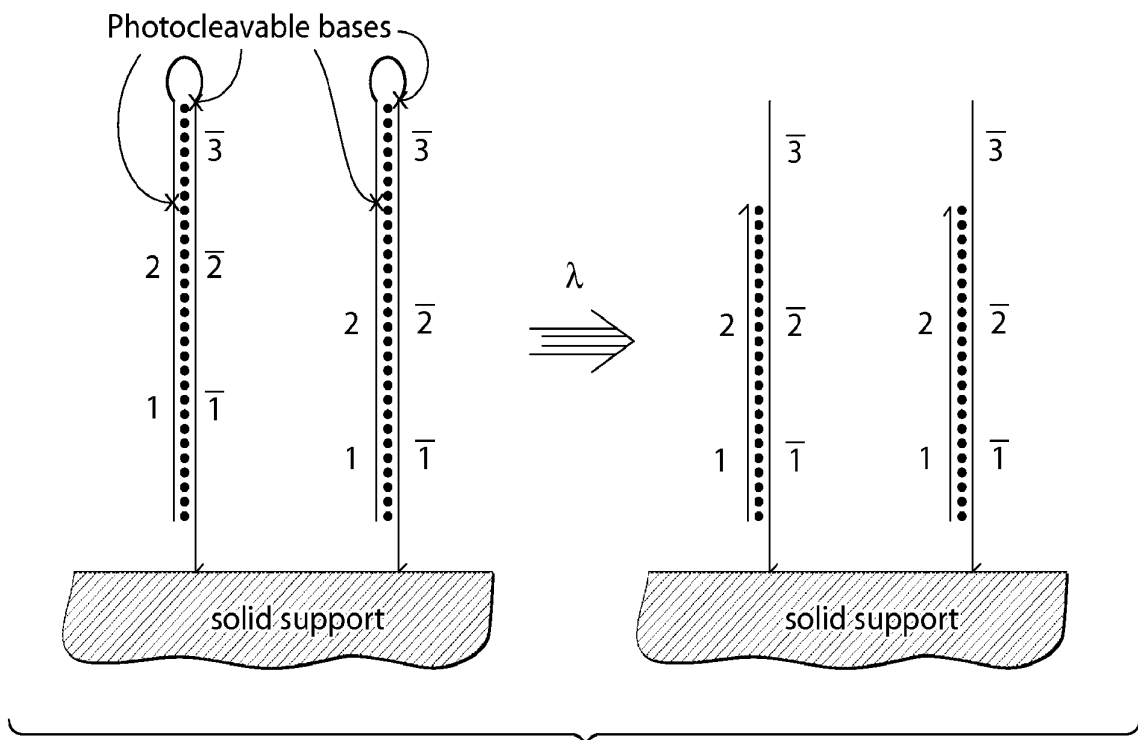


Fig. 12B

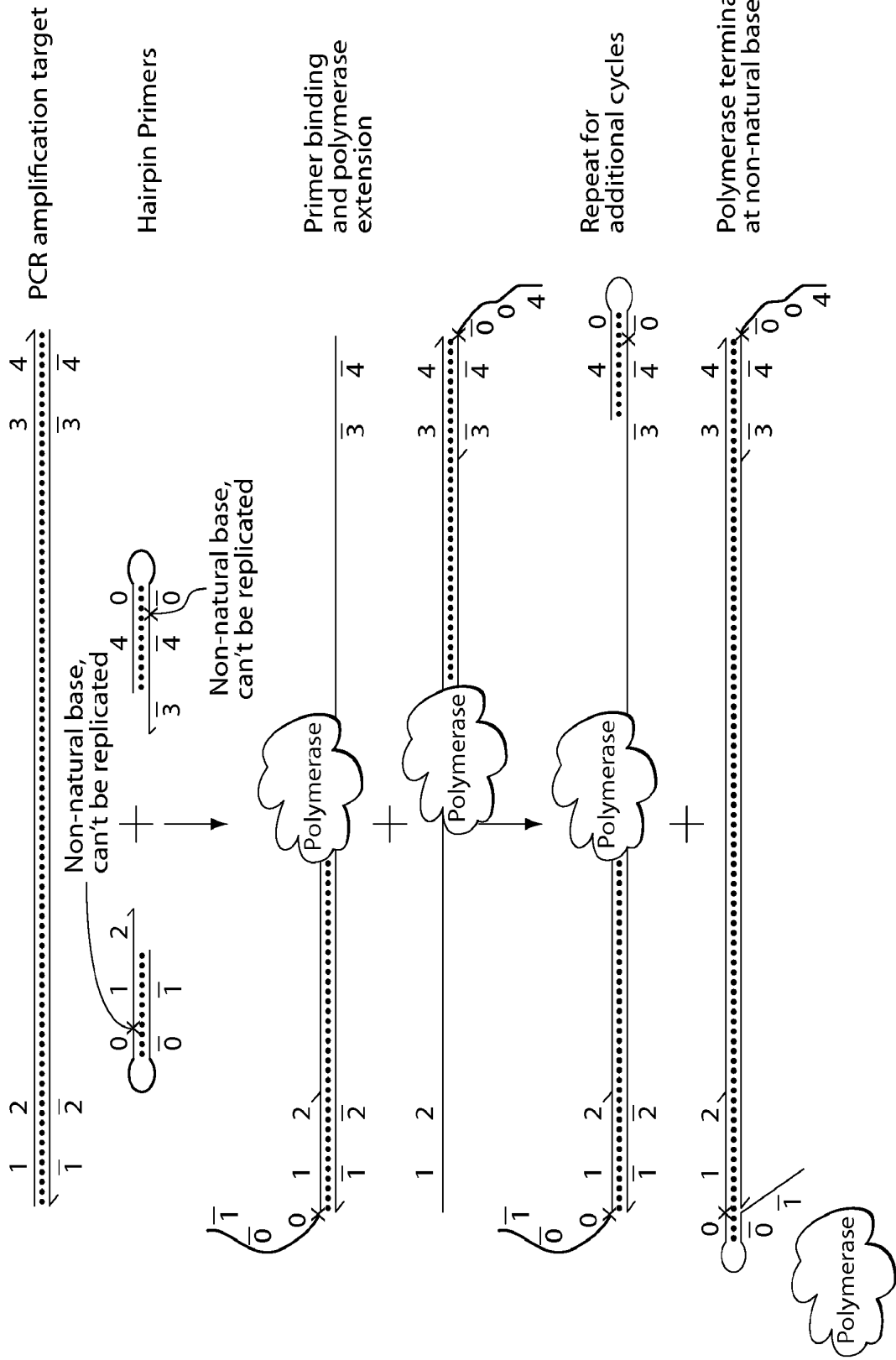


Fig. 13

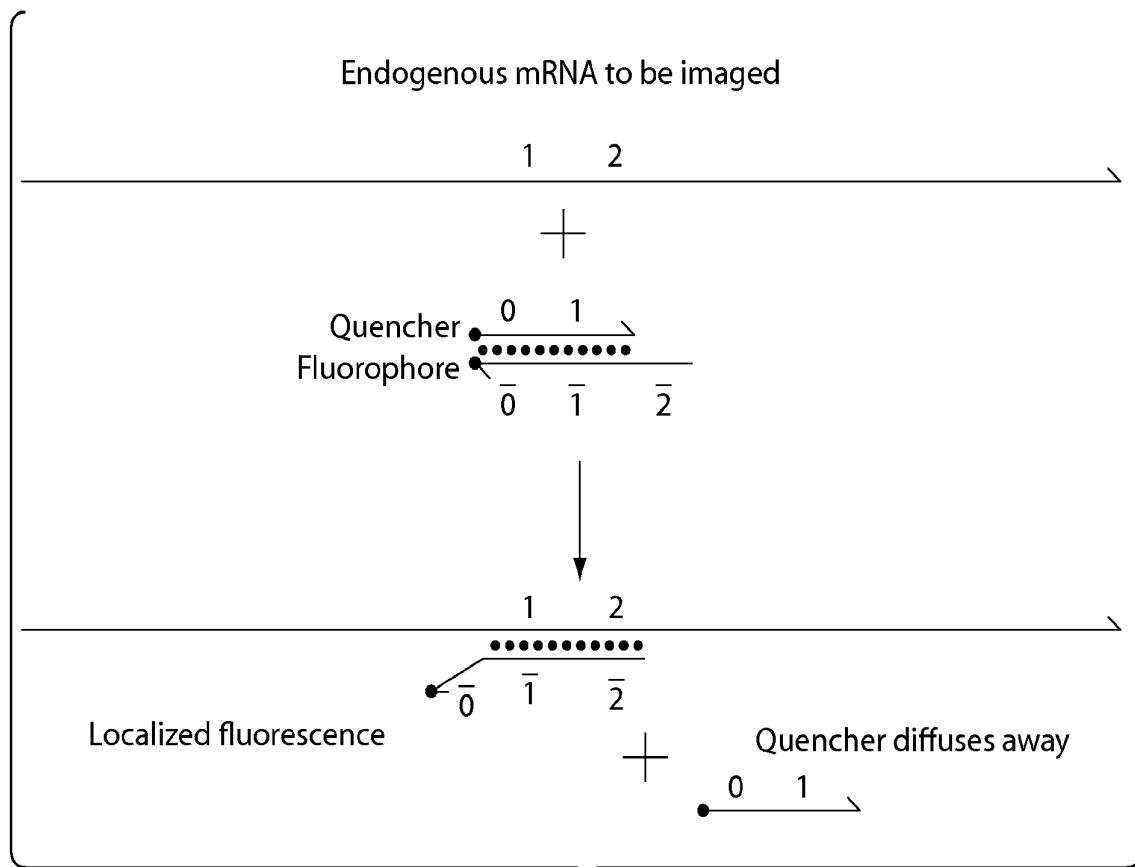


Fig. 14

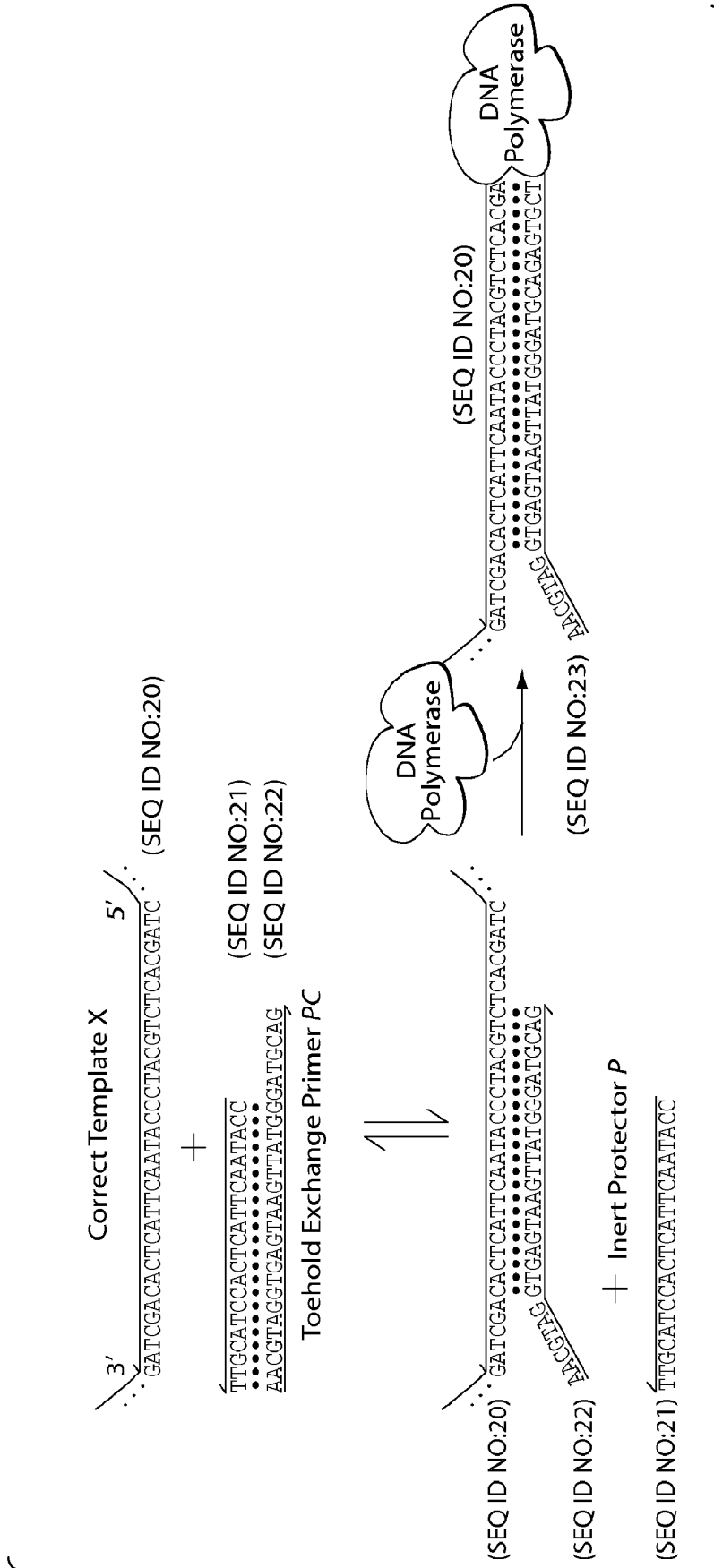


Fig. 15A

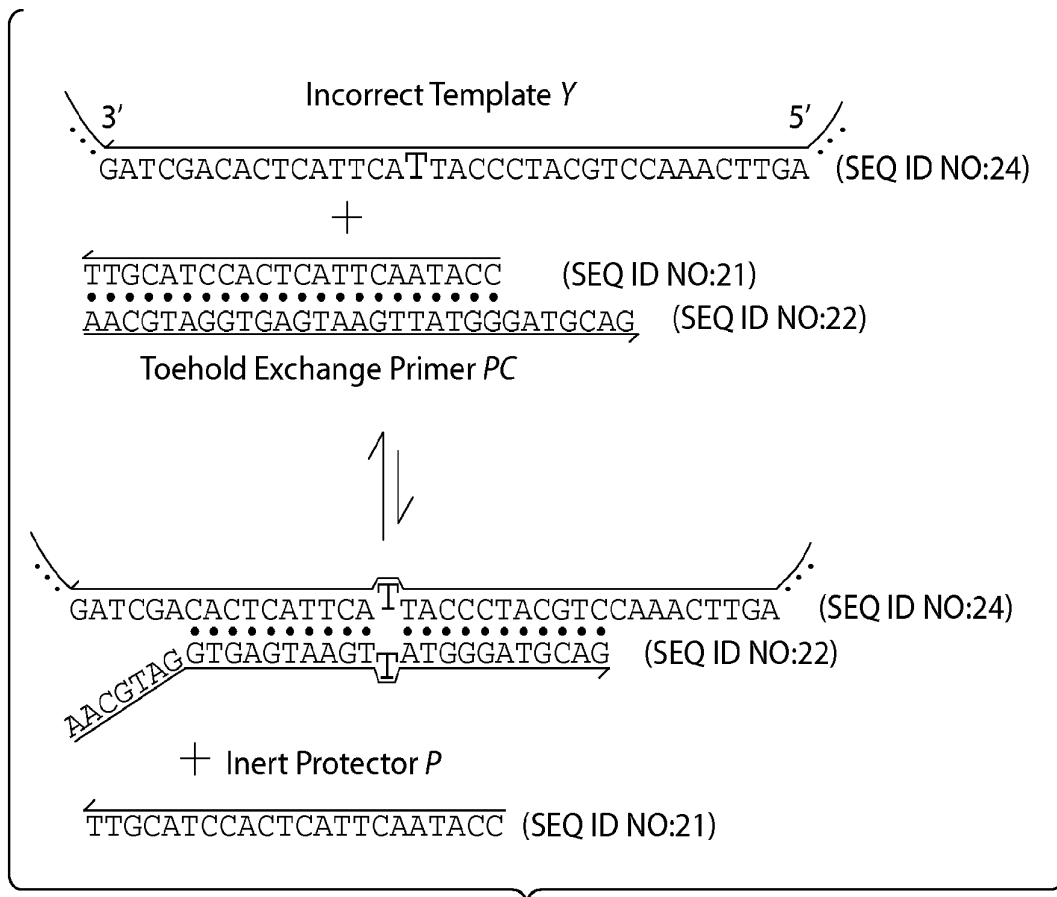


Fig. 15B

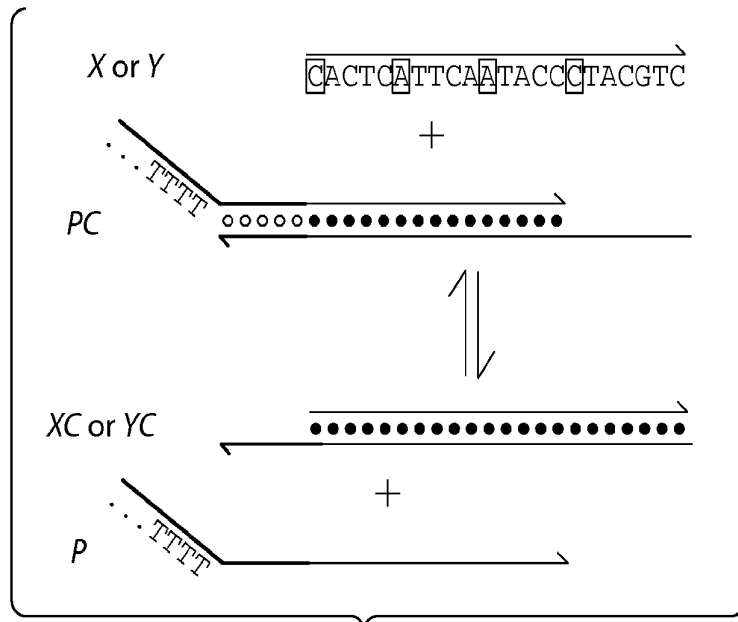


Fig. 16A

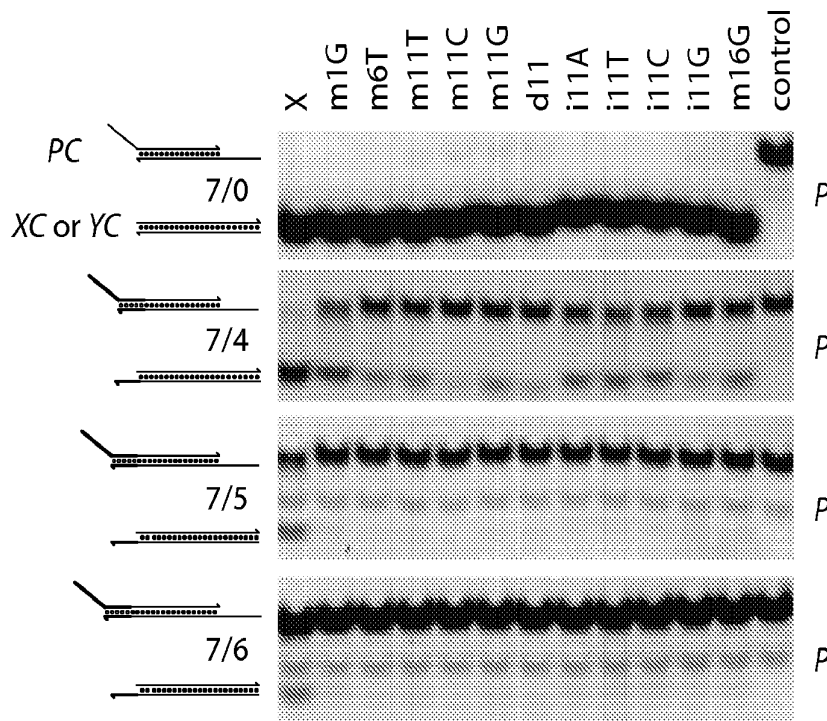


Fig. 16B

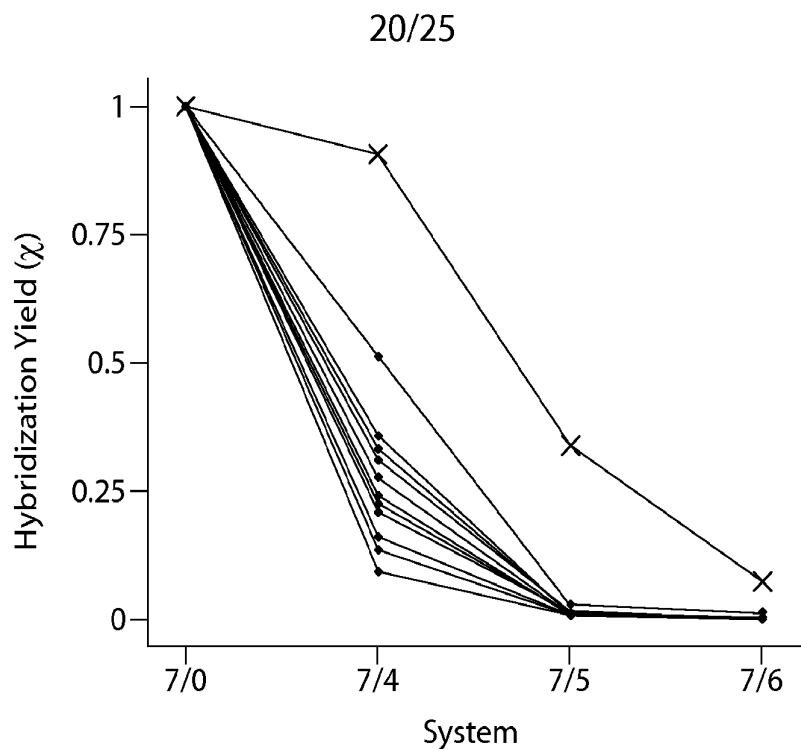


Fig. 16C

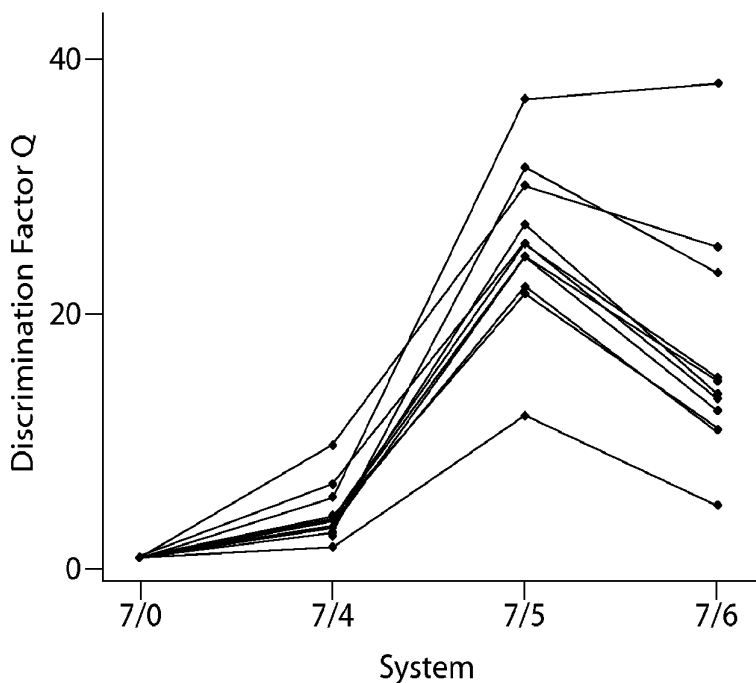


Fig. 16D

21/25

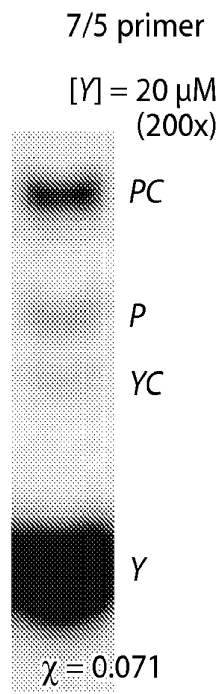


Fig. 16E

22/25

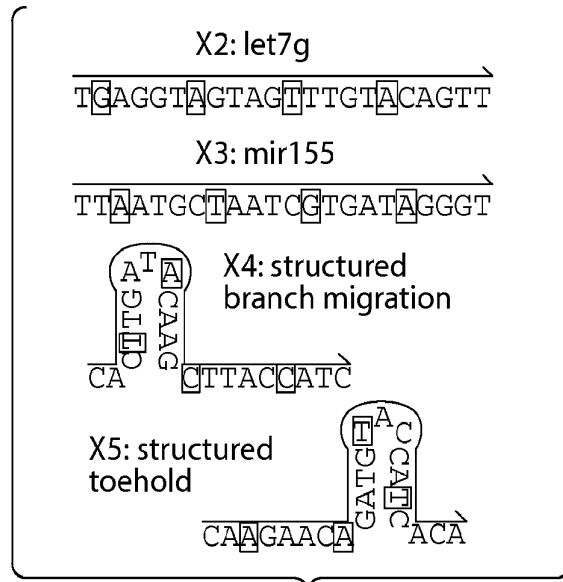


Fig. 17A

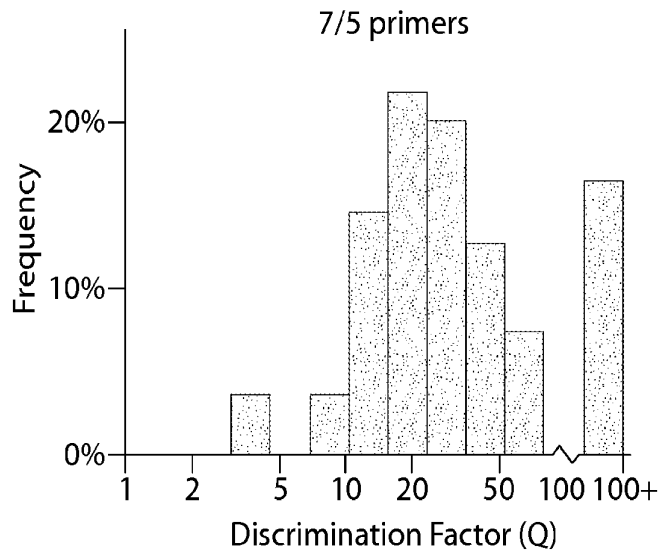


Fig. 17B

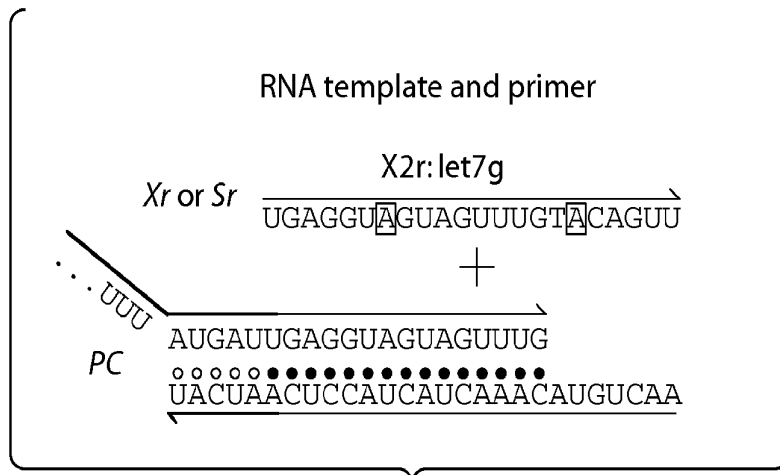


Fig. 17C

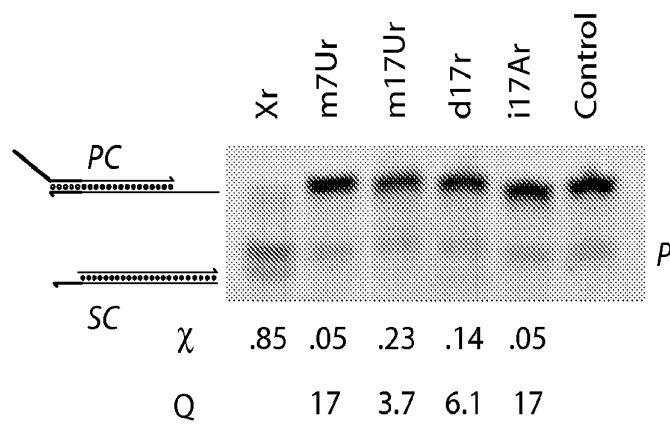


Fig. 17D

24/25

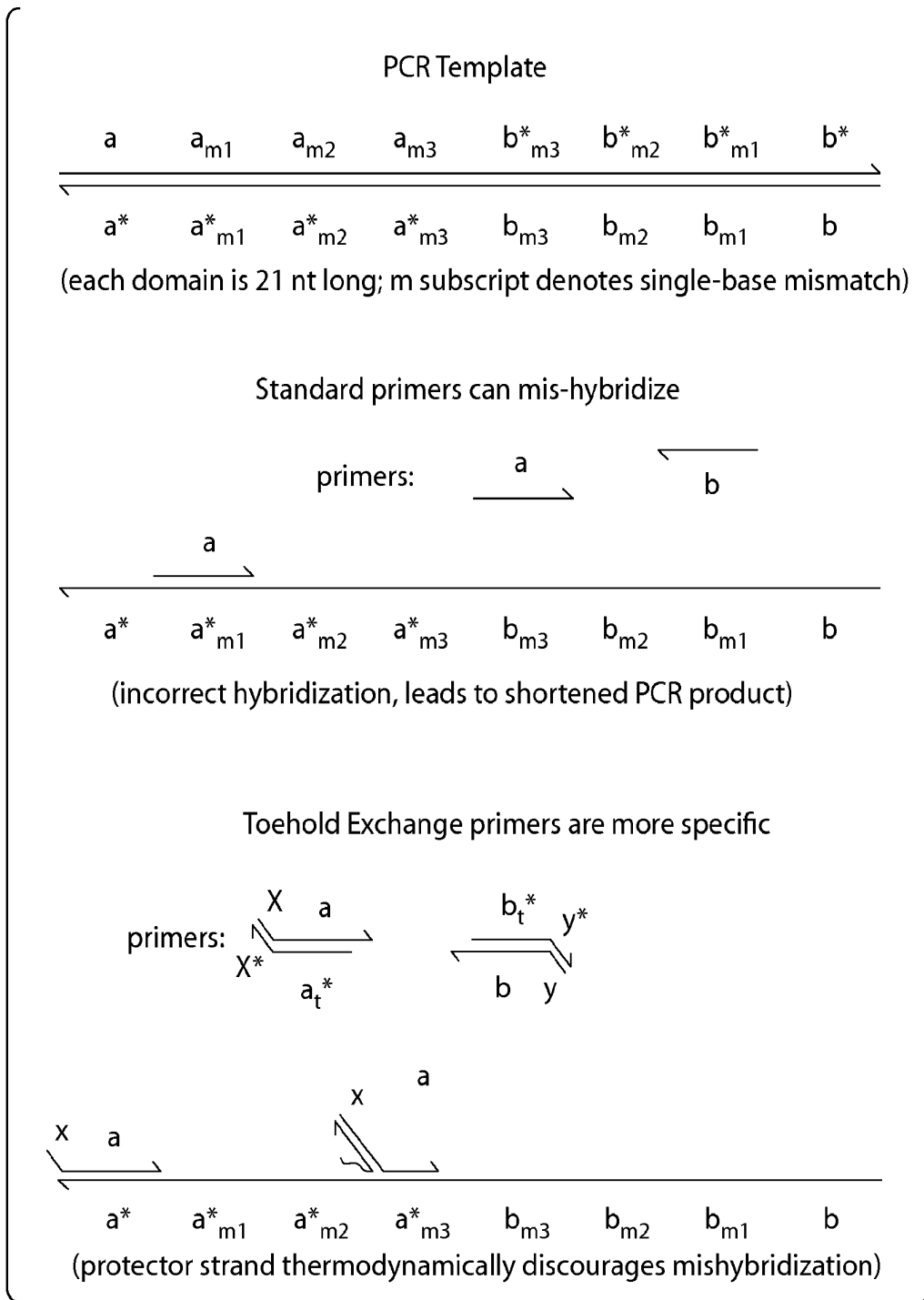


Fig. 18A

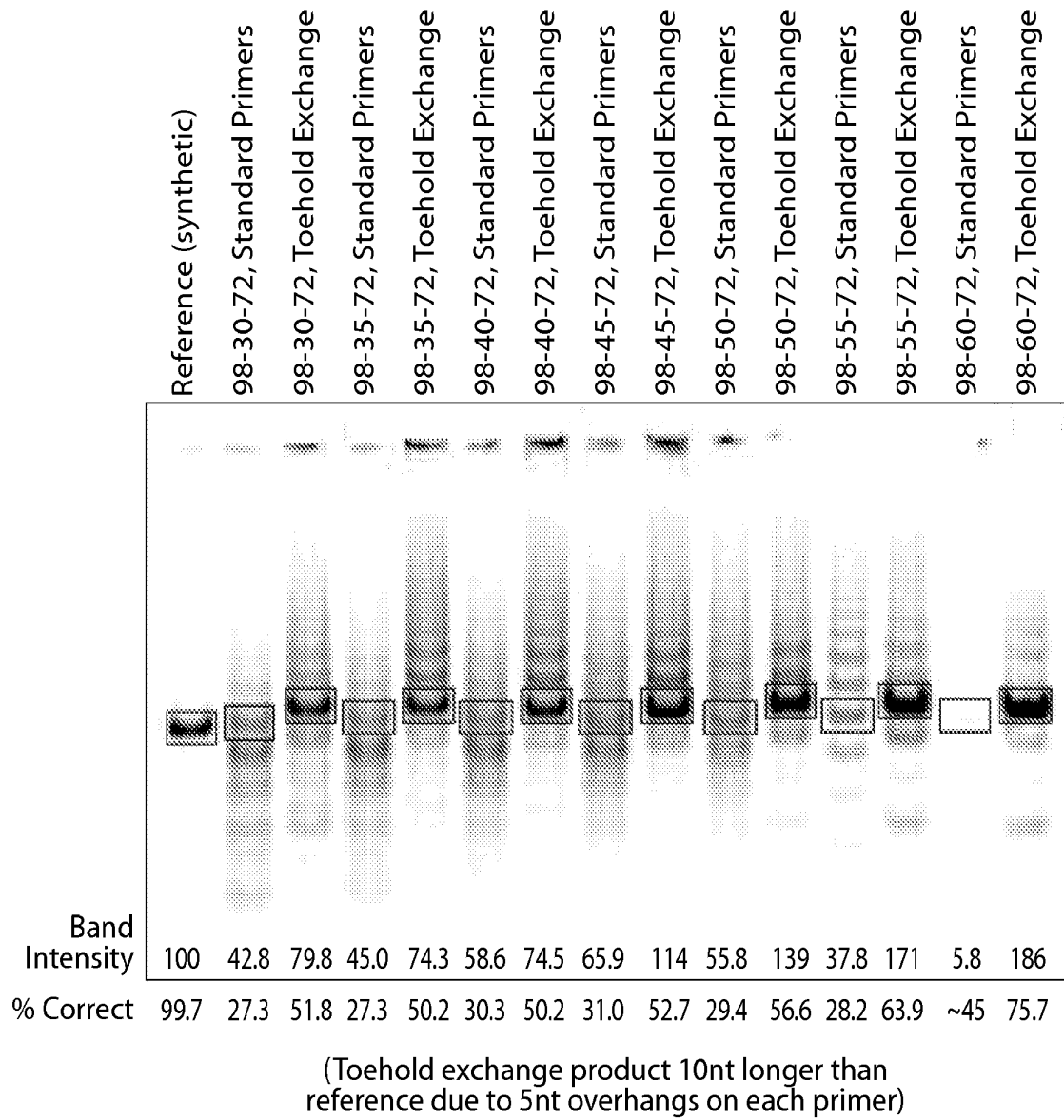


Fig. 18B

H049870420W000-SEQUENCE LISTING.TXT
SEQUENCE LISTING

<110> Harvard University
<120> COMPOSITIONS OF TOEHOLD PRIMER DUPLEXES AND METHODS OF USE
<130> H0498.70420W000
<140> TBD
<141> 2011-10-27
<150> US 61/407,291
<151> 2010-10-27
<160> 149
<170> Patent In version 3.5
<210> 1
<211> 29
<212> DNA
<213> Artificial Sequence
<220>
<223> Synthetic Oligonucleotide
<400> 1
attcagacat tcaatacct acgtctcca 29

<210> 2
<211> 22
<212> DNA
<213> Artificial Sequence
<220>
<223> Synthetic Oligonucleotide
<400> 2
tcgtcacat tcaatacct ac 22

<210> 3
<211> 29
<212> DNA
<213> Artificial Sequence
<220>
<223> Synthetic Oligonucleotide
<400> 3
agcagtggta agttatggga tgcagaggt 29

<210> 4
<211> 22
<212> DNA
<213> Artificial Sequence
<220>
<223> Synthetic Oligonucleotide
<400> 4
cattcaatac cctactagat gt 22

<210> 5
<211> 29

<212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 5
 taagtcggta agttatggga tgatctaca 29

 <210> 6
 <211> 49
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 6
 tggagacgta gggattgaa tgaggTTTTT tcctcattca ataccctac 49

 <210> 7
 <211> 49
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 7
 cattcaatac cctactagtt ttttctagta gggattgaa tgtctgaat 49

 <210> 8
 <211> 28
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 8
 ttacattcaa taccctacgt ctccatga 28

 <210> 9
 <211> 61
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 9
 cgctccaac cTTTTTGGT tggagacgta gggattgaa tgaggTTTTT tcctcattca 60
 a 61

 <210> 10
 <211> 11
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

<400> 10
cgtctccaac c 11

<210> 11
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 11
ggttgagac gtaggtatt gaatgaggtt ttttcctcat tcaa 44

<210> 12
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 12
cctcattcaa 10

<210> 13
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 13
cgtctccaac ctttttgggt tgaagacgt aggtattga atgagg 46

<210> 14
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 14
ggagtaagtt atgggatgca gaggttgg 28

<210> 15
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 15
cgtctccatt ttttggaga cgtagggtat tgaatgtttt ttcattcaa 49

<210> 16
<211> 22
<212> DNA
<213> Artificial Sequence

H049870420W000-SEQUENCELISTING.TXT

<220>
 <223> Synthetic Oligonucleotide
 <400> 16
 cattcaatac cctacgtctc ca 22

<210> 17
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 17
 cattcaatac cctacgtatc ca 22

<210> 18
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 18
 tcgacacat tcaatacct ac 22

<210> 19
 <211> 29
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 19
 agctgtggta agttatggga tgcagaggt 29

<210> 20
 <211> 36
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 20
 gatcgacact cattcaatac cctacgtctc acgatc 36

<210> 21
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 21
 ttgcatccac tcattcaata cc 22

<210> 22
 <211> 29
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 22
 aacgtaggtg agtaagttat gggatgcag 29

 <210> 23
 <211> 35
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 23
 aacgtaggtg agtaagttat gggatgcaga gtgct 35

 <210> 24
 <211> 36
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 24
 gatcgacact cattcattac cctacgtcaa acttga 36

 <210> 25
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 25
 cactcattca ataccctag tc 22

 <210> 26
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 26
 tgaggtagta gttgtacag tt 22

 <210> 27
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

<400> 27
ttaatgctaa tcgtgatagg gt 22

<210> 28
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 28
cacttgatac aagcttacca tc 22

<210> 29
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 29
caagaacaga tgtacatca ca 22

<210> 30
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 30
ugagguagua guuugacagu u 21

<210> 31
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 31
uuuau gauug agguaguagu uug 23

<210> 32
<211> 27
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 32
uacuaacucc aucaucaaac augucaa 27

<210> 33
<211> 21
<212> DNA
<213> Artificial Sequence

H049870420W000-SEQUENCELISTING.TXT

<220>
 <223> Synthetic Oligonucleotide
 <400> 33
 tgcattcact cattcaatac c 21

<210> 34
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 34
 gcatccactc attcaatacc 20

<210> 35
 <211> 19
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 35
 catccactca ttcaatacc 19

<210> 36
 <211> 15
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 36
 cactcattca atacc 15

<210> 37
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 37
 atgattgagg tagtagtttg 20

<210> 38
 <211> 19
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 38
 tgattgaggt agtagtttg 19

<210> 39
 <211> 20
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 39
 aggatttaat gctaatcgtg 20

<210> 40
 <211> 19
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 40
 ggatttaatg ctaatcgtg 19

<210> 41
 <211> 20
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 41
 ctcatcactt gatacaagct 20

<210> 42
 <211> 19
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 42
 tcatcacttg atacaagct 19

<210> 43
 <211> 20
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 43
 cgttccaaga acagatgtac 20

<210> 44
 <211> 19
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

<400> 44
gttccaagaa cagatgtac 19

<210> 45
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 45
gactcattca ataccctacg tc 22

<210> 46
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 46
cactctttca ataccctacg tc 22

<210> 47
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 47
cactcattca ttaccctacg tc 22

<210> 48
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 48
cactcattca ctaccctacg tc 22

<210> 49
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 49
cactcattca gtaccctacg tc 22

<210> 50
<211> 21
<212> DNA
<213> Artificial Sequence

H049870420W000-SEQUENCELISTING.TXT

<220>
 <223> Synthetic Oligonucleotide
 <400> 50
 cactcattca taccctacgt c 21

<210> 51
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 51
 cactcattca aataccctac gtc 23

<210> 52
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 52
 cactcattca tataccctac gtc 23

<210> 53
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 53
 cactcattca cataccctac gtc 23

<210> 54
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 54
 cactcattca gataccctac gtc 23

<210> 55
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 55
 cactcattca ataccgtacg tc 22

<210> 56
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 56
 tcaggtagta gttgtacag tt 22

<210> 57
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 57
 tgaggttgta gttgtacag tt 22

<210> 58
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 58
 tgaggtagta gattgtacag tt 22

<210> 59
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 59
 tgaggtagta gttgttcag tt 22

<210> 60
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 60
 tgaggtagta gttgtccag tt 22

<210> 61
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

<400> 61
tgaggtagta gtttgtgcag tt 22

<210> 62
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 62
tgaggtagta gtttgtcagt t 21

<210> 63
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 63
tgaggtagta gtttgaaca gtt 23

<210> 64
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 64
tgaggtagta gtttgttaca gtt 23

<210> 65
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 65
tgaggtagta gtttgtgaca gtt 23

<210> 66
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 66
tgaggtagta gtttgtcaca gtt 23

<210> 67
<211> 22
<212> DNA
<213> Artificial Sequence

H049870420W000-SEQUENCELISTING.TXT

<220>
 <223> Synthetic Oligonucleotide
 <400> 67
 tttatgctaa tcgtgatagg gt 22

<210> 68
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 68
 ttaatgcaaa tcgtgatagg gt 22

<210> 69
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 69
 ttaatgcgaa tcgtgatagg gt 22

<210> 70
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 70
 ttaatgccaa tcgtgatagg gt 22

<210> 71
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 71
 ttaatgcaat cgtgatagg t 21

<210> 72
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 72
 ttaatgcata atcgtgatag ggt 23

<210> 73
 <211> 23
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 73
 ttaatgctta atcgtgatag ggt 23

<210> 74
 <211> 23
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 74
 ttaatgctga atcgtgatag ggt 23

<210> 75
 <211> 23
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 75
 ttaatgccta atcgtgatag ggt 23

<210> 76
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 76
 ttaatgctaa tcctgatagg gt 22

<210> 77
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 77
 ttaatgctaa tcgtgattgg gt 22

<210> 78
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

<400> 78
cacatgatac aagcttacca tc 22

<210> 79
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 79
cacttgattc aagcttacca tc 22

<210> 80
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 80
cacttgatac aaggttacca tc 22

<210> 81
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 81
cacttgatac aagcttacga tc 22

<210> 82
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 82
cacttgatac aagcttaca tc 22

<210> 83
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 83
cacttgatac aagcttacta tc 22

<210> 84
<211> 21
<212> DNA
<213> Artificial Sequence

H049870420W000-SEQUENCELISTING.TXT

<220>
 <223> Synthetic Oligonucleotide
 <400> 84
 cacttgatc aagcttacat c 21

<210> 85
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 85
 cacttgatc aagcttacc atc 23

<210> 86
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 86
 cacttgatc aagcttacgc atc 23

<210> 87
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 87
 cacttgatc aagcttacac atc 23

<210> 88
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 88
 cacttgatc aagcttactc atc 23

<210> 89
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 89
 catgaacaga tgtaccatca ca 22

<210> 90
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 90
 caagaactga tgtacatca ca 22

<210> 91
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 91
 caagaacaga tgaacatca ca 22

<210> 92
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 92
 caagaacaga tgcacatca ca 22

<210> 93
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 93
 caagaacaga tggacatca ca 22

<210> 94
 <211> 21
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 94
 caagaacaga tgacatcac a 21

<210> 95
 <211> 23
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

<400> 95
caagaacaga tgataccatc aca 23

<210> 96
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 96
caagaacaga tgttaccatc aca 23

<210> 97
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 97
caagaacaga tgctaccatc aca 23

<210> 98
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 98
caagaacaga tggtagcatc aca 23

<210> 99
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 99
caagaacaga tgtaccaaca ca 22

<210> 100
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 100
cactcattca ataccctacg tctttt 26

<210> 101
<211> 22
<212> DNA
<213> Artificial Sequence

H049870420W000-SEQUENCELISTING.TXT

<220>
 <223> Synthetic Oligonucleotide
 <400> 101
 cactcattca ataccctacg tc 22

<210> 102
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 102
 tttatgattg aggtagtagt ttg 23

<210> 103
 <211> 27
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 103
 tactaactcc atcatcaaac atgtcaa 27

<210> 104
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 104
 tttaggattt aatgctaatac gtg 23

<210> 105
 <211> 27
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 105
 tcctaaatta cgattagcac tatccca 27

<210> 106
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 106
 tttcgtccca agaacagatg tac 23

<210> 107
 <211> 27
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 107
 gcagggttct tgtctacatg gtagtgt 27

<210> 108
 <211> 24
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 108
 tttgcatcc actcattcaa tacc 24

<210> 109
 <211> 28
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 109
 acgtagtgta gtaagttatg ggatgcag 28

<210> 110
 <211> 23
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 110
 tttgcatcca ctattcaat acc 23

<210> 111
 <211> 27
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 111
 cgtagtgag taagttatgg gatgcag 27

<210> 112
 <211> 36
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

H049870420W000-SEQUENCELISTING.TXT

<400> 112
ggacgttgat atgggacgta gggattgaa tgagtg 36

<210> 113
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 113
gcttaagcct tcactattca atgtatgaga ataaaccctc ctctgtaga ggctaagtac 60
t 61

<210> 114
<211> 60
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 114
cgtcgcctcgg tttgcttctg catggagggt gaggactttt atacacaggg aagcgagctc 60

<210> 115
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 115
tgagtagta gttgtacag ttgagaagtt aggttg 36

<210> 116
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 116
caacctaact tctcaactgt acaactact acctca 36

<210> 117
<211> 60
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 117
gagctcgctt ccctgtgtat aaaagtcctc accctcatg cagaagcaaa ccgagcgacg 60

<210> 118
<211> 61

H049870420W000-SEQUENCELISTING.TXT

<212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide

<400> 118
 agtacttagc ctctaacaga ggagggttta ttctcataca ttgaatagtg aaggcttaag 60
 c 61

<210> 119
 <211> 36
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide

<400> 119
 cactcattca ataccctagc tcccatatca acgtcc 36

<210> 120
 <211> 36
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide

<400> 120
 ggacgttgat atgggacgtc gggattgaa tgagtg 36

<210> 121
 <211> 35
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide

<400> 121
 tgacctaag ctaactcgtg tcgctctgcg gcttc 35

<210> 122
 <211> 35
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide

<400> 122
 tctacttgac gtatacggtc tagctgccac ggagg 35

<210> 123
 <211> 36
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide

<400> 123
 tggtagta gttgtacag ttgagaagtt aggttg 36

<210> 124
 <211> 36
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide

<400> 124
 caacctaact tctcaactgt acaactact accaca 36

<210> 125
 <211> 35
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide

<400> 125
 cctccgtggc agctagaccg tatacgtcaa gtaga 35

<210> 126
 <211> 35
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide

<400> 126
 gaagccgcag agcgacacga gtagcatta ggtca 35

<210> 127
 <211> 36
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide

<400> 127
 cactcattca atacccgacg tccatatca acgtcc 36

<210> 128
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide

<400> 128
 gacgtagggt attgaatgag tg 22

<210> 129
 <211> 22
 <212> DNA
 <213> Artificial Sequence

H049870420W000-SEQUENCELISTING.TXT

<220>
 <223> Synthetic Oligonucleotide
 <400> 129
 aactgtacaa actactacct ca 22

<210> 130
 <211> 27
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 130
 ggtcagacgt aggtattga atgagtg 27

<210> 131
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 131
 tcaataccct acgtctgacc 20

<210> 132
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 132
 tcaataccct acgtctgacc 20

<210> 133
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 133
 gtagtttgta cagttagtcg 20

<210> 134
 <211> 49
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 134
 gagtggacgt aggtattga atgagtgac gtcgggtatt gaatgagtg 49

<210> 135
 <211> 44
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 135
 gacgtagggt attaatgag tggacgtagg gtactgaatg agtg 44

<210> 136
 <211> 44
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 136
 tgaggtagta gttgtatag tttgaggtag tactttgtac agtt 44

<210> 137
 <211> 49
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 137
 tgggtagta gttgtacag tttgaggtag tagtttgtac agtttgagg 49

<210> 138
 <211> 49
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 138
 cctcaaactg tacaactac tacctcaaac tgtacaact actaccaca 49

<210> 139
 <211> 44
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

 <400> 139
 aactgtacaa agtactacct caactatac aaactactac ctca 44

<210> 140
 <211> 44
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Oligonucleotide

<400> 140
cactcattca gtaccctagc tccactcatt taatacccta cgtc 44

<210> 141
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 141
cactcattca ataccgagc tccactcatt caatacccta cgtccactc 49

<210> 142
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 142
gagtgagct cgggtattga atgagtgac gtaggtatt gaatgagtg 49

<210> 143
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 143
gacgtagggt attaatgag tggacgtagg gtactgaatg agtg 44

<210> 144
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 144
tgaggtagta gttgtatag tttgaggtag tactttgtac agtt 44

<210> 145
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Oligonucleotide

<400> 145
tgaggtagta gttgtacag tttgtggtag tagtttgtac agtttgagg 49

<210> 146
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 146
 cctcaaactg taaaactac tacctcaaac tgtaaaact actaccaca 49

<210> 147
 <211> 44
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 147
 aactgtacaa agtactacct caactatac aaactactac ctca 44

<210> 148
 <211> 44
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 148
 cactcattca gtaccctacg tccactcatt taatacccta cgtc 44

<210> 149
 <211> 49
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Oligonucleotide
 <400> 149
 cactcattca atacccgacg tccactcatt caatacccta cgccactc 49