US 20190285620A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2019/0285620 A1

Green et al. (43) Pub. Date: Sep. 19, 2019

(54) **LOW-COST DETECTION OF NOROVIRUS USING PAPER-BASED CELL-FREE SYSTEMS AND SYNBODY-BASED VIRAL ENRICHMENT**

(71) Applicant: **Arizona Board of Regents on behalf of Arizona State University**, Scottsdale, AZ (US)

(72) Inventors: **Alexander Green**, Scottsdale, AZ (US); **Duo Ma**, Tempe, AZ (US); **Luhui Shen**, Tempe, AZ (US); **Chris Diehnelt**, Chandler, AZ (US)

(21) Appl. No.: **16/279,045**

(22) Filed: **Feb. 19, 2019**

Related U.S. Application Data

(60) Provisional application No. 62/632,792, filed on Feb. 20, 2018.

Publication Classification

(51) **Int. Cl.**

G01N 33/543	(2006.01)
G01N 33/569	(2006.01)
C07K 19/00	(2006.01)
G01N 33/53	(2006.01)

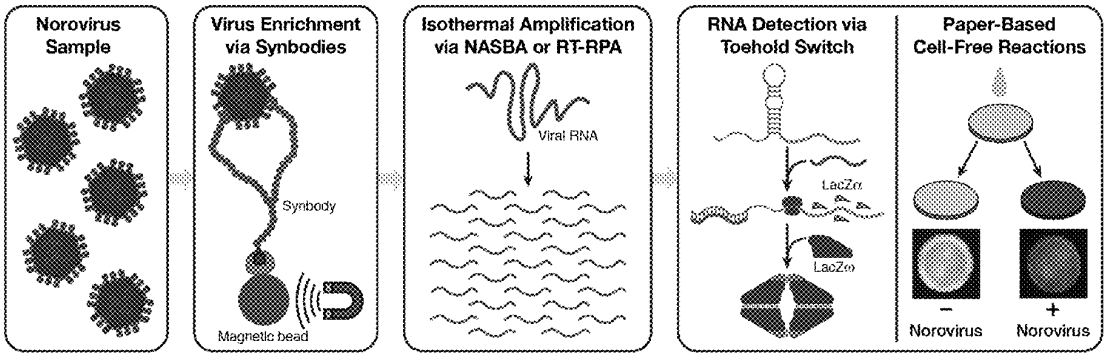
(52) **U.S. Cl.**

CPC ***G01N 33/54306*** (2013.01); ***G01N 33/569*** (2013.01); ***C07K 19/00*** (2013.01); *C07K 2317/569* (2013.01); ***G01N 33/5434*** (2013.01); *C07K 2318/00* (2013.01); *C07K 2317/60* (2013.01); ***G01N 33/5308*** (2013.01)

(57) **ABSTRACT**

Provided herein are methods and systems for low-cost, low-equipment detection of pathogens in biological sample. In particular, provided herein is a low-cost method for detecting norovirus that provides reliable, visible test with femtomolar, attomolar, and zeptomolar detection limits and that uses materials suitable for deployment of the methods in the field.

Specification includes a Sequence Listing.



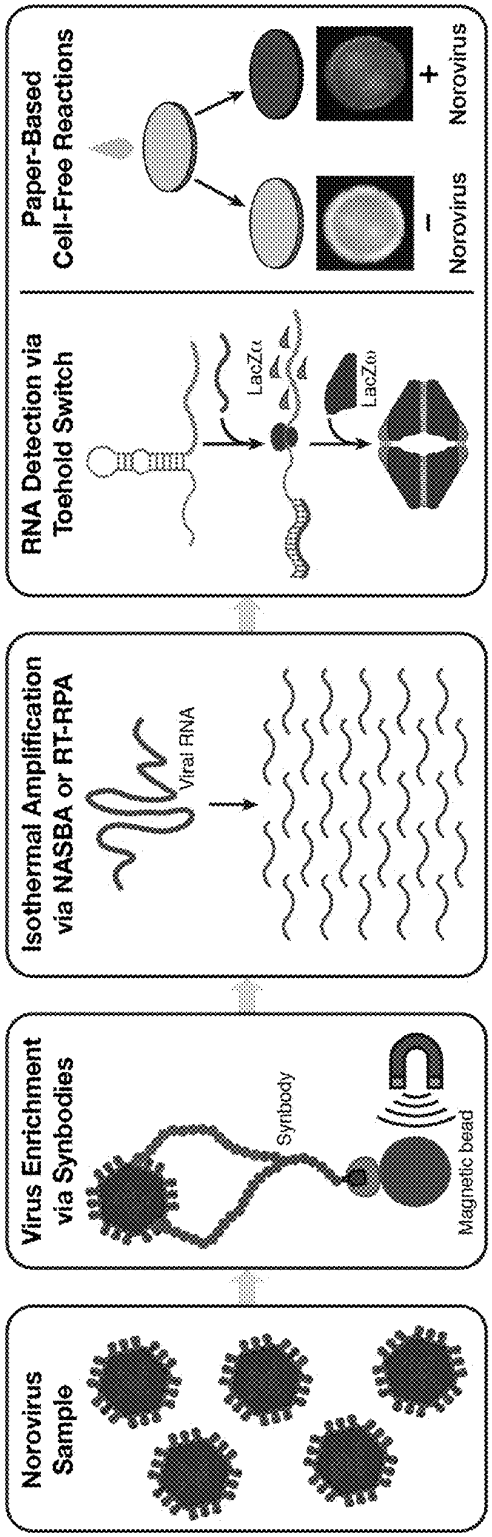
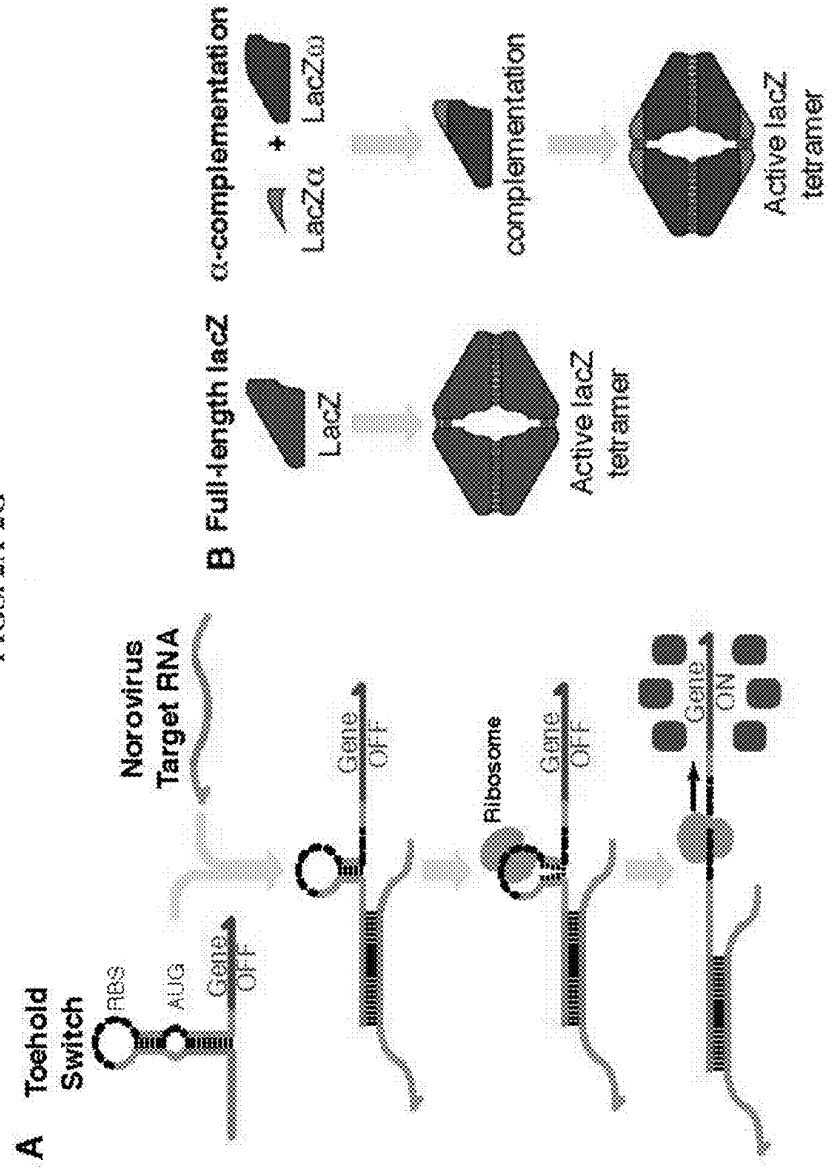
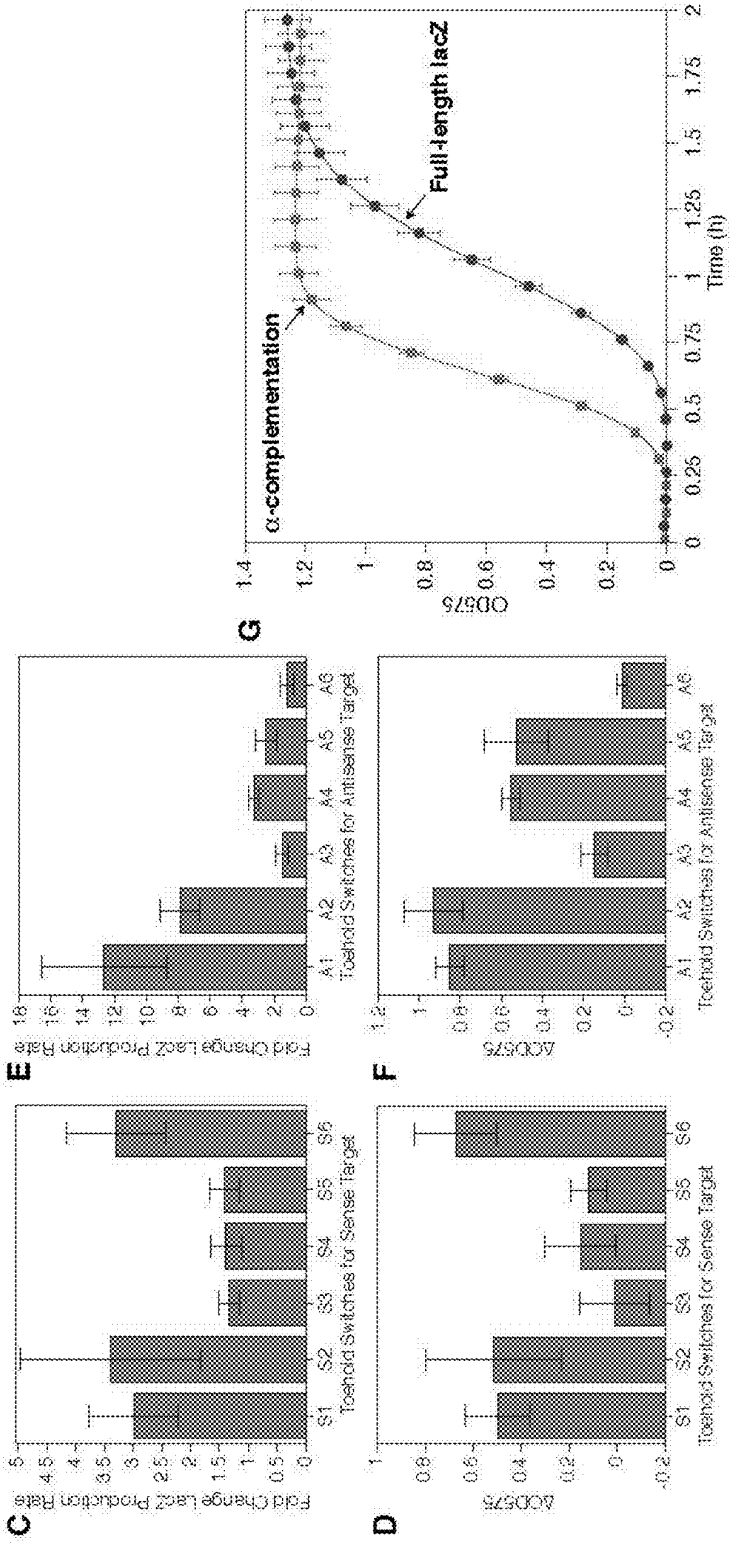

Norovirus Sample
Virus Enrichment via Synbodies
Synbody
Magnetic bead
Isothermal Amplification via NASBA or RT-RPA
Viral RNA
RNA Detection via Toehold Switch
LacZα
LacZω
Paper-Based Cell-Free Reactions
+
Norovirus
–
Norovirus

FIG. 1

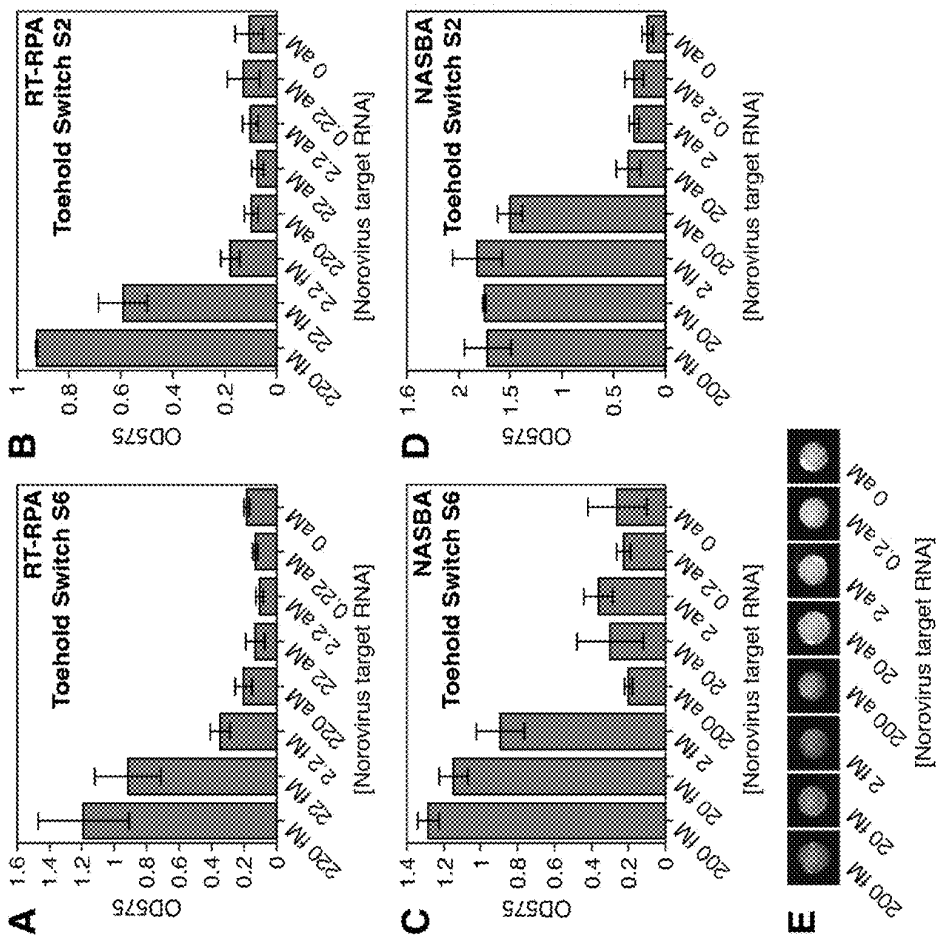
FIGS. 2A-2G

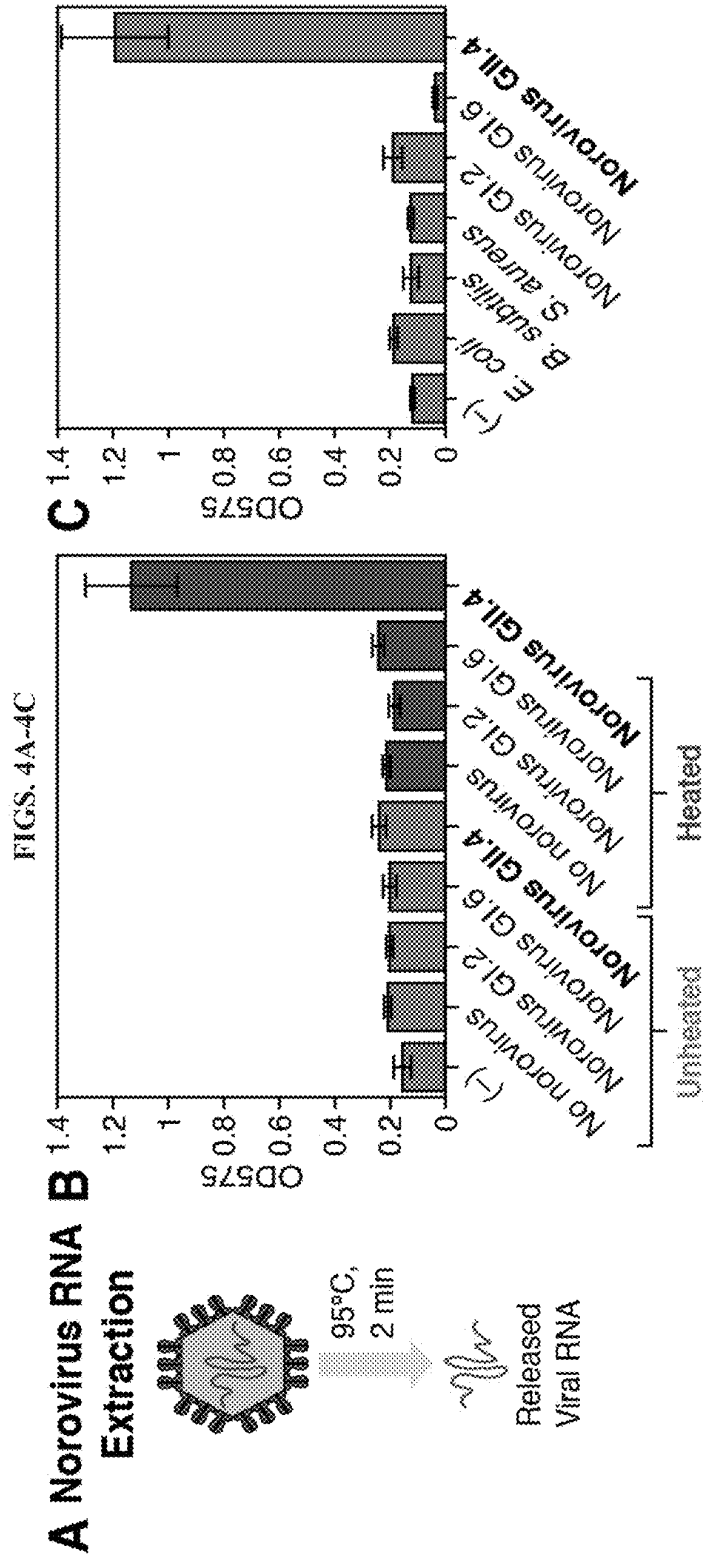

A
Toehold Switch
RBS
AUG
Gene OFF
Norovirus Target RNA
Gene OFF
Ribosome
Gene OFF
Gene ON
B Full-length lacZ
LacZ
Active lacZ tetramer
α-complementation
LacZα
+
LacZω
complementation
Active lacZ tetramer

FIGS. 2A-2G, CONTINUED



FIGS. 3A-3E


A
RT-RPA
Toehold Switch S6
OD575
[Norovirus target RNA]
220 fM
22 fM
2.2 fM
220 aM
22 aM
2.2 aM
0.22 aM
0 aM
B
RT-RPA
Toehold Switch S2
C
NASBA
Toehold Switch S6
D
NASBA
Toehold Switch S2
E
200 fM
20 fM
2 fM
200 aM
20 aM
2 aM
0.2 aM
0 aM


A Norovirus RNA Extraction
95°C, 2 min
Released Viral RNA
B
OD575
1.4
1.2
1
0.8
0.6
0.4
0.2
0
(-)
No norovirus
Norovirus GI.2
Norovirus GI.6
Norovirus GII.4
Unheated
Heated
C
OD575
(-)
E. coli
B. subtilis
S. aureus
Norovirus GI.2
Norovirus GI.6
Norovirus GII.4

FIGS. 4A-4C

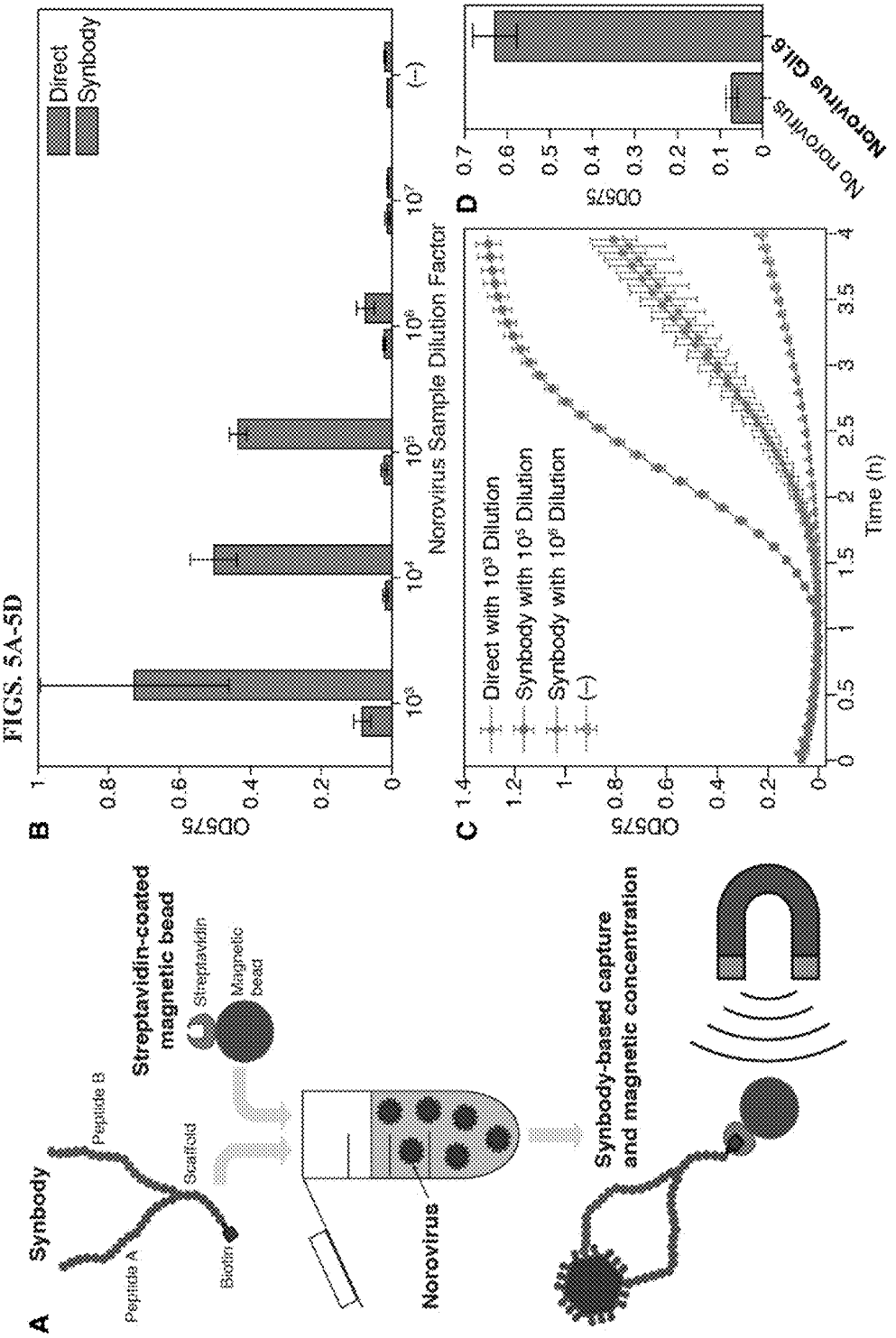
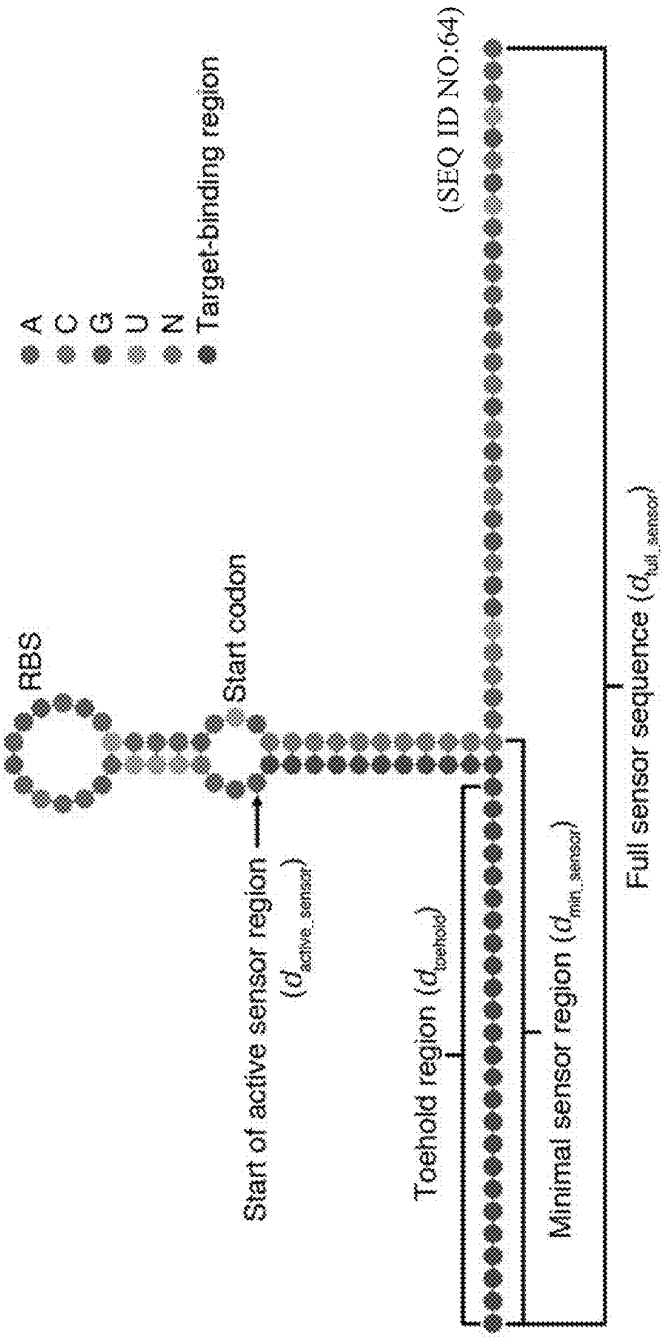
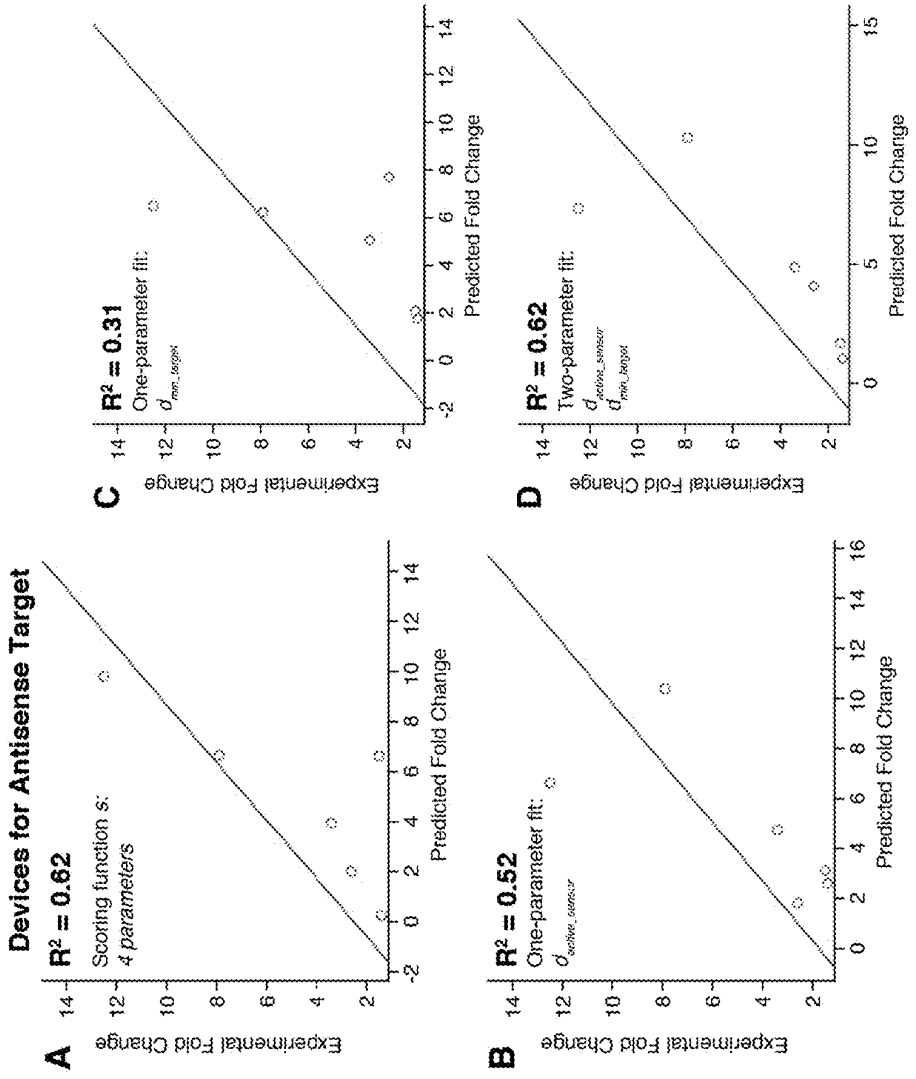

FIGS. 5A-5D
A
Synbody
Peptide B
Scaffold
Peptide A
Biotin
Streptavidin-coated magnetic bead
Streptavidin
Magnetic bead
Norovirus
Synbody-based capture and magnetic concentration
B
Direct
Synbody
OD575
Norovirus Sample Dilution Factor
(-)
C
Direct with 10³ Dilution
Synbody with 10⁵ Dilution
Synbody with 10⁶ Dilution
(-)
OD575
Time (h)
D
OD575
No norovirus
Norovirus GII.6

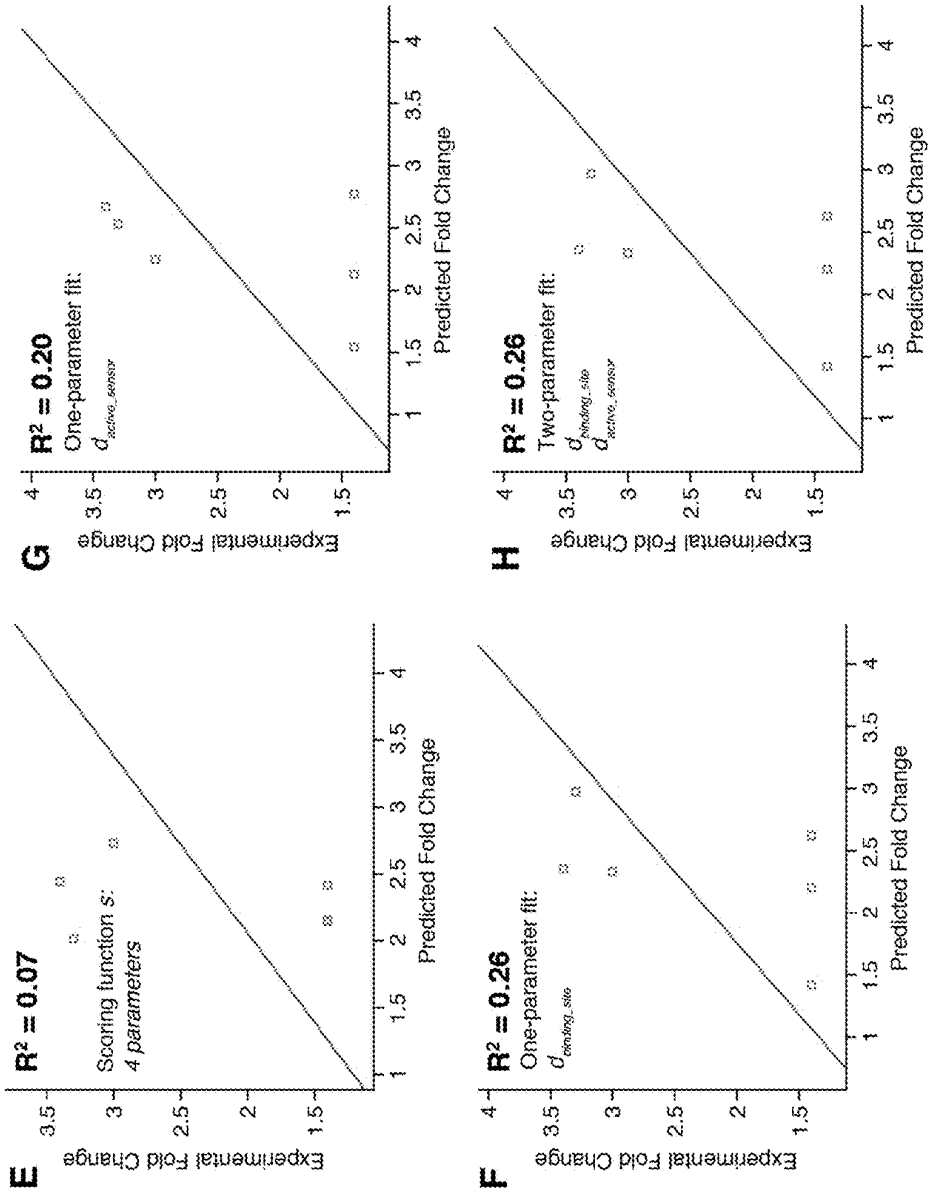
FIG. 6


A
C
G
U
N
Target-binding region
(SEQ ID NO:64)
RBS
Start codon
Start of active sensor region (d_{active_sensor})
Toehold region ($d_{toehold}$)
Minimal sensor region (d_{min_sensor})
Full sensor sequence (d_{full_sensor})

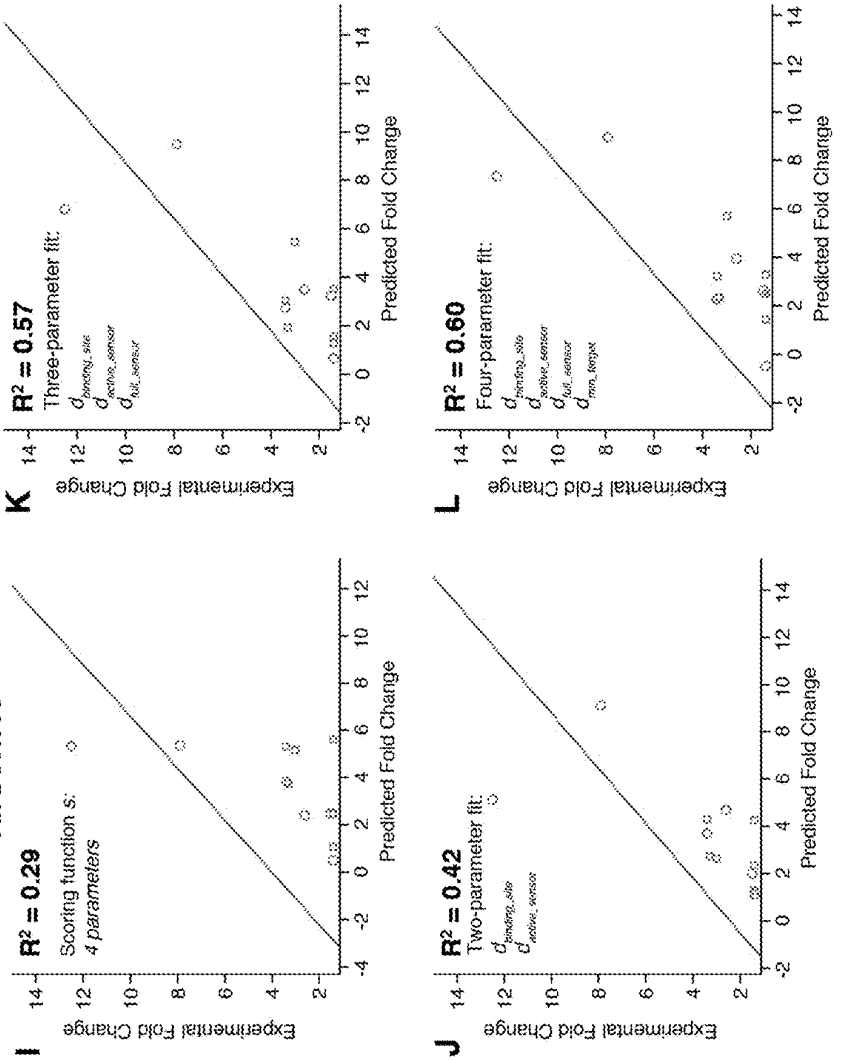
FIGS. 7A-7L


Devices for Antisense Target
A
R² = 0.62
Scoring function s:
4 parameters
Experimental Fold Change
Predicted Fold Change
B
R² = 0.52
One-parameter fit:
d_active_sensor
C
R² = 0.31
One-parameter fit:
d_min_target
D
R² = 0.62
Two-parameter fit:
d_active_sensor
d_min_target

FIGS. 7A-7L, CONTINUED



FIGS. 7A-7L, CONTINUED



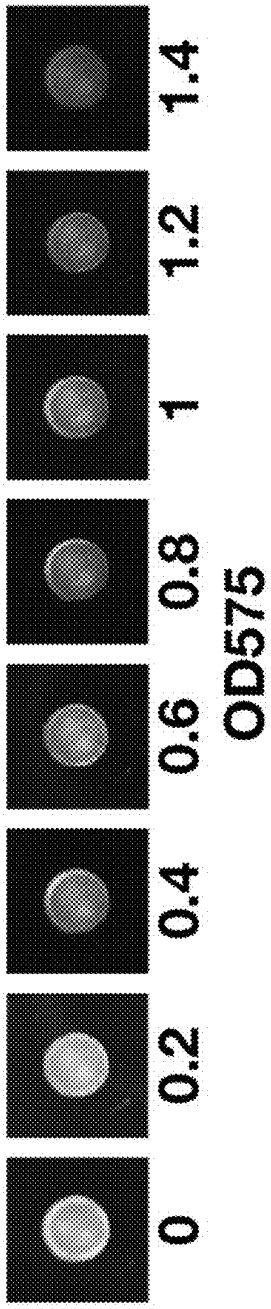


FIG. 8

FIGS. 9A-9D

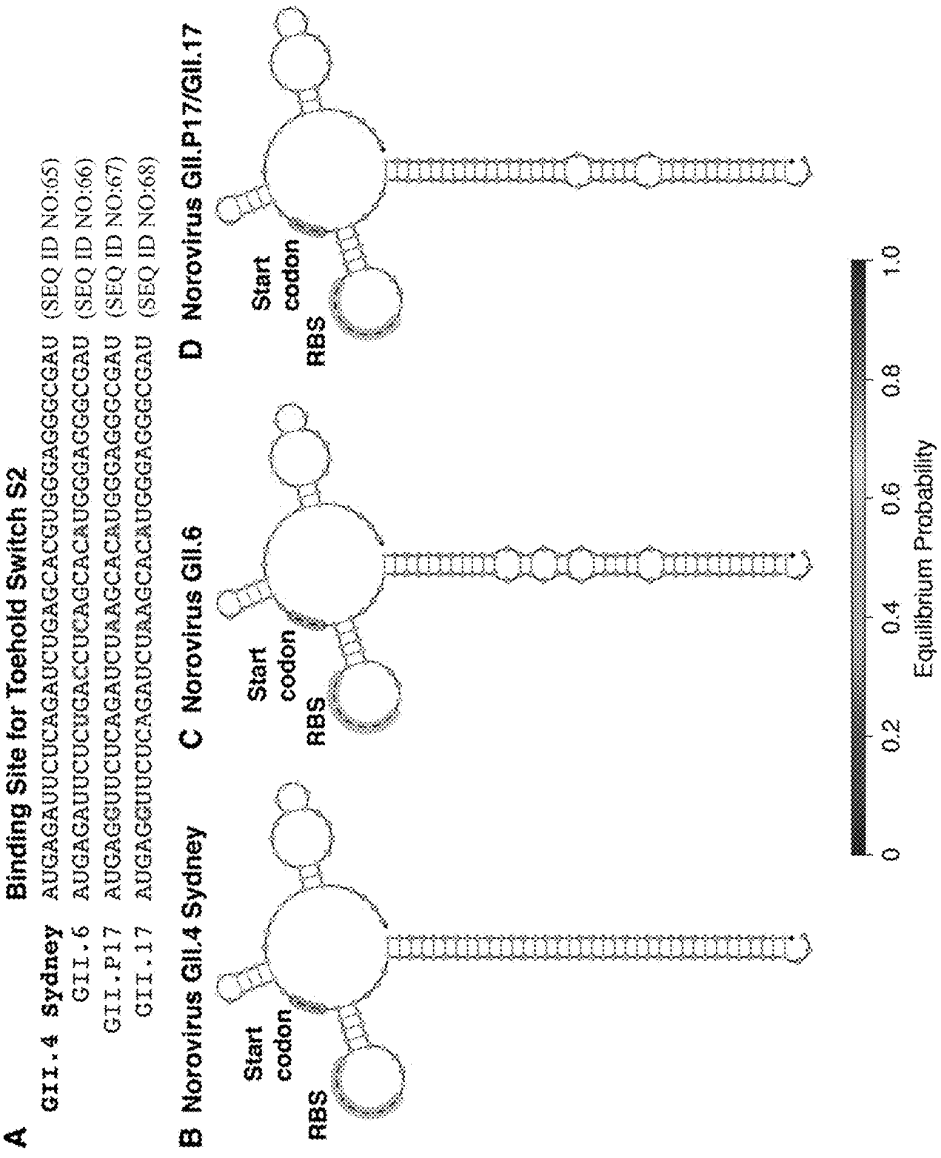

A
Binding Site for Toehold Switch S2
GII.4 Sydney AUGAGAUUCUCAGAUCUGAGCACGUGGGAGGGCGAU (SEQ ID NO:65)
GII.6 AUGAGAUUCUCUGACCUCAGCACACAUGGGAGGGCGAU (SEQ ID NO:66)
GII.P17 AUGAGGUUCUCAGAUCUAAGCACACAUGGGAGGGCGAU (SEQ ID NO:67)
GII.17 AUGAGGUUCUCAGAUCUAAGCACACAUGGGAGGGCGAU (SEQ ID NO:68)
B Norovirus GII.4 Sydney
C Norovirus GII.6
D Norovirus GII.P17/GII.17
Start codon
RBS
0
0.2
0.4
0.6
0.8
1.0
Equilibrium Probability

FIG. 10

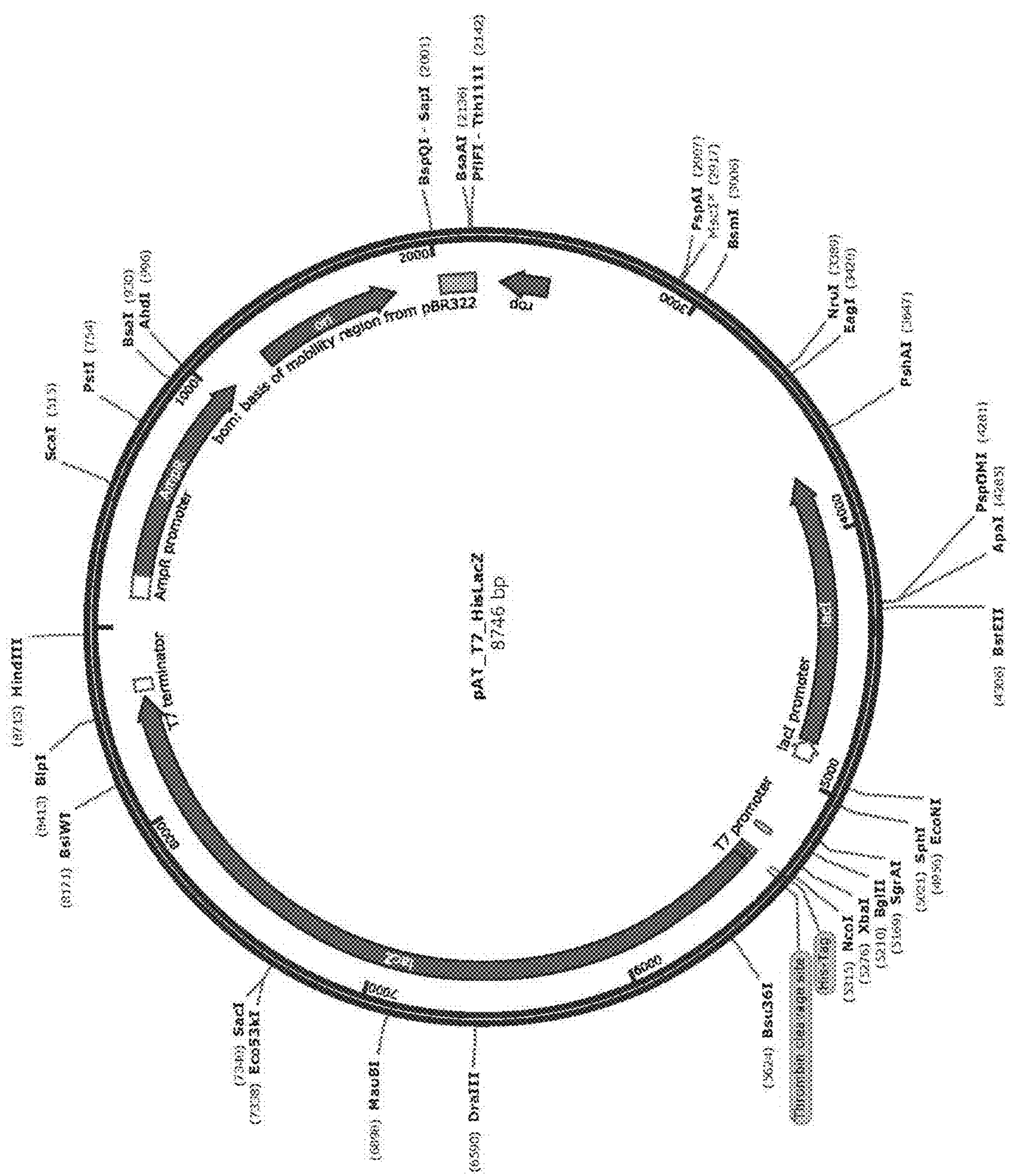

pAT_T7_HisLacZ
8746 bp
bom: basis of mobility region from pBR322
AmpR promoter
T7 terminator
T7 promoter
lacI promoter
rop

FIG. 11

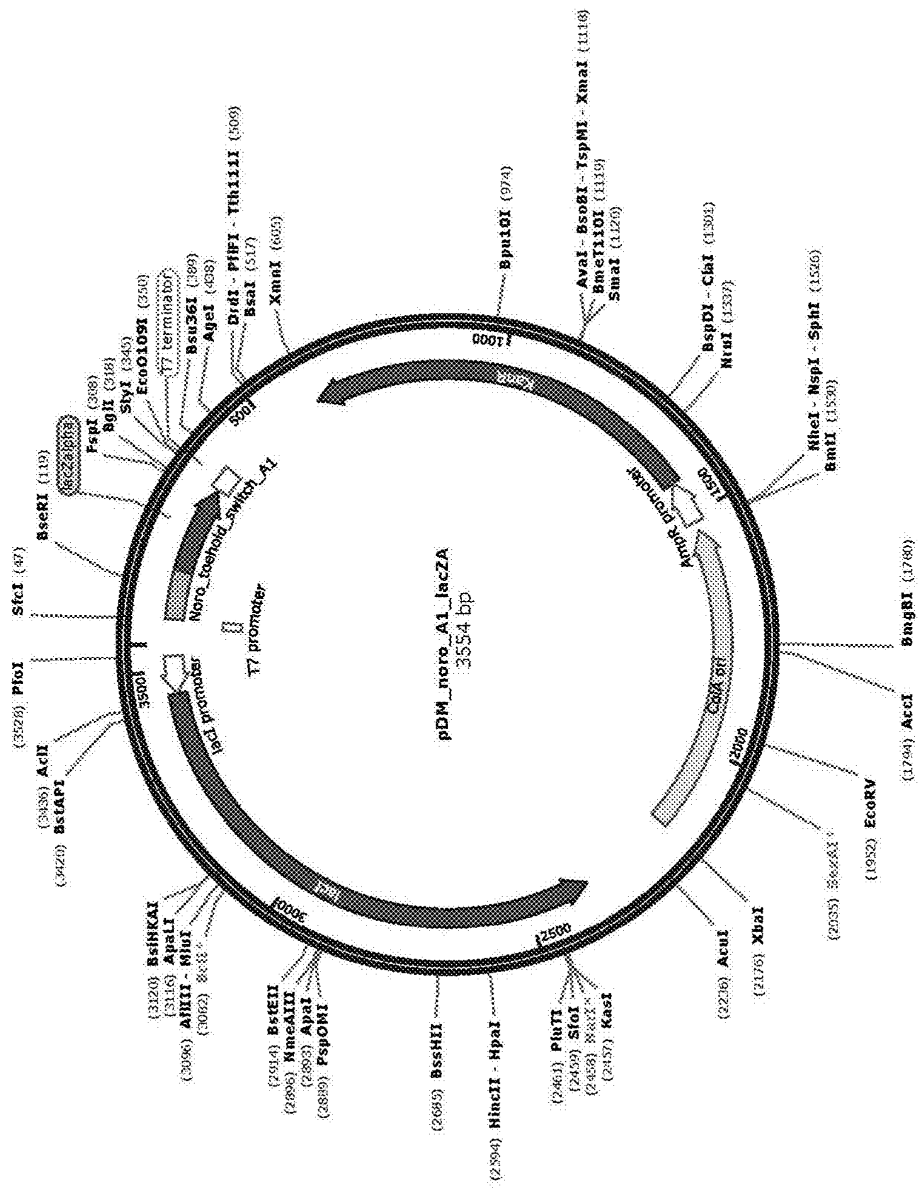

pDM_noro_A1_lacZA
3554 bp
Noro_toehold_switch_A1
T7 promoter
lacI promoter
AmpR promoter
ColA ori
T7 terminator
lacZalpha
500
1000
1500
2000
2500
3000
3500

FIG. 12

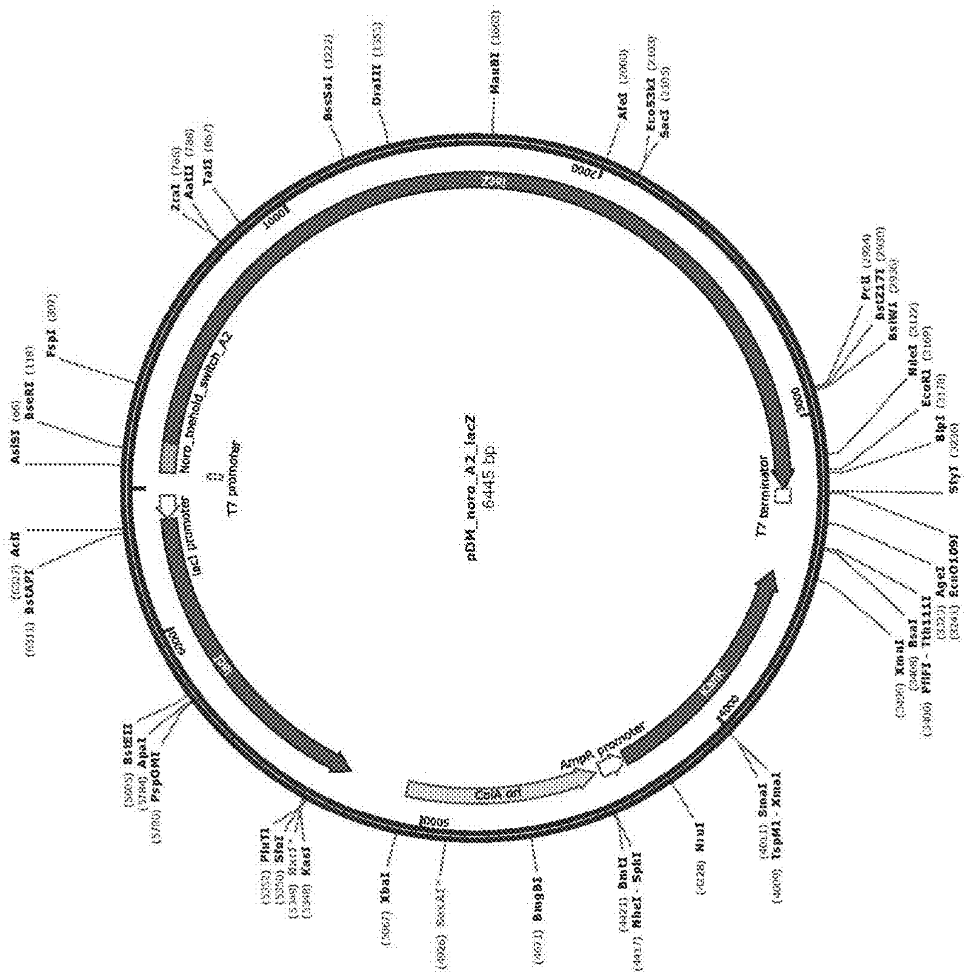

pDM_noro_A2_lacZ
6445 bp
T7 promoter
lacI promoter
T7 terminator
AmpR promoter
Noro_toehold_switch_A2

FIG. 13

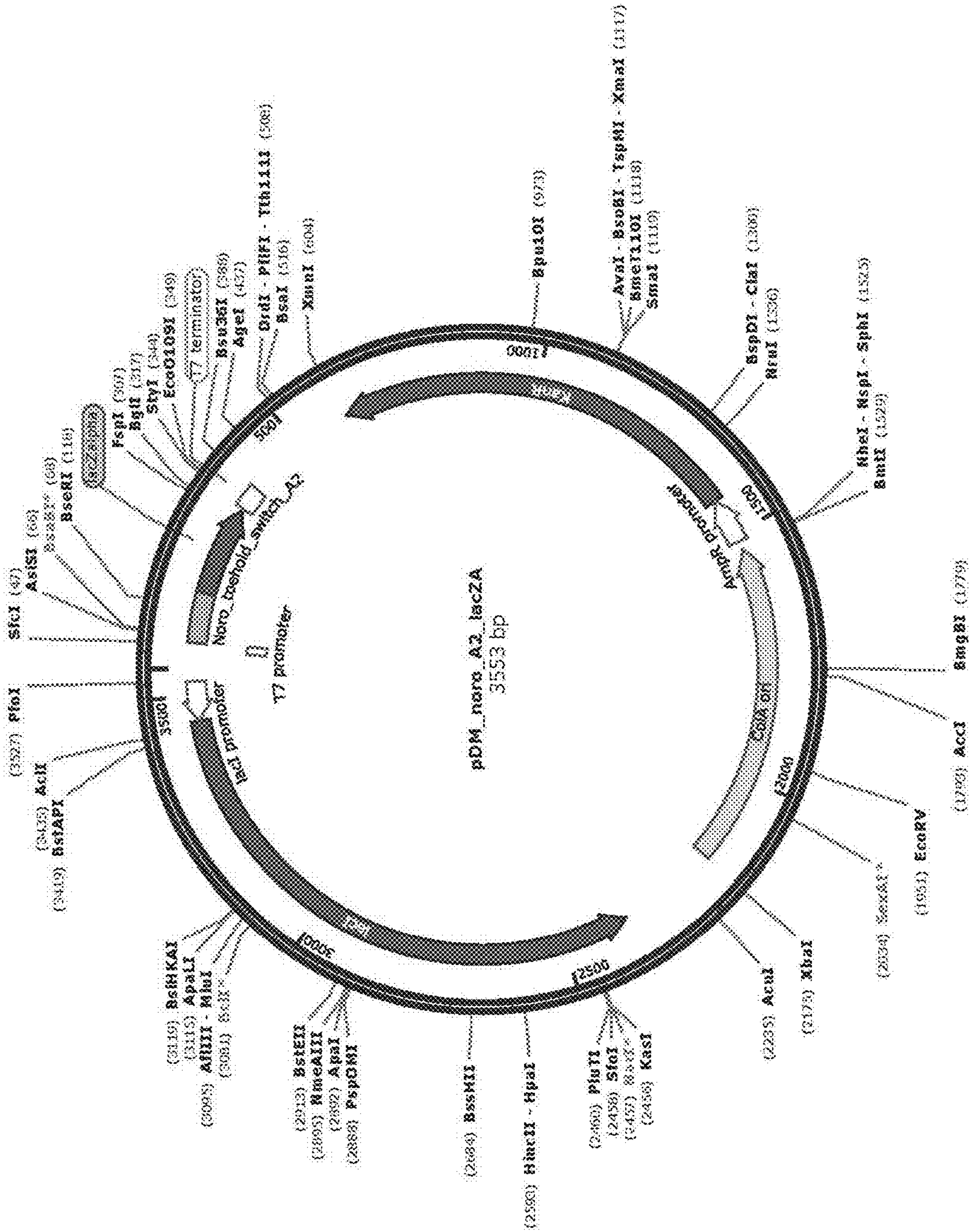

pDM_noro_A2_lacZA
3553 bp
Noro_toehold_switch_A2
T7 promoter
T7 terminator
lacI promoter
lacI
KanR
AmpR promoter
ColE1 ori
SfcI (47)
AsiSI (66)
BsaBI* (68)
BseRI (118)
FspI (307)
BglI (317)
StyI (344)
EcoO109I (349)
Bsu36I (388)
AgeI (457)
DrdI - PflFI - Tth111I (508)
BsaI (516)
XmnI (604)
Bpu10I (973)
AvaI - BsoBI - TspMI - XmaI (1117)
SmaI (1119)
BspDI - ClaI (1306)
NruI (1336)
NheI - NspI - SphI (1525)
BmtI (1529)
BsmBI (1779)
AccI (1793)
EcoRV (1951)
XbaI (2175)
AcuI (2235)
PluTI (2460)
SfoI (2458)
KasI (2456)
BssHII (2684)
HincII - HpaI (2593)
BstEII (2913)
RmeAIII (2895)
ApaI (2892)
PspOMI (2888)
BciHKAI (3119)
ApaLI (3115)
AflIII - MluI (3095)
BstAPI (3419)
AclI (3435)
PfoI (3527)
500
1000
1500
2000
2500
3000
3500

FIG. 14

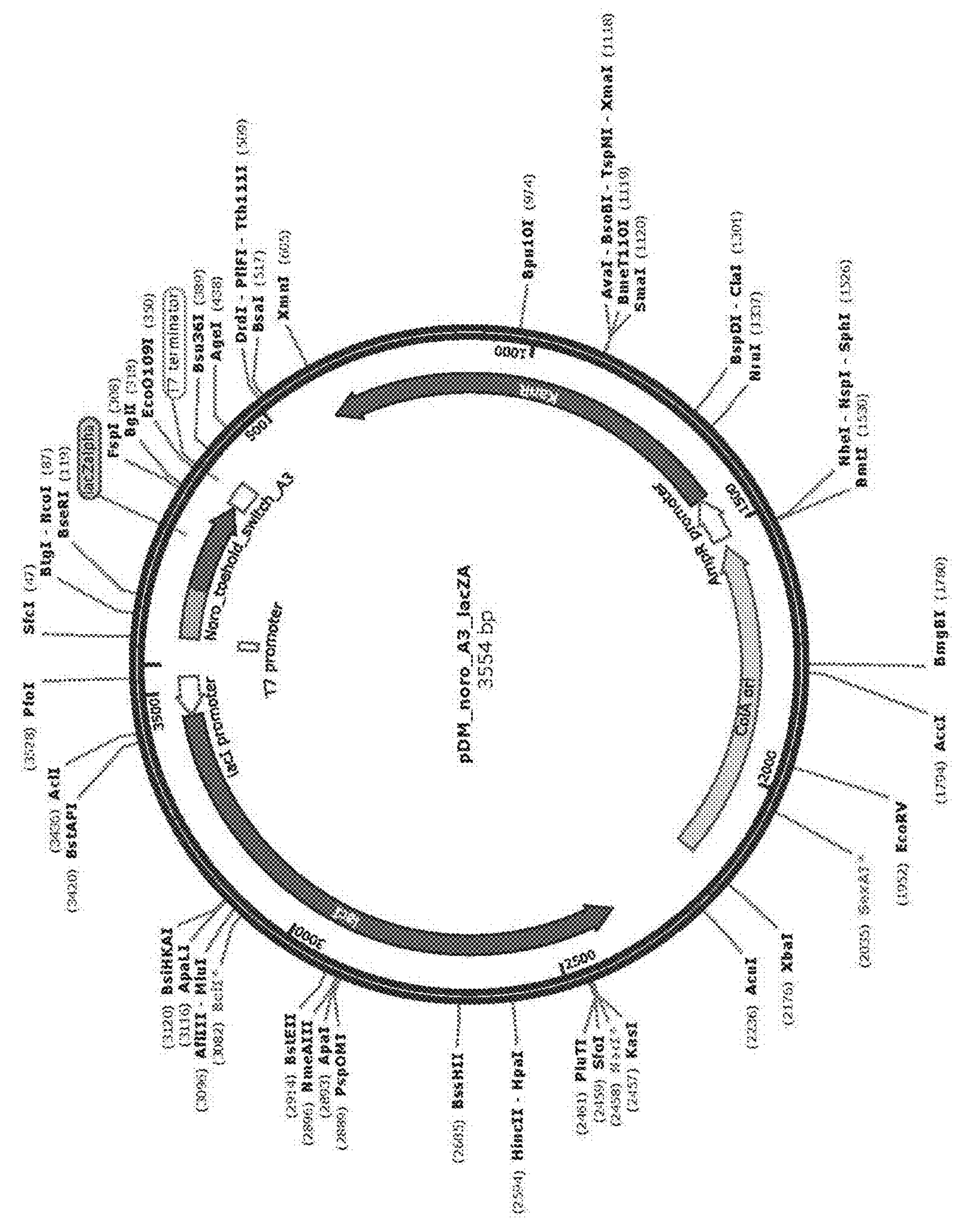

pDM_noro_A3_lacZA
3554 bp
T7 promoter
lacI promoter
AmpR promoter
ColE1 ori
T7 terminator

FIG. 15

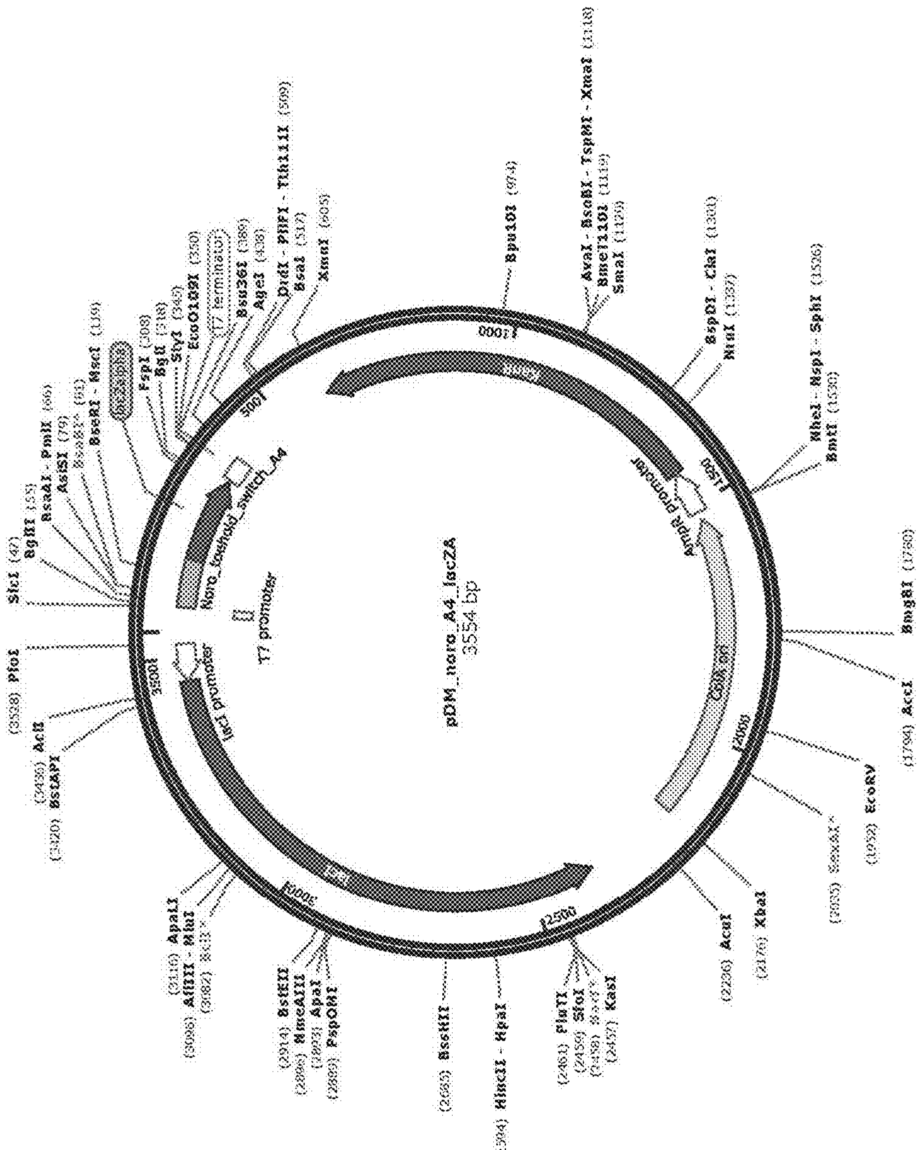

pDM_noro_A4_lacZA
3554 bp
T7 promoter
lacI promoter
AmpR promoter
Noro_toehold_switch_A4
ColE ori
Bpu10I (974)
BmgBI
AccI
EcoRV
XbaI
HincII - HpaI
BstAPI
PfoI
SfcI (42)

FIG. 16

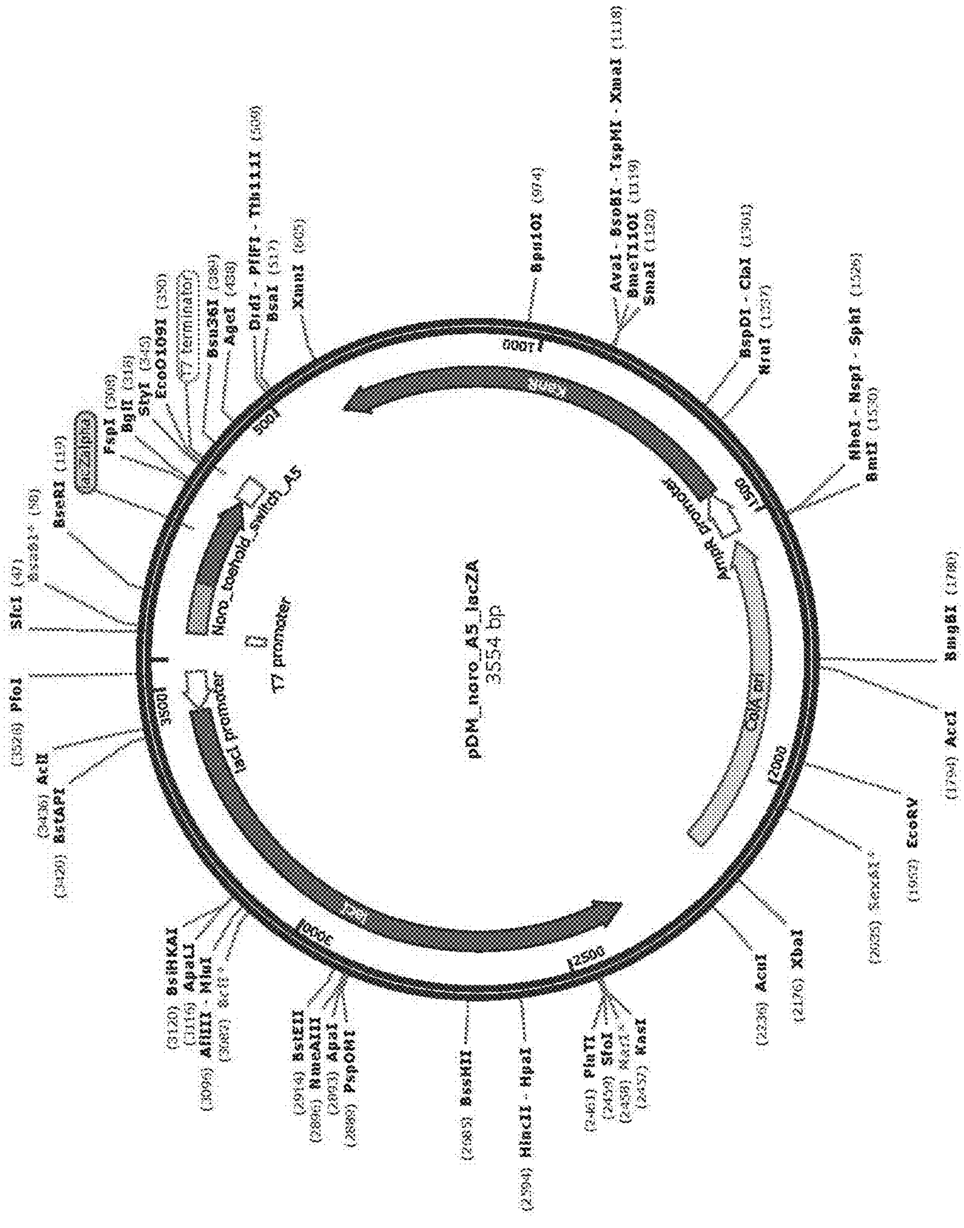

pDM_noro_A5_lacZA
3554 bp
T7 promoter
Noro_toehold_switch_A5
T7 terminator
KanR
AmpR promoter
ColA ori
lacI
lacI promoter
500
1000
1500
2000
2500
3000
3500
SfcI (47)
BsaBI* (58)
BseRI (113)
FspI (308)
BglI (318)
StyI (345)
EcoO109I (350)
Bsu36I (389)
AgeI (438)
DrdI - PflFI - Tth111I (509)
BsaI (517)
XmnI (605)
Bpu10I (974)
AvaI - BsoBI - TspMI - XmaI (1118)
BmeT110I (1119)
SmaI (1120)
BspDI - ClaI (1301)
NruI (1337)
NheI - NspI - SphI (1526)
BmtI (1530)
BsmBI (1780)
AccI (1794)
EcoRV (1962)
SexAI* (2025)
XbaI (2176)
AcuI (2336)
PluTI (2461)
SfoI (2459)
NarI* (2458)
KasI (2457)
HincII - HpaI (2594)
BssHII (2685)
BstEII (2914)
NmeAIII (2896)
ApaI (2893)
PspOMI (2889)
BsiHKAI (3120)
ApaLI (3116)
AflIII - MluI (3096)
BclI* (3082)
BstAPI (3429)
AclI (3436)
PfoI (3526)

FIG. 17

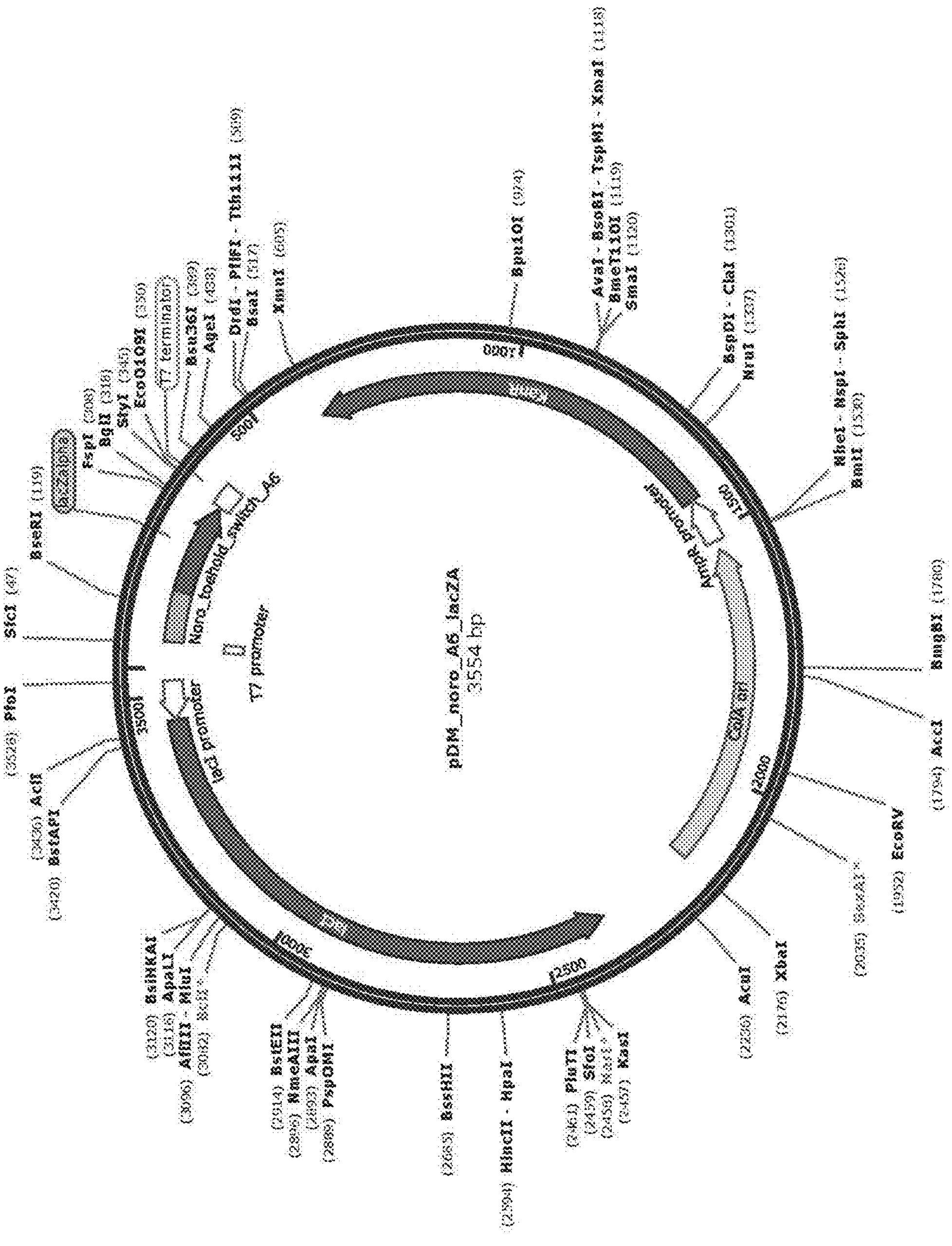

pDM_noro_A6_lacZA
3554 bp
T7 promoter
Noro_toehold_switch_A6
lacZalpha
T7 terminator
lacI promoter
AmpR promoter
ColA ori
SfcI (47)
BseRI (119)
FspI (308)
BglI (318)
StyI (345)
AgeI (438)
BsaI (517)
XmnI (605)
Bpu10I (974)
SmaI (1120)
BspDI - ClaI (1301)
NruI (1337)
NheI - NspI - SphI (1526)
BmtI (1530)
BsmBI (1786)
AccI (1794)
EcoRV (1952)
AciI (2226)
XbaI (2176)
KasI (2457)
SfoI (2459)
PluTI (2461)
BssHII (2685)
HincII - HpaI (2794)
ApaI (2893)
PspOMI (2889)
BstEII (2914)
AflIII - MluI (3096)
ApaLI (3118)
BsiHKAI (3120)
BstAPI (3420)
AclI (3436)
PfoI (3526)

FIG. 18

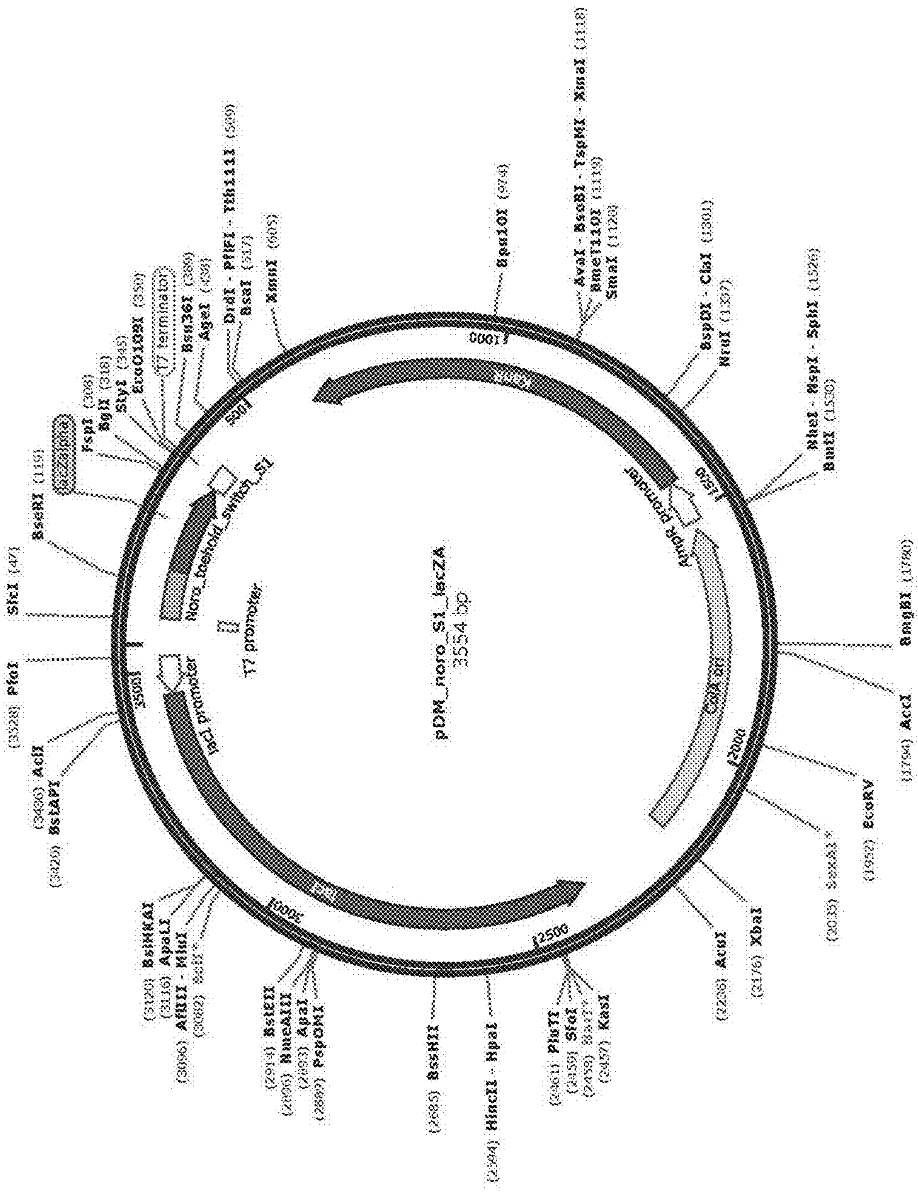

pDM_noro_S1_lacZA
3554 bp
T7 promoter
Noro_toehold_switch_S1
T7 terminator
lacI promoter
AmpR promoter
ColE1 ori
BssHI (139)
SfcI (47)
PfoI (3528)
AclI (3436)
BstAPI (3420)
FspI (308)
BglI (318)
StyI (348)
EcoO109I (350)
Bsu36I (389)
AgeI (430)
DrdI - PflFI - Tth111I (589)
BsaI (587)
XmnI (605)
Bpu10I (974)
AvaI - BsoBI - TspMI - XmaI (1118)
BmeT110I (1119)
SmaI (1120)
BspDI - ClaI (1301)
NruI (1307)
NheI - NspI - SphI (1526)
BmtI (1530)
BmgBI (1790)
AccI (1794)
EcoRV (1952)
XbaI (2176)
AcuI (2236)
PluTI (2461)
SfoI (2459)
KasI (2457)
HincII - HpaI (2564)
BssHII (2685)
BstEII (2914)
NmeAIII (2896)
ApaI (2893)
PspOMI (2889)
BsiHKAI (3120)
ApaLI (3116)
AflIII - MluI (3096)

FIG. 19

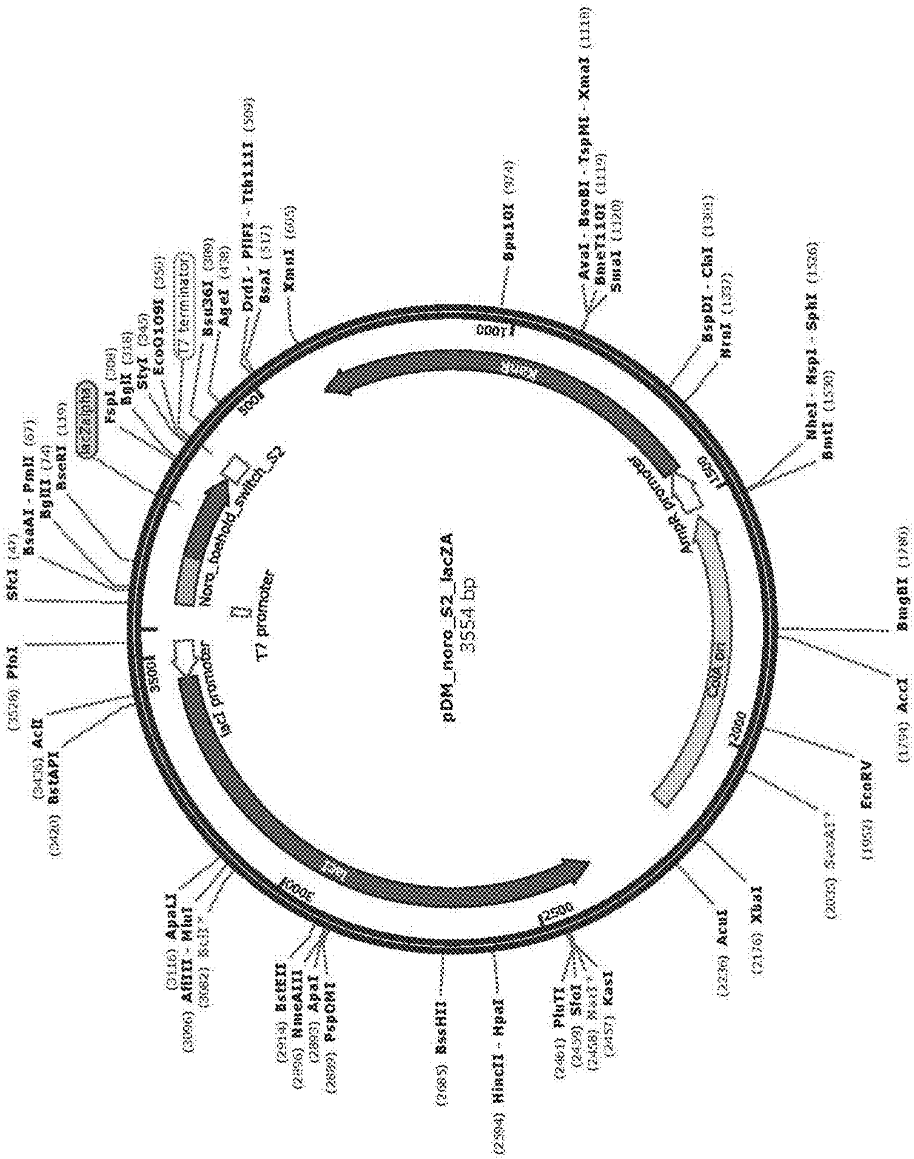

pDM_noro_S2_lacZA
3554 bp
T7 promoter
Noro_toehold_switch_S2
T7 terminator
lacI promoter
lacI
AmpR promoter
XbaI
EcoRV
AccI
KasI
HpaI
BssHII
ApaI
PfoI
SfcI
FspI
StyI
AgeI
XmnI
Bpu10I
SmaI
ClaI
NruI
SphI
BmgBI

FIG. 20

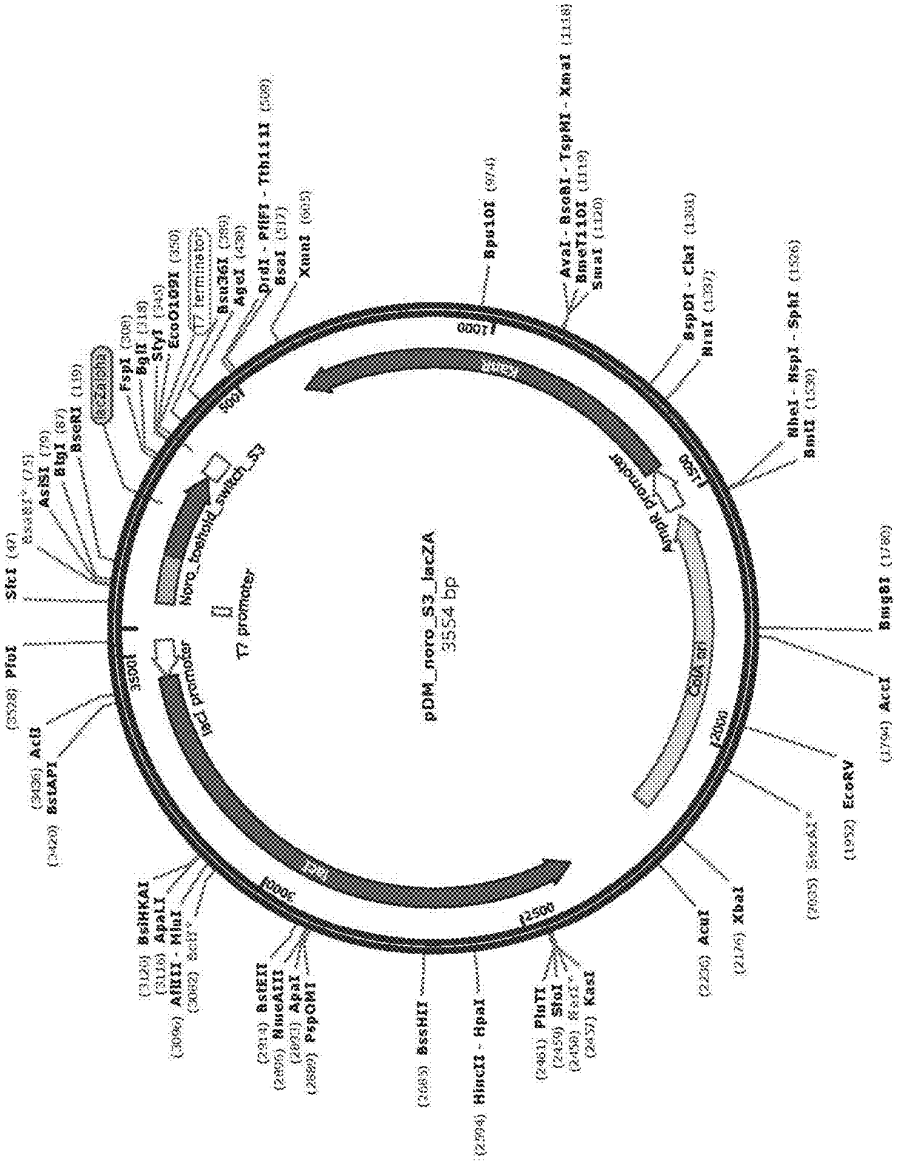
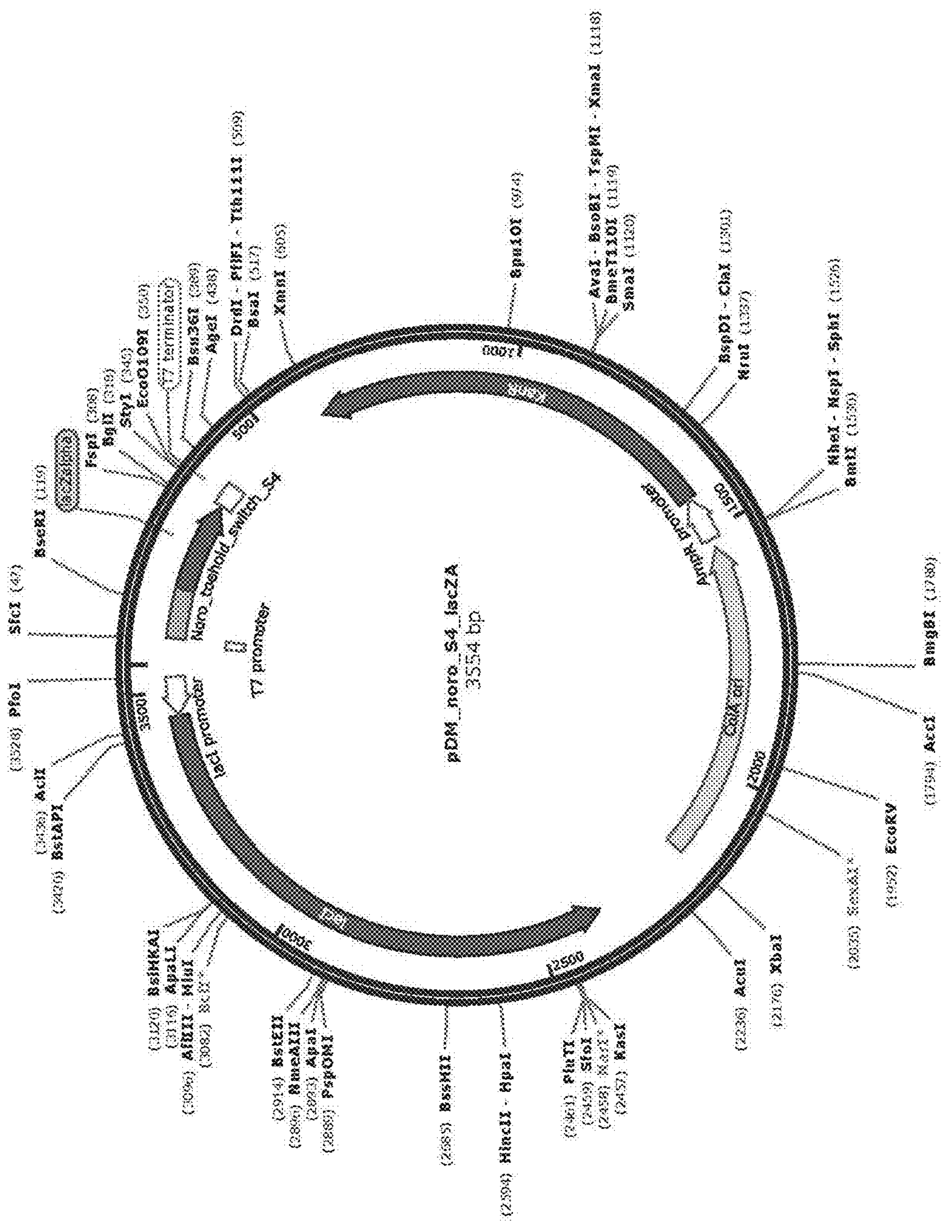

pDM_noro_S3_lacZA
3554 bp
T7 promoter
Noro_bestmold_switch_S3
lacI promoter
lacI
AmpR promoter
AmpR
ColE1 ori
EcoRV
AccI
BmgBI
NheI - NspI - SphI
BmtI
BspDI - ClaI
NruI
AvaI - BsoBI - TspMI - XmaI
BmeT110I
SmaI
BpuIOI
XmnI
DrdI - PflFI - Tth111I
BsaI
AgeI
Bsu36I
T7 terminator
Eco0109I
StyI
BglII
FspI
BseRI
BtgI
AsiSI
SbfI
PfoI
ActI
BstAPI
BsiWKAI
ApaLI
AflIII - MluI
BstEII
BsmAI
ApaI
PspOMI
BssHII
HincII - HpaI
PluTI
SbfI
KasI
AccI
XbaI
EcoRV
1000
2000
2500
3000
3500

FIG. 21



pDM_noro_S4_lacZA
3554 bp
T7 promoter
Noro_toehold_switch_S4
T7 terminator
AmpR promoter
lacI promoter
lacI
SfcI (47)
BseRI (119)
FspI (308)
BglI (319)
StyI (340)
EcoO109I (358)
Bsu36I (389)
AgeI (438)
DrdI - PflFI - Tth111I (509)
BsaI (517)
XmnI (805)
AvaI - BsoBI - TspMI - XmaI (1118)
BmeT110I (1119)
SmaI (1120)
BspDI - ClaI (1301)
NruI (1337)
NheI - NspI - SphI (1526)
BmtI (1530)
BmgBI (1780)
AccI (1794)
EcoRV (1952)
XbaI (2176)
AcuI (2236)
KasI (2457)
PluTI (2461)
HincII - HpaI (2594)
BstEII (2914)
AflIII - MluI (3096)
ApaLI (3116)
BstAPI (3426)
AclI (3436)

FIG. 22

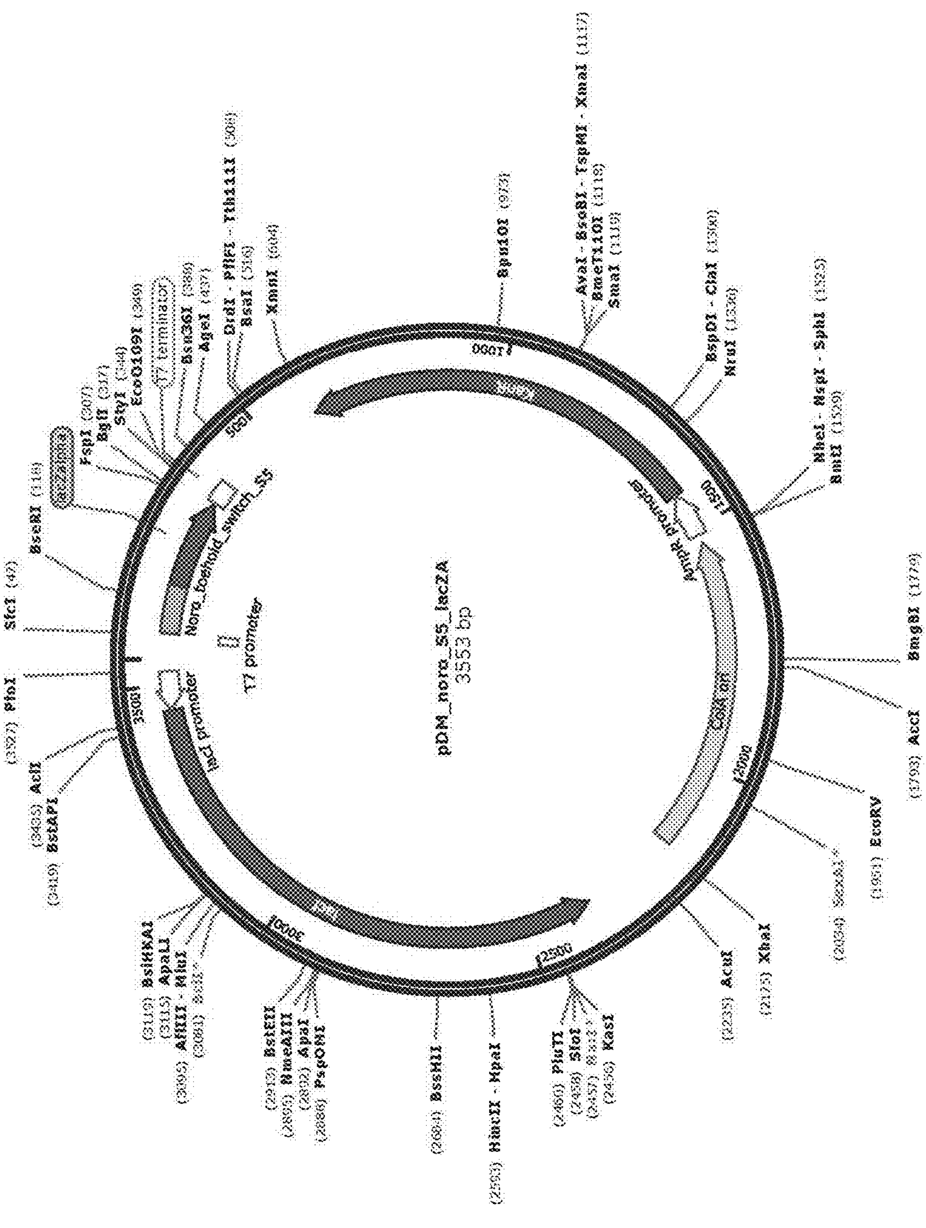

pDM_nora_S5_lacZA
3553 bp
T7 promoter
lacI promoter
Nora_toehold_switch_S5
AmpR promoter
T7 terminator
SfcI (49)
BseRI (118)
AccI (1793)
EcoRV
XbaI
BssHII
SpoI
ClaI
NruI

FIG. 23

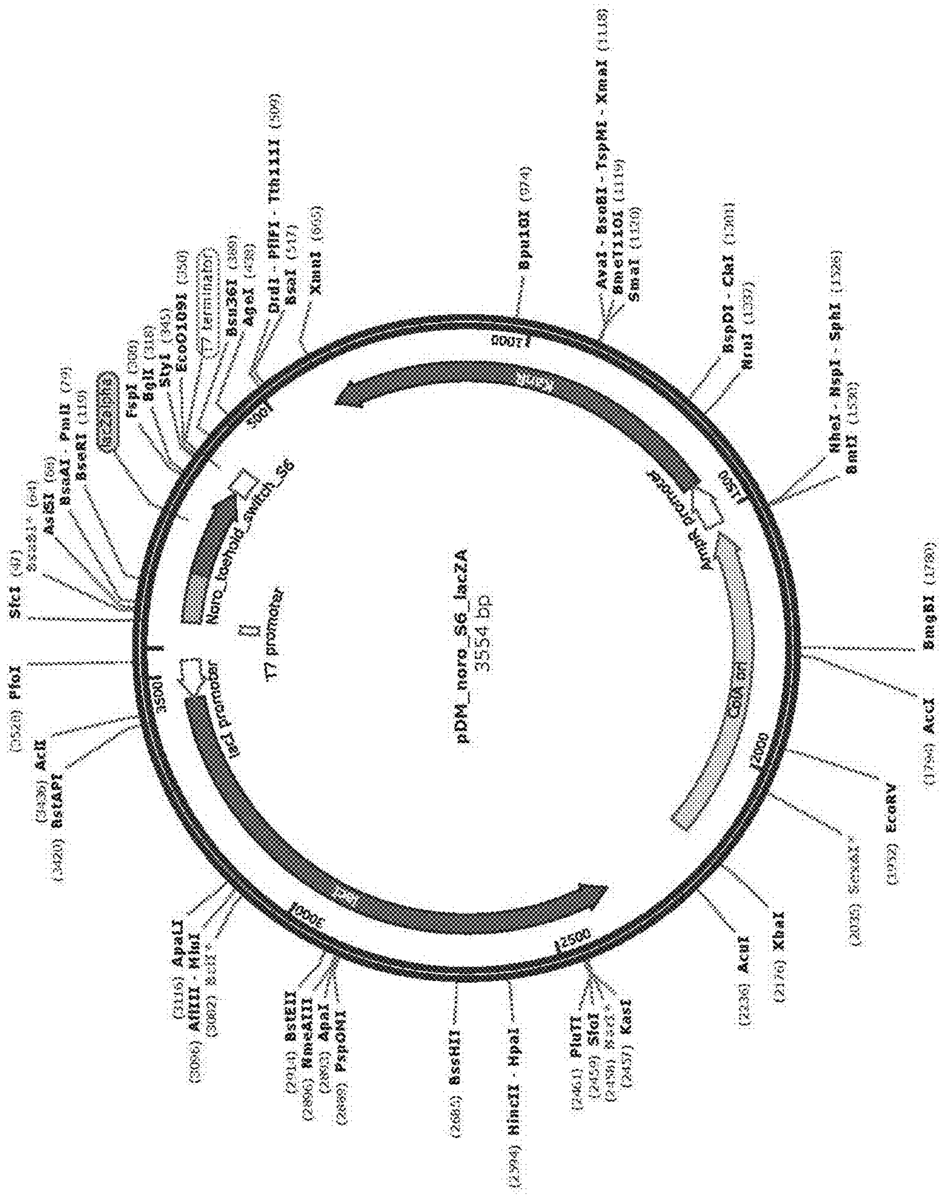
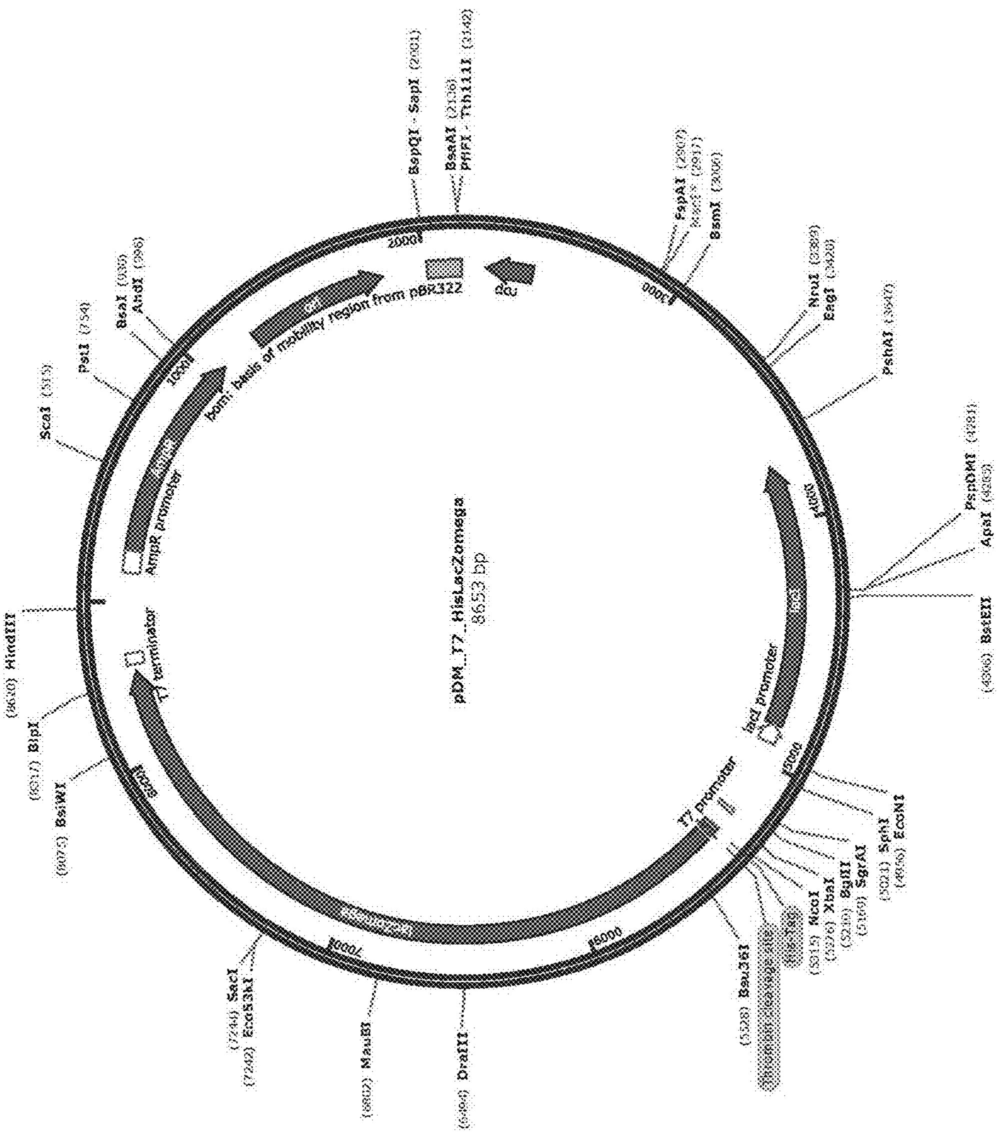

pDM_noro_S6_lacZA
3554 bp
T7 promoter
lacI promoter
Noro_toehold_switch_S6
AmpR promoter

FIG. 24


pDM_T7_HisLacZomega
8653 bp
AmpR promoter
bom: basis of mobility region from pBR322
T7 terminator
T7 promoter
lacI promoter
BspQI - SapI (2101)
BsaAI (2136)
PflFI - Tth111I (2142)
FspAI (2907)
BsmI (3006)
NruI (3389)
EagI (3420)
PshAI (3947)
PspOMI (4281)
ApaI (4283)
BstEII (4306)
EcoNI (4956)
SphI (5021)
SgrAI (5169)
BglII (5239)
XbaI (5276)
NcoI (5315)
Bsu36I (5328)
DraIII (6494)
MauBI (6802)
SacI (7244)
Eco53kI (7242)
BsiWI (8075)
BlpI (8217)
HindIII (8620)
ScaI (515)
PstI (754)
BsaI (935)
AhdI (996)

LOW-COST DETECTION OF NOROVIRUS USING PAPER-BASED CELL-FREE SYSTEMS AND SYNBODY-BASED VIRAL ENRICHMENT

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 62/632,792, filed Feb. 20, 2018, which is incorporated in its entirety herein by reference.

STATEMENT REGARDING FEDERALLY FUNDED RESEARCH OR DEVELOPMENT

[0002] This invention was made with government support under GM126892 awarded by the National Institutes of Health and ADHS16-162400 awarded by USDA/NIFA. The government has certain rights in the invention.

BACKGROUND

[0003] Noroviruses are a primary cause of gastroenteritis and foodborne illness with cases that affect millions of people worldwide each year. At present, the most commonly used methods to detect norovirus are immunochromatographic lateral flow assays and PCR-based tests. Immunochromatographic lateral flow assays employ antibodies that recognize viral surface proteins and can be advantageous since they do not require specialized equipment and provide test results in 15 minutes. However, these tests provide limited sensitivity and their results can be strongly genotype dependent. Real-time quantitative reverse transcriptase PCR (qRT-PCR) is currently the gold standard for detection of norovirus. qRT-PCR assays can be targeted to conserved regions of the norovirus genome and they provide high specificity and sensitivity. These assays, however, require expensive thermal cycling equipment and are typically run in centralized laboratories. Shipment of samples can delay test results and specialized laboratory equipment is often not available in developing countries or in remote settings, such as ships at sea, where outbreaks frequently occur. Highly automated commercial instruments such as the Cepheid GeneXpert have been developed for decentralized use. However, these instruments are expensive. Even with negotiated prices for low- and middle-income countries, the GeneXpert instrument, for instance, costs $17,000 and has cartridges available at a concessional price of $9.98. These factors lead to an overall per test cost of $14.93 once labor, consumable, and other costs are included. In the absence of discounts, GeneXpert costs rise substantially to $30.26-$155.44 per test depending on the country. Accordingly, there remains a need in the art for inexpensive tests for pathogens such as norovirus that do not require sophisticated laboratory equipment and that provide timely results for disease containment.

SUMMARY

[0004] This disclosure is related to methods of detecting pathogen infection using paper-based cell-free transcription-translation reactions. More particularly, the embodiments provided herein relate to methods in which a sample is enriched for the pathogen using synbodies and isothermal amplification followed by detection of pathogen nucleic acids using sequence-specific toehold switches and cell-free transcription and translation reactions.

[0005] In a first aspect, provided herein is a method of detecting a target pathogen nucleic acid in a sample. The method can comprise or consist essentially of the steps of (a) contacting a biological sample obtained from a subject to a pathogen detection agent under conditions that promote binding of the pathogen detection agent to the target pathogen if present in the sample; (b) isolating nucleic acids from pathogen bound by the pathogen detection agent; (c) amplifying the isolated nucleic acids using isothermal amplification; and (d) contacting the amplified nucleic acid to a toehold switch, where the toehold switch encodes at least a portion of a reporter protein and comprises one or more single-stranded toehold sequence domains that are complementary to a target pathogen nucleic acid or the reverse complement thereof, where the contacting occurs under conditions that allow translation of the coding domain in the presence of the target nucleic acid but not in the absence of the target nucleic acid, and detecting the reporter protein as an indicator that the target pathogen nucleic acid is present in the amplified nucleic acids. The pathogen detection agent can be a norovirus detection agent and the target pathogen nucleic acid is norovirus RNA. The norovirus detection agent can be a synbody. The synbody can comprise biotin. The biotin-containing synbody can be bound to a streptavidin-coated magnetic bead. Isolating can comprise a magnetic capture assay. The toehold switch can encode at least a portion of lacZ. The toehold switch can encode lacZα and the amplified nucleic acids are contacted under conditions which promote formation of a lacZ tetramer. LacZω can be provided on a substrate to which the amplified nucleic acids are contacted. Target pathogen nucleic acid can be detected at concentrations in a range of zeptomoles/liter (zM). Target pathogen nucleic acid can be detected at concentration between about 270 zM to about 270 aM.

[0006] In another aspect, provided herein is a device for identifying a pathogen-associated nucleic acid, comprising a preserved paper test article, where a method provided herein is performed using the preserved paper test article. The paper test article can be preserved by freeze-drying.

[0007] In another aspect, provided herein is a synthetic norovirus-specific toehold switch sensor comprising a fully or partially double-stranded stem domain, a loop domain, a toehold domain, and at least a portion of a coding sequence of a reporter gene, where the toehold domain and at least a portion of the stem domain are complementary to a target norovirus RNA sequence. The sensor can comprise a RNA sequence selected from SEQ ID NOs:1-12.

[0008] In a further aspect, provided herein is a kit for detecting a pathogen-associated nucleic acid, comprising a plurality of preserved paper test articles, a pathogen detection agent, a plurality of toehold switches that encode at least a portion of a reporter protein and comprise one or more single-stranded toehold sequence domains that are complementary to a target pathogen nucleic acid or the reverse complement thereof, and an electronic optical reader. The pathogen detection agent can be a synbody.

[0009] In another aspect, provided herein is a kit for detecting a pathogen-associated nucleic acid, comprising a plurality of preserved test tube test articles, a pathogen detection agent, a plurality of toehold switches that encode at least a portion of a reporter protein and comprise one or more single-stranded toehold sequence domains that are complementary to a target pathogen nucleic acid or the reverse complement thereof, and an electronic optical

reader. The pathogen detection agent can be a synbody. In some cases, the kit further comprises instructions for performing a method as provided herein.

INCORPORATION BY REFERENCE

[0010] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference in their entirety as if each individual publication, patent, and patent application was specifically and individually indicated to be incorporated by reference.

[0011] This application includes a sequence listing in computer readable form (a "txt" file) that is submitted herewith. This sequence listing is incorporated by reference herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] This patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

[0013] FIG. **1** is a schematic illustrating an embodiment of a norovirus detection assay using paper-based cell-free transcription-translation reactions. A norovirus sample is first enriched using synbodies and viral RNA amplified isothermally using nucleic acid sequence-based amplification (NASBA) or reverse transcriptase recombinase polymerase amplification (RT-RPA). The amplified nucleic acids are added to paper-based cell-free reactions where norovirus RNAs are detected by sequence-specific toehold switches. The toehold switches generate the lacZα peptide, which produces a purple-colored product after complementation with lacZω. Samples positive for norovirus can be identified by their purple color following the assay.

[0014] FIGS. **2**A-**2**G demonstrate detection of norovirus target RNA using toehold switches and α-complementation. A, Schematic of toehold switch operation in response to the target RNA. A weak stem containing the ribosomal binding site (RBS) is retained after target binding. This stem unwinds during binding of the ribosome to enable translation of the output gene. B, Enzymatically active lacZ tetramer formation occurs directly for full-length lacZ, while lacZα and lacZω must first assemble via a-complementation prior to tetramer formation. C, D, Measurements of the fold change in lacZ production rate (C) and ΔOD575 (D) of six toehold switches targeting the sense orientation of the norovirus target RNA after two hours of the cell-free reaction. E, F, Measurements of the fold change in lacZ production rate (E) and ΔOD575 (F) of six toehold switches targeting the antisense orientation of the norovirus target RNA. Change in lacZ production rate was measured after 1 hour of cell-free reaction (C, E) and ΔOD575 was measured after 2 hours of cell-free reaction (D, F). G, OD575 for toehold switch A2 as function of cell-free reaction time when outputting full-length lacZ compared to lacZα in a reaction supplemented with pre-synthesized lacZω.

[0015] FIGS. **3**A-**3**E demonstrate detection limit measurements for synthetic norovirus GII.4 target RNAs subject to isothermal amplification and detection using toehold switches. A,B, OD575 after amplification using RT-RPA and detection using toehold switches S6 (A) and S2 (B) in two-hour cell-free reactions. C,D, OD575 after amplification using NASBA and detection using toehold switch S6 (C) and S2 (D) in two-hour cell-free reactions. (E) Photographs of paper-based reactions using NASBA for amplification and toehold switch S2 for detection. Photographs were taken after 1 hour of the cell-free reactions.

[0016] FIGS. **4**A-**4**C demonstrate detection of live norovirus GII.4 Sydney and cross-reactivity testing. A, Norovirus RNA was extracted by diluting a stool sample 1:50 into PBS and briefly heating to 95° C. for 2 minutes. B, Measurement of OD575 after a two-hour paper-based reaction for a water-only negative control (−) and stool samples with and without norovirus particles before and after the brief heating treatment. All samples were subject to amplification via NASBA and detection with toehold switch S2. Only the heated norovirus GII.4 Sydney sample activates the toehold switch. C, Cross-reactivity testing of the assay against RNA from multiple bacteria, norovirus genotypes, and a water-only negative control. All samples were subject to NASBA and toehold switch S2 detection. OD575 was measured after two hours of the cell-free reaction.

[0017] FIGS. **5**A-**5**D illustrate implementation of a synbody-based capture and concentration method for norovirus detection. A, Illustration of the synbody enrichment technique. Biotin-labelled synbodies engineered to recognize diverse norovirus genotypes are used to bind to virus particles in a dilute solution and are in turn captured by streptavidin-coated magnetic beads. Magnetic capture enables concentration of the captured norovirus particles. B, Measurement of OD575 after two-hour cell-free reactions with toehold switch S2. Samples subject to synbody-based concentration and samples used directly without concentration were amplified by NASBA. The negative control (−) is a water-only sample. C, Time-course measurements of OD575 for synbody-concentrated samples compared to samples used directly. OD575 for a sample used directly after 1000-fold dilution is comparable to a concentrated sample initially diluted by 10^6-fold. D, Detection of norovirus GII.6 from a stool sample using toehold switch S2 and updated NASBA primers for the GII.6 genome. OD575 measurements were taken after two hours of the paper-based cell-free reaction and using a norovirus-negative stool sample as comparison.

[0018] FIG. **6** is a schematic of toehold switch regions used for calculation of design ensemble defect parameters. The full sensor sequence used for calculation of d_{full_sensor} is shown along with subsequences spanning the toehold region ($d_{toehold}$) and the toehold and hairpin regions (d_{min_sensor}). The start of the active sensor region is indicated by the arrow and begins at the base that falls 1 nt after the target binding region. The red target-binding region and the gray N bases are the only ones that change for each device.

[0019] FIGS. **7**A-**7**L are a series of linear regressions to detect correlations between toehold switch design parameters and their fold change performance for norovirus detection. (A-D) Linear regressions applied only to the set of antisense RNA sensors that provided overall better dynamic range. A two-parameter fit (D) is able to match the correlation with experimental results provided by a four-parameter fit. (E-H) Linear regressions applied only to the set of sense RNA sensors. No correlation is observed with the design scoring function s (E). One- and two-parameter fits also do not show much correlation with the experimental results. (I-L) Linear regressions applied to the full set of 12 toehold switches. Use of the same set of predictor variables in the scoring function s show limited agreement. However, linear

regressions with increasing numbers of parameters provide much stronger correlations (J-L), ultimately reaching R^2=0.60 for a four-parameter regression.

[0020] FIG. **8** is a series of photographs of paper-based toehold switch reactions using lacZ as the output protein. The colorimetric test results can be easily seen by eye with OD575 down to at least 0.4.

[0021] FIGS. **9**A-**9**D present a comparison of toehold switch interactions with different norovirus GII genotypes. (A) Sequence alignment for strains GII.4 Sydney, GII.6, GII.P17, and GII.17 within the target binding region of the toehold switch. Three or four mutations are present within the 36-nt target domain. Binding sites for GII.P17 and GII.17 are identical. (B-D) Predicted secondary structures of target-switch complexes formed between toehold switch S2 and the three unique target RNAs: GII.4 Sydney (B), GII.6 (C), and GII.P17/GII.17 (D). All targets provide strong binding to the toehold switch. The active sensor region of the switch RNA has the same secondary structure across all three targets, which suggests that translational efficiency will be sufficient to report on target binding.

[0022] FIG. **10** is a map of plasmid pAT_T7_HisLacZ (SEQ ID NO:69).

[0023] FIG. **11** is a map of plasmid pDM_noro_A1_lacZA (SEQ ID NO:70).

[0024] FIG. **12** is a map of plasmid pDM_noro_A2_lacZ (SEQ ID NO:71).

[0025] FIG. **13** is a map of plasmid pDM_noro_A2_lacZA (SEQ ID NO:72).

[0026] FIG. **14** is a map of plasmid pDM_noro_A3_lacZA (SEQ ID NO:73).

[0027] FIG. **15** is a map of plasmid pDM_noro_A4_lacZA (SEQ ID NO:74).

[0028] FIG. **16** is a map of plasmid pDM_noro_A5_lacZA (SEQ ID NO:75).

[0029] FIG. **17** is a map of plasmid pDM_noro_A6_lacZA (SEQ ID NO:76).

[0030] FIG. **18** is a map of plasmid pDM_noro_S1_lacZA (SEQ ID NO:77).

[0031] FIG. **19** is a map of plasmid pDM_noro_S2_lacZA (SEQ ID NO:78).

[0032] FIG. **20** is a map of plasmid pDM_noro_S3_lacZA (SEQ ID NO:79).

[0033] FIG. **21** is a map of plasmid pDM_noro_S4_lacZA (SEQ ID NO:80).

[0034] FIG. **22** is a map of plasmid pDM_noro_S5_lacZA (SEQ ID NO:81).

[0035] FIG. **23** is a map of plasmid pDM_noro_S6_lacZA (SEQ ID NO:82).

[0036] FIG. **24** is a map of plasmid pDM_T7_HisLacZomega.

DETAILED DESCRIPTION

[0037] The methods and compositions provided herein are based at least in part on the inventors' development of a highly sensitive assay for detecting pathogen nucleic acids down to concentrations of 200 attomoles/liter (aM) in reactions that can be directly read by eye. Advantages of the methods and compositions provided herein are multifold and include, for example, low-cost identification of infectious agents (e.g., foodborne pathogens) in a versatile diagnostic assay that provides visible test results, does not require expensive thermal cycling and other laboratory equipment, and can be easily deployed for rapid diagnosis in the field. Moreover, the methods and compositions provided herein obviate the need for expensive equipment and facilitate decentralized assays.

[0038] Accordingly, in a first aspect, provided herein is a method of detecting a target pathogen nucleic acid in a sample. The method can comprise or consist essentially of the following steps: (a) contacting a biological sample obtained from a subject to a pathogen detection agent under conditions that promote binding of the pathogen detection agent to the target pathogen if present in the sample; (b) isolating nucleic acids from pathogen bound by the pathogen detection agent; (b) amplifying the isolated nucleic acids using isothermal amplification; and (c) contacting the amplified nucleic acid to a toehold switch, wherein the toehold switch encodes at least a portion of a reporter protein and comprises one or more single-stranded toehold sequence domains that are complementary to a target nucleic acid or the reverse complement thereof, where the contacting occurs under conditions that allow translation of the coding domain in the presence of the target nucleic acid but not in the absence of the target nucleic acid, and detecting the reporter protein as an indicator that the target pathogen nucleic acid is present in the amplified nucleic acids. By enriching for and amplifying nucleic acids of the target pathogen, the method allows for detection of a pathogen in a biological sample that is dilute or contains few copies of the target pathogen. In some cases, the methods permit target nucleic acid detection with femtomolar, attomolar, and zeptomolar detection limits. As demonstrated in the Examples section, the method enables detection of norovirus GII.4 Sydney from a stool sample down to concentrations of 270 aM without the use of a concentration step. The SI prefix "atto" represents a factor of 10^{18}, or in exponential notation, 1E-18. The Examples also demonstrate that synbody-based enrichment of the virus can lower the detection limit by 1000-fold to 270 zeptomoles/liter (zM). The SI prefix "zepto" represents a factor of 10^{-21}, or in exponential notation, 1E-21.

[0039] As used herein, the term "pathogen" refers to any infectious agent and includes viruses, parasites, bacteria, fungi, and prions. By way of non-limiting example, pathogens may comprise viruses including, without limitation, noroviruses (e.g., Norwalk virus), flaviruses, human immunodeficiency virus (HIV), Ebola virus, single stranded RNA viruses, single stranded DNA viruses, double-stranded RNA viruses, double-stranded DNA viruses. Other pathogens include but are not limited to parasites (e.g., malaria parasites and other protozoan and metazoan pathogens (*Plasmodia* species, *Leishmania* species, *Schistosoma* species, *Trypanosoma* species)), bacteria (e.g., *Mycobacteria,* in particular, *M. tuberculosis, Salmonella, Streptococci, E. coli, Staphylococci*), fungi (e.g., *Candida* species, *Aspergillus* species, *Pneumocystis jirovecii, Pneumocystis carinii,* and other *Pneumocystis* species), and prions. In some cases, the pathogenic microorganism, e.g. pathogenic bacteria, may be one which causes cancer in certain human cell types. An advantage of the methods described herein is that they can be applied for the detection and identification of essentially any nucleic acid-containing organism. Accordingly, the pathogen can be virtually any pathogen or infectious agent for which genetic information (e.g., gene sequences) is available. In other cases, the target nucleic acid is human in origin. In such cases, the methods can be employed to detect

one or more target nucleic acids in a biological sample such as a biological sample obtained for forensic analysis, for genotyping, and the like.

[0040] Pathogen detection agents include, without limitation, antibodies, synbodies, peptides, polypeptides, and aptamers. Referring to FIG. **1**, synbodies are useful pathogen detection agents for the methods provided herein. As used herein, the terms "synbody" and "synbodies" refer to synthetic peptide affinity ligands. In some cases, the synbody is a synthetic bivalent affinity ligand comprising or consisting essentially of two or more compounds such as peptides, joined by a linker, and identified as having affinity for the same target molecule (e.g., a protein of interest). Synbodies, can be developed by linking two low affinity 15-20 amino acid (aa) long peptides to produce a high affinity synbody for a target protein (e.g., a viral coat protein) or bacteria. Synbodies have affinities and specificities similar to antibodies. Unlike antibodies, however, which often lose their affinity as a pathogen strain evolves, synbodies have broad cross-affinity for multiple pathogen genotypes, which enables them to recognize and specifically bind to a range of pathogen genotypes. In some cases, the synbody is a norovirus synbody such as, for example, the norovirus synbodies disclosed in U.S. Pat. No. 9,766,239, which is incorporated by reference herein.

[0041] Specific binding refers to the binding of a compound to a target (e.g., a component of a sample) that is detectably higher in magnitude and distinguishable from non-specific binding occurring to at least one unrelated target. Specific binding can be the result of formation of bonds between particular functional groups or particular spatial fit (e.g., lock and key type) whereas nonspecific binding is usually the result of van der Waals forces. Specific binding does not however imply that a compound binds one and only one target. Thus, a compound can and often does show specific binding of different strengths to several different targets and only nonspecific binding to other targets. Preferably, different degrees of specific binding can be distinguished from one another as can specific binding from nonspecific binding.

[0042] In certain embodiments, the synbody is tagged with biotin and then contacted to streptavidin-coated magnetic beads. In this manner, pathogen protein (e.g., viral particles) bound to the synbody-bead can be captured and concentrated using magnets. Referring to FIG. **5**A, a synbody-based magnetic bead capture assay can be used to concentrate pathogen (e.g., norovirus) from dilute solutions. In this example, captured norovirus was heated to 95° C. to release norovirus RNA, and the released RNA was subjected to isothermal amplification and applied to paper-based cell-free systems containing a norovirus-specific toehold switch.

[0043] In certain embodiments, the method employs programmable riboregulators known as toehold switches. As used herein, the term "toehold switch" generally refers to a nucleic acid-based regulator of gene expression, configured to repress or activate translation of an open reading frame and thus production of a protein. Toehold switches, which are a type of prokaryotic riboregulator, activate gene expression in response to cognate RNAs with essentially arbitrary sequences. Gene regulation is achieved through the presence of a regulatory nucleic acid element (the cis-repressive RNA or crRNA) within the 5' untranslated region (5' UTR) of an mRNA molecule. The cis-repressive nucleic acid element (crRNA) forms a hairpin structure comprising a stem domain and a loop domain through complementary base pairing. The hairpin structure blocks access to the mRNA transcript by the ribosome, thereby preventing translation. In some embodiments, the stem domain of the hairpin structure sequesters the ribosome binding site (RBS). In some embodiments, including, for example, embodiments involving eukaryotic cells, the stem domain of the hairpin structure is positioned upstream of the start (or initiation) codon. As described in the Examples, that follow, toehold switches particularly useful for the methods provided herein are configured for lower leakage relative to previously described riboregulators. As illustrated in FIG. **2**A, binding of a cognate target RNA to the updated toehold switch unwinds the lower half of the switch RNA hairpin and leaves the conserved upper stem-loop intact. This upper stem-loop is sufficiently weak to expose the RBS to enable translation to occur. Unlike earlier toehold switch mRNA sensors, the updated systems do not employ an RNA refolding domain downstream of the start codon, which could hamper translation of the output gene.

[0044] In some cases, toehold switches are synthetic (engineered) molecules. In other cases, toehold switches comprise endogenous, naturally occurring RNAs or regions thereof. See, for example, U.S. 2015/0275203. The stem domain can be as small as 12 bps, but in some cases will be longer than 12 bps, including 13, 14, 15, 16, 17, 18, 19, 20, or more base pairs in length. In other cases, the loop domain is complementary to a non-naturally occurring RNA. The toehold domain can be 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more nucleotides in length.

[0045] The toehold switch further comprises a fully or partially double-stranded stem domain comprising an initiation codon, a loop domain comprising a RBS, and a coding domain. The unpaired region upstream of the RBS in a toehold switch can be shortened or lengthened to modulate protein output and, in turn, device dynamic range. In some cases, the toehold and stem domains are complementary in sequence to a naturally occurring RNA. In other cases, the sequence detected can also be the complement of the naturally occurring RNA. For example, after isothermal amplification, it is possible to transcribe the antisense of the RNA rather than the sense.

[0046] The toehold switch can further comprise a thermodynamically stable double-stranded stem domain, a loop domain comprising a ribosome binding site, and a coding domain. Preferably, the loop domain is 11 nucleotides or 12 nucleotides in length. In some cases, the length of loop domains can be increased or decreased, for example, to alter reaction thermodynamics.

[0047] In certain embodiments, the toehold switch is configured to detect a portion of a pathogen genome that is conserved among two or more species or strains of the pathogen. For example, the Examples that follow describe identifying conserved sequence regions of a norovirus GII genome suitable for isothermal amplification and toehold-switch-based detection. In some cases, toehold switches useful for the methods provided herein include, without limitation, synthetic norovirus-specific toehold switches that comprise a fully or partially double-stranded stem domain, a loop domain, a toehold domain, and at least a portion of a coding sequence of a reporter gene, wherein the toehold domain and at least a portion of the stem domain are complementary to a target norovirus RNA sequence. In

some cases, synthetic norovirus-specific toehold switches comprise an RNA sequence selected from SEQ ID NOs:1-12 set forth in Table 1.

[0048] As shown in FIGS. **2**A and **2**B, the toehold switch can be operably linked to a reporter element (e.g., at least a portion of an *E. coli* lacZ reporter element encoding β-galactosidase) that is 3' to the hairpin structure. As used herein, the term "operably linked" refers to a relationship between two nucleic acid sequences wherein the production or expression of one of the nucleic acid sequences is controlled by, regulated by, modulated by, etc., the other nucleic acid sequence. Reporter proteins appropriate for the methods provided herein include, without limitation, enzymatic reporters (e.g., β-galactosidase, alkaline phosphatase, DHFR, CAT), fluorescent or chemiluminescent reporters (e.g., GFP variants, mCherry, luciferase, e.g., luciferase derived from the firefly (*Photinus pyralis*) or the sea pansy (*Renilla reniformis*) and mutants thereof), etc.

[0049] Any isothermal amplification protocol can be used according to the methods provided herein. Exemplary types of isothermal amplification include, without limitation, nucleic acid sequence-based amplification (NASBA), reverse transcriptase recombinase polymerase amplification (RT-RPA), loop-mediated isothermal amplification (LAMP), strand displacement amplification (SDA), helicase-dependent amplification (HDA), nicking enzyme amplification reaction (NEAR), signal mediated amplification of RNA technology (SMART), rolling circle amplification (RCA), isothermal multiple displacement amplification (IMDA), single primer isothermal amplification (SPIA), recombinase polymerase amplification (RPA), and polymerase spiral reaction (PSR, available at nature.com/articles/srep12723 on the World Wide Web).

[0050] In some cases, it may be advantageous to adapt the methods described herein for high-throughput, reproducible, and rapid detection, for example in a clinical setting. When output from the toehold switch is coupled to a reporter element, such as a LacZ reporter element or portion thereof, the riboregulator acts as a genetically encodable sensor and detectable probe for endogenous DNA or RNA (e.g., endogenous pathogen DNA, endogenous pathogen RNA) in a sample. For example, such toehold switches can be provided in a device configured for rapid, reproducible detection in a non-laboratory setting (e.g., clinical setting). In some cases, the device comprises a preserved paper test article, upon which any step(s) of the method provided herein can be performed. In preferred embodiments, the paper test article is preserved by freeze-drying. The reporter element can be a reporter protein, e.g., a polypeptide with an easily assayed enzymatic activity or detectable signal that is naturally absent from the host cell. Exemplary but non-limiting reporter proteins include lacZ, catalase, β-glucoronidase, xylE, GFP, RFP, YFP, CFP, neomycin phosphotransferase, luciferase, mCherry, and derivatives or variants thereof. In some embodiments of any of the aspects, the reporter protein is suitable for use in a colorimetric assay. Examples of genes encoding fluorescent proteins that may be used in accordance with the invention include, without limitation, those proteins provided in U.S. Patent Application No. 2012/0003630 (see Table 59 therein), incorporated herein by reference.

[0051] In certain embodiments, alpha-complementation is employed to decrease assay times and strengthen output from the cell-free transcription-translation reactions. As shown in FIG. **2**B, it may be advantageous to divide the lacZ enzyme into two peptides termed α and ω. The lacZ α-peptide (lacZα) consists of the first 50 to 59 residues from the N terminus of lacZ and the ω-peptide (lacZω) comprises the remaining ~970 lacZ residues. The complete lacZ must form a tetramer before it becomes catalytically active; however, lacZω cannot form a tetramer on its own as it lacks residues critical for assembly. As a result, both lacZα and lacZω must be expressed before complementation occurs and an active lacZ tetramer can assemble. As demonstrated in the Examples section, the use of a-complementation, in which the lacZα and lacZω peptides complement to form the active lacZ enzyme, can reduce the time to detection of the paper-based assay by up to 23 minutes or 43% compared to experiments employing the full-length lacZ as the toehold switch output.

[0052] In some cases, DNA encoding the norovirus-specific toehold switches may be cloned into vectors upstream of the lacZα open reading frame. Such constructs can be used in paper-based cell-free reactions when supplemented with lacZω. In some cases, lacZω is provided as a pre-synthesized compound on the reaction substrate (e.g., a paper-based cell-free reaction substrate). For example, the toehold switch can encode at least a portion of lacZ such as lacZα. The amplified nucleic acids are contacted to cell-free reaction substrate to which lacZω is provided as a pre-synthesized peptide and under conditions which promote formation of a lacZ tetramer.

[0053] Other complementation reporter systems can be used for the methods described herein. For example, Green Fluorescent Protein (GFP) can be split by removing a single beta strand from its barrel structures to generate a large molecular weight GFP1-10 peptide, comprising beta strands one through 10, and a small molecular weight GFP11 peptide, comprising the 11^{th} beta strand. When both peptides are present, they spontaneously reassemble into a fluorescently active combined protein. Split GFP systems are described, for example, at nature.com/articles/ncomms11046 and nature.com/articles/s41467-017-00494-8 on the World Wide Web. sfCherry, an improved folding version of mCherry, and mNeonGreen2 can also be split in similar ways and provide analogous fluorescence readout via complementation. There are multiple versions of split cas9 that can also be activated through complementation. See, for example, nature.com/articles/nbt.3149 and pnas.org/content/112/10/2984 on the World Wide Web.

[0054] Any appropriate sample can be used according to the methods provided herein. In some cases, the sample is a biological sample obtained from an individual (e.g., a human subject, a non-human mammal). The sample is, in some cases, a diagnostic sample. The sample type will vary depending on the target pathogen. For example, norovirus, including human forms of norovirus (i.e., Norwalk virus), can be detected in stool specimens, sputum, blood or vomitus of diseased individuals. Norovirus can also be present in body tissues, such as brain tissue, in an infected mammalian organism. Accordingly, a diagnostic sample for detecting norovirus can be a stool sample, a sputum sample, a vomitus sample, a tissue sample, or a blood sample. Samples appropriate for use according to the methods provided herein can also include, without limitation, food samples, drinking water, environmental samples, and agricultural products. In some cases, samples appropriate for use according to the methods provided herein are "non-biological" in whole or in

part. Non-biological samples include, without limitation, plastic and packaging materials, paper, clothing fibers, and metal surfaces. In certain embodiments, the methods provided herein are used in food safety and food biosecurity applications, such as screening food products and materials used in food processing or packaging for the presence of pathogens in biological and/or non-biological samples.

[0055] Other applications for which the methods provided herein include, without limitation, profiling species in an environment (e.g., water); profiling species in an human or animal microbiome; food safety applications (e.g., detecting the presence of a pathogenic species, determining or confirming food source/origin such as type of animal or crop plant); obtaining patient expression profiles (e.g., detecting expression of a gene or panel of genes (e.g., biomarkers)) to monitor the patient's response to a therapeutic regimen, to select a therapeutic regimen suitable for the patient, or to detect exposure of the patient to a toxin or environmental agent that affects expression of the gene or a panel of genes.

[0056] In some cases, the device is used with a portable electronic reader. In this manner, the electronic reader serves as companion technology that provides robust and quantitative measurements of device outputs. In some embodiments, the electronic reader comprises readily available consumer components, open-source code, and laser-cut acrylic housing, and is powered by a rechargeable lithium ion battery. The electronic reader can further comprise an onboard data storage unit. In some cases, to achieve sensitive detection of toehold switch signal output, an acrylic chip that holds the freeze-dried, paper-based reactions is placed into the reader between a light source (e.g., to read optical density at excitation and emission wavelengths of light appropriate for and characteristic of a particular detectable reporter) and electronic sensors. In some cases, the light source is a light emitting diode (LED) light source. Samples can be read using onboard electronics. In this manner, a portable electronic reader can provide low-noise measurements of changes associated with the reporter element including changes in light transmission due to LacZ-mediated color change.

[0057] In certain embodiments, provided herein is a device for identifying a pathogen-associated nucleic acid, comprising a preserved paper test article, wherein the methods described herein are performed using the preserved paper test article. In some cases, the paper test article is preserved by freeze-drying.

[0058] Articles of Manufacture

[0059] In another aspect, the present invention provides articles of manufacture useful for detecting a pathogen in a sample according to the methods provided herein. In certain embodiments, the article of manufacture is a kit for detecting norovirus, where the kit comprises a norovirus detecting agent, a plurality of preserved paper test articles as described herein, and an electronic optical reader. Optionally, a kit can further include instructions for performing the pathogen detection methods provided herein.

[0060] In certain embodiments, provided herein is a kit for detecting a pathogen-associated nucleic acid, where the kit comprises a plurality of preserved paper test articles, a pathogen detection agent, a plurality of toehold switches that encode at least a portion of a reporter protein and comprise one or more single-stranded toehold sequence domains that are complementary to a target pathogen nucleic acid or the reverse complement thereof, and an electronic optical reader. In some cases, the kit also comprises instructions for performing the pathogen detection methods provided herein.

[0061] In other embodiments, provided herein is a kit for detecting a pathogen-associated nucleic acid, where the kit comprises a plurality of preserved test tube test articles, a pathogen detection agent, a plurality of toehold switches that encode at least a portion of a reporter protein and comprise one or more single-stranded toehold sequence domains that are complementary to a target pathogen nucleic acid or the reverse complement thereof, and an electronic optical reader. In some cases, the kit also comprises instructions for performing the pathogen detection methods provided herein.

[0062] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

[0063] The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

[0064] The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

[0065] As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

[0066] As used herein, the terms "approximately" or "about" in reference to a number are generally taken to include numbers that fall within a range of 5% in either direction (greater than or less than) the number unless otherwise stated or otherwise evident from the context (except where such number would exceed 100% of a possible value). Where ranges are stated, the endpoints are

included within the range unless otherwise stated or otherwise evident from the context.

[0067] The present invention has been described in terms of one or more preferred embodiments, and it should be appreciated that many equivalents, alternatives, variations, and modifications, aside from those expressly stated, are possible and within the scope of the invention.

EXAMPLES

[0068] This section demonstrates a paper-based, cell-free platform for the detection of the prevalent GII.4 Sydney norovirus genotype (FIG. **1**). Beginning from fecal samples or a dilute solution containing the virus, the assay employs biotin-labelled synthetic peptide affinity ligands known as synbodies to capture norovirus particles and concentrate the captured particles using streptavidin-coated magnetic beads. A brief heating step is used to release the norovirus RNA, and either NASBA or reverse transcriptase RPA (RT-RPA) is employed to amplify the viral RNA. The amplification products are then added to paper-based cell-free reactions where norovirus-specific toehold switches are used to verify their sequences and produce the lacZα peptide, which provides a visual reaction readout. We demonstrate that this assay enables detection of norovirus target RNAs down to concentrations of 200 aM without the use of a concentration step, and further show that synbody-based enrichment of the virus can lower the detection limit by 1000-fold when applied to a clinical fecal sample. We also demonstrate that the use of a-complementation, in which the lacZα and lacZω peptides complement to form the active lacZ enzyme, can reduce the time to detection of the paper-based assay by up to 23 minutes or 43% compared to experiments employing the full-length lacZ as the toehold switch output. These results expand the range of sample types and viruses that can be analyzed using paper-based cell-free systems and provide new strategies to improve the sensitivity and reduce the time of these inexpensive diagnostic assays.

[0069] Materials and Methods

[0070] Norovirus samples and bacterial strains: Stool samples positive for the norovirus GII.4 Sydney genotype and the norovirus GI.2 genotype were generously provided by Jan Vinjé from the National Calicivirus Laboratory at the Centers for Disease Control and Prevention (CDC). *Escherichia coli* MG1655 (ATCC, 700926), methicillin-resistant *Staphylococcus aureus* MRSA252 (ATCC, BAA-1720), and *Bacillus subtilis* 168 (ATCC, 23857) were used for assay cross-reactivity experiments. For these experiments, RNA from the bacteria was extracted using a Quick-RNA Fungal/Bacterial Miniprep Kit (Zymo Research) following the manufacturer's instructions. To obtain purified viral RNA for cross-reactivity experiments, 5 μL of GII.4, GI.2, and GI.6 positive stool samples were suspended in 140 μL RNase-free water. The viral RNA was extracted by using QIAamp DSP Viral RNA Mini Kit (Qiagen, U.S.A.) according to the manufacturer's instructions. RNAs were eluted with 50 μL RNase-free water and stored at −80° C. *E. coli* DH5α (ThermoFisher Scientific) was used for cloning of toehold switch plasmids.

[0071] In silico selection of toehold switch designs: An updated version of the selection algorithm described previously[32] was used to identify toehold switches for detection of norovirus RNA. The algorithm facilitated selection six promising designs from a set of over 100 candidate toehold switches generated from each norovirus target RNA. Candidate devices were designed to bind to a 36-nt continuous region of the norovirus target RNA. Putative toehold switches were generated at 1-nt increments along the norovirus target RNA and multiple ensemble defect levels were computed for each sensor based on its deviation from the ideal secondary structure of the toehold switch. Ensemble defects were calculated for the toehold switch 5' end through to the 3' end of the hairpin (d_{min_sensor}), the toehold domain of the toehold switch ($d_{toehold}$), the binding site of the toehold switch within the target RNA ($d_{binding\_site}$), and the toehold switch region starting with the base immediately 3' of the target RNA binding site and extending 31 nts beyond the last base on the 3' end of the hairpin (d_{active_sensor}).

[0072] For the latter two parameters, the ensemble defect was calculated based on a completely single-stranded ideal secondary structure. The parameter d_{active_sensor} was intended to provide a measure of any secondary structures in the activated toehold switch that could interfere with translation after binding to the target RNA. In addition to ensemble defects, the equilibrium fraction f of target/toehold switch complexes in a system with equimolar concentrations of target and toehold switch RNAs was calculated as a measure of the affinity of the two RNAs. In practice, this parameter was almost always equal to 1. Designs that produced in-frame stop codons in the output gene were eliminated from further consideration. Each of the parameters was then normalized such that their maximum value across the set of putative designs for a given target RNA was equal to 1. These normalized parameters, designated by an overscore, were then inserted into a scoring function s:

$$s=5\bar{d}_{toehold}+4\bar{d}_{active\_sensor}+2\bar{d}_{min\_sensor}+2\bar{d}_{binding\_site}+(1-f)$$

Toehold switches displaying the lowest values of s and screened to have f>0.9 were selected for experimental testing. Sequences of the toehold switches generated by the algorithm are provided in Table 1 along with those of the norovirus target regions. The weighting coefficients used in the scoring function were determined empirically based on testing of earlier toehold switch mRNA sensor designs[30, 32].

TABLE 1

Toehold switch and norovirus target RNA sequences

Name	RNA Sequence
Toehold switch S1 RNA	GGGCCAUCUUCAUUCACAAAACUGGGAGCCAGAUUGCGAGGACUUUA GAACAGAGGAGAUAAAGAUGUCGCAAUCUGGAACCUGGCGGCAGCGC AAGAAGAUG (SEQ ID NO: i)
Toehold switch S2 RNA	GGGAUCGCCCUCCCACGUGCUCAGAUCUGAGAAUCUCAUGGACUUUA GAACAGAGGAGAUAAAGAUGAUGAGAUUCUCAACCUGGCGGCAGCGC AAGAAGAUG (SEQ ID NO: 2)

TABLE 1 -continued

Toehold switch and norovirus target RNA sequences

Name	RNA Sequence
Toehold switch S3 RNA	GGGACAAAACUGGGAGCCAGAUUGCGAUCGCCCUCCCACGGACUUUA GAACAGAGGAGAUAAAGAUGGUGGGAGGGCGAACCUGGCGGCAGCGC AAGAAGAUG (SEQ ID NO: 3)
Toehold switch S4 RNA	GGGCUGGGACGAGGUUGGCUGCGGACCCAUCAGAUGGGUGGACUUUA GAACAGAGGAGAUAAAGAUGACCCAUCUGAUAACCUGGCGGCAGCGC AAGAAGAUG (SEQ ID NO: 4)
Toehold switch S5 RNA	GGGUCAUUCGACGCCAUCUUCAUUCACAAAACUGGGAGGGACUUUAG AACAGAGGAGAUAAAGAUGCUCCCAGUUUUAACCUGGCGGCAGCGCA AGAAGAUG (SEQ ID NO: 5)
Toehold switch S6 RNA	GGGAGCCAGAUUGCGAUCGCCCUCCCACGUGCUCAGAUCGGACUUUA GAACAGAGGAGAUAAAGAUGGAUCUGAGCACAACCUGGCGGCAGCGC AAGAAGAUG (SEQ ID NO: 6)
Toehold switch A1 RNA	GGGUCUGAUGGGUCCGCAGCCAACCUCGUCCCAGAGGUCGGACUUUA GAACAGAGGAGAUAAAGAUGGACCUCUGGGAAACCUGGCGGCAGCGC AAGAAGAUG (SEQ ID NO: 7)
Toehold switch A2 RNA	GGGUGGGAGGGCGAUCGCAAUCUGGCUCCCAGUUUUGUGGACUUUAG AACAGAGGAGAUAAAGAUGACAAAACUGGGAACCUGGCGGCAGCGCA AGAAGAUG (SEQ ID NO: 8)
Toehold switch A3 RNA	GGGUGUGAAUGAAGAUGGCGUCGAAUGACGCCAACCCAUGGACUUUA GAACAGAGGAGAUAAAGAUGAUGGGUUGGCGAACCUGGCGGCAGCGC AAGAAGAUG (SEQ ID NO: 9)
Toehold switch A4 RNA	GGGAGAUCUGAGCACGUGGGAGGGCGAUCGCAAUCUGGCGGACUUUA GAACAGAGGAGAUAAAGAUGGCCAGAUUGCGAACCUGGCGGCAGCGC AAGAAGAUG (SEQ ID NO: 10)
Toehold switch A5 RNA	GGGAUCGCAAUCUGGCUCCCAGUUUUGUGAAUGAAGAUGGGACUUUA GAACAGAGGAGAUAAAGAUGCAUCUUCAUUCAACCUGGCGGCAGCGC AAGAAGAUG (SEQ ID NO: 11)
Toehold switch A6 RNA	GGGUCGAAUGACGCCAACCCAUCUGAUGGGUCCGCAGCCGGACUUUA GAACAGAGGAGAUAAAGAUGGGCUGCGGACCAACCUGGCGGCAGCGC AAGAAGAUG (SEQ ID NO: 12)
Norovirus GII.4 sense target	AUGGAUUUUUACGUGCCCAGGCAAGAGCCAAUGUUCAGAUGGAUGAG AUUCUCAGAUCUGAGCACGUGGGAGGGCGAUCGCAAUCUGGCUCCCA GUUUUGUGAAUGAAGAUGGCGUCGAAUGACGCCAACCCAUCUGAUGG GUCCGCAGCCAACCUCGUCCCAGAGGUCAACAAUGAGGUUAUGGCUU UGGAGCCCGU (SEQ ID NO: 13)
Norovirus GII.4 antisense target	ACGGGCUCCAAAGCCAUAACCUCAUUGUUGACCUCUGGGACGAGGUU GGCUGCGGACCCAUCAGAUGGGUUGGCGUCAUUCGACGCCAUCUUCA UUCACAAAACUGGGAGCCAGAUUGCGAUCGCCCUCCCACGUGCUCAGA UCUGAGAAUCUCAUCCAUCUGAACAUUGGCUCUUGCCUGGGCACGUA AAAAUCCAU (SEQ ID NO: 14)
Norovirus GII.P17 sense target	AUGGAUUUUUAUGUGCCCAGACAAGAGUCAAUGUUCAGAUGGAUGAG GUUCUCAGAUCUAAGCACAUGGGAGGGCGAUCGCAAUCUGGCUCCCA GUUUUGUGAAUGAAGAUGGCGUCGAAUGACGCCGCUCCAUCUAAUGA UGGUGCUGCUGGUCUCGUACCAGAGGGCAACAACGAG (SEQ ID NO: 15)
Norovirus GII.17 sense target	AUGGAUUUUUAUGUGCCCAGACAAGAGUCAAUGUUCAGAUGGAUGAG GUUCUCAGAUCUAAGCACAUGGGAGGGCGAUCGCAAUCUGGCUCCCA GUUUUGUGAAUGAAGAUGGCGUCGAAUGACGCCGCUCCAUCUAAUGA UGGUGCUGCUGGUCUCGUACCAGAGGGCAACAACGAG (SEQ ID NO: 16)
Norovirus GII.6 antisense target	UGGAGUUUUAUGUGCCCAGACAAGAGGCCAUGUUCAGGUGGAUGAGA UUCUCUGACCUCAGCACAUGGGAGGGCGAUCGCAAUCUUGCUCCCGA GGGUGUGAAUGAAGAUGGCGUCGAAUGACGCUGCUCCAUCGAAUGAU GGUGCUGCCAACCUCGUACCAGAGGCCAACAAUGAGGUUAUGGC (SEQ ID NO: 17)

[0073] Calculation of Ensemble Defects for Toehold Switch Designs: Ensemble defect levels were calculated using NUPACK for toehold switch designs over the regions specified in the main text and indicated in FIG. **6**. The toehold switch designs for norovirus contained a conserved upper hairpin domain with the sequence GGACUUUAGAAC **AGAGGAGA** UAAAG **AUG** (SEQ ID NO:18), with the RBS and start codon shown in bold, and a 31-nt linker between the sensor and the output gene with the sequence AACCUGGCGGCAGCGCAAGAAGAUGCGUAAA (SEQ ID NO:19). The parameters d_{min_sensor} and $d_{toehold}$ were calculated by first computing the pairwise binding probabilities for the toehold switch sequence from the 5' end through to the 31st base beyond the 3' end of the hairpin (i.e., the full sequence shown in FIG. **6**). These binding probabilities were then used to compute the ensemble defect from the specified sequence regions of d_{min_sensor} and $d_{toehold}$ using the target secondary structures shown in FIG. **6**. Calculating the ensemble defect in this way enabled the effect of sequences outside the main region of interest to be considered for design purposes. For $d_{binding\_site}$, the ensemble defect was calculated in an analogous manner using the pairwise binding probabilities of the complete target RNA sequence and specifying a completely single-stranded ideal secondary structure in the binding site region. For d_{active_sensor}, the ensemble defect was calculated directly from the sequence region starting from the base indicated in FIG. **6**. A completely single-stranded secondary structure was used for assessing design quality for d_{active_sensor}.

[0074] In addition to the four terms above, we calculated two additional ensemble defect parameters during the design process. The term d_{full_sensor} was generated by computing the ensemble defect for the full toehold switch sequence and structure shown in FIG. **6**. The term d_{min_target} was generated by taking the 36-nt sequence targeted by the toehold switch and computing its ensemble defect with a completely single-stranded ideal structure.

[0075] Assessment and Further Optimization of Toehold Switch Selection Algorithm: To determine the effectiveness of the described toehold switch selection method, we have taken experimental fold change in lacZ production data (FIGS. **7**C, **7**E) and investigated if these data display significant correlations with the selection scoring function and the six computed ensemble defect parameters. Correlations between these parameters and the experimental results were assessed using R^2 values and regression coefficients generated from the Matlab multiple linear regression using least squares function regress. For these regressions, a column of ones was appended to the matrix of predictor variables to allow the model to include a constant term or offset.

[0076] FIG. **7**A shows the linear regression obtained using experimental fold change from the toehold switches for the antisense target and the design scoring function s. We found that the scoring function, which made use of four ensemble defect parameters, provided a fairly strong correlation with the experimental data with R^2=0.62. We then examined the correlations between the antisense experimental data and the set of six ensemble defect parameters. The two parameters that showed the strongest correlations were d_{active_sensor} and d_{min_target}, which yielded R^2 of 0.52 and 0.31, respectively (see FIGS. **7**B-**7**C). A two-parameter linear regression combining both these terms provided R^2=0.62, which matched the correlation observed for the four-parameter scoring function.

[0077] We applied the same series of analyses to the set of toehold switches for the sense norovirus target. However, these devices showed much weaker correlations between design parameters and experimental results (see FIGS. **7**E-**7**H). The scoring function did not display any correlation with the fold change in lacZ (R^2=0.07) and the top two single-parameter fits did not display strong correlations, R^2=0.26 and R^2=0.20 for $d_{binding\_site}$ and d_{active_sensor}, respectively. Furthermore, no improvements were observed by combining the two parameters into the same regression.

[0078] Since the terms used in the scoring function were originally normalized for each target RNA, we could not use the scoring function directly to determine if it was highly correlated with the experimental results from all 12 devices since they bound to different target RNAs. Instead, we took the fold change experimental results and supplied the regression with the set of the four predictor variables used by the scoring function but in non-normalized form: d_{min_sensor}, $d_{toehold}$, $d_{binding\_site}$, and d_{active_sensor}. (Note: The fifth predictor variable f, the equilibrium fraction, used in the scoring function was equal to one for all devices tested.). In this case, the linear regression provided limited correlation with R^2=0.29 (see FIG. **7**I). To determine if other combinations of ensemble defect parameters could provide a stronger correlation with the experimental data, we computed linear regressions for all two-, three-, and four-parameter combinations as shown in FIGS. **7**J-**7**L. We found that the combination of $d_{binding\_site}$ and d_{active_sensor} were most effective for the two-parameter fits, yielding R^2=0.42, a substantial increase over the scoring function combination. Addition of d_{full_sensor} to the pair provided the best three-parameter fit with another sizeable increase in fit quality to R^2=0.57. Finally, the optimal four-parameter fit, which added d_{min_target} to the trio of predictor variables, provided a small increase in R^2 to 0.60.

[0079] The three- and four-parameter linear regressions generated the following equations for predicting the fold change for the toehold switch sensors:

Three Parameter Fit (R^2=0.57):

[0080]

$$\text{Fold change} = -71.7\, d_{full\_sensor} - 49.1\, d_{active\_sensor} - 22.6\, d_{binding\_site} + 54.3$$

Four-Parameter Fit (R^2=0.60):

[0081]

$$\text{Fold change} = -93.2\, d_{full\_sensor} - 43.3\, d_{active\_sensor} - 22.1\, d_{binding\_site} - 9.4\, d_{min\_target} + 61.3$$

[0082] For both these linear fitting functions, negative coefficients are used in front of all of the ensemble defect parameters as expected, since lower defect levels should lead to higher toehold switch performance (i.e. fold change). In addition, the parameters d_{full_sensor}, d_{active_sensor}, and $d_{binding\_site}$ are listed from highest to lowest fitting function weighting factor. These three parameters define the most critical functional elements of the toehold switch devices. A properly folded secondary structure of the full sensor is required to provide a toehold region for target binding and a strong hairpin structure to repress translational leakage.

The active sensor requires a translation start site with low secondary structure to promote rapid production of the output gene. Lastly, a binding site with low secondary structure helps ensure that the site is accessible for sensor binding. We expect that design selection algorithms can be further improved using strategies similar to the one described here and using much larger libraries of toehold switches to probe a wider range of sensor and target sequences experimentally.

[0083] Toehold switch plasmid construction: Synthetic DNA (Integrated DNA Technologies) encoding the norovirus-specific toehold switch sensors was amplified by PCR and inserted into plasmids using Gibson assembly with 30-bp overlap regions. Sequences of the toehold switches and the norovirus targets are provided in Table 1. Plasmids and DNA templates for transcription were constructed using conventional molecular biology techniques. The sequences of the plasmids were confirmed using Sanger sequencing (DNASU Sequencing Core, Tempe). The list of plasmids used in this work are provided in Table 2. Maps of these plasmids are presented in FIG. **10**-FIG. **24**. Sequences of the primers used for plasmid construction are listed Table 3. This table lists the source template amplified by each primer pair and indicates what plasmid was produced following Gibson assembly of the resulting PCR products. The synthetic DNA sequences used to generate toehold switches for insertion into plasmids are listed in Table 4. This table also contains the primers used for Sanger sequencing of the plasmids.

TABLE 2

List of Plasmids

Name	Marker	Description
pAT_T7_HisLacZ (SEQ ID NO: 69)	Amp	T7 RNAP-driven expression of N-terminal His-tagged lacZ. pET15b backbone.
ZIKV_Sensor_27B_LacZ (Addgene #: 75006)	Kan	T7 RNAP-driven expression of Zika virus sensing toehold switch with lacZ reporter. pCOLAduet backbone.
pDM_T7_HisLacZomega	Amp	T7 RNAP-drive expression of N-terminal His-tagged lacZω. pET15b backbone.
pDM_noro_S1_lacZA (SEQ ID NO: 77) pDM_noro_S2_lacZA (SEQ ID NO: 78) pDM_noro_S3_lacZA (SEQ ID NO: 79) pDM_noro_S4_lacZA (SEQ ID NO: 80) pDM_noro_S5_lacZA (SEQ ID NO: 81) pDM_noro_S6_lacZA (SEQ ID NO: 82)	Kan	T7 RNAP-driven expression of norovirus sense orientation toehold switches (S1 to S6) with a lacZα reporter. pCOLAduet backbone.
pDM_noro_A1_lacZA (SEQ ID NO: 70) pDM_noro_A2_lacZA (SEQ ID NO: 72) pDM_noro_A3_lacZA (SEQ ID NO: 73) pDM_noro_A4_lacZA (SEQ ID NO: 74) pDM_noro_A5_lacZA (SEQ ID NO: 75) pDM_noro_A6_lacZA (SEQ ID NO: 76)	Kan	T7 RNAP-driven expression of norovirus antisense orientation toehold switches (A1 to A6) with a lacZα reporter. pCOLAduet backbone.
pDM_noro_A2_lacZ (SEQ ID NO: 71)	Kan	T7 RNAP-driven expression of norovirus antisense orientation toehold switch A2 with full-length lacZ reporter. pCOLAduet backbone.

TABLE 3

List of PCR Primers Used for Plasmid Construction

Primer Name	Sequence	template	Destination plasmid(s)
lacZ_pET15b_fwd	TAACTAGCATAACCC CTTGGGG (SEQ ID NO: 20)	pET15b	pAT_T7_HisLacZ (SEQ ID NO: 69)
lacZ_pET15b_rev	CATATGGCTGCCGCG CGG (SEQ ID NO: 21)		
lacZ_insert_fwd	AGCGGCCTGGTGCCG CGCGGCAGCCATATG CGTAAAATGACCATG ATTACGGATTCACT (SEQ ID NO: 22)	*E. coli* MG1655 genome	

TABLE 3 -continued

List of PCR Primers Used for Plasmid Construction

Primer Name	Sequence	template	Destination plasmid(s)
lacZ_insert_rev	TTTAGAGGCCCCAAG GGGTTATGCTAGTTAT TTTTGACACCAGACCA ACTGGT (SEQ ID NO: 23)		
lacZomega_BB_fwd	AACAGTTGCGCAGCC TGA (SEQ ID NO: 24) CCAGTGAATCCGTAA	pAT_T7_HisLacZ	pDM_T7_ HisLacZomega
lacZomega_BB_rev	TCATGGTCAT (SEQ ID NO: 25)		
lacZomega_insert_L	ATGACCATGATTACG GATTCACTGGCCGTCG CCCGCACCGA (SEQ ID NO: 26)	lacZomega_insert_R	
lacZomega_insert_R	TCAGGCTGCGCAACT GTTGGGAAGGGCGAT CGGTGCGGGC (SEQ ID NO: 27)	lacZomega_insert_L	
lacZalpha_BB_fwd	TAGCATAACCCCTTGG GGC (SEQ ID NO: 28)	pDM_noro_A2_lacZA (SEQ ID NO: 72)	pDM_noro_A2_lacZA (SEQ ID NO: 72)
lacZalpha_BB_rev	GCGCAACTGTTGGGA AGG (SEQ ID NO: 29)		
lacZalpha_insert_L	CGCACCGATCGCCCTT CCCAACAGTTGCGCA GCCTGAATGGCGAAT GGTAAT (SEQ ID NO: 30)	lacZalpha_insert_R	
lacZalpha_insert_R	CCCGTTTAGAGGCCCC AAGGGGTTATGCTATT ATTACCATTCGCCATT CAGG (SEQ ID NO: 31)	lacZalpha_insert_L	
lacZ_BB_fwd	ATGACCATGATTACG GATTCACTGGCCGTC (SEQ ID NO: 32)	ZIKV_Sensor_27B_LacZ, pDM_noro_A2_lacZA	pDM_noro_A#_lacZ, pDM_noro_A#_lacZA, pDM_noro_S#_lacZ,
Dstar_lacZ_BB_rev	CCGGCTACCGTAGAA ACGCGAATTTACTAG CATAAGGGAGAGCGT CGAGATC (SEQ ID NO: 33)		pDM_noro_S#_ lacZA, where # = {1, 2, 3, 4, 5, 6}
Dnorm_TS_insert_fwd	CTAGTAAATTCGCGTT TCTACGGTAGCCGGG CGCTAATACGACTCA CTATAGGG (SEQ ID NO: 34)	Toehold switch DNA strands	
TS_insert_linker_rev	GACGGCCAGTGAATC CGTAATCATGGTCATC TTCTTGCGCTGCCGCC AGGTT (SEQ ID NO: 35)		

TABLE 4

List of DNA Strands Used for Toehold Switches and Sequencing

Primer Name	Sequence	Description
Toehold switch S1 DNA	GCGCTAATACGACTCACTATAGGGCCATC TTCATTCACAAAACTGGGAGCCAGATTGC GAGGACTTTAGAACAGAGGAGATAAAGAT GTCGCAATCTGGAACCTGGCGGCAGCGCA AGAAGATG (SEQ ID NO: 36)	Toehold switch DNA templates
Toehold switch S2 DNA	GCGCTAATACGACTCACTATAGGGATCGC CCTCCCACGTGCTCAGATCTGAGAATCTCA TGGACTTTAGAACAGAGGAGATAAAGATG ATGAGATTCTCAACCTGGCGGCAGCGCAA GAAGATG (SEQ ID NO: 37)	

TABLE 4 -continued

List of DNA Strands Used for Toehold Switches and Sequencing

Primer Name	Sequence	Description
Toehold switch S3 DNA	GCGCTAATACGACTCACTATAGGGACAAA ACTGGGAGCCAGATTGCGATCGCCCTCCC ACGGACTTTAGAACAGAGGAGATAAAGAT GGTGGGAGGGCGAACCTGGCGGCAGCGC AAGAAGATG (SEQ ID NO: 38)	
Toehold switch S4 DNA	GCGCTAATACGACTCACTATAGGGCTGGG ACGAGGTTGGCTGCGGACCCATCAGATGG GTGGACTTTAGAACAGAGGAGATAAAGAT GACCCATCTGATAACCTGGCGGCAGCGCA AGAAGATG (SEQ ID NO: 39)	
Toehold switch S5 DNA	GCGCTAATACGACTCACTATAGGGTCATT CGACGCCATCTTCATTCACAAAACTGGGA GGGACTTTAGAACAGAGGAGATAAAGATG CTCCCAGTTTTAACCTGGCGGCAGCGCAA GAAGATG (SEQ ID NO: 40)	
Toehold switch S6 DNA	GCGCTAATACGACTCACTATAGGGAGCCA GATTGCGATCGCCCTCCCACGTGCTCAGAT CGGACTTTAGAACAGAGGAGATAAAGATG GATCTGAGCACAACCTGGCGGCAGCGCAA GAAGATG (SEQ ID NO: 41)	
Toehold switch A1 DNA	GCGCTAATACGACTCACTATAGGGTCTGA TGGGTCCGCAGCCAACCTCGTCCCAGAGG TCGGACTTTAGAACAGAGGAGATAAAGAT GGACCTCTGGGAAACCTGGCGGCAGCGCA AGAAGATG (SEQ ID NO: 42)	
Toehold switch A2 DNA	GCGCTAATACGACTCACTATAGGGTGGGA GGGCGATCGCAATCTGGCTCCCAGTTTTGT GGACTTTAGAACAGAGGAGATAAAGATGA CAAAACTGGGAACCTGGCGGCAGCGCAAG AAGATG (SEQ ID NO: 43)	
Toehold switch A3 DNA	GCGCTAATACGACTCACTATAGGGTGTGA ATGAAGATGGCGTCGAATGACGCCAACCC ATGGACTTTAGAACAGAGGAGATAAAGAT GATGGGTTGGCGAACCTGGCGGCAGCGCA AGAAGATG (SEQ ID NO: 44)	
Toehold switch A4 DNA	GCGCTAATACGACTCACTATAGGGAGATC TGAGCACGTGGGAGGGCGATCGCAATCTG GCGGACTTTAGAACAGAGGAGATAAAGAT GGCCAGATTGCGAACCTGGCGGCAGCGCA AGAAGATG (SEQ ID NO: 45)	
Toehold switch A5 DNA	GCGCTAATACGACTCACTATAGGGATCGC AATCTGGCTCCCAGTTTTGTGAATGAAGAT GGGACTTTAGAACAGAGGAGATAAAGATG CATCTTCATTCAACCTGGCGGCAGCGCAA GAAGATG (SEQ ID NO: 46)	
Toehold switch A6 DNA	GCGCTAATACGACTCACTATAGGGTCGAA TGACGCCAACCCATCTGATGGGTCCGCAG CCGGACTTTAGAACAGAGGAGATAAAGAT GGGCTGCGGACCAACCTGGCGGCAGCGCA AGAAGATG (SEQ ID NO: 47)	
pCOLA_seq_fwd	CGTTACTGGTTTCACATTCACCACCC (SEQ ID NO: 48)	Sequencing primer used for confirming sequence of toehold switch sensors inserted in pCOLA, pCDF, pACYC expression vectors.
pET15b_seq_fwd1	CCTGCCACCATACCCACGC (SEQ ID NO: 49)	Sequencing primer used for confirming sequence of genes inserted into pET15b vectors within the multiple cloning site region.

[0084] Preparation of paper-based cell-free systems: Cell-free transcription-translation systems (NEB, PURExpress) were prepared for freeze-drying with the following components by volume: cell-free solution A, 40%; cell-free solution B, 30%; RNase Inhibitor (Roche, 03335402001, distributed by MilliporeSigma), 2%; chlorophenol red-b-D-galactopyranoside (Roche, 10884308001, distributed by MilliporeSigma, 24 mg/ml), 2.5%; with the remaining volume reserved for toehold switch DNA, water, and lacZω peptide added to a final concentration of 2 μM. When testing the toehold switches expressed from a plasmid, the plasmid DNA was added to the cell-free reaction mix to a final concentration of 30 ng/μL. When testing toehold switches

expressed from linear DNA, the DNA was added to the cell-free reaction mix to a final concentration of 33 nM.

[0085] Filter paper (Whatman, 1442-042) for housing the cell-free reactions was first blocked with 5% bovine serum albumin (BSA) overnight. After blocking, the paper was washed three times in water for 5 to 10 minutes. The paper was then heated to 50° C. for drying and cut into 2-mm diameter paper disks using a biopsy punch. The disks were transferred into 200-μL, PCR tubes and 1.8 μL of the cell-free reaction mix was applied to each disk. PCR tubes containing the paper disks were then flash frozen in liquid nitrogen and transferred into a lyophilizer to dry overnight. Measurements were performed on the resulting paper disks two to four days after the freeze-drying process was completed. The paper disks remained active for at least a month of room-temperature storage using conditions described previously[30], with the systems stored under nitrogen, shielded from light, and in the presence of silica gel desiccation packages.

[0086] Screening of norovirus-specific toehold switches: Norovirus target RNA was produced using T7 RNA polymerase-based transcription (Epicenter, ASF3257) from linearized DNA templates. 1.8 μL of a 5 μM solution of the target RNA was applied to a paper disk containing the embedded cell-free system and DNA for the toehold switch. The progress of the cell-free reaction was then monitored in a plate reader (Biotek, H1MF) at 37° C. in triplicate. The relative absorbance of the paper-based reactions at 575 nm wavelength or OD575 was calculated by taking the absorbance at 575 nm and subtracting from it the absorbance at 575 nm measured at the start of the reaction. This relative absorbance thus removes any absorbance contribution from the paper disk and the lacZ substrate chlorophenol red-b-D-galactopyranoside. The fold change in lacZ production rate was calculated by computing the rate of change in OD575 and dividing the rate obtained for the toehold switch in the presence of the target RNA by that obtained in the absence of the target RNA. The fold change in lacZ production rate was measured after one hour of cell-free reaction for assessment of the toehold switches. The change in OD575 or ΔOD575 was calculated by taking the OD575 for the reaction with the toehold switch and the target RNA and subtracting from it the OD575 for the reaction of the toehold switch without the target RNA. ΔOD575 was computed after two hours of cell-free reaction. Errors in OD575 were determined from the standard deviation of triplicate measurements. Errors in fold change lacZ production rate and ΔOD575 were determined by adding the relative and absolute errors of OD575 in quadrature, respectively. Welch's unequal variances t-test was used to calculate p-values for plate reader detection experiments with $p<0.05$ used as the cutoff to define a statistically significant result.

[0087] Isothermal amplification of norovirus RNA: For NASBA experiments, reaction buffer (Life Sciences, NECB-24; 33.5%), nucleotide mix (Life Sciences NECN-24; 16.5%), RNase inhibitor (Roche, 03335402001; 0.5%), 12.5 mM of each DNA primer (2%), nuclease free water (2.5%), and RNA amplicon (20%) were assembled at 4° C. and incubated at 65° C. for 2 min, followed by a 10 min incubation at 41° C. Enzyme Mix (Life Sciences NEC-1-24; 25%) was then added to the reaction (for a final volume of 5 μL), and the mixture was incubated at 41° C. for 2 hr. The amplified product was then diluted 1:6 in water and applied to paper disks containing the cell-free system and DNA for the toehold switch. Sequences of the primers used for NASBA and RT-RPA are provided in Table 5.

Norovirus isothermal amplification primers

Norovirus Genotype	Toehold Switch	Forward Primer	Reverse Primer
GI1.4	S1	AATTCTAATACGACTCACTATAG GGAGAAGGATTCTCAGATCTGAG CACGTGGGA (SEQ ID NO: 50)	ATTGTTGACCTCTGGGACGA (SEQ ID NO: 51)
	S2	AATTCTAATACGACTCACTATAG GGAGAAGGCAGGCAAGAGCCAA TGTTCAGA (SEQ ID NO: 52)	CTCATTGTTGACCTCTGGGA (SEQ ID NO: 53)
	S6	AATTCTAATACGACTCACTATAG GGAGAAGGGCAAGAGCCAATGTT CAGATGGA (SEQ ID NO: 54)	CTCATTGTTGACCTCTGGGA (SEQ ID NO: 55)
	A1	AATTCTAATACGACTCACTATAG GGAGAAGGGCTCCAAAGCCATAA CCTCA (SEQ ID NO: 56)	GCAAGAGCCAATGTTCAGATGG A (SEQ ID NO: 57)
	A2	AATTCTAATACGACTCACTATAG GGAGAAGGCTCATTGTTGACCTC TGGGA (SEQ ID NO: 58)	GATGGATGAGATTCTCAGATCT GA (SEQ ID NO: 59)
	A4	AATTCTAATACGACTCACTATAG GGAGAAGGCTCATTGTTGACCTC TGGGA (SEQ ID NO: 60)	CAAGAGCCAATGTTCAGATGGA (SEQ ID NO: 61)
GII.6	S2	AATTCTAATACGACTCACTATAG GGAGAAGGCAGACAAGAGGCCA TGTTCA (SEQ ID NO: 62)	TCATTGTTGGCCTCTGGTACGA (SEQ ID NO: 63)

[0088] RT-RPA experiments used the commercial TwistAmp Basic RT kit (TwistDx). Reactions were prepared by combining 10 μM forward primer (4.8%), 10 μM reverse primer (4.8%), rehydration buffer, RNase Inhibitor (Roche, 03335402001; 4.4%), and RNA amplicon (22%) at room temperature and transferring the mixture to the freeze-dried reaction pellet. After mixing, 2.5 μL of 280 mM magnesium acetate (5%) was added to start the reaction and it was incubated at 41° C. for 5-7 minutes. The reaction tube was then inverted vigorously 8-10 times, spun down briefly, and returned to incubation at 41° C. for 2 hours. The amplified

product was then diluted 1:6 in water and applied to paper disks containing the cell-free system and DNA for the toehold switch.

[0089] For determination of assay detection limits, NASBA and RPA reactions were run in triplicate for each concentration of the target RNA or virus and applied to the paper-based toehold switch reactions as described above.

[0090] Synbody-based virus enrichment: A 30-μL volume of MyOne Streptavidin C1 streptavidin-coated magnetic beads (Life Technologies, U.S.A.), corresponding to 2.1×10^8 to 3.6×10^8 total beads, was added to Protein LowBind tubes (Eppendorf, U.S.A.). The bead storage solution was removed and the beads were washed three times with 1 mL of PBST (0.05% Tween 20 in 1× phosphate-buffered saline). The beads were then blocked with 3% BSA in PBST overnight at 4° C. The following day, the beads were suspended in fresh 3% BSA in PBST and blocked for an additional 2 hours. The beads were then washed three times with PBST and suspended in 30 μL of 1× PBS (phosphate-buffered saline) to yield a final suspension of blocked magnetic beads.

[0091] A dilution series of virus particles ranging from $1:10^3$ to $1:10^7$ was prepared by first taking a 1-μL aliquot of a norovirus GII.4 Sydney positive stool sample and diluting it into 1 mL of PBS. The resulting $1:10^3$ sample was serially diluted by factors of ten into PBS to generate the rest of the dilution series. Biotin-labelled synbody ASU1052 (described in Gupta et al., (2017) *Anal. Chem.* 89:7174-7181, which is incorporated herein by reference) was then added to a concentration of 1 μM into each diluted sample and incubated with shaking for 1 hour at room temperature. The solutions were then added to the blocked streptavidin-coated magnetic beads and shaken for an additional 15 minutes at room temperature. The beads were washed three times with PBST and one time with PBS and then suspended with 50 μL water. The beads were incubated for 2 min at 95° C. to release the viral RNA for analysis. 50 μL of each stool dilution was also incubated for 2 minutes at 95° C. and used for comparison.

[0092] For cross-reactivity testing and tests of the assay against the GII.6 genotype, 1 μL of GII.4, GII.6, and GI.6 positive stool samples, as well as a norovirus-negative stool sample, were diluted into 1 ml of PBS and followed by the synbody enrichment procedure described above.

[0093] Results and Discussion

[0094] Design of Toehold Switches for Norovirus GII Detection

[0095] We first identified conserved sequence regions of the norovirus GII genome suitable for isothermal amplification and toehold-switch-based detection. Over 400 norovirus GII complete and partial genome sequences were downloaded from the NCBI database and aligned. A 200-nt target sequence that was highly conserved across the norovirus GII genomes was identified for subsequent amplification and detection experiments. This conserved sequence ran from the C-terminal region of the viral RNA-dependent RNA polymerase through to the N-terminal region of VP1, the major capsid protein.

[0096] Toehold switches for detection of the target sequence were then generated based on an updated design first applied to the detection of the Zika virus. The updated toehold switch design provided lower leakage compared to earlier toehold switches and was originally developed for evaluating AND logic expressions in *E. coli*. As illustrated in FIG. **2**A, binding of a cognate target RNA to the updated toehold switch unwinds the lower half of the switch RNA hairpin and leaves the conserved upper stem-loop intact. This upper stem-loop is sufficiently weak to expose the ribosomal binding site (RBS) to enable translation to occur. Unlike earlier toehold switch mRNA sensors, the updated systems do not employ an RNA refolding domain downstream of the start codon, which could hamper translation of the output gene.

[0097] Based on the modified operating mechanism of the toehold switches, we implemented an updated design selection algorithm to identify the toehold switches most likely to be effective at detecting the target RNA. This algorithm modelled the interaction of a series of toehold switches designed to bind along the target RNA in 1-nt increments using the NUPACK software package. Ensemble defect levels and the affinity of the toehold switch for the target RNA were used to select designs most likely to perform well. Since the target RNA can be transcribed in either the sense or antisense direction following amplification, the top six toehold switches for the sense and antisense target RNAs were selected for experimental testing (see Table 1).

[0098] Faster RNA Detection with Toehold Switches Using α-Complementation of lacZ

[0099] In previous work using paper-based cell-free systems, the lacZ enzyme has been used as the output gene for the toehold switch to produce a visible test result through cleavage of a chromogenic substrate. LacZ, however, at 3.1 kb in length is a relatively long reporter gene compared to alternatives such as GFP (0.75 kb) and mCherry (0.72 kb), which leads to several drawbacks. In particular, the longer length of lacZ means that a greater fraction of the cell-free system resources is consumed during transcription and translation, which weakens the output from the assay, and longer times are required for the protein to be synthesized and fold, which increases the time required for the test.

[0100] In response to the above limitations, we investigated using α-complementation of lacZ to decrease assay times and strengthen output from the cell-free transcription-translation reactions. Alpha-complementation is a widely applied technique often used for screening cloning vectors. It works by dividing the lacZ enzyme into two peptides termed α and ω (FIG. **2**B). The lacZ α-peptide (lacZα) consists of the first 50 to 59 residues from the N terminus of lacZ and the ω-peptide (lacZω) comprises the remaining ~970 lacZ residues. The complete lacZ must form a tetramer before it becomes catalytically active; however, lacZω cannot form a tetramer on its own as it lacks residues critical for assembly. As a result, both lacZα and lacZω must be expressed before complementation occurs and an active lacZ tetramer can assemble.

[0101] We thus implemented toehold switches that used lacZα as the output protein and added the much larger lacZω peptide as a pre-synthesized component to the paper-based cell-free reactions. Since lacZα is encoded in 180 bp, which is only ~6% of the length of the full lacZ gene, transcription and translation of each lacZα molecule should occur faster compared to lacZ and could in principle impose a substantially smaller burden on the cell-free system for each active lacZ tetramer formed. DNA encoding the norovirus-specific toehold switches was cloned into vectors upstream of the lacZα open reading frame. Following sequence confirmation, the resulting plasmids were tested in paper-based cell-free reactions supplemented with lacZω, and cleavage

of the chromogenic substrate chlorophenol red-b-D-galactopyranoside was monitored using a plate reader. FIGS. **2**C-**2**F shows the results of these experiments with six toehold switches named S1, S2, etc., for the sense orientation of the target RNA and six toehold switches named A1, A2, etc., for the antisense target orientation. All of the toehold switches were tested in parallel with reactions in which no target RNA was present. These experiments were then used to determine the fold change in the lacZ production rate and the ΔOD575 for each sensor. Three of the sense toehold switches provided ON/OFF ratios of approximately three or more (FIG. **2**C) and displayed a change in absorbance at 575 nm (ΔOD575) of at least 0.4 (FIG. **2**D), which can be discerned by eye. The toehold switches for the antisense target provided better performance overall with ON/OFF ratios up to 12.6-fold for A1 (FIG. **2**E) and ΔOD575 up to 0.92 for A2 (FIG. **2**F). Although the in silico selection algorithm successfully generated functional toehold switches for the two norovirus targets, we only detected appreciable correlations between the scoring function and the toehold switches for the antisense target. The sense target devices showed no correlations with the scoring function. Analysis of the experimental data indicate that other combinations of ensemble defect parameters coupled with different weighting factors can provide more accurate predictions of device performance (see FIGS. **7**A-**7**L).

[0102] To determine the effect of a-complementation on detection speed, we took one of the better performing toehold switches, A2, and inserted it into a plasmid upstream of the full lacZ open reading frame. PCR was then used to amplify linear DNA fragments from both lacZα and full-length lacZ plasmids and equal concentrations of the two DNA products were tested in paper-based cell-free reactions in the presence of the norovirus target RNA. We observed a substantial increase in the speed of the colorimetric reaction for the lacZα systems compared to full-length lacZ (FIG. **2**G). Applying OD575=0.4 as the detection threshold, the lacZα reporter reached a positive result in 33 minutes compared to 56 minutes for the complete lacZ, which corresponds to a 40% reduction in detection time (see FIG. **8** for photographs of the paper-based reactions at different OD575 values). Since both reactions reach saturation and completely cleave the substrate within the two-hour measurement shown in FIG. **2**G, we attribute the increased speed of the reaction in these conditions to the faster folding time of lacZα compared to lacZ, rather than to any decrease in the burden on the cell-free reaction caused by the shorter reporter protein.

[0103] Isothermal Amplification Using NASBA and RT-RPA

[0104] Since the concentrations of norovirus in stool samples from symptomatic patients range from ~30 attomoles/liter (aM) to ~3 picomoles/liter (pM), the toehold switches cannot be efficiently activated by viral nucleic acids without an amplification step. We investigated the NASBA and RT-RPA isothermal amplification techniques to determine which provided the lowest limit of detection against the norovirus GII target RNA. The six toehold switches providing the highest ON/OFF ratios were selected for testing with amplified RNA. Since each sensor targeted different regions within the conserved target sequence, we evaluated different amplification primers for each sensor. One primer from each pair contained a 5' T7 promoter sequence so that the resulting amplicon could be transcribed into RNA for optimal detection using the corresponding toehold switch.

[0105] Toehold switches S2 and S6 provided the lowest detection limits in the amplification tests. Two-hour amplification reactions were run with synthetic norovirus GII target RNAs ranging in concentration from 220 femtomoles/liter (fM) to 0.2 aM. The SI prefix "femto" represents a factor of 10^{-15}, or in exponential notation, 1E-15. The amplified products were then diluted seven-fold and applied to the toehold switch reactions. For the RT-RPA reactions, both S2 and S6 toehold switches could detect down to 22 fM of the norovirus RNA with colorimetric outputs that could be readily discerned by eye (FIGS. **3**A, **3**B). Statistically significant concentrations as low as 2.2 fM could be detected from quantitative plate reader absorbance measurements for toehold switch S2 after 3 hours and toehold switch S6 after 1 hour.

[0106] NASBA tests provided improved detection limits compared to RPA. For toehold switch S6, we could discern concentrations down to 2 fM by eye within 2 hours and by plate reader within 1 hour (FIG. **3**C). Although toehold switch S2 was not one of the very top performers in the initial screen (FIG. **2**), it provided the lowest detection limit when coupled with NASBA. Experiments showed this sensor could detect down to 200 aM concentrations of the synthetic norovirus transcript (FIG. **3**D). In addition, the sensor enabled detection by eye in 60 minutes at the 200 am detection limit as shown in FIG. **3**E and by plate reader in 28 minutes. A concentration of 200 aM corresponds to 600 copies of the RNA template in the 5 μL NASBA reaction.

[0107] Diagnostic Validation with Active Norovirus

[0108] To validate the detection platform, we performed experiments with active norovirus samples and tested the assay for cross-reactivity against other potential pathogens. Following previous reports on norovirus and our earlier work on the Zika virus, we first evaluated a simple method for extracting viral RNA from infected stool samples using a brief heating step. A norovirus GII.4 Sydney positive stool sample was diluted 1:50 in PBS and heated for two minutes at 95° C. (FIG. **4**A). The same procedure was applied to a stool sample not infected with the virus and two additional stool samples containing norovirus GI.2 and GI.6. These heated samples, along with comparison unheated samples and a water-only negative control, were both amplified by NASBA over 2 hours and applied to a paper-based reaction with toehold switch S2. The unheated samples all yielded minimal changes in toehold switch output compared to the negative control. The OD575 of the heated sample with norovirus GII.4 Sydney increased to 1.13, while the OD575 of the other heated samples remained below 0.25 (FIG. **4**B). Thus, the simple heating method was effective at releasing RNA from norovirus particles and the assay was specific for norovirus GII.4 Sydney.

[0109] To further evaluate cross-reactivity, we extracted RNA from *E. coli, B. subtilis,* and a methicillin-resistant *S. aureus* (MRSA) strain and added the RNA at masses of 80.6 ng, 123.5 ng, and 100.8 ng, respectively, to the NASBA reaction. RNA was also extracted from stool samples containing norovirus GII.4 Sydney, GI.2, and GI.6 and added to the NASBA reaction at a concentration of approximately 20 fM. None of these samples of bacterial RNA nor the GI.2 and GI.6 norovirus genotypes were able to activate toehold

switch S2 for visual detection. The system was strongly activated by norovirus GII.4 Sydney RNA (FIG. **4**C).

[0110] Norovirus Enrichment Using a Synbody-Based Magnetic Bead Technique

[0111] The ability to identify norovirus in dilute solutions or from large solution volumes is valuable for improving diagnostic sensitivity and for confirming complete decontamination of an area following an outbreak. For instance, dilute liquids, such as cleaning solutions from kitchen and bathroom surfaces, can be tested for residual virus following cleanup. To this end, we employed a synbody-based magnetic bead capture assay to concentrate norovirus from dilute solutions (FIG. **5**A). Synbodies are synthetic bivalent affinity ligands composed of two 15- to 20-mer peptides screened to bind to the surface of a protein of interest. Synbodies have affinities and specificities similar to antibodies[41, 42]. Unlike antibodies, however, which often lose their affinity as norovirus strains evolve[1], synbodies have broad cross-affinity for multiple norovirus genotypes, which enables them to recognize a range of norovirus genotypes within both the GI and GII genogroups[35].

[0112] To capture and concentrate the virus, we took stool samples positive for norovirus GII.4 Sydney at a concentration of 270 fM as determined by qRT-PCR and prepared a series of higher dilutions ranging from $1:10^3$ to $1:10^7$ in PBS. Biotin-labelled synbody ASU1052, which was previously validated against multiple norovirus strains[35], and streptavidin-coated magnetic beads were added sequentially to the diluted samples with shaking at room temperature for 75 minutes total. After magnetic capture and washing, the beads were suspended with 50 μL of water and heated to 95° C. for 2 min to release the virus RNA. These virus samples, along with comparison ones heated but not subjected to synbody capture, were then amplified using NASBA and applied to paper-based cell-free systems containing toehold switch S2.

[0113] FIG. **5**B displays the absorbance change produced from the reactions after two hours with the two different sets of samples. For synbody-concentrated samples, norovirus could be detected by eye with dilution factors up to 10^5, which corresponds to a concentration of 2.7 aM. In contrast, none of the samples used directly and not subjected to concentration could be detected within two hours by eye. To further compare the two preparation methods, FIG. **5**C shows the absorbance change over time for several virus samples. The synbody-concentrated sample prepared from a 10^5 dilution crosses the eye-based detection threshold of OD575=0.4 in under two hours and provides a statistically significant positive signal in the plate reader after 66 minutes. The profile of the non-concentrated sample diluted 1000-fold nearly matches that of the synbody-concentrated sample diluted 10^6-fold over the full 4-hour measurement. Both samples cross the visual detection threshold after 3 hours and provide positive results from quantitative plate reader measurements in approximately 2 hours, which correspond to norovirus GII.4 Sydney detection limits of 270 aM and 270 zM for the non-concentrated and synbody-concentrated samples, respectively. The synbody-based concentration technique thus enables a 1000-fold improvement in the detection limit of the norovirus assay.

[0114] To determine if the assay could also be applied to closely related norovirus genotypes, we also tested the systems against a stool sample with the norovirus GII.6 genotype. Virus particles were enriched using the ASU1052 synbodies and subject to NASBA using the primers optimized for GII.4 Sydney amplification. Unfortunately, these primers were not effective for this genotype. Primers modified to match the GII.6 genome, however, enabled successful amplification. Despite the presence of some mismatches between toehold switch S2 and its binding site on the GII.6 amplicon (see FIGS. **9**A-**9**D), a visible OD575 signal was observed from paper-based reactions within two hours (FIG. **4**D). Thus, toehold switch S2 is capable of detecting amplicons from both the norovirus GII.4 Sydney and GII.4 genotypes.

[0115] We have demonstrated a paper-based assay for detection of norovirus that does not require expensive thermal cycling equipment, provides test results that can be read directly by eye, and employs toehold switch riboregulators to eliminate false positives caused by non-specific amplification. The assay enables visual detection of norovirus down to a concentration of 270 aM from clinical stool samples containing live norovirus particles from the GII.4 Sydney genotype. The addition of a virus capture and concentration step using synbodies enables a further 1000-fold improvement in the sensitivity of the assay, allowing concentrations as low as 270 zM to be detected by eye after a three-hour paper-based reaction. This work also demonstrates that paper-based transcription-translation systems can remain active upon exposure to samples diluted from stool and confirms that RPA products can be successfully detected in the cell-free reactions, albeit with a higher detection limit than comparison NASBA products.

[0116] The norovirus assay provides significant improvements in sensitivity compared to our previously reported diagnostic assay for the Zika virus[31]. The Zika virus test provided a 1 fM detection limit against synthetic target RNAs and detected the virus from plasma at a concentration of 2.8 fM. In contrast, the norovirus assay demonstrated a 5-fold lower detection limit of 200 aM against a synthetic target and was successfully applied to a stool sample with a 270 aM concentration of norovirus. Addition of the synbody concentration step thus yielded an overall 5000-fold improvement in the detection limit. The Zika virus is known to be present at very low levels in symptomatic patients, with serum concentrations ranging from 8 zM to 6.1 fM with an average of 160 aM[43]. These concentrations are 10- to 100-fold lower than those observed for patients with the related dengue and chikungunya viruses[44]. Accordingly, our synbody-based concentration methods could prove valuable for extending the existing Zika test to more carriers of the virus. While the Zika diagnostic was only applied to a plasma sample from a viremic rhesus macaque, we have also demonstrated in this work that the diagnostic platform can be used on human stool samples, which can be used to identify many other causes of acute gastrointestinal illness beyond norovirus.

[0117] Although our norovirus assay provides sufficient sensitivity for detection from clinical samples, at present it requires 3-6 hours of processing time to reach a test result, which is substantially longer than many other diagnostics that employ isothermal amplification. We expect that large reductions in assay time can be obtained by further optimization of the synbody-based enrichment technique, by designing toehold switches optimized for quicker and stronger output, and by implementing new reporter proteins with faster activation. Indeed, the substantial decrease in reaction time that we observed using a-complementation of lacZ

suggests that there is ample room for improvement using alternative reporters. Moreover, use of faster amplification techniques such as RT-RPA with improved primers or strand-displacement amplification (SDA) could further decrease the time to detection for the technique. We also expect that toehold switch dynamic range against pathogen RNAs can be improved with continued refinement of in silico selection algorithms. In particular, screening experiments examining larger numbers of toehold switches against diverse target RNAs will be essential for generating in silico design scoring functions that are able to accurately predict their performance when deployed in cell-free transcription-translation systems.

[0118] The assay can also be improved by reducing its cost. In addition to the ~$1/test price of the paper-based component of the assay[30], the per test costs of NASBA, streptavidin-coated magnetic beads, and biotinylated synbodies are $2.25, $5.38, and $0.10, respectively. The total cost in materials for the assay is thus $8.73 and the overall assay requires approximately 35 minutes of hands on time. A previous study in South Africa to assess GeneXpert cartridge costs has reported an average lab technician salary of $9.07/hr,[14] which brings the total assay cost to $14.02 with labor included. Materials costs for this estimate are based on retail prices for the components. It is likely that the quantities of magnetic beads used in the assay can be reduced substantially with further refinement of the experimental procedures, and materials costs can decrease with purchases at larger scales. Even without optimization of the assay toward reduced price, the total cost per assay remains lower than the $14.93 calculated for GeneXpert cartridges in South Africa where concessional pricing is in effect[14]. Furthermore, our assay does not require large initial expenditures for purchasing expensive equipment.

[0119] The continual emergence of new variants of norovirus means that our paper-based assay will need to be updated as other strains replace GII.4 Sydney to ensure that false negatives do not occur. For instance, the GII.P17-GII.17 norovirus strain has recently become predominant in Asia[10] and immunochromatographic tests, which were developed for the GII.4 strain, have demonstrated 1000-fold poorer detection limits against the emergent strain[45]. To reduce the probability of false negatives, our assay employs a target sequence that is well conserved across different GII strains, including GII.P17 and GII.17. The toehold switch S2 sensor is predicted by NUPACK simulations to tolerate several mismatches in the target RNA, particularly within the toehold region, and still expose the RBS and start codon to enable translation of the reporter gene (see FIGS. **9**A-**9**D). This resiliency against sequence variations is evidenced by the ability of device S2 to activate against the GII.6 strain (FIG. **5**D). In cases where there is larger sequence divergence, sensor mRNAs that employ multiple toehold switch hairpins upstream of a single output gene can be used to detect different norovirus strains or to compensate for locations with higher sequence variability to avoid false negatives. We have demonstrated that such OR logic systems can be used to detect six completely sequence-independent target RNAs using a single sensor mRNA in *E. coli*[36]. We expect that similar approaches can be used in the paper-based reactions and prove more parsimonious with cell-free systems resources than other implementations employing multiple independent mRNAs. Like other nucleic acid tests that employ amplification, false negatives can also occur when the amplification primers do not have sufficient homology with the target amplicon. Such sequence variability can be addressed using primers with degenerate bases at positions known to have high probability of sequence divergence.

[0120] Despite these areas for improvement, the reasonably low cost of the assay and its reliance on only inexpensive equipment enables it to be implemented in decentralized contexts such as remote clinics or cruise ships with trained operators. Furthermore, coupling the validated molecular components of the assay with companion hardware for incubation and readout[31] or liquid handling[46] has the potential to substantially reduce operator training requirements and lead to more widespread deployment in the future. Lastly, the demonstrated ability of synbodies and toehold switches to bind to proteins and nucleic acids, respectively, from a variety of different pathogens[30-32, 41, 42] indicates that our combined concentration and detection approach can be successfully applied to a diverse range of infectious agents.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 82

<210> SEQ ID NO 1
<211> LENGTH: 103
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: sythetic

<400> SEQUENCE: 1

gggccaucuu cauucacaaa acugggagcc agauugcgag gacuuuagaa cagaggagau 60

aaagaugucg caaucuggaa ccuggcggca gcgcaagaag aug 103

<210> SEQ ID NO 2
<211> LENGTH: 103
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

-continued

<400> SEQUENCE: 2

gggaucgccc ucccacgugc ucagaucuga gaaucucaug gacuuuagaa cagaggagau 60

aaagaugaug agauucucaa ccuggcggca gcgcaagaag aug 103

<210> SEQ ID NO 3
<211> LENGTH: 103
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 3

gggacaaaac ugggagccag auugcgaucg cccucccacg gacuuuagaa cagaggagau 60

aaagauggug ggagggcgaa ccuggcggca gcgcaagaag aug 103

<210> SEQ ID NO 4
<211> LENGTH: 103
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 4

gggcugggac gagguuggcu gcggacccau cagaugggug gacuuuagaa cagaggagau 60

aaagaugacc caucugauaa ccuggcggca gcgcaagaag aug 103

<210> SEQ ID NO 5
<211> LENGTH: 102
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 5

gggucauucg acgccaucuu cauucacaaa acugggaggg acuuuagaac agaggagaua 60

aagaugcucc caguuuuaac cuggcggcag cgcaagaaga ug 102

<210> SEQ ID NO 6
<211> LENGTH: 103
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 6

gggagccaga uugcgaucgc ccucccacgu gcucagaucg gacuuuagaa cagaggagau 60

aaagauggau cugagcacaa ccuggcggca gcgcaagaag aug 103

<210> SEQ ID NO 7
<211> LENGTH: 103
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 7

gggucugaug gguccgcagc caaccucguc ccagaggucg gacuuuagaa cagaggagau 60

aaagauggac cucugggaaa ccuggcggca gcgcaagaag aug 103

-continued

<210> SEQ ID NO 8
<211> LENGTH: 102
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 8

gggugggagg gcgaucgcaa ucuggcuccc aguuuugugg acuuuagaac agaggagaua 60

aagaugacaa aacugggaac cuggcggcag cgcaagaaga ug 102

<210> SEQ ID NO 9
<211> LENGTH: 103
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 9

gggugugaau gaagauggcg ucgaaugacg ccaacccaug gacuuuagaa cagaggagau 60

aaagaugaug gguuggcgaa ccuggcggca gcgcaagaag aug 103

<210> SEQ ID NO 10
<211> LENGTH: 103
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 10

gggagaucug agcacguggg agggcgaucg caaucuggcg gacuuuagaa cagaggagau 60

aaagauggcc agauugcgaa ccuggcggca gcgcaagaag aug 103

<210> SEQ ID NO 11
<211> LENGTH: 103
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 11

gggaucgcaa ucuggcuccc aguuuuguga augaagaugg gacuuuagaa cagaggagau 60

aaagaugcau cuucauucaa ccuggcggca gcgcaagaag aug 103

<210> SEQ ID NO 12
<211> LENGTH: 103
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 12

gggucgaaug acgccaaccc aucugauggg uccgcagccg gacuuuagaa cagaggagau 60

aaagaugggc ugcggaccaa ccuggcggca gcgcaagaag aug 103

<210> SEQ ID NO 13
<211> LENGTH: 198
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 13

auggauuuuu acgugcccag gcaagagcca auguucagau ggaugagauu cucagaucug 60

agcacguggg agggcgaucg caaucuggcu cccaguuuug ugaaugaaga uggcgucgaa 120

ugacgccaac ccaucugaug gguccgcagc caaccucguc ccagagguca acaaugaggu 180

uauggcuuug gagcccgu 198

<210> SEQ ID NO 14
<211> LENGTH: 198
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 14

acgggcucca aagccauaac cucauuguug accucuggga cgagguuggc ugcggaccca 60

ucagaugggu uggcgucauu cgacgccauc uucauucaca aaacugggag ccagauugcg 120

aucgcccucc cacgugcuca gaucugagaa ucucauccau cugaacauug gcucuugccu 180

gggcacguaa aaauccau 198

<210> SEQ ID NO 15
<211> LENGTH: 178
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 15

auggauuuuu augugcccag acaagaguca auguucagau ggaugagguu cucagaucua 60

agcacauggg agggcgaucg caaucuggcu cccaguuuug ugaaugaaga uggcgucgaa 120

ugacgccgcu ccaucuaaug auggugcugc uggucucgua ccagagggca acaacgag 178

<210> SEQ ID NO 16
<211> LENGTH: 178
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 16

auggauuuuu augugcccag acaagaguca auguucagau ggaugagguu cucagaucua 60

agcacauggg agggcgaucg caaucuggcu cccaguuuug ugaaugaaga uggcgucgaa 120

ugacgccgcu ccaucuaaug auggugcugc uggucucgua ccagagggca acaacgag 178

<210> SEQ ID NO 17
<211> LENGTH: 185
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 17

uggaguuuua ugugcccaga caagaggcca uguucaggug gaugagauuc ucugaccuca 60

gcacauggga gggcgaucgc aaucuugcuc ccgagggugu gaaugaagau ggcgucgaau 120

-continued

gacgcugcuc caucgaauga uggugcugcc aaccucguac cagaggccaa caaugagguu 180

auggc 185

<210> SEQ ID NO 18
<211> LENGTH: 28
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 18

ggacuuuaga acagaggaga uaaagaug 28

<210> SEQ ID NO 19
<211> LENGTH: 31
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 19

aaccuggcgg cagcgcaaga agaugcguaa a 31

<210> SEQ ID NO 20
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 20

taactagcat aacccttgg gg 22

<210> SEQ ID NO 21
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 21

catatggctg ccgcgcgg 18

<210> SEQ ID NO 22
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 22

agcggcctgg tgccgcgcgg cagccatatg cgtaaaatga ccatgattac ggattcact 59

<210> SEQ ID NO 23
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 23

tttagaggcc ccaaggggtt atgctagtta tttttgacac cagaccaact ggt 53

-continued

<210> SEQ ID NO 24
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 24

aacagttgcg cagcctga 18

<210> SEQ ID NO 25
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 25

ccagtgaatc cgtaatcatg gtcat 25

<210> SEQ ID NO 26
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 26

atgaccatga ttacggattc actggccgtc gcccgcaccg a 41

<210> SEQ ID NO 27
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 27

tcaggctgcg caactgttgg gaagggcgat cggtgcgggc 40

<210> SEQ ID NO 28
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 28

tagcataacc ccttggggc 19

<210> SEQ ID NO 29
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 29

gcgcaactgt tgggaagg 18

<210> SEQ ID NO 30
<211> LENGTH: 52

-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 30

cgcaccgatc gcccttccca acagttgcgc agcctgaatg gcgaatggta at 52

<210> SEQ ID NO 31
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 31

cccgtttaga ggccccaagg ggttatgcta ttattaccat tcgccattca gg 52

<210> SEQ ID NO 32
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 32

atgaccatga ttacggattc actggccgtc 30

<210> SEQ ID NO 33
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 33

ccggctaccg tagaaacgcg aatttactag cataagggag agcgtcgaga tc 52

<210> SEQ ID NO 34
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 34

ctagtaaatt cgcgtttcta cggtagccgg gcgctaatac gactcactat aggg 54

<210> SEQ ID NO 35
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 35

gacggccagt gaatccgtaa tcatggtcat cttcttgcgc tgccgccagg tt 52

<210> SEQ ID NO 36
<211> LENGTH: 124
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

-continued

<400> SEQUENCE: 36

gcgctaatac gactcactat agggccatct tcattcacaa aactgggagc cagattgcga 60

ggactttaga acagaggaga taaagatgtc gcaatctgga acctggcggc agcgcaagaa 120

gatg 124

<210> SEQ ID NO 37
<211> LENGTH: 124
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 37

gcgctaatac gactcactat agggatcgcc ctcccacgtg ctcagatctg agaatctcat 60

ggactttaga acagaggaga taaagatgat gagattctca acctggcggc agcgcaagaa 120

gatg 124

<210> SEQ ID NO 38
<211> LENGTH: 124
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 38

gcgctaatac gactcactat agggacaaaa ctgggagcca gattgcgatc gccctcccac 60

ggactttaga acagaggaga taaagatggt gggagggcga acctggcggc agcgcaagaa 120

gatg 124

<210> SEQ ID NO 39
<211> LENGTH: 124
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 39

gcgctaatac gactcactat agggctggga cgaggttggc tgcggaccca tcagatgggt 60

ggactttaga acagaggaga taaagatgac ccatctgata acctggcggc agcgcaagaa 120

gatg 124

<210> SEQ ID NO 40
<211> LENGTH: 123
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 40

gcgctaatac gactcactat agggtcattc gacgccatct tcattcacaa aactgggagg 60

gactttagaa cagaggagat aaagatgctc ccagttttaa cctggcggca gcgcaagaag 120

atg 123

<210> SEQ ID NO 41
<211> LENGTH: 124

-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 41

gcgctaatac gactcactat agggagccag attgcgatcg ccctcccacg tgctcagatc 60

ggactttaga acagaggaga taaagatgga tctgagcaca acctggcggc agcgcaagaa 120

gatg 124

<210> SEQ ID NO 42
<211> LENGTH: 124
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 42

gcgctaatac gactcactat agggtctgat gggtccgcag ccaacctcgt cccagaggtc 60

ggactttaga acagaggaga taaagatgga cctctgggaa acctggcggc agcgcaagaa 120

gatg 124

<210> SEQ ID NO 43
<211> LENGTH: 123
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 43

gcgctaatac gactcactat agggtgggag ggcgatcgca atctggctcc cagttttgtg 60

gactttagaa cagaggagat aaagatgaca aaactgggaa cctggcggca gcgcaagaag 120

atg 123

<210> SEQ ID NO 44
<211> LENGTH: 124
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 44

gcgctaatac gactcactat agggtgtgaa tgaagatggc gtcgaatgac gccaacccat 60

ggactttaga acagaggaga taaagatgat gggttggcga acctggcggc agcgcaagaa 120

gatg 124

<210> SEQ ID NO 45
<211> LENGTH: 124
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 45

gcgctaatac gactcactat agggagatct gagcacgtgg gagggcgatc gcaatctggc 60

ggactttaga acagaggaga taaagatggc cagattgcga acctggcggc agcgcaagaa 120

gatg 124

-continued

<210> SEQ ID NO 46
<211> LENGTH: 124
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 46

gcgctaatac gactcactat agggatcgca atctggctcc cagttttgtg aatgaagatg 60

ggactttaga acagaggaga taaagatgca tcttcattca acctggcggc agcgcaagaa 120

gatg 124

<210> SEQ ID NO 47
<211> LENGTH: 124
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 47

gcgctaatac gactcactat agggtcgaat gacgccaacc catctgatgg gtccgcagcc 60

ggactttaga acagaggaga taaagatggg ctgcggacca acctggcggc agcgcaagaa 120

gatg 124

<210> SEQ ID NO 48
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 48

cgttactggt ttcacattca ccaccc 26

<210> SEQ ID NO 49
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 49

cctgccacca tacccacgc 19

<210> SEQ ID NO 50
<211> LENGTH: 55
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 50

aattctaata cgactcacta tagggagaag gattctcaga tctgagcacg tggga 55

<210> SEQ ID NO 51
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

-continued

<400> SEQUENCE: 51

attgttgacc tctgggacga 20

<210> SEQ ID NO 52
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 52

aattctaata cgactcacta tagggagaag gcaggcaaga gccaatgttc aga 53

<210> SEQ ID NO 53
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 53

ctcattgttg acctctggga 20

<210> SEQ ID NO 54
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 54

aattctaata cgactcacta tagggagaag ggcaagagcc aatgttcaga tgga 54

<210> SEQ ID NO 55
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 55

ctcattgttg acctctggga 20

<210> SEQ ID NO 56
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 56

aattctaata cgactcacta tagggagaag ggctccaaag ccataacctc a 51

<210> SEQ ID NO 57
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 57

gcaagagcca atgttcagat gga 23

-continued

<210> SEQ ID NO 58
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 58

aattctaata cgactcacta tagggagaag gctcattgtt gacctctggg a 51

<210> SEQ ID NO 59
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 59

gatggatgag attctcagat ctga 24

<210> SEQ ID NO 60
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 60

aattctaata cgactcacta tagggagaag gctcattgtt gacctctggg a 51

<210> SEQ ID NO 61
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 61

caagagccaa tgttcagatg ga 22

<210> SEQ ID NO 62
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 62

aattctaata cgactcacta tagggagaag gcagacaaga ggccatgttc a 51

<210> SEQ ID NO 63
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 63

tcattgttgg cctctggtac ga 22

<210> SEQ ID NO 64
<211> LENGTH: 70

-continued

<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (29)..(39)
<223> OTHER INFORMATION: n is a, c, g, or u

<400> SEQUENCE: 64

ggacuuuaga acagaggaga uaaagaugnn nnnnnnnnna accuggcggc agcgcaagaa 60

gaugcguaaa 70

<210> SEQ ID NO 65
<211> LENGTH: 36
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 65

augagauucu cagaucugag cacgugggag ggcgau 36

<210> SEQ ID NO 66
<211> LENGTH: 36
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 66

augagauucu cugaccucag cacaugggag ggcgau 36

<210> SEQ ID NO 67
<211> LENGTH: 36
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 67

augagguucu cagaucuaag cacaugggag ggcgau 36

<210> SEQ ID NO 68
<211> LENGTH: 36
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 68

augagguucu cagaucuaag cacaugggag ggcgau 36

<210> SEQ ID NO 69
<211> LENGTH: 8746
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 69

ttcttgaaga cgaaagggcc tcgtgatacg cctatttta taggttaatg tcatgataat 60

aatggtttct tagacgtcag gtggcacttt tcggggaaat gtgcgcggaa cccctatttg 120

-continued

tttatttttc taaatacatt caaatatgta tccgctcatg agacaataac cctgataaat 180

gcttcaataa tattgaaaaa ggaagagtat gagtattcaa catttccgtg tcgcccttat 240

tccctttttt gcggcatttt gccttcctgt ttttgctcac ccagaaacgc tggtgaaagt 300

aaaagatgct gaagatcagt tgggtgcacg agtgggttac atcgaactgg atctcaacag 360

cggtaagatc cttgagagtt ttcgccccga agaacgtttt ccaatgatga gcacttttaa 420

agttctgcta tgtggcgcgg tattatcccg tgttgacgcc gggcaagagc aactcggtcg 480

ccgcatacac tattctcaga atgacttggt tgagtactca ccagtcacag aaaagcatct 540

tacggatggc atgacagtaa gagaattatg cagtgctgcc ataaccatga gtgataacac 600

tgcggccaac ttacttctga caacgatcgg aggaccgaag gagctaaccg ctttttgca 660

caacatgggg gatcatgtaa ctcgccttga tcgttgggaa ccggagctga atgaagccat 720

accaaacgac gagcgtgaca ccacgatgcc tgcagcaatg gcaacaacgt tgcgcaaact 780

attaactggc gaactactta ctctagcttc ccggcaacaa ttaatagact ggatggaggc 840

ggataaagtt gcaggaccac ttctgcgctc ggcccttccg gctggctggt ttattgctga 900

taaatctgga gccggtgagc gtgggtctcg cggtatcatt gcagcactgg ggccagatgg 960

taagccctcc cgtatcgtag ttatctacac gacggggagt caggcaacta tggatgaacg 1020

aaatagacag atcgctgaga taggtgcctc actgattaag cattggtaac tgtcagacca 1080

agtttactca tatatacttt agattgattt aaaacttcat ttttaattta aaaggatcta 1140

ggtgaagatc ctttttgata atctcatgac caaaatccct taacgtgagt tttcgttcca 1200

ctgagcgtca gaccccgtag aaaagatcaa aggatcttct tgagatcctt ttttctgcg 1260

cgtaatctgc tgcttgcaaa caaaaaaacc accgctacca gcggtggttt gtttgccgga 1320

tcaagagcta ccaactcttt ttccgaaggt aactggcttc agcagagcgc agataccaaa 1380

tactgtcctt ctagtgtagc cgtagttagg ccaccacttc aagaactctg tagcaccgcc 1440

tacatacctc gctctgctaa tcctgttacc agtggctgct gccagtggcg ataagtcgtg 1500

tcttaccggg ttggactcaa gacgatagtt accggataag gcgcagcggt cgggctgaac 1560

gggggttcg tgcacacagc ccagcttgga gcgaacgacc tacaccgaac tgagatacct 1620

acagcgtgag ctatgagaaa gcgccacgct tcccgaaggg agaaaggcgg acaggtatcc 1680

ggtaagcggc agggtcggaa caggagagcg cacgagggag cttccagggg gaaacgcctg 1740

gtatctttat agtcctgtcg ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg 1800

ctcgtcaggg gggcggagcc tatggaaaaa cgccagcaac gcggcctttt tacggttcct 1860

ggccttttgc tggccttttg ctcacatgtt ctttcctgcg ttatcccctg attctgtgga 1920

taaccgtatt accgccttttg agtgagctga taccgctcgc cgcagccgaa cgaccgagcg 1980

cagcgagtca gtgagcgagg aagcggaaga gcgcctgatg cggtattttc tccttacgca 2040

tctgtgcggt atttcacacc gcatatatgg tgcactctca gtacaatctg ctctgatgcc 2100

gcatagttaa gccagtatac actccgctat cgctacgtga ctgggtcatg gctgcgcccc 2160

gacacccgcc aacacccgct gacgcgccct gacgggcttg tctgctcccg gcatccgctt 2220

acagacaagc tgtgaccgtc tccgggagct gcatgtgtca gaggttttca ccgtcatcac 2280

cgaaacgcgc gaggcagctg cggtaaagct catcagcgtg gtcgtgaagc gattcacaga 2340

tgtctgcctg ttcatccgcg tccagctcgt tgagtttctc cagaagcgtt aatgtctggc 2400

-continued

```
ttctgataaa gcgggccatg ttaagggcgg ttttttcctg tttggtcact gatgcctccg   2460

tgtaaggggg atttctgttc atgggggtaa tgataccgat gaaacgagag aggatgctca   2520

cgatacgggt tactgatgat gaacatgccc ggttactgga acgttgtgag ggtaaacaac   2580

tggcggtatg gatgcggcgg gaccagagaa aaatcactca gggtcaatgc cagcgcttcg   2640

ttaatacaga tgtaggtgtt ccacagggta gccagcagca tcctgcgatg cagatccgga   2700

acataatggt gcagggcgct gacttccgcg tttccagact ttacgaaaca cggaaaccga   2760

agaccattca tgttgttgct caggtcgcag acgttttgca gcagcagtcg cttcacgttc   2820

gctcgcgtat cggtgattca ttctgctaac cagtaaggca accccgccag cctagccggg   2880

tcctcaacga caggagcacg atcatgcgca cccgtggcca ggacccaacg ctgcccgaga   2940

tgcgccgcgt gcggctgctg gagatggcgg acgcgatgga tatgttctgc caagggttgg   3000

tttgcgcatt cacagttctc cgcaagaatt gattggctcc aattcttgga gtggtgaatc   3060

cgttagcgag gtgccgccgg cttccattca ggtcgaggtg gcccggctcc atgcaccgcg   3120

acgcaacgcg gggaggcaga caaggtatag ggcggcgcct acaatccatg ccaacccgtt   3180

ccatgtgctc gccgaggcgg cataaatcgc cgtgacgatc agcggtccag tgatcgaagt   3240

taggctggta agagccgcga gcgatccttg aagctgtccc tgatggtcgt catctacctg   3300

cctggacagc atggcctgca acgcgggcat cccgatgccg ccggaagcga gaagaatcat   3360

aatggggaag gccatccagc ctcgcgtcgc gaacgccagc aagacgtagc ccagcgcgtc   3420

ggccgccatg ccggcgataa tggcctgctt ctcgccgaaa cgtttggtgg cgggaccagt   3480

gacgaaggct tgagcgaggg cgtgcaagat tccgaatacc gcaagcgaca ggccgatcat   3540

cgtcgcgctc cagcgaaagc ggtcctcgcc gaaaatgacc cagagcgctg ccggcacctg   3600

tcctacgagt tgcatgataa agaagacagt cataagtgcg gcgacgatag tcatgccccg   3660

cgcccaccgg aaggagctga ctgggttgaa ggctctcaag ggcatcggtc gagatcccgg   3720

tgcctaatga gtgagctaac ttacattaat tgcgttgcgc tcactgcccg ctttccagtc   3780

gggaaacctg tcgtgccagc tgcattaatg aatcggccaa cgcgcgggga gaggcggttt   3840

gcgtattggg cgccagggtg gtttttcttt tcaccagtga gacgggcaac agctgattgc   3900

ccttcaccgc ctggccctga gagagttgca gcaagcggtc cacgctggtt tgccccagca   3960

ggcgaaaatc ctgtttgatg gtggttaacg gcgggatata acatgagctg tcttcggtat   4020

cgtcgtatcc cactaccgag atatccgcac caacgcgcag cccggactcg gtaatggcgc   4080

gcattgcgcc cagcgccatc tgatcgttgg caaccagcat cgcagtggga acgatgccct   4140

cattcagcat ttgcatggtt tgttgaaaac cggacatggc actccagtcg ccttcccgtt   4200

ccgctatcgg ctgaatttga ttgcgagtga gatatttatg ccagccagcc agacgcagac   4260

gcgccgagac agaacttaat gggcccgcta acagcgcgat ttgctggtga cccaatgcga   4320

ccagatgctc cacgcccagt cgcgtaccgt cttcatggga gaaaataata ctgttgatgg   4380

gtgtctggtc agagacatca agaaataacg ccggaacatt agtgcaggca gcttccacag   4440

caatggcatc ctggtcatcc agcggatagt taatgatcag cccactgacg cgttgcgcga   4500

gaagattgtg caccgccgct ttacaggctt cgacgccgct tcgttctacc atcgacacca   4560

ccacgctggc acccagttga tcggcgcgag atttaatcgc cgcgacaatt tgcgacggcg   4620

cgtgcagggc cagactggag gtggcaacgc caatcagcaa cgactgtttg cccgccagtt   4680
```

-continued

gttgtgccac gcggttggga atgtaattca gctccgccat cgccgcttcc acttttccc 4740

gcgttttcgc agaaacgtgg ctggcctggt tcaccacgcg ggaaacggtc tgataagaga 4800

caccggcata ctctgcgaca tcgtataacg ttactggttt cacattcacc accctgaatt 4860

gactctcttc cgggcgctat catgccatac cgcgaaaggt tttgcgccat tcgatggtgt 4920

ccgggatctc gacgctctcc cttatgcgac tcctgcatta ggaagcagcc cagtagtagg 4980

ttgaggccgt tgagcaccgc cgccgcaagg aatggtgcat gcaaggagat ggcgcccaac 5040

agtcccccgg ccacggggcc tgccaccata cccacgccga aacaagcgct catgagcccg 5100

aagtggcgag cccgatcttc cccatcggtg atgtcggcga tataggcgcc agcaaccgca 5160

cctgtggcgc cggtgatgcc ggccacgatg cgtccggcgt agaggatcga gatctcgatc 5220

ccgcgaaatt aatacgactc actatagggg aattgtgagc ggataacaat tcccctctag 5280

aaataatttt gtttaacttt aagaaggaga tataccatgg gcagcagcca tcatcatcat 5340

catcacagca gcggcctggt gccgcgcggc agccatatgc gtaaaatgac catgattacg 5400

gattcactgg ccgtcgtttt acaacgtcgt gactgggaaa accctggcgt tacccaactt 5460

aatcgccttg cagcacatcc ccctttcgcc agctggcgta atagcgaaga ggcccgcacc 5520

gatcgccctt cccaacagtt gcgcagcctg aatggcgaat ggcgctttgc ctggtttccg 5580

gcaccagaag cggtgccgga aagctggctg gagtgcgatc ttcctgaggc cgatactgtc 5640

gtcgtccccc caaactggca gatgcacggt tacgatgcgc ccatctacac caacgtgacc 5700

tatcccatta cggtcaatcc gccgtttgtt cccacggaga atccgacggg ttgttactcg 5760

ctcacattta atgttgatga aagctggcta caggaaggcc agacgcgaat tatttttgat 5820

ggcgttaact cggcgtttca tctgtggtgc aacgggcgct gggtcggtta cggccaggac 5880

agtcgtttgc cgtctgaatt tgacctgagc gcatttttac gcgccggaga aaaccgcctc 5940

gcggtgatgg tgctgcgctg gagtgacggc agttatctgg aagatcagga tatgtggcgg 6000

atgagcggca ttttccgtga cgtctcgttg ctgcataaac cgactacaca aatcagcgat 6060

ttccatgttg ccactcgctt taatgatgat ttcagccgcg ctgtactgga ggctgaagtt 6120

cagatgtgcg gcgagttgcg tgactaccta cgggtaacag tttctttatg gcagggtgaa 6180

acgcaggtcg ccagcggcac cgcgcctttc ggcggtgaaa ttatcgatga gcgtggtggt 6240

tatgccgatc gcgtcacact acgtctgaac gtcgaaaacc cgaaactgtg gagcgccgaa 6300

atcccgaatc tctatcgtgc ggtggttgaa ctgcacaccg ccgacggcac gctgattgaa 6360

gcagaagcct gcgatgtcgg tttccgcgag gtgcggattg aaaatggtct gctgctgctg 6420

aacggcaagc cgttgctgat tcgaggcgtt aaccgtcacg agcatcatcc tctgcatggt 6480

caggtcatgg atgagcagac gatggtgcag gatatcctgc tgatgaagca gaacaacttt 6540

aacgccgtgc gctgttcgca ttatccgaac catccgctgt ggtacacgct gtgcgaccgc 6600

tacggcctgt atgtggtgga tgaagccaat attgaaaccc acggcatggt gccaatgaat 6660

cgtctgaccg atgatccgcg ctggctaccg gcgatgagcg aacgcgtaac gcgaatggtg 6720

cagcgcgatc gtaatcaccc gagtgtgatc atctggtcgc tggggaatga atcaggccac 6780

ggcgctaatc acgacgcgct gtatcgctgg atcaaatctg tcgatccttc ccgcccggtg 6840

cagtatgaag gcggcggagc cgacaccacg gccaccgata ttatttgccc gatgtacgcg 6900

cgcgtggatg aagaccagcc cttcccggct gtgccgaaat ggtccatcaa aaaatggctt 6960

-continued

tcgctacctg gagagacgcg cccgctgatc ctttgcgaat acgcccacgc gatgggtaac 7020

agtcttggcg gtttcgctaa atactggcag gcgtttcgtc agtatccccg tttacagggc 7080

ggcttcgtct gggactgggt ggatcagtcg ctgattaaat atgatgaaaa cggcaacccg 7140

tggtcggctt acggcggtga ttttggcgat acgccgaacg atcgccagtt ctgtatgaac 7200

ggtctggtct ttgccgaccg cacgccgcat ccagcgctga cggaagcaaa acaccagcag 7260

cagtttttcc agttccgttt atccgggcaa accatcgaag tgaccagcga atacctgttc 7320

cgtcatagcg ataacgagct cctgcactgg atggtggcgc tggatggtaa gccgctggca 7380

agcggtgaag tgcctctgga tgtcgctcca caaggtaaac agttgattga actgcctgaa 7440

ctaccgcagc cggagagcgc cgggcaactc tggctcacag tacgcgtagt gcaaccgaac 7500

gcgaccgcat ggtcagaagc cgggcacatc agcgcctggc agcagtggcg tctggcggaa 7560

aacctcagtg tgacgctccc cgccgcgtcc cacgccatcc cgcatctgac caccagcgaa 7620

atggattttt gcatcgagct gggtaataag cgttggcaat ttaaccgcca gtcaggcttt 7680

ctttcacaga tgtggattgg cgataaaaaa caactgctga cgccgctgcg cgatcagttc 7740

acccgtgcac cgctggataa cgacattggc gtaagtgaag cgacccgcat tgaccctaac 7800

gcctgggtcg aacgctggaa ggcggcgggc cattaccagg ccgaagcagc gttgttgcag 7860

tgcacggcag atacacttgc tgatgcggtg ctgattacga ccgctcacgc gtggcagcat 7920

caggggaaaa ccttatttat cagccggaaa acctaccgga ttgatggtag tggtcaaatg 7980

gcgattaccg ttgatgttga agtggcgagc gatacaccgc atccggcgcg gattggcctg 8040

aactgccagc tggcgcaggt agcagagcgg gtaaactggc tcggattagg gccgcaagaa 8100

aactatcccg accgccttac tgccgcctgt tttgaccgct gggatctgcc attgtcagac 8160

atgtataccc cgtacgtctt cccgagcgaa aacggtctgc gctgcgggac gcgcgaattg 8220

aattatggcc cacaccagtg gcgcggcgac ttccagttca acatcagccg ctacagtcaa 8280

cagcaactga tggaaaccag ccatcgccat ctgctgcacg cggaagaagg cacatggctg 8340

aatatcgacg gtttccatat ggggattggt ggcgacgact cctggagccc gtcagtatcg 8400

gcggaattcc agctgagcgc cggtcgctac cattaccagt tggtctggtg tcaaaaataa 8460

ctagcataac cccttggggc ctctaaatag cataacccct tggggcctct aaacgggtct 8520

tgaggggttt tttgctgaaa ggaggaacta tatccggata tcccgcaaga ggcccggcag 8580

taccggcata accaagccta tgcctacagc atccagggtg acggtgccga ggatgacgat 8640

gagcgcattg ttagatttca tacacggtgc ctgactgcgt tagcaattta actgtgataa 8700

actaccgcat taaagcttat cgatgataag ctgtcaaaca tgagaa 8746

<210> SEQ ID NO 70
<211> LENGTH: 3554
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 70

ctagtaaatt cgcgtttcta cggtagccgg gcgctaatac gactcactat agggtctgat 60

gggtccgcag ccaacctcgt cccagaggtc ggactttaga acagaggaga taaagatgga 120

cctctgggaa acctggcggc agcgcaagaa gatgaccatg attacggatt cactggccgt 180

-continued

```
cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc caacttaatc gccttgcagc    240
acatcccccc ttcgccagct ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca    300
acagttgcgc agcctgaatg gcgaatggta ataatagcat aacccccttgg ggcctctaaa    360
cgggtcttga ggggtttttt gctgaaacct caggcatttg agaagcacac ggtcacactg    420
cttccggtag tcaataaacc ggtaaaccag caatagacat aagcggctat ttaacgaccc    480
tgccctgaac cgacgacaag ctgacgaccg ggtctccgca agtggcactt ttcggggaaa    540
tgtgcgcgga acccctattt gtttatttttt ctaaatacat tcaaatatgt atccgctcat    600
gaattaattc ttagaaaaac tcatcgagca tcaaatgaaa ctgcaattta ttcatatcag    660
gattatcaat accatatttt tgaaaaagcc gtttctgtaa tgaaggagaa aactcaccga    720
ggcagttcca taggatggca agatcctggt atcggtctgc gattccgact cgtccaacat    780
caatacaacc tattaatttc ccctcgtcaa aaataaggtt atcaagtgag aaatcaccat    840
gagtgacgac tgaatccggt gagaatggca aaagtttatg catttctttc cagacttgtt    900
caacaggcca gccattacgc tcgtcatcaa aatcactcgc atcaaccaaa ccgttattca    960
ttcgtgattg cgcctgagcg agacgaaata cgcggtcgct gttaaaagga caattacaaa   1020
caggaatcga atgcaaccgg cgcaggaaca ctgccagcgc atcaacaata ttttcacctg   1080
aatcaggata ttcttctaat acctggaatg ctgttttccc ggggatcgca gtggtgagta   1140
accatgcatc atcaggagta cggataaaat gcttgatggt cggaagaggc ataaattccg   1200
tcagccagtt tagtctgacc atctcatctg taacatcatt ggcaacgcta cctttgccat   1260
gtttcagaaa caactctggc gcatcgggct tcccatacaa tcgatagatt gtcgcacctg   1320
attgcccgac attatcgcga gcccatttat acccatataa atcagcatcc atgttggaat   1380
ttaatcgcgg cctagagcaa gacgtttccc gttgaatatg gctcatactc ttccttttttc   1440
aatattattg aagcatttat cagggttatt gtctcatgag cggatacata tttgaatgta   1500
tttagaaaaa taaacaaata ggcatgctag cgcagaaacg tcctagaaga tgccaggagg   1560
atacttagca gagagacaat aaggccggag cgaagccgtt tttccatagg ctccgccccc   1620
ctgacgaaca tcacgaaatc tgacgctcaa atcagtggtg gcgaaacccg acaggactat   1680
aaagatacca ggcgtttccc cctgatggct ccctcttgcg ctctcctgtt cccgtcctgc   1740
ggcgtccgtg ttgtggtgga ggctttaccc aaatcaccac gtcccgttcc gtgtagacag   1800
ttcgctccaa gctgggctgt gtgcaagaac cccccgttca gcccgactgc tgcgccttat   1860
ccggtaacta tcatcttgag tccaacccgg aaagacacga caaaacgcca ctggcagcag   1920
ccattggtaa ctgagaatta gtggatttag atatcgagag tcttgaagtg gtggcctaac   1980
agaggctaca ctgaaaggac agtatttggt atctgcgctc cactaaagcc agttaccagg   2040
ttaagcagtt ccccaactga cttaaccttc gatcaaaccg cctccccagg cggttttttc   2100
gtttacagag caggagatta cgacgatcgt aaaaggatct caagaagatc ctttacggat   2160
tcccgacacc atcactctag atttcagtgc aatttatctc ttcaaatgta gcacctgaag   2220
tcagccccat acgatataag ttgtaattct catgttagtc atgccccgcg cccaccggaa   2280
ggagctgact gggttgaagg ctctcaaggg catcggtcga gatcccggtg cctaatgagt   2340
gagctaactt acattaattg cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc   2400
gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg   2460
```

-continued

ccagggtggt ttttcttttc accagtgaga cgggcaacag ctgattgccc ttcaccgcct 2520

ggccctgaga gagttgcagc aagcggtcca cgctggtttg ccccagcagg cgaaaatcct 2580

gtttgatggt ggttaacggc gggatataac atgagctgtc ttcggtatcg tcgtatccca 2640

ctaccgagat gtccgcacca acgcgcagcc cggactcggt aatggcgcgc attgcgccca 2700

gcgccatctg atcgttggca accagcatcg cagtgggaac gatgccctca ttcagcattt 2760

gcatggtttg ttgaaaaccg gacatggcac tccagtcgcc ttcccgttcc gctatcggct 2820

gaatttgatt gcgagtgaga tatttatgcc agccagccag acgcagacgc gccgagacag 2880

aacttaatgg gcccgctaac agcgcgattt gctggtgacc caatgcgacc agatgctcca 2940

cgcccagtcg cgtaccgtct tcatgggaga aaataatact gttgatgggt gtctggtcag 3000

agacatcaag aaataacgcc ggaacattag tgcaggcagc ttccacagca atggcatcct 3060

ggtcatccag cggatagtta atgatcagcc cactgacgcg ttgcgcgaga agattgtgca 3120

ccgccgcttt acaggcttcg acgccgcttc gttctaccat cgacaccacc acgctggcac 3180

ccagttgatc ggcgcgagat ttaatcgccg cgacaatttg cgacggcgcg tgcagggcca 3240

gactggaggt ggcaacgcca atcagcaacg actgtttgcc cgccagttgt tgtgccacgc 3300

ggttgggaat gtaattcagc tccgccatcg ccgcttccac tttttcccgc gttttcgcag 3360

aaacgtggct ggcctggttc accacgcggg aaacggtctg ataagagaca ccggcatact 3420

ctgcgacatc gtataacgtt actggtttca cattcaccac cctgaattga ctctcttccg 3480

ggcgctatca tgccataccg cgaaaggttt tgcgccattc gatggtgtcc gggatctcga 3540

cgctctccct tatg 3554

<210> SEQ ID NO 71
<211> LENGTH: 6445
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 71

ctagtaaatt cgcgtttcta cggtagccgg gcgctaatac gactcactat agggtgggag 60

ggcgatcgca atctggctcc cagttttgtg gactttagaa cagaggagat aaagatgaca 120

aaactgggaa cctggcggca gcgcaagaag atgaccatga ttacggattc actggccgtc 180

gttttacaac gtcgtgactg ggaaaaccct ggcgttaccc aacttaatcg ccttgcagca 240

catccccctt tcgccagctg gcgtaatagc gaagaggccc gcaccgatcg cccttcccaa 300

cagttgcgca gcctgaatgg cgaatggcgc tttgcctggt ttccggcacc agaagcggtg 360

ccggaaagct ggctggagtg cgatcttcct gaggccgata ctgtcgtcgt cccctcaaac 420

tggcagatgc acggttacga tgcgcccatc tacaccaacg tgacctatcc cattacggtc 480

aatccgccgt ttgttcccac ggagaatccg acgggttgtt actcgctcac atttaatgtt 540

gatgaaagct ggctacagga aggccagacg cgaattattt ttgatggcgt taactcggcg 600

tttcatctgt ggtgcaacgg gcgctgggtc ggttacggcc aggacagtcg tttgccgtct 660

gaatttgacc tgagcgcatt tttacgcgcc ggagaaaacc gcctcgcggt gatggtgctg 720

cgctggagtg acggcagtta tctggaagat caggatatgt ggcggatgag cggcattttc 780

cgtgacgtct cgttgctgca taaaccgact acacaaatca gcgatttcca tgttgccact 840

-continued

cgctttaatg atgatttcag ccgcgctgta ctggaggctg aagttcagat gtgcggcgag 900

ttgcgtgact acctacgggt aacagtttct ttatggcagg gtgaaacgca ggtcgccagc 960

ggcaccgcgc ctttcggcgg tgaaattatc gatgagcgtg gtggttatgc cgatcgcgtc 1020

acactacgtc tgaacgtcga aaacccgaaa ctgtggagcg ccgaaatccc gaatctctat 1080

cgtgcggtgg ttgaactgca caccgccgac ggcacgctga ttgaagcaga agcctgcgat 1140

gtcggtttcc gcgaggtgcg gattgaaaat ggtctgctgc tgctgaacgg caagccgttg 1200

ctgattcgag gcgttaaccg tcacgagcat catcctctgc atggtcaggt catggatgag 1260

cagacgatgg tgcaggatat cctgctgatg aagcagaaca actttaacgc cgtgcgctgt 1320

tcgcattatc cgaaccatcc gctgtggtac acgctgtgcg accgctacgg cctgtatgtg 1380

gtggatgaag ccaatattga aacccacggc atggtgccaa tgaatcgtct gaccgatgat 1440

ccgcgctggc taccggcgat gagcgaacgc gtaacgcgaa tggtgcagcg cgatcgtaat 1500

cacccgagtg tgatcatctg gtcgctgggg aatgaatcag gccacggcgc taatcacgac 1560

gcgctgtatc gctggatcaa atctgtcgat ccttcccgcc cggtgcagta tgaaggcggc 1620

ggagccgaca ccacggccac cgatattatt tgcccgatgt acgcgcgcgt ggatgaagac 1680

cagcccttcc cggctgtgcc gaaatggtcc atcaaaaaat ggctttcgct acctggagag 1740

acgcgcccgc tgatcctttg cgaatacgcc cacgcgatgg gtaacagtct tggcggtttc 1800

gctaaatact ggcaggcgtt tcgtcagtat ccccgtttac agggcggctt cgtctgggac 1860

tgggtggatc agtcgctgat taaatatgat gaaaacggca acccgtggtc ggcttacggc 1920

ggtgattttg gcgatacgcc gaacgatcgc cagttctgta tgaacggtct ggtctttgcc 1980

gaccgcacgc cgcatccagc gctgacggaa gcaaaacacc agcagcagtt tttccagttc 2040

cgtttatccg ggcaaaccat cgaagtgacc agcgaatacc tgttccgtca tagcgataac 2100

gagctcctgc actggatggt ggcgctggat ggtaagccgc tggcaagcgg tgaagtgcct 2160

ctggatgtcg ctccacaagg taaacagttg attgaactgc ctgaactacc gcagccggag 2220

agcgccgggc aactctggct cacagtacgc gtagtgcaac cgaacgcgac cgcatggtca 2280

gaagccgggc acatcagcgc ctggcagcag tggcgtctgg cggaaaacct cagtgtgacg 2340

ctccccgccg cgtcccacgc catcccgcat ctgaccacca gcgaaatgga tttttgcatc 2400

gagctgggta ataagcgttg gcaatttaac cgccagtcag gctttctttc acagatgtgg 2460

attggcgata aaaaacaact gctgacgccg ctgcgcgatc agttcacccg tgcaccgctg 2520

gataacgaca ttggcgtaag tgaagcgacc cgcattgacc ctaacgcctg ggtcgaacgc 2580

tggaaggcgg cgggccatta ccaggccgaa gcagcgttgt tgcagtgcac ggcagataca 2640

cttgctgatg cggtgctgat tacgaccgct cacgcgtggc agcatcaggg gaaaacctta 2700

tttatcagcc ggaaaaccta ccggattgat ggtagtggtc aaatggcgat taccgttgat 2760

gttgaagtgg cgagcgatac accgcatccg gcgcggattg gcctgaactg ccagctggcg 2820

caggtagcag agcgggtaaa ctggctcgga ttagggccgc aagaaaacta tcccgaccgc 2880

cttactgccg cctgttttga ccgctgggat ctgccattgt cagacatgta taccccgtac 2940

gtcttcccga gcgaaaacgg tctgcgctgc gggacgcgcg aattgaatta tggcccacac 3000

cagtggcgcg gcgacttcca gttcaacatc agccgctaca gtcaacagca actgatggaa 3060

accagccatc gccatctgct gcacgcggaa gaaggcacat ggctgaatat cgacggtttc 3120

-continued

catatgggga ttggtggcga cgactcctgg agcccgtcag tatcggcgga attccagctg 3180

agcgccggtc gctaccatta ccagttggtc tggtgtcaaa aataatagca taacccctttg 3240

gggcctctaa acgggtcttg aggggttttt tgctgaaacc tcaggcattt gagaagcaca 3300

cggtcacact gcttccggta gtcaataaac cggtaaacca gcaatagaca taagcggcta 3360

tttaacgacc ctgccctgaa ccgacgacaa gctgacgacc gggtctccgc aagtggcact 3420

tttcggggaa atgtgcgcgg aacccctatt tgtttatttt tctaaataca ttcaaatatg 3480

tatccgctca tgaattaatt cttagaaaaa ctcatcgagc atcaaatgaa actgcaattt 3540

attcatatca ggattatcaa taccatattt ttgaaaaagc cgtttctgta atgaaggaga 3600

aaactcaccg aggcagttcc ataggatggc aagatcctgg tatcggtctg cgattccgac 3660

tcgtccaaca tcaatacaac ctattaattt cccctcgtca aaaataaggt tatcaagtga 3720

gaaatcacca tgagtgacga ctgaatccgg tgagaatggc aaaagtttat gcatttcttt 3780

ccagacttgt tcaacaggcc agccattacg ctcgtcatca aaatcactcg catcaaccaa 3840

accgttattc attcgtgatt gcgcctgagc gagacgaaat acgcggtcgc tgttaaaagg 3900

acaattacaa acaggaatcg aatgcaaccg gcgcaggaac actgccagcg catcaacaat 3960

attttcacct gaatcaggat attcttctaa tacctggaat gctgttttcc cggggatcgc 4020

agtggtgagt aaccatgcat catcaggagt acggataaaa tgcttgatgg tcggaagagg 4080

cataaattcc gtcagccagt ttagtctgac catctcatct gtaacatcat tggcaacgct 4140

accttttgcca tgtttcagaa acaactctgg cgcatcgggc ttcccataca atcgatagat 4200

tgtcgcacct gattgcccga cattatcgcg agcccattta tacccatata aatcagcatc 4260

catgttggaa tttaatcgcg gcctagagca agacgtttcc cgttgaatat ggctcatact 4320

cttccttttt caatattatt gaagcattta tcagggttat tgtctcatga gcggatacat 4380

atttgaatgt atttagaaaa ataaacaaat aggcatgcta gcgcagaaac gtcctagaag 4440

atgccaggag gatacttagc agagagacaa taaggccgga gcgaagccgt ttttccatag 4500

gctccgcccc cctgacgaac atcacgaaat ctgacgctca aatcagtggt ggcgaaaccc 4560

gacaggacta taaagatacc aggcgtttcc ccctgatggc tccctcttgc gctctcctgt 4620

tcccgtcctg cggcgtccgt gttgtggtgg aggctttacc caaatcacca cgtcccgttc 4680

cgtgtagaca gttcgctcca agctgggctg tgtgcaagaa ccccccgttc agcccgactg 4740

ctgcgcctta tccggtaact atcatcttga gtccaacccg gaaagacacg acaaaacgcc 4800

actggcagca gccattggta actgagaatt agtggattta gatatcgaga gtcttgaagt 4860

ggtggcctaa cagaggctac actgaaagga cagtatttgg tatctgcgct ccactaaagc 4920

cagttaccag gttaagcagt tccccaactg acttaacctt cgatcaaacc gcctccccag 4980

gcggtttttt cgtttacaga gcaggagatt acgacgatcg taaaaggatc tcaagaagat 5040

cctttacgga ttcccgacac catcactcta gatttcagtg caatttatct cttcaaatgt 5100

agcacctgaa gtcagcccca tacgatataa gttgtaattc tcatgttagt catgccccgc 5160

gcccaccgga aggagctgac tgggttgaag gctctcaagg gcatcggtcg agatcccggt 5220

gcctaatgag tgagctaact tacattaatt gcgttgcgct cactgcccgc tttccagtcg 5280

ggaaacctgt cgtgccagct gcattaatga atcggccaac gcgcggggag aggcggtttg 5340

cgtattgggc gccagggtgg tttttctttt caccagtgag acgggcaaca gctgattgcc 5400

-continued

cttcaccgcc tggccctgag agagttgcag caagcggtcc acgctggttt gccccagcag 5460

gcgaaaatcc tgtttgatgg tggttaacgg cgggatataa catgagctgt cttcggtatc 5520

gtcgtatccc actaccgaga tgtccgcacc aacgcgcagc ccggactcgg taatggcgcg 5580

cattgcgccc agcgccatct gatcgttggc aaccagcatc gcagtgggaa cgatgccctc 5640

attcagcatt tgcatggttt gttgaaaacc ggacatggca ctccagtcgc cttcccgttc 5700

cgctatcggc tgaatttgat tgcgagtgag atatttatgc cagccagcca gacgcagacg 5760

cgccgagaca gaacttaatg ggcccgctaa cagcgcgatt tgctggtgac ccaatgcgac 5820

cagatgctcc acgcccagtc gcgtaccgtc ttcatgggag aaaataatac tgttgatggg 5880

tgtctggtca gagacatcaa gaaataacgc cggaacatta gtgcaggcag cttccacagc 5940

aatggcatcc tggtcatcca gcggatagtt aatgatcagc ccactgacgc gttgcgcgag 6000

aagattgtgc accgccgctt tacaggcttc gacgccgctt cgttctacca tcgacaccac 6060

cacgctggca cccagttgat cggcgcgaga tttaatcgcc gcgacaattt gcgacggcgc 6120

gtgcagggcc agactggagg tggcaacgcc aatcagcaac gactgtttgc ccgccagttg 6180

ttgtgccacg cggttgggaa tgtaattcag ctccgccatc gccgcttcca ctttttcccg 6240

cgttttcgca gaaacgtggc tggcctggtt caccacgcgg gaaacggtct gataagagac 6300

accggcatac tctgcgacat cgtataacgt tactggtttc acattcacca ccctgaattg 6360

actctcttcc gggcgctatc atgccatacc gcgaaaggtt ttgcgccatt cgatggtgtc 6420

cgggatctcg acgctctccc ttatg 6445

<210> SEQ ID NO 72
<211> LENGTH: 3553
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 72

ctagtaaatt cgcgtttcta cggtagccgg gcgctaatac gactcactat agggtgggag 60

ggcgatcgca atctggctcc cagttttgtg gactttagaa cagaggagat aaagatgaca 120

aaactgggaa cctggcggca gcgcaagaag atgaccatga ttacggattc actggccgtc 180

gttttacaac gtcgtgactg ggaaaaccct ggcgttaccc aacttaatcg ccttgcagca 240

catccccctt tcgccagctg gcgtaatagc gaagaggccc gcaccgatcg cccttcccaa 300

cagttgcgca gcctgaatgg cgaatggtaa taatagcata acccccttggg gcctctaaac 360

gggtcttgag gggttttttg ctgaaacctc aggcatttga gaagcacacg gtcacactgc 420

ttccggtagt caataaaccg gtaaaccagc aatagacata agcggctatt taacgaccct 480

gccctgaacc gacgacaagc tgacgaccgg gtctccgcaa gtggcacttt tcggggaaat 540

gtgcgcggaa cccctatttg tttatttttc taaatacatt caaatatgta tccgctcatg 600

aattaattct tagaaaaact catcgagcat caaatgaaac tgcaatttat tcatatcagg 660

attatcaata ccatattttt gaaaaagccg tttctgtaat gaaggagaaa actcaccgag 720

gcagttccat aggatggcaa gatcctggta tcggtctgcg attccgactc gtccaacatc 780

aatacaacct attaatttcc cctcgtcaaa aataaggtta tcaagtgaga aatcaccatg 840

agtgacgact gaatccggtg agaatggcaa aagtttatgc atttctttcc agacttgttc 900

-continued

aacaggccag ccattacgct cgtcatcaaa atcactcgca tcaaccaaac cgttattcat 960

tcgtgattgc gcctgagcga gacgaaatac gcggtcgctg ttaaaaggac aattacaaac 1020

aggaatcgaa tgcaaccggc gcaggaacac tgccagcgca tcaacaatat tttcacctga 1080

atcaggatat tcttctaata cctggaatgc tgttttcccg gggatcgcag tggtgagtaa 1140

ccatgcatca tcaggagtac ggataaaatg cttgatggtc ggaagaggca taaattccgt 1200

cagccagttt agtctgacca tctcatctgt aacatcattg gcaacgctac ctttgccatg 1260

tttcagaaac aactctggcg catcgggctt cccatacaat cgatagattg tcgcacctga 1320

ttgcccgaca ttatcgcgag cccatttata cccatataaa tcagcatcca tgttggaatt 1380

taatcgcggc ctagagcaag acgtttcccg ttgaatatgg ctcatactct tccttttca 1440

atattattga agcatttatc agggttattg tctcatgagc ggatacatat ttgaatgtat 1500

ttagaaaaat aaacaaatag gcatgctagc gcagaaacgt cctagaagat gccaggagga 1560

tacttagcag agagacaata aggccggagc gaagccgttt ttccataggc tccgcccccc 1620

tgacgaacat cacgaaatct gacgctcaaa tcagtggtgg cgaaacccga caggactata 1680

aagataccag gcgtttcccc ctgatggctc cctcttgcgc tctcctgttc ccgtcctgcg 1740

gcgtccgtgt tgtggtggag gctttaccca aatcaccacg tcccgttccg tgtagacagt 1800

tcgctccaag ctgggctgtg tgcaagaacc ccccgttcag cccgactgct gcgccttatc 1860

cggtaactat catcttgagt ccaacccgga aagacacgac aaaacgccac tggcagcagc 1920

cattggtaac tgagaattag tggatttaga tatcgagagt cttgaagtgg tggcctaaca 1980

gaggctacac tgaaaggaca gtatttggta tctgcgctcc actaaagcca gttaccaggt 2040

taagcagttc cccaactgac ttaaccttcg atcaaaccgc ctccccaggc ggtttttttcg 2100

tttacagagc aggagattac gacgatcgta aaaggatctc aagaagatcc tttacggatt 2160

cccgacacca tcactctaga tttcagtgca atttatctct tcaaatgtag cacctgaagt 2220

cagccccata cgatataagt tgtaattctc atgttagtca tgccccgcgc ccaccggaag 2280

gagctgactg ggttgaaggc tctcaagggc atcggtcgag atcccggtgc ctaatgagtg 2340

agctaactta cattaattgc gttgcgctca ctgcccgctt tccagtcggg aaacctgtcg 2400

tgccagctgc attaatgaat cggccaacgc gcggggagag gcggtttgcg tattgggcgc 2460

cagggtggtt tttcttttca ccagtgagac gggcaacagc tgattgccct tcaccgcctg 2520

gccctgagag agttgcagca agcggtccac gctggtttgc cccagcaggc gaaaatcctg 2580

tttgatggtg gttaacggcg ggatataaca tgagctgtct tcggtatcgt cgtatcccac 2640

taccgagatg tccgcaccaa cgcgcagccc ggactcggta atggcgcgca ttgcgcccag 2700

cgccatctga tcgttggcaa ccagcatcgc agtgggaacg atgccctcat tcagcatttg 2760

catggtttgt tgaaaaccgg acatggcact ccagtcgcct tcccgttccg ctatcggctg 2820

aatttgattg cgagtgagat atttatgcca gccagccaga cgcagacgcg ccgagacaga 2880

acttaatggg cccgctaaca gcgcgatttg ctggtgaccc aatgcgacca gatgctccac 2940

gcccagtcgc gtaccgtctt catgggagaa aataatactg ttgatgggtg tctggtcaga 3000

gacatcaaga aataacgccg gaacattagt gcaggcagct tccacagcaa tggcatcctg 3060

gtcatccagc ggatagttaa tgatcagccc actgacgcgt tgcgcgagaa gattgtgcac 3120

cgccgcttta caggcttcga cgccgcttcg ttctaccatc gacaccacca cgctggcacc 3180

-continued

cagttgatcg gcgcgagatt taatcgccgc gacaatttgc gacggcgcgt gcagggccag 3240

actggaggtg gcaacgccaa tcagcaacga ctgtttgccc gccagttgtt gtgccacgcg 3300

gttgggaatg taattcagct ccgccatcgc cgcttccact ttttcccgcg ttttcgcaga 3360

aacgtggctg gcctggttca ccacgcggga aacggtctga taagagacac cggcatactc 3420

tgcgacatcg tataacgtta ctggtttcac attcaccacc ctgaattgac tctcttccgg 3480

gcgctatcat gccataccgc gaaaggtttt gcgccattcg atggtgtccg ggatctcgac 3540

gctctccctt atg 3553

<210> SEQ ID NO 73
<211> LENGTH: 3554
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 73

ctagtaaatt cgcgtttcta cggtagccgg gcgctaatac gactcactat agggtgtgaa 60

tgaagatggc gtcgaatgac gccaacccat ggactttaga acagaggaga taaagatgat 120

gggttggcga acctggcggc agcgcaagaa gatgaccatg attacggatt cactggccgt 180

cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc caacttaatc gccttgcagc 240

acatcccccc ttcgccagct ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca 300

acagttgcgc agcctgaatg gcgaatggta ataatagcat aacccccttgg ggcctctaaa 360

cgggtcttga ggggtttttt gctgaaacct caggcatttg agaagcacac ggtcacactg 420

cttccggtag tcaataaacc ggtaaaccag caatagacat aagcggctat ttaacgaccc 480

tgccctgaac cgacgacaag ctgacgaccg ggtctccgca agtggcactt ttcggggaaa 540

tgtgcgcgga acccctattt gtttatttttt ctaaatacat tcaaatatgt atccgctcat 600

gaattaattc ttagaaaaac tcatcgagca tcaaatgaaa ctgcaattta ttcatatcag 660

gattatcaat accatatttt tgaaaaagcc gtttctgtaa tgaaggagaa aactcaccga 720

ggcagttcca taggatggca agatcctggt atcggtctgc gattccgact cgtccaacat 780

caatacaacc tattaatttc ccctcgtcaa aaataaggtt atcaagtgag aaatcaccat 840

gagtgacgac tgaatccggt gagaatggca aaagtttatg catttctttc cagacttgtt 900

caacaggcca gccattacgc tcgtcatcaa aatcactcgc atcaaccaaa ccgttattca 960

ttcgtgattg cgcctgagcg agacgaaata cgcggtcgct gttaaaagga caattacaaa 1020

caggaatcga atgcaaccgg cgcaggaaca ctgccagcgc atcaacaata ttttcacctg 1080

aatcaggata ttcttctaat acctggaatg ctgttttccc ggggatcgca gtggtgagta 1140

accatgcatc atcaggagta cggataaaat gcttgatggt cggaagaggc ataaattccg 1200

tcagccagtt tagtctgacc atctcatctg taacatcatt ggcaacgcta cctttgccat 1260

gtttcagaaa caactctggc gcatcgggct tcccatacaa tcgatagatt gtcgcacctg 1320

attgcccgac attatcgcga gcccatttat acccatataa atcagcatcc atgttggaat 1380

ttaatcgcgg cctagagcaa gacgtttccc gttgaatatg gctcatactc ttccttttttc 1440

aatattattg aagcatttat cagggttatt gtctcatgag cggatacata tttgaatgta 1500

tttagaaaaa taaacaaata ggcatgctag cgcagaaacg tcctagaaga tgccaggagg 1560

-continued

atacttagca gagagacaat aaggccggag cgaagccgtt tttccatagg ctccgccccc 1620

ctgacgaaca tcacgaaatc tgacgctcaa atcagtggtg gcgaaacccg acaggactat 1680

aaagatacca ggcgtttccc cctgatggct ccctcttgcg ctctcctgtt cccgtcctgc 1740

ggcgtccgtg ttgtggtgga ggctttaccc aaatcaccac gtcccgttcc gtgtagacag 1800

ttcgctccaa gctgggctgt gtgcaagaac cccccgttca gcccgactgc tgcgccttat 1860

ccggtaacta tcatcttgag tccaacccgg aaagacacga caaaacgcca ctggcagcag 1920

ccattggtaa ctgagaatta gtggatttag atatcgagag tcttgaagtg gtggcctaac 1980

agaggctaca ctgaaaggac agtatttggt atctgcgctc cactaaagcc agttaccagg 2040

ttaagcagtt ccccaactga cttaaccttc gatcaaaccg cctccccagg cggttttttc 2100

gtttacagag caggagatta cgacgatcgt aaaaggatct caagaagatc ctttacggat 2160

tcccgacacc atcactctag atttcagtgc aatttatctc ttcaaatgta gcacctgaag 2220

tcagccccat acgatataag ttgtaattct catgttagtc atgccccgcg cccaccggaa 2280

ggagctgact gggttgaagg ctctcaaggg catcggtcga gatcccggtg cctaatgagt 2340

gagctaactt acattaattg cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc 2400

gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg 2460

ccagggtggt ttttcttttc accagtgaga cgggcaacag ctgattgccc ttcaccgcct 2520

ggccctgaga gagttgcagc aagcggtcca cgctggtttg ccccagcagg cgaaaatcct 2580

gtttgatggt ggttaacggc gggatataac atgagctgtc ttcggtatcg tcgtatccca 2640

ctaccgagat gtccgcacca acgcgcagcc cggactcggt aatggcgcgc attgcgccca 2700

gcgccatctg atcgttggca accagcatcg cagtgggaac gatgccctca ttcagcattt 2760

gcatggtttg ttgaaaaccg gacatggcac tccagtcgcc ttcccgttcc gctatcggct 2820

gaatttgatt gcgagtgaga tatttatgcc agccagccag acgcagacgc gccgagacag 2880

aacttaatgg gcccgctaac agcgcgattt gctggtgacc caatgcgacc agatgctcca 2940

cgcccagtcg cgtaccgtct tcatgggaga aaataatact gttgatgggt gtctggtcag 3000

agacatcaag aaataacgcc ggaacattag tgcaggcagc ttccacagca atggcatcct 3060

ggtcatccag cggatagtta atgatcagcc cactgacgcg ttgcgcgaga agattgtgca 3120

ccgccgcttt acaggcttcg acgccgcttc gttctaccat cgacaccacc acgctggcac 3180

ccagttgatc ggcgcgagat ttaatcgccg cgacaatttg cgacggcgcg tgcagggcca 3240

gactggaggt ggcaacgcca atcagcaacg actgtttgcc cgccagttgt tgtgccacgc 3300

ggttgggaat gtaattcagc tccgccatcg ccgcttccac tttttcccgc gttttcgcag 3360

aaacgtggct ggcctggttc accacgcggg aaacggtctg ataagagaca ccggcatact 3420

ctgcgacatc gtataacgtt actggtttca cattcaccac cctgaattga ctctcttccg 3480

ggcgctatca tgccataccg cgaaaggttt tgcgccattc gatggtgtcc gggatctcga 3540

cgctctccct tatg 3554

<210> SEQ ID NO 74
<211> LENGTH: 3554
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

-continued

<400> SEQUENCE: 74

ctagtaaatt cgcgtttcta cggtagccgg gcgctaatac gactcactat agggagatct 60

gagcacgtgg gagggcgatc gcaatctggc ggactttaga acagaggaga taaagatggc 120

cagattgcga acctggcggc agcgcaagaa gatgaccatg attacggatt cactggccgt 180

cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc caacttaatc gccttgcagc 240

acatcccccc ttcgccagct ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca 300

acagttgcgc agcctgaatg gcgaatggta ataatagcat aacccccttgg ggcctctaaa 360

cgggtcttga ggggtttttt gctgaaacct caggcatttg agaagcacac ggtcacactg 420

cttccggtag tcaataaacc ggtaaaccag caatagacat aagcggctat ttaacgaccc 480

tgccctgaac cgacgacaag ctgacgaccg ggtctccgca agtggcactt ttcggggaaa 540

tgtgcgcgga acccctattt gtttatttttt ctaaatacat tcaaatatgt atccgctcat 600

gaattaattc ttagaaaaac tcatcgagca tcaaatgaaa ctgcaattta ttcatatcag 660

gattatcaat accatatttt tgaaaaagcc gtttctgtaa tgaaggagaa aactcaccga 720

ggcagttcca taggatggca agatcctggt atcggtctgc gattccgact cgtccaacat 780

caatacaacc tattaatttc ccctcgtcaa aaataaggtt atcaagtgag aaatcaccat 840

gagtgacgac tgaatccggt gagaatggca aaagtttatg catttctttc cagacttgtt 900

caacaggcca gccattacgc tcgtcatcaa aatcactcgc atcaaccaaa ccgttattca 960

ttcgtgattg cgcctgagcg agacgaaata cgcggtcgct gttaaaagga caattacaaa 1020

caggaatcga atgcaaccgg cgcaggaaca ctgccagcgc atcaacaata ttttcacctg 1080

aatcaggata ttcttctaat acctggaatg ctgttttccc ggggatcgca gtggtgagta 1140

accatgcatc atcaggagta cggataaaat gcttgatggt cggaagaggc ataaattccg 1200

tcagccagtt tagtctgacc atctcatctg taacatcatt ggcaacgcta cctttgccat 1260

gtttcagaaa caactctggc gcatcgggct tcccatacaa tcgatagatt gtcgcacctg 1320

attgcccgac attatcgcga gcccatttat acccatataa atcagcatcc atgttggaat 1380

ttaatcgcgg cctagagcaa gacgtttccc gttgaatatg gctcatactc ttccttttc 1440

aatattattg aagcatttat cagggttatt gtctcatgag cggatacata tttgaatgta 1500

tttagaaaaa taaacaaata ggcatgctag cgcagaaacg tcctagaaga tgccaggagg 1560

atacttagca gagagacaat aaggccggag cgaagccgtt tttccatagg ctccgccccc 1620

ctgacgaaca tcacgaaatc tgacgctcaa atcagtggtg gcgaaacccg acaggactat 1680

aaagatacca ggcgtttccc cctgatggct ccctcttgcg ctctcctgtt cccgtcctgc 1740

ggcgtccgtg ttgtggtgga ggctttaccc aaatcaccac gtcccgttcc gtgtagacag 1800

ttcgctccaa gctgggctgt gtgcaagaac cccccgttca gcccgactgc tgcgccttat 1860

ccggtaacta tcatcttgag tccaacccgg aaagacacga caaaacgcca ctggcagcag 1920

ccattggtaa ctgagaatta gtggatttag atatcgagag tcttgaagtg gtggcctaac 1980

agaggctaca ctgaaaggac agtatttggt atctgcgctc cactaaagcc agttaccagg 2040

ttaagcagtt ccccaactga cttaaccttc gatcaaaccg cctccccagg cggtttttc 2100

gtttacagag caggagatta cgacgatcgt aaaaggatct caagaagatc ctttacggat 2160

tcccgacacc atcactctag atttcagtgc aatttatctc ttcaaatgta gcacctgaag 2220

-continued

tcagccccat acgatataag ttgtaattct catgttagtc atgccccgcg cccaccggaa 2280

ggagctgact gggttgaagg ctctcaaggg catcggtcga gatcccggtg cctaatgagt 2340

gagctaactt acattaattg cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc 2400

gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg 2460

ccagggtggt ttttcttttc accagtgaga cgggcaacag ctgattgccc ttcaccgcct 2520

ggccctgaga gagttgcagc aagcggtcca cgctggtttg ccccagcagg cgaaaatcct 2580

gtttgatggt ggttaacggc gggatataac atgagctgtc ttcggtatcg tcgtatccca 2640

ctaccgagat gtccgcacca acgcgcagcc cggactcggt aatggcgcgc attgcgccca 2700

gcgccatctg atcgttggca accagcatcg cagtgggaac gatgccctca ttcagcattt 2760

gcatggtttg ttgaaaaccg gacatggcac tccagtcgcc ttcccgttcc gctatcggct 2820

gaatttgatt gcgagtgaga tatttatgcc agccagccag acgcagacgc gccgagacag 2880

aacttaatgg gcccgctaac agcgcgattt gctggtgacc caatgcgacc agatgctcca 2940

cgcccagtcg cgtaccgtct tcatgggaga aaataatact gttgatgggt gtctggtcag 3000

agacatcaag aaataacgcc ggaacattag tgcaggcagc ttccacagca atggcatcct 3060

ggtcatccag cggatagtta atgatcagcc cactgacgcg ttgcgcgaga agattgtgca 3120

ccgccgcttt acaggcttcg acgccgcttc gttctaccat cgacaccacc acgctggcac 3180

ccagttgatc ggcgcgagat ttaatcgccg cgacaatttg cgacggcgcg tgcagggcca 3240

gactggaggt ggcaacgcca atcagcaacg actgtttgcc cgccagttgt tgtgccacgc 3300

ggttgggaat gtaattcagc tccgccatcg ccgcttccac tttttcccgc gttttcgcag 3360

aaacgtggct ggcctggttc accacgcggg aaacggtctg ataagagaca ccggcatact 3420

ctgcgacatc gtataacgtt actggtttca cattcaccac cctgaattga ctctcttccg 3480

ggcgctatca tgccataccg cgaaaggttt tgcgccattc gatggtgtcc gggatctcga 3540

cgctctccct tatg 3554

<210> SEQ ID NO 75
<211> LENGTH: 3554
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 75

ctagtaaatt cgcgtttcta cggtagccgg gcgctaatac gactcactat agggatcgca 60

atctggctcc cagttttgtg aatgaagatg ggactttaga acagaggaga taaagatgca 120

tcttcattca acctggcggc agcgcaagaa gatgaccatg attacggatt cactggccgt 180

cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc caacttaatc gccttgcagc 240

acatcccccc ttcgccagct ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca 300

acagttgcgc agcctgaatg gcgaatggta ataatagcat aacccccttgg ggcctctaaa 360

cgggtcttga ggggtttttt gctgaaacct caggcatttg agaagcacac ggtcacactg 420

cttccggtag tcaataaacc ggtaaaccag caatagacat aagcggctat ttaacgaccc 480

tgccctgaac cgacgacaag ctgacgaccg ggtctccgca agtggcactt ttcggggaaa 540

tgtgcgcgga acccctattt gtttatttttt ctaaatacat tcaaatatgt atccgctcat 600

-continued

gaattaattc ttagaaaaac tcatcgagca tcaaatgaaa ctgcaattta ttcatatcag 660

gattatcaat accatatttt tgaaaaagcc gtttctgtaa tgaaggagaa aactcaccga 720

ggcagttcca taggatggca agatcctggt atcggtctgc gattccgact cgtccaacat 780

caatacaacc tattaatttc ccctcgtcaa aaataaggtt atcaagtgag aaatcaccat 840

gagtgacgac tgaatccggt gagaatggca aaagtttatg catttctttc cagacttgtt 900

caacaggcca gccattacgc tcgtcatcaa aatcactcgc atcaaccaaa ccgttattca 960

ttcgtgattg cgcctgagcg agacgaaata cgcggtcgct gttaaaagga caattacaaa 1020

caggaatcga atgcaaccgg cgcaggaaca ctgccagcgc atcaacaata ttttcacctg 1080

aatcaggata ttcttctaat acctggaatg ctgttttccc ggggatcgca gtggtgagta 1140

accatgcatc atcaggagta cggataaaat gcttgatggt cggaagaggc ataaattccg 1200

tcagccagtt tagtctgacc atctcatctg taacatcatt ggcaacgcta cctttgccat 1260

gtttcagaaa caactctggc gcatcgggct tcccatacaa tcgatagatt gtcgcacctg 1320

attgcccgac attatcgcga gcccatttat acccatataa atcagcatcc atgttggaat 1380

ttaatcgcgg cctagagcaa gacgtttccc gttgaatatg gctcatactc ttccttttc 1440

aatattattg aagcatttat cagggttatt gtctcatgag cggatacata tttgaatgta 1500

tttagaaaaa taaacaaata ggcatgctag cgcagaaacg tcctagaaga tgccaggagg 1560

atacttagca gagagacaat aaggccggag cgaagccgtt tttccatagg ctccgccccc 1620

ctgacgaaca tcacgaaatc tgacgctcaa atcagtggtg gcgaaacccg acaggactat 1680

aaagatacca ggcgtttccc cctgatggct ccctcttgcg ctctcctgtt cccgtcctgc 1740

ggcgtccgtg ttgtggtgga ggctttaccc aaatcaccac gtcccgttcc gtgtagacag 1800

ttcgctccaa gctgggctgt gtgcaagaac cccccgttca gcccgactgc tgcgccttat 1860

ccggtaacta tcatcttgag tccaacccgg aaagacacga caaaacgcca ctggcagcag 1920

ccattggtaa ctgagaatta gtggatttag atatcgagag tcttgaagtg gtggcctaac 1980

agaggctaca ctgaaaggac agtatttggt atctgcgctc cactaaagcc agttaccagg 2040

ttaagcagtt ccccaactga cttaaccttc gatcaaaccg cctccccagg cggttttttc 2100

gtttacagag caggagatta cgacgatcgt aaaaggatct caagaagatc ctttacggat 2160

tcccgacacc atcactctag atttcagtgc aatttatctc ttcaaatgta gcacctgaag 2220

tcagccccat acgatataag ttgtaattct catgttagtc atgccccgcg cccaccggaa 2280

ggagctgact gggttgaagg ctctcaaggg catcggtcga gatcccggtg cctaatgagt 2340

gagctaactt acattaattg cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc 2400

gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg 2460

ccagggtggt ttttcttttc accagtgaga cgggcaacag ctgattgccc ttcaccgcct 2520

ggccctgaga gagttgcagc aagcggtcca cgctggtttg ccccagcagg cgaaaatcct 2580

gtttgatggt ggttaacggc gggatataac atgagctgtc ttcggtatcg tcgtatccca 2640

ctaccgagat gtccgcacca acgcgcagcc cggactcggt aatggcgcgc attgcgccca 2700

gcgccatctg atcgttggca accagcatcg cagtgggaac gatgccctca ttcagcattt 2760

gcatggtttg ttgaaaaccg gacatggcac tccagtcgcc ttcccgttcc gctatcggct 2820

gaatttgatt gcgagtgaga tatttatgcc agccagccag acgcagacgc gccgagacag 2880

-continued

aacttaatgg gcccgctaac agcgcgattt gctggtgacc caatgcgacc agatgctcca 2940

cgcccagtcg cgtaccgtct tcatgggaga aaataatact gttgatgggt gtctggtcag 3000

agacatcaag aaataacgcc ggaacattag tgcaggcagc ttccacagca atggcatcct 3060

ggtcatccag cggatagtta atgatcagcc cactgacgcg ttgcgcgaga agattgtgca 3120

ccgccgcttt acaggcttcg acgccgcttc gttctaccat cgacaccacc acgctggcac 3180

ccagttgatc ggcgcgagat ttaatcgccg cgacaatttg cgacggcgcg tgcagggcca 3240

gactggaggt ggcaacgcca atcagcaacg actgtttgcc cgccagttgt tgtgccacgc 3300

ggttgggaat gtaattcagc tccgccatcg ccgcttccac tttttcccgc gttttcgcag 3360

aaacgtggct ggcctggttc accacgcggg aaacggtctg ataagagaca ccggcatact 3420

ctgcgacatc gtataacgtt actggtttca cattcaccac cctgaattga ctctcttccg 3480

ggcgctatca tgccataccg cgaaaggttt tgcgccattc gatggtgtcc gggatctcga 3540

cgctctccct tatg 3554

<210> SEQ ID NO 76
<211> LENGTH: 3554
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 76

ctagtaaatt cgcgtttcta cggtagccgg gcgctaatac gactcactat agggtcgaat 60

gacgccaacc catctgatgg gtccgcagcc ggactttaga acagaggaga taaagatggg 120

ctgcggacca acctggcggc agcgcaagaa gatgaccatg attacggatt cactggccgt 180

cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc caacttaatc gccttgcagc 240

acatcccccc ttcgccagct ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca 300

acagttgcgc agcctgaatg gcgaatggta ataatagcat aacccccttgg ggcctctaaa 360

cgggtcttga ggggtttttt gctgaaacct caggcatttg agaagcacac ggtcacactg 420

cttccggtag tcaataaacc ggtaaaccag caatagacat aagcggctat ttaacgaccc 480

tgccctgaac cgacgacaag ctgacgaccg ggtctccgca agtggcactt ttcggggaaa 540

tgtgcgcgga acccctattt gtttatttttt ctaaatacat tcaaatatgt atccgctcat 600

gaattaattc ttagaaaaac tcatcgagca tcaaatgaaa ctgcaattta ttcatatcag 660

gattatcaat accatatttt tgaaaaagcc gtttctgtaa tgaaggagaa aactcaccga 720

ggcagttcca taggatggca agatcctggt atcggtctgc gattccgact cgtccaacat 780

caatacaacc tattaatttc ccctcgtcaa aaataaggtt atcaagtgag aaatcaccat 840

gagtgacgac tgaatccggt gagaatggca aaagtttatg catttctttc cagacttgtt 900

caacaggcca gccattacgc tcgtcatcaa aatcactcgc atcaaccaaa ccgttattca 960

ttcgtgattg cgcctgagcg agacgaaata cgcggtcgct gttaaaagga caattacaaa 1020

caggaatcga atgcaaccgg cgcaggaaca ctgccagcgc atcaacaata ttttcacctg 1080

aatcaggata ttcttctaat acctggaatg ctgttttccc ggggatcgca gtggtgagta 1140

accatgcatc atcaggagta cggataaaat gcttgatggt cggaagaggc ataaattccg 1200

tcagccagtt tagtctgacc atctcatctg taacatcatt ggcaacgcta cctttgccat 1260

-continued

gtttcagaaa caactctggc gcatcgggct tcccatacaa tcgatagatt gtcgcacctg 1320

attgcccgac attatcgcga gcccatttat acccatataa atcagcatcc atgttggaat 1380

ttaatcgcgg cctagagcaa gacgtttccc gttgaatatg gctcatactc ttccttttc 1440

aatattattg aagcatttat cagggttatt gtctcatgag cggatacata tttgaatgta 1500

tttagaaaaa taaacaaata ggcatgctag cgcagaaacg tcctagaaga tgccaggagg 1560

atacttagca gagagacaat aaggccggag cgaagccgtt tttccatagg ctccgccccc 1620

ctgacgaaca tcacgaaatc tgacgctcaa atcagtggtg gcgaaacccg acaggactat 1680

aaagatacca ggcgtttccc cctgatggct ccctcttgcg ctctcctgtt cccgtcctgc 1740

ggcgtccgtg ttgtggtgga ggctttaccc aaatcaccac gtcccgttcc gtgtagacag 1800

ttcgctccaa gctgggctgt gtgcaagaac cccccgttca gcccgactgc tgcgccttat 1860

ccggtaacta tcatcttgag tccaacccgg aaagacacga caaaacgcca ctggcagcag 1920

ccattggtaa ctgagaatta gtggatttag atatcgagag tcttgaagtg gtggcctaac 1980

agaggctaca ctgaaaggac agtatttggt atctgcgctc cactaaagcc agttaccagg 2040

ttaagcagtt ccccaactga cttaaccttc gatcaaaccg cctccccagg cggttttttc 2100

gtttacagag caggagatta cgacgatcgt aaaaggatct caagaagatc ctttacggat 2160

tcccgacacc atcactctag atttcagtgc aatttatctc ttcaaatgta gcacctgaag 2220

tcagccccat acgatataag ttgtaattct catgttagtc atgccccgcg cccaccggaa 2280

ggagctgact gggttgaagg ctctcaaggg catcggtcga gatcccggtg cctaatgagt 2340

gagctaactt acattaattg cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc 2400

gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg 2460

ccagggtggt ttttcttttc accagtgaga cgggcaacag ctgattgccc ttcaccgcct 2520

ggccctgaga gagttgcagc aagcggtcca cgctggtttg ccccagcagg cgaaaatcct 2580

gtttgatggt ggttaacggc gggatataac atgagctgtc ttcggtatcg tcgtatccca 2640

ctaccgagat gtccgcacca acgcgcagcc cggactcggt aatggcgcgc attgcgccca 2700

gcgccatctg atcgttggca accagcatcg cagtgggaac gatgccctca ttcagcattt 2760

gcatggtttg ttgaaaaccg gacatggcac tccagtcgcc ttcccgttcc gctatcggct 2820

gaatttgatt gcgagtgaga tatttatgcc agccagccag acgcagacgc gccgagacag 2880

aacttaatgg gcccgctaac agcgcgattt gctggtgacc caatgcgacc agatgctcca 2940

cgcccagtcg cgtaccgtct tcatgggaga aaataatact gttgatgggt gtctggtcag 3000

agacatcaag aaataacgcc ggaacattag tgcaggcagc ttccacagca atggcatcct 3060

ggtcatccag cggatagtta atgatcagcc cactgacgcg ttgcgcgaga agattgtgca 3120

ccgccgcttt acaggcttcg acgccgcttc gttctaccat cgacaccacc acgctggcac 3180

ccagttgatc ggcgcgagat ttaatcgccg cgacaatttg cgacggcgcg tgcagggcca 3240

gactggaggt ggcaacgcca atcagcaacg actgtttgcc cgccagttgt tgtgccacgc 3300

ggttgggaat gtaattcagc tccgccatcg ccgcttccac tttttcccgc gttttcgcag 3360

aaacgtggct ggcctggttc accacgcggg aaacggtctg ataagagaca ccggcatact 3420

ctgcgacatc gtataacgtt actggtttca cattcaccac cctgaattga ctctcttccg 3480

ggcgctatca tgccataccg cgaaaggttt tgcgccattc gatggtgtcc gggatctcga 3540

-continued

cgctctccct tatg 3554

<210> SEQ ID NO 77
<211> LENGTH: 3554
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 77

ctagtaaatt cgcgtttcta cggtagccgg gcgctaatac gactcactat agggccatct 60

tcattcacaa aactgggagc cagattgcga ggactttaga acagaggaga taaagatgtc 120

gcaatctgga acctggcggc agcgcaagaa gatgaccatg attacggatt cactggccgt 180

cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc caacttaatc gccttgcagc 240

acatcccccc ttcgccagct ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca 300

acagttgcgc agcctgaatg gcgaatggta ataatagcat aacccccttgg ggcctctaaa 360

cgggtcttga ggggtttttt gctgaaacct caggcatttg agaagcacac ggtcacactg 420

cttccggtag tcaataaacc ggtaaaccag caatagacat aagcggctat ttaacgaccc 480

tgccctgaac cgacgacaag ctgacgaccg ggtctccgca agtggcactt ttcggggaaa 540

tgtgcgcgga acccctattt gtttatttttt ctaaatacat tcaaatatgt atccgctcat 600

gaattaattc ttagaaaaac tcatcgagca tcaaatgaaa ctgcaattta ttcatatcag 660

gattatcaat accatatttt tgaaaaagcc gtttctgtaa tgaaggagaa aactcaccga 720

ggcagttcca taggatggca agatcctggt atcggtctgc gattccgact cgtccaacat 780

caatacaacc tattaatttc ccctcgtcaa aaataaggtt atcaagtgag aaatcaccat 840

gagtgacgac tgaatccggt gagaatggca aaagtttatg catttctttc cagacttgtt 900

caacaggcca gccattacgc tcgtcatcaa aatcactcgc atcaaccaaa ccgttattca 960

ttcgtgattg cgcctgagcg agacgaaata cgcggtcgct gttaaaagga caattacaaa 1020

caggaatcga atgcaaccgg cgcaggaaca ctgccagcgc atcaacaata ttttcacctg 1080

aatcaggata ttcttctaat acctggaatg ctgttttccc ggggatcgca gtggtgagta 1140

accatgcatc atcaggagta cggataaaat gcttgatggt cggaagaggc ataaattccg 1200

tcagccagtt tagtctgacc atctcatctg taacatcatt ggcaacgcta cctttgccat 1260

gtttcagaaa caactctggc gcatcgggct tcccatacaa tcgatagatt gtcgcacctg 1320

attgcccgac attatcgcga gcccatttat acccatataa atcagcatcc atgttggaat 1380

ttaatcgcgg cctagagcaa gacgtttccc gttgaatatg gctcatactc ttccttttttc 1440

aatattattg aagcatttat cagggttatt gtctcatgag cggatacata tttgaatgta 1500

tttagaaaaa taaacaaata ggcatgctag cgcagaaacg tcctagaaga tgccaggagg 1560

atacttagca gagagacaat aaggccggag cgaagccgtt tttccatagg ctccgccccc 1620

ctgacgaaca tcacgaaatc tgacgctcaa atcagtggtg gcgaaacccg acaggactat 1680

aaagatacca ggcgtttccc cctgatggct ccctcttgcg ctctcctgtt cccgtcctgc 1740

ggcgtccgtg ttgtggtgga ggctttaccc aaatcaccac gtcccgttcc gtgtagacag 1800

ttcgctccaa gctgggctgt gtgcaagaac cccccgttca gcccgactgc tgcgccttat 1860

ccggtaacta tcatcttgag tccaacccgg aaagacacga caaaacgcca ctggcagcag 1920

-continued

```
ccattggtaa ctgagaatta gtggatttag atatcgagag tcttgaagtg gtggcctaac   1980

agaggctaca ctgaaaggac agtatttggt atctgcgctc cactaaagcc agttaccagg   2040

ttaagcagtt ccccaactga cttaaccttc gatcaaaccg cctccccagg cggttttttc   2100

gtttacagag caggagatta cgacgatcgt aaaaggatct caagaagatc ctttacggat   2160

tcccgacacc atcactctag atttcagtgc aatttatctc ttcaaatgta gcacctgaag   2220

tcagccccat acgatataag ttgtaattct catgttagtc atgccccgcg cccaccggaa   2280

ggagctgact gggttgaagg ctctcaaggg catcggtcga gatcccggtg cctaatgagt   2340

gagctaactt acattaattg cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc   2400

gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg   2460

ccagggtggt ttttcttttc accagtgaga cgggcaacag ctgattgccc ttcaccgcct   2520

ggccctgaga gagttgcagc aagcggtcca cgctggtttg ccccagcagg cgaaaatcct   2580

gtttgatggt ggttaacggc gggatataac atgagctgtc ttcggtatcg tcgtatccca   2640

ctaccgagat gtccgcacca acgcgcagcc cggactcggt aatggcgcgc attgcgccca   2700

gcgccatctg atcgttggca accagcatcg cagtgggaac gatgccctca ttcagcattt   2760

gcatggtttg ttgaaaaccg gacatggcac tccagtcgcc ttcccgttcc gctatcggct   2820

gaatttgatt gcgagtgaga tatttatgcc agccagccag acgcagacgc gccgagacag   2880

aacttaatgg gcccgctaac agcgcgattt gctggtgacc caatgcgacc agatgctcca   2940

cgcccagtcg cgtaccgtct tcatgggaga aaataatact gttgatgggt gtctggtcag   3000

agacatcaag aaataacgcc ggaacattag tgcaggcagc ttccacagca atggcatcct   3060

ggtcatccag cggatagtta atgatcagcc cactgacgcg ttgcgcgaga agattgtgca   3120

ccgccgcttt acaggcttcg acgccgcttc gttctaccat cgacaccacc acgctggcac   3180

ccagttgatc ggcgcgagat ttaatcgccg cgacaatttg cgacggcgcg tgcagggcca   3240

gactggaggt ggcaacgcca atcagcaacg actgtttgcc cgccagttgt tgtgccacgc   3300

ggttgggaat gtaattcagc tccgccatcg ccgcttccac tttttcccgc gttttcgcag   3360

aaacgtggct ggcctggttc accacgcggg aaacggtctg ataagagaca ccggcatact   3420

ctgcgacatc gtataacgtt actggtttca cattcaccac cctgaattga ctctcttccg   3480

ggcgctatca tgccataccg cgaaaggttt tgcgccattc gatggtgtcc gggatctcga   3540

cgctctccct tatg                                                     3554
```

<210> SEQ ID NO 78
<211> LENGTH: 3554
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 78

```
ctagtaaatt cgcgtttcta cggtagccgg gcgctaatac gactcactat agggatcgcc     60

ctcccacgtg ctcagatctg agaatctcat ggactttaga acagaggaga taaagatgat    120

gagattctca acctggcggc agcgcaagaa gatgaccatg attacggatt cactggccgt    180

cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc caacttaatc gccttgcagc    240

acatcccccc ttcgccagct ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca    300
```

-continued

acagttgcgc agcctgaatg gcgaatggta ataatagcat aacccttgg ggcctctaaa 360

cgggtcttga ggggtttttt gctgaaacct caggcatttg agaagcacac ggtcacactg 420

cttccggtag tcaataaacc ggtaaaccag caatagacat aagcggctat ttaacgaccc 480

tgccctgaac cgacgacaag ctgacgaccg ggtctccgca agtggcactt ttcggggaaa 540

tgtgcgcgga acccctattt gtttatttttt ctaaatacat tcaaatatgt atccgctcat 600

gaattaattc ttagaaaaac tcatcgagca tcaaatgaaa ctgcaattta ttcatatcag 660

gattatcaat accatatttt tgaaaaagcc gtttctgtaa tgaaggagaa aactcaccga 720

ggcagttcca taggatggca agatcctggt atcggtctgc gattccgact cgtccaacat 780

caatacaacc tattaatttc ccctcgtcaa aaataaggtt atcaagtgag aaatcaccat 840

gagtgacgac tgaatccggt gagaatggca aaagtttatg catttctttc cagacttgtt 900

caacaggcca gccattacgc tcgtcatcaa aatcactcgc atcaaccaaa ccgttattca 960

ttcgtgattg cgcctgagcg agacgaaata cgcggtcgct gttaaaagga caattacaaa 1020

caggaatcga atgcaaccgg cgcaggaaca ctgccagcgc atcaacaata ttttcacctg 1080

aatcaggata ttcttctaat acctggaatg ctgttttccc ggggatcgca gtggtgagta 1140

accatgcatc atcaggagta cggataaaat gcttgatggt cggaagaggc ataaattccg 1200

tcagccagtt tagtctgacc atctcatctg taacatcatt ggcaacgcta cctttgccat 1260

gtttcagaaa caactctggc gcatcgggct tcccatacaa tcgatagatt gtcgcacctg 1320

attgcccgac attatcgcga gcccatttat acccatataa atcagcatcc atgttggaat 1380

ttaatcgcgg cctagagcaa gacgtttccc gttgaatatg gctcatactc ttccttttc 1440

aatattattg aagcatttat cagggttatt gtctcatgag cggatacata tttgaatgta 1500

tttagaaaaa taaacaaata ggcatgctag cgcagaaacg tcctagaaga tgccaggagg 1560

atacttagca gagagacaat aaggccggag cgaagccgtt tttccatagg ctccgccccc 1620

ctgacgaaca tcacgaaatc tgacgctcaa atcagtggtg gcgaaacccg acaggactat 1680

aaagatacca ggcgtttccc cctgatggct ccctcttgcg ctctcctgtt cccgtcctgc 1740

ggcgtccgtg ttgtggtgga ggctttaccc aaatcaccac gtcccgttcc gtgtagacag 1800

ttcgctccaa gctgggctgt gtgcaagaac cccccgttca gcccgactgc tgcgccttat 1860

ccggtaacta tcatcttgag tccaacccgg aaagacacga caaaacgcca ctggcagcag 1920

ccattggtaa ctgagaatta gtggatttag atatcgagag tcttgaagtg gtggcctaac 1980

agaggctaca ctgaaaggac agtatttggt atctgcgctc cactaaagcc agttaccagg 2040

ttaagcagtt ccccaactga cttaaccttc gatcaaaccg cctccccagg cggtttttc 2100

gtttacagag caggagatta cgacgatcgt aaaaggatct caagaagatc ctttacggat 2160

tcccgacacc atcactctag atttcagtgc aatttatctc ttcaaatgta gcacctgaag 2220

tcagccccat acgatataag ttgtaattct catgttagtc atgccccgcg cccaccggaa 2280

ggagctgact gggttgaagg ctctcaaggg catcggtcga gatcccggtg cctaatgagt 2340

gagctaactt acattaattg cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc 2400

gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg 2460

ccagggtggt ttttcttttc accagtgaga cgggcaacag ctgattgccc ttcaccgcct 2520

ggccctgaga gagttgcagc aagcggtcca cgctggtttg ccccagcagg cgaaaatcct 2580

-continued

gtttgatggt ggttaacggc gggatataac atgagctgtc ttcggtatcg tcgtatccca 2640

ctaccgagat gtccgcacca acgcgcagcc cggactcggt aatggcgcgc attgcgccca 2700

gcgccatctg atcgttggca accagcatcg cagtgggaac gatgccctca ttcagcattt 2760

gcatggtttg ttgaaaaccg gacatggcac tccagtcgcc ttcccgttcc gctatcggct 2820

gaatttgatt gcgagtgaga tatttatgcc agccagccag acgcagacgc gccgagacag 2880

aacttaatgg gcccgctaac agcgcgattt gctggtgacc caatgcgacc agatgctcca 2940

cgcccagtcg cgtaccgtct tcatgggaga aaataatact gttgatgggt gtctggtcag 3000

agacatcaag aaataacgcc ggaacattag tgcaggcagc ttccacagca atggcatcct 3060

ggtcatccag cggatagtta atgatcagcc cactgacgcg ttgcgcgaga agattgtgca 3120

ccgccgcttt acaggcttcg acgccgcttc gttctaccat cgacaccacc acgctggcac 3180

ccagttgatc ggcgcgagat ttaatcgccg cgacaatttg cgacggcgcg tgcagggcca 3240

gactggaggt ggcaacgcca atcagcaacg actgtttgcc cgccagttgt tgtgccacgc 3300

ggttgggaat gtaattcagc tccgccatcg ccgcttccac tttttcccgc gttttcgcag 3360

aaacgtggct ggcctggttc accacgcggg aaacggtctg ataagagaca ccggcatact 3420

ctgcgacatc gtataacgtt actggtttca cattcaccac cctgaattga ctctcttccg 3480

ggcgctatca tgccataccg cgaaaggttt tgcgccattc gatggtgtcc gggatctcga 3540

cgctctccct tatg 3554

<210> SEQ ID NO 79
<211> LENGTH: 3554
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 79

ctagtaaatt cgcgtttcta cggtagccgg gcgctaatac gactcactat agggacaaaa 60

ctgggagcca gattgcgatc gccctcccac ggactttaga acagaggaga taaagatggt 120

gggagggcga acctggcggc agcgcaagaa gatgaccatg attacggatt cactggccgt 180

cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc caacttaatc gccttgcagc 240

acatcccccc ttcgccagct ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca 300

acagttgcgc agcctgaatg gcgaatggta ataatagcat aacccccttgg ggcctctaaa 360

cgggtcttga ggggtttttt gctgaaacct caggcatttg agaagcacac ggtcacactg 420

cttccggtag tcaataaacc ggtaaaccag caatagacat aagcggctat ttaacgaccc 480

tgccctgaac cgacgacaag ctgacgaccg ggtctccgca agtggcactt ttcggggaaa 540

tgtgcgcgga acccctattt gtttatttttt ctaaatacat tcaaatatgt atccgctcat 600

gaattaattc ttagaaaaac tcatcgagca tcaaatgaaa ctgcaattta ttcatatcag 660

gattatcaat accatatttt tgaaaaagcc gtttctgtaa tgaaggagaa aactcaccga 720

ggcagttcca taggatggca agatcctggt atcggtctgc gattccgact cgtccaacat 780

caatacaacc tattaatttc ccctcgtcaa aaataaggtt atcaagtgag aaatcaccat 840

gagtgacgac tgaatccggt gagaatggca aaagtttatg catttctttc cagacttgtt 900

caacaggcca gccattacgc tcgtcatcaa aatcactcgc atcaaccaaa ccgttattca 960

-continued

ttcgtgattg cgcctgagcg agacgaaata cgcggtcgct gttaaaagga caattacaaa 1020

caggaatcga atgcaaccgg cgcaggaaca ctgccagcgc atcaacaata ttttcacctg 1080

aatcaggata ttcttctaat acctggaatg ctgttttccc ggggatcgca gtggtgagta 1140

accatgcatc atcaggagta cggataaaat gcttgatggt cggaagaggc ataaattccg 1200

tcagccagtt tagtctgacc atctcatctg taacatcatt ggcaacgcta cctttgccat 1260

gtttcagaaa caactctggc gcatcgggct tcccatacaa tcgatagatt gtcgcacctg 1320

attgcccgac attatcgcga gcccatttat acccatataa atcagcatcc atgttggaat 1380

ttaatcgcgg cctagagcaa gacgtttccc gttgaatatg gctcatactc ttccttttc 1440

aatattattg aagcatttat cagggttatt gtctcatgag cggatacata tttgaatgta 1500

tttagaaaaa taaacaaata ggcatgctag cgcagaaacg tcctagaaga tgccaggagg 1560

atacttagca gagagacaat aaggccggag cgaagccgtt tttccatagg ctccgccccc 1620

ctgacgaaca tcacgaaatc tgacgctcaa atcagtggtg gcgaaacccg acaggactat 1680

aaagatacca ggcgtttccc cctgatggct ccctcttgcg ctctcctgtt cccgtcctgc 1740

ggcgtccgtg ttgtggtgga ggctttaccc aaatcaccac gtcccgttcc gtgtagacag 1800

ttcgctccaa gctgggctgt gtgcaagaac cccccgttca gcccgactgc tgcgccttat 1860

ccggtaacta tcatcttgag tccaacccgg aaagacacga caaaacgcca ctggcagcag 1920

ccattggtaa ctgagaatta gtggatttag atatcgagag tcttgaagtg gtggcctaac 1980

agaggctaca ctgaaaggac agtatttggt atctgcgctc cactaaagcc agttaccagg 2040

ttaagcagtt ccccaactga cttaaccttc gatcaaaccg cctccccagg cggtttttc 2100

gtttacagag caggagatta cgacgatcgt aaaaggatct caagaagatc ctttacggat 2160

tcccgacacc atcactctag atttcagtgc aatttatctc ttcaaatgta gcacctgaag 2220

tcagccccat acgatataag ttgtaattct catgttagtc atgccccgcg cccaccggaa 2280

ggagctgact gggttgaagg ctctcaaggg catcggtcga gatcccggtg cctaatgagt 2340

gagctaactt acattaattg cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc 2400

gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg 2460

ccagggtggt ttttcttttc accagtgaga cgggcaacag ctgattgccc ttcaccgcct 2520

ggccctgaga gagttgcagc aagcggtcca cgctggtttg ccccagcagg cgaaaatcct 2580

gtttgatggt ggttaacggc gggatataac atgagctgtc ttcggtatcg tcgtatccca 2640

ctaccgagat gtccgcacca acgcgcagcc cggactcggt aatggcgcgc attgcgccca 2700

gcgccatctg atcgttggca accagcatcg cagtgggaac gatgccctca ttcagcattt 2760

gcatggtttg ttgaaaaccg gacatggcac tccagtcgcc ttcccgttcc gctatcggct 2820

gaatttgatt gcgagtgaga tatttatgcc agccagccag acgcagacgc gccgagacag 2880

aacttaatgg gcccgctaac agcgcgattt gctggtgacc caatgcgacc agatgctcca 2940

cgcccagtcg cgtaccgtct tcatgggaga aaataatact gttgatgggt gtctggtcag 3000

agacatcaag aaataacgcc ggaacattag tgcaggcagc ttccacagca atggcatcct 3060

ggtcatccag cggatagtta atgatcagcc cactgacgcg ttgcgcgaga agattgtgca 3120

ccgccgcttt acaggcttcg acgccgcttc gttctaccat cgacaccacc acgctggcac 3180

ccagttgatc ggcgcgagat ttaatcgccg cgacaatttg cgacggcgcg tgcagggcca 3240

-continued

gactggaggt ggcaacgcca atcagcaacg actgtttgcc cgccagttgt tgtgccacgc 3300

ggttgggaat gtaattcagc tccgccatcg ccgcttccac tttttcccgc gttttcgcag 3360

aaacgtggct ggcctggttc accacgcggg aaacggtctg ataagagaca ccggcatact 3420

ctgcgacatc gtataacgtt actggtttca cattcaccac cctgaattga ctctcttccg 3480

ggcgctatca tgccataccg cgaaaggttt tgcgccattc gatggtgtcc gggatctcga 3540

cgctctccct tatg 3554

<210> SEQ ID NO 80
<211> LENGTH: 3554
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 80

ctagtaaatt cgcgtttcta cggtagccgg gcgctaatac gactcactat agggctggga 60

cgaggttggc tgcggaccca tcagatgggt ggactttaga acagaggaga taaagatgac 120

ccatctgata acctggcggc agcgcaagaa gatgaccatg attacggatt cactggccgt 180

cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc caacttaatc gccttgcagc 240

acatcccccc ttcgccagct ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca 300

acagttgcgc agcctgaatg gcgaatggta ataatagcat aacccccttgg ggcctctaaa 360

cgggtcttga ggggtttttt gctgaaacct caggcatttg agaagcacac ggtcacactg 420

cttccggtag tcaataaacc ggtaaaccag caatagacat aagcggctat ttaacgaccc 480

tgccctgaac cgacgacaag ctgacgaccg ggtctccgca agtggcactt ttcggggaaa 540

tgtgcgcgga acccctattt gtttatttttt ctaaatacat tcaaatatgt atccgctcat 600

gaattaattc ttagaaaaac tcatcgagca tcaaatgaaa ctgcaattta ttcatatcag 660

gattatcaat accatatttt tgaaaaagcc gtttctgtaa tgaaggagaa aactcaccga 720

ggcagttcca taggatggca agatcctggt atcggtctgc gattccgact cgtccaacat 780

caatacaacc tattaatttc ccctcgtcaa aaataaggtt atcaagtgag aaatcaccat 840

gagtgacgac tgaatccggt gagaatggca aaagtttatg catttctttc cagacttgtt 900

caacaggcca gccattacgc tcgtcatcaa aatcactcgc atcaaccaaa ccgttattca 960

ttcgtgattg cgcctgagcg agacgaaata cgcggtcgct gttaaaagga caattacaaa 1020

caggaatcga atgcaaccgg cgcaggaaca ctgccagcgc atcaacaata ttttcacctg 1080

aatcaggata ttcttctaat acctggaatg ctgttttccc ggggatcgca gtggtgagta 1140

accatgcatc atcaggagta cggataaaat gcttgatggt cggaagaggc ataaattccg 1200

tcagccagtt tagtctgacc atctcatctg taacatcatt ggcaacgcta cctttgccat 1260

gtttcagaaa caactctggc gcatcgggct tcccatacaa tcgatagatt gtcgcacctg 1320

attgcccgac attatcgcga gcccatttat acccatataa atcagcatcc atgttggaat 1380

ttaatcgcgg cctagagcaa gacgtttccc gttgaatatg gctcatactc ttccttttttc 1440

aatattattg aagcatttat cagggttatt gtctcatgag cggatacata tttgaatgta 1500

tttagaaaaa taaacaaata ggcatgctag cgcagaaacg tcctagaaga tgccaggagg 1560

atacttagca gagagacaat aaggccggag cgaagccgtt tttccatagg ctccgccccc 1620

-continued

ctgacgaaca tcacgaaatc tgacgctcaa atcagtggtg gcgaaacccg acaggactat 1680

aaagatacca ggcgtttccc cctgatggct ccctcttgcg ctctcctgtt cccgtcctgc 1740

ggcgtccgtg ttgtggtgga ggctttaccc aaatcaccac gtcccgttcc gtgtagacag 1800

ttcgctccaa gctgggctgt gtgcaagaac cccccgttca gcccgactgc tgcgccttat 1860

ccggtaacta tcatcttgag tccaacccgg aaagacacga caaaacgcca ctggcagcag 1920

ccattggtaa ctgagaatta gtggatttag atatcgagag tcttgaagtg gtggcctaac 1980

agaggctaca ctgaaaggac agtatttggt atctgcgctc cactaaagcc agttaccagg 2040

ttaagcagtt ccccaactga cttaaccttc gatcaaaccg cctccccagg cggttttttc 2100

gtttacagag caggagatta cgacgatcgt aaaaggatct caagaagatc ctttacggat 2160

tcccgacacc atcactctag atttcagtgc aatttatctc ttcaaatgta gcacctgaag 2220

tcagccccat acgatataag ttgtaattct catgttagtc atgccccgcg cccaccggaa 2280

ggagctgact gggttgaagg ctctcaaggg catcggtcga gatcccggtg cctaatgagt 2340

gagctaactt acattaattg cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc 2400

gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg 2460

ccagggtggt ttttcttttc accagtgaga cgggcaacag ctgattgccc ttcaccgcct 2520

ggccctgaga gagttgcagc aagcggtcca cgctggtttg ccccagcagg cgaaaatcct 2580

gtttgatggt ggttaacggc gggatataac atgagctgtc ttcggtatcg tcgtatccca 2640

ctaccgagat gtccgcacca acgcgcagcc cggactcggt aatggcgcgc attgcgccca 2700

gcgccatctg atcgttggca accagcatcg cagtgggaac gatgccctca ttcagcattt 2760

gcatggtttg ttgaaaaccg gacatggcac tccagtcgcc ttcccgttcc gctatcggct 2820

gaatttgatt gcgagtgaga tatttatgcc agccagccag acgcagacgc gccgagacag 2880

aacttaatgg gcccgctaac agcgcgattt gctggtgacc caatgcgacc agatgctcca 2940

cgcccagtcg cgtaccgtct tcatgggaga aaataatact gttgatgggt gtctggtcag 3000

agacatcaag aaataacgcc ggaacattag tgcaggcagc ttccacagca atggcatcct 3060

ggtcatccag cggatagtta atgatcagcc cactgacgcg ttgcgcgaga agattgtgca 3120

ccgccgcttt acaggcttcg acgccgcttc gttctaccat cgacaccacc acgctggcac 3180

ccagttgatc ggcgcgagat ttaatcgccg cgacaatttg cgacggcgcg tgcagggcca 3240

gactggaggt ggcaacgcca atcagcaacg actgtttgcc cgccagttgt tgtgccacgc 3300

ggttgggaat gtaattcagc tccgccatcg ccgcttccac tttttcccgc gttttcgcag 3360

aaacgtggct ggcctggttc accacgcggg aaacggtctg ataagagaca ccggcatact 3420

ctgcgacatc gtataacgtt actggtttca cattcaccac cctgaattga ctctcttccg 3480

ggcgctatca tgccataccg cgaaaggttt tgcgccattc gatggtgtcc gggatctcga 3540

cgctctccct tatg 3554

<210> SEQ ID NO 81
<211> LENGTH: 3553
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 81

-continued

ctagtaaatt cgcgtttcta cggtagccgg gcgctaatac gactcactat agggtcattc 60

gacgccatct tcattcacaa aactgggagg gactttagaa cagaggagat aaagatgctc 120

ccagttttaa cctggcggca gcgcaagaag atgaccatga ttacggattc actggccgtc 180

gttttacaac gtcgtgactg ggaaaaccct ggcgttaccc aacttaatcg ccttgcagca 240

catccccctt tcgccagctg gcgtaatagc gaagaggccc gcaccgatcg cccttcccaa 300

cagttgcgca gcctgaatgg cgaatggtaa taatagcata acccccttggg gcctctaaac 360

gggtcttgag gggttttttg ctgaaacctc aggcatttga gaagcacacg gtcacactgc 420

ttccggtagt caataaaccg gtaaaccagc aatagacata agcggctatt taacgaccct 480

gccctgaacc gacgacaagc tgacgaccgg gtctccgcaa gtggcacttt tcggggaaat 540

gtgcgcggaa cccctatttg tttatttttc taaatacatt caaatatgta tccgctcatg 600

aattaattct tagaaaaact catcgagcat caaatgaaac tgcaatttat tcatatcagg 660

attatcaata ccatattttt gaaaaagccg tttctgtaat gaaggagaaa actcaccgag 720

gcagttccat aggatggcaa gatcctggta tcggtctgcg attccgactc gtccaacatc 780

aatacaacct attaatttcc cctcgtcaaa aataaggtta tcaagtgaga aatcaccatg 840

agtgacgact gaatccggtg agaatggcaa aagtttatgc atttctttcc agacttgttc 900

aacaggccag ccattacgct cgtcatcaaa atcactcgca tcaaccaaac cgttattcat 960

tcgtgattgc gcctgagcga gacgaaatac gcggtcgctg ttaaaaggac aattacaaac 1020

aggaatcgaa tgcaaccggc gcaggaacac tgccagcgca tcaacaatat tttcacctga 1080

atcaggatat tcttctaata cctggaatgc tgttttcccg gggatcgcag tggtgagtaa 1140

ccatgcatca tcaggagtac ggataaaatg cttgatggtc ggaagaggca taaattccgt 1200

cagccagttt agtctgacca tctcatctgt aacatcattg gcaacgctac ctttgccatg 1260

tttcagaaac aactctggcg catcgggctt cccatacaat cgatagattg tcgcacctga 1320

ttgcccgaca ttatcgcgag cccatttata cccatataaa tcagcatcca tgttggaatt 1380

taatcgcggc ctagagcaag acgtttcccg ttgaatatgg ctcatactct tccttttca 1440

atattattga agcatttatc agggttattg tctcatgagc ggatacatat ttgaatgtat 1500

ttagaaaaat aaacaaatag gcatgctagc gcagaaacgt cctagaagat gccaggagga 1560

tacttagcag agagacaata aggccggagc gaagccgttt ttccataggc tccgcccccc 1620

tgacgaacat cacgaaatct gacgctcaaa tcagtggtgg cgaaacccga caggactata 1680

aagataccag gcgtttcccc ctgatggctc cctcttgcgc tctcctgttc ccgtcctgcg 1740

gcgtccgtgt tgtggtggag gctttaccca aatcaccacg tcccgttccg tgtagacagt 1800

tcgctccaag ctgggctgtg tgcaagaacc ccccgttcag cccgactgct gcgccttatc 1860

cggtaactat catcttgagt ccaacccgga aagacacgac aaaacgccac tggcagcagc 1920

cattggtaac tgagaattag tggatttaga tatcgagagt cttgaagtgg tggcctaaca 1980

gaggctacac tgaaaggaca gtatttggta tctgcgctcc actaaagcca gttaccaggt 2040

taagcagttc cccaactgac ttaaccttcg atcaaaccgc ctccccaggc ggtttttttcg 2100

tttacagagc aggagattac gacgatcgta aaaggatctc aagaagatcc tttacggatt 2160

cccgacacca tcactctaga tttcagtgca atttatctct tcaaatgtag cacctgaagt 2220

cagccccata cgatataagt tgtaattctc atgttagtca tgccccgcgc ccaccggaag 2280

-continued

gagctgactg ggttgaaggc tctcaagggc atcggtcgag atcccggtgc ctaatgagtg 2340

agctaactta cattaattgc gttgcgctca ctgcccgctt tccagtcggg aaacctgtcg 2400

tgccagctgc attaatgaat cggccaacgc gcggggagag gcggtttgcg tattgggcgc 2460

cagggtggtt tttcttttca ccagtgagac gggcaacagc tgattgccct tcaccgcctg 2520

gccctgagag agttgcagca agcggtccac gctggtttgc cccagcaggc gaaaatcctg 2580

tttgatggtg gttaacggcg ggatataaca tgagctgtct tcggtatcgt cgtatcccac 2640

taccgagatg tccgcaccaa cgcgcagccc ggactcggta atggcgcgca ttgcgcccag 2700

cgccatctga tcgttggcaa ccagcatcgc agtgggaacg atgccctcat tcagcatttg 2760

catggtttgt tgaaaaccgg acatggcact ccagtcgcct tcccgttccg ctatcggctg 2820

aatttgattg cgagtgagat atttatgcca gccagccaga cgcagacgcg ccgagacaga 2880

acttaatggg cccgctaaca gcgcgatttg ctggtgaccc aatgcgacca gatgctccac 2940

gcccagtcgc gtaccgtctt catgggagaa aataatactg ttgatgggtg tctggtcaga 3000

gacatcaaga aataacgccg gaacattagt gcaggcagct tccacagcaa tggcatcctg 3060

gtcatccagc ggatagttaa tgatcagccc actgacgcgt tgcgcgagaa gattgtgcac 3120

cgccgcttta caggcttcga cgccgcttcg ttctaccatc gacaccacca cgctggcacc 3180

cagttgatcg gcgcgagatt taatcgccgc gacaatttgc gacggcgcgt gcagggccag 3240

actggaggtg gcaacgccaa tcagcaacga ctgtttgccc gccagttgtt gtgccacgcg 3300

gttgggaatg taattcagct ccgccatcgc cgcttccact ttttcccgcg ttttcgcaga 3360

aacgtggctg gcctggttca ccacgcggga aacggtctga taagagacac cggcatactc 3420

tgcgacatcg tataacgtta ctggtttcac attcaccacc ctgaattgac tctcttccgg 3480

gcgctatcat gccataccgc gaaaggtttt gcgccattcg atggtgtccg ggatctcgac 3540

gctctcccct atg 3553

<210> SEQ ID NO 82
<211> LENGTH: 3554
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 82

ctagtaaatt cgcgtttcta cggtagccgg gcgctaatac gactcactat agggagccag 60

attgcgatcg ccctcccacg tgctcagatc ggactttaga acagaggaga taaagatgga 120

tctgagcaca acctggcggc agcgcaagaa gatgaccatg attacggatt cactggccgt 180

cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc caacttaatc gccttgcagc 240

acatcccccc ttcgccagct ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca 300

acagttgcgc agcctgaatg gcgaatggta ataatagcat aacccccttgg ggcctctaaa 360

cgggtcttga ggggtttttt gctgaaacct caggcatttg agaagcacac ggtcacactg 420

cttccggtag tcaataaacc ggtaaaccag caatagacat aagcggctat ttaacgaccc 480

tgccctgaac cgacgacaag ctgacgaccg ggtctccgca agtggcactt ttcggggaaa 540

tgtgcgcgga acccctattt gtttatttttt ctaaatacat tcaaatatgt atccgctcat 600

gaattaattc ttagaaaaac tcatcgagca tcaaatgaaa ctgcaattta ttcatatcag 660

-continued

gattatcaat accatatttt tgaaaaagcc gtttctgtaa tgaaggagaa aactcaccga 720

ggcagttcca taggatggca agatcctggt atcggtctgc gattccgact cgtccaacat 780

caatacaacc tattaatttc ccctcgtcaa aaataaggtt atcaagtgag aaatcaccat 840

gagtgacgac tgaatccggt gagaatggca aaagtttatg catttctttc cagacttgtt 900

caacaggcca gccattacgc tcgtcatcaa aatcactcgc atcaaccaaa ccgttattca 960

ttcgtgattg cgcctgagcg agacgaaata cgcggtcgct gttaaaagga caattacaaa 1020

caggaatcga atgcaaccgg cgcaggaaca ctgccagcgc atcaacaata ttttcacctg 1080

aatcaggata ttcttctaat acctggaatg ctgttttccc ggggatcgca gtggtgagta 1140

accatgcatc atcaggagta cggataaaat gcttgatggt cggaagaggc ataaattccg 1200

tcagccagtt tagtctgacc atctcatctg taacatcatt ggcaacgcta cctttgccat 1260

gtttcagaaa caactctggc gcatcgggct tcccatacaa tcgatagatt gtcgcacctg 1320

attgcccgac attatcgcga gcccatttat acccatataa atcagcatcc atgttggaat 1380

ttaatcgcgg cctagagcaa gacgtttccc gttgaatatg gctcatactc ttccttttc 1440

aatattattg aagcatttat cagggttatt gtctcatgag cggatacata tttgaatgta 1500

tttagaaaaa taaacaaata ggcatgctag cgcagaaacg tcctagaaga tgccaggagg 1560

atacttagca gagagacaat aaggccggag cgaagccgtt tttccatagg ctccgccccc 1620

ctgacgaaca tcacgaaatc tgacgctcaa atcagtggtg gcgaaacccg acaggactat 1680

aaagatacca ggcgtttccc cctgatggct ccctcttgcg ctctcctgtt cccgtcctgc 1740

ggcgtccgtg ttgtggtgga ggctttaccc aaatcaccac gtcccgttcc gtgtagacag 1800

ttcgctccaa gctgggctgt gtgcaagaac cccccgttca gcccgactgc tgcgccttat 1860

ccggtaacta tcatcttgag tccaacccgg aaagacacga caaaacgcca ctggcagcag 1920

ccattggtaa ctgagaatta gtggatttag atatcgagag tcttgaagtg gtggcctaac 1980

agaggctaca ctgaaaggac agtatttggt atctgcgctc cactaaagcc agttaccagg 2040

ttaagcagtt ccccaactga cttaaccttc gatcaaaccg cctccccagg cggttttttc 2100

gtttacagag caggagatta cgacgatcgt aaaaggatct caagaagatc ctttacggat 2160

tcccgacacc atcactctag atttcagtgc aatttatctc ttcaaatgta gcacctgaag 2220

tcagccccat acgatataag ttgtaattct catgttagtc atgccccgcg cccaccggaa 2280

ggagctgact gggttgaagg ctctcaaggg catcggtcga gatcccggtg cctaatgagt 2340

gagctaactt acattaattg cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc 2400

gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg 2460

ccagggtggt ttttcttttc accagtgaga cgggcaacag ctgattgccc ttcaccgcct 2520

ggccctgaga gagttgcagc aagcggtcca cgctggtttg ccccagcagg cgaaaatcct 2580

gtttgatggt ggttaacggc gggatataac atgagctgtc ttcggtatcg tcgtatccca 2640

ctaccgagat gtccgcacca acgcgcagcc cggactcggt aatggcgcgc attgcgccca 2700

gcgccatctg atcgttggca accagcatcg cagtgggaac gatgccctca ttcagcattt 2760

gcatggtttg ttgaaaaccg gacatggcac tccagtcgcc ttcccgttcc gctatcggct 2820

gaatttgatt gcgagtgaga tatttatgcc agccagccag acgcagacgc gccgagacag 2880

aacttaatgg gcccgctaac agcgcgattt gctggtgacc caatgcgacc agatgctcca 2940

-continued

```
cgcccagtcg cgtaccgtct tcatgggaga aaataatact gttgatgggt gtctggtcag   3000

agacatcaag aaataacgcc ggaacattag tgcaggcagc ttccacagca atggcatcct   3060

ggtcatccag cggatagtta atgatcagcc cactgacgcg ttgcgcgaga agattgtgca   3120

ccgccgcttt acaggcttcg acgccgcttc gttctaccat cgacaccacc acgctggcac   3180

ccagttgatc ggcgcgagat ttaatcgccg cgacaatttg cgacggcgcg tgcagggcca   3240

gactggaggt ggcaacgcca atcagcaacg actgtttgcc cgccagttgt tgtgccacgc   3300

ggttgggaat gtaattcagc tccgccatcg ccgcttccac tttttcccgc gttttcgcag   3360

aaacgtggct ggcctggttc accacgcggg aaacggtctg ataagagaca ccggcatact   3420

ctgcgacatc gtataacgtt actggtttca cattcaccac cctgaattga ctctcttccg   3480

ggcgctatca tgccataccg cgaaaggttt tgcgccattc gatggtgtcc gggatctcga   3540

cgctctccct tatg                                                     3554
```

We claim:

1. A method of detecting a target pathogen nucleic acid in a sample, the method comprising the steps of:

(a) contacting a biological sample obtained from a subject to a pathogen detection agent under conditions that promote binding of the pathogen detection agent to the target pathogen if present in the sample;

(b) isolating nucleic acids from pathogen bound by the pathogen detection agent;

(c) amplifying the isolated nucleic acids using isothermal amplification; and

(d) contacting the amplified nucleic acid to a toehold switch, wherein the toehold switch encodes at least a portion of a reporter protein and comprises one or more single-stranded toehold sequence domains that are complementary to a target pathogen nucleic acid or the reverse complement thereof, wherein the contacting occurs under conditions that allow translation of the coding domain in the presence of the target nucleic acid but not in the absence of the target nucleic acid, and detecting the reporter protein as an indicator that the target pathogen nucleic acid is present in the amplified nucleic acids.

2. The method of claim **1**, wherein the pathogen detection agent is a norovirus detection agent and the target pathogen nucleic acid is norovirus RNA.

3. The method of claim **1**, wherein target pathogen nucleic acid is detected at concentrations in a range of zeptomoles/liter (zM).

4. The method of claim **1**, wherein target pathogen nucleic acid is detected at concentration between about 270 zM to about 270 aM.

5. The method of claim **1**, wherein the pathogen detection agent is a synbody.

6. The method of claim **5**, wherein the synbody comprises biotin.

7. The method of claim **6**, wherein the biotin-containing synbody is bound to a streptavidin-coated magnetic bead.

8. The method of claim **6**, wherein isolating comprises a magnetic capture assay.

9. The method of claim **1**, wherein the toehold switch encodes at least a portion of lacZ.

10. The method of claim **1**, wherein the toehold switch encodes lacZα and the amplified nucleic acids are contacted under conditions which promote formation of a lacZ tetramer.

11. The method of claim **10**, wherein lacZω is provided on a substrate to which the amplified nucleic acids are contacted.

12. A synthetic norovirus-specific toehold switch sensor comprising a fully or partially double-stranded stem domain, a loop domain, a toehold domain, and at least a portion of a coding sequence of a reporter gene, wherein the toehold domain and at least a portion of the stem domain are complementary to a target norovirus RNA sequence.

13. The toehold switch sensor of claim **12**, comprising a RNA sequence selected from SEQ ID NOs:1-12.

14. A kit for detecting a pathogen-associated nucleic acid, comprising a plurality of preserved test articles, a pathogen detection agent, a plurality of toehold switches that encode at least a portion of a reporter protein and comprise one or more single-stranded toehold sequence domains that are complementary to a target pathogen nucleic acid or the reverse complement thereof, and an electronic optical reader.

15. The kit of claim **14**, wherein the pathogen detection agent is a synbody.

16. The kit of claim **14**, wherein the plurality of preserved test articles comprises one or more selected from preserved paper test articles and preserved test tube articles.

17. The kit of claim **14**, further comprising instructions for performing the method of claim **1**.

* * * * *