



(86) Date de dépôt PCT/PCT Filing Date: 2010/03/04

(87) Date publication PCT/PCT Publication Date: 2010/09/10

(85) Entrée phase nationale/National Entry: 2011/09/02

(86) N° demande PCT/PCT Application No.: US 2010/026165

(87) N° publication PCT/PCT Publication No.: 2010/102085

(30) Priorité/Priority: 2009/03/04 (US61/157,332)

(51) Cl.Int./Int.Cl. *C12N 15/54* (2006.01),
C12N 9/12 (2006.01), *C12N 9/96* (2006.01),
C12P 19/34 (2006.01)

(71) Demandeur/Applicant:
BOARD OF REGENTS, THE UNIVERSITY OF TEXAS
SYSTEM, US

(72) Inventeurs/Inventors:
LAMBOWITZ, ALAN M., US;
MOHR, SABINE, US;
MOHR, GEORG, US;
GHANEM, EMAN, US

(74) Agent: NORTON ROSE CANADA
S.E.N.C.R.L., S.R.L./LLP

(54) Titre : PROTEINES DE FUSION DE TRANSCRIPTASE INVERSE STABILISEE

(54) Title: STABILIZED REVERSE TRANSCRIPTASE FUSION PROTEINS

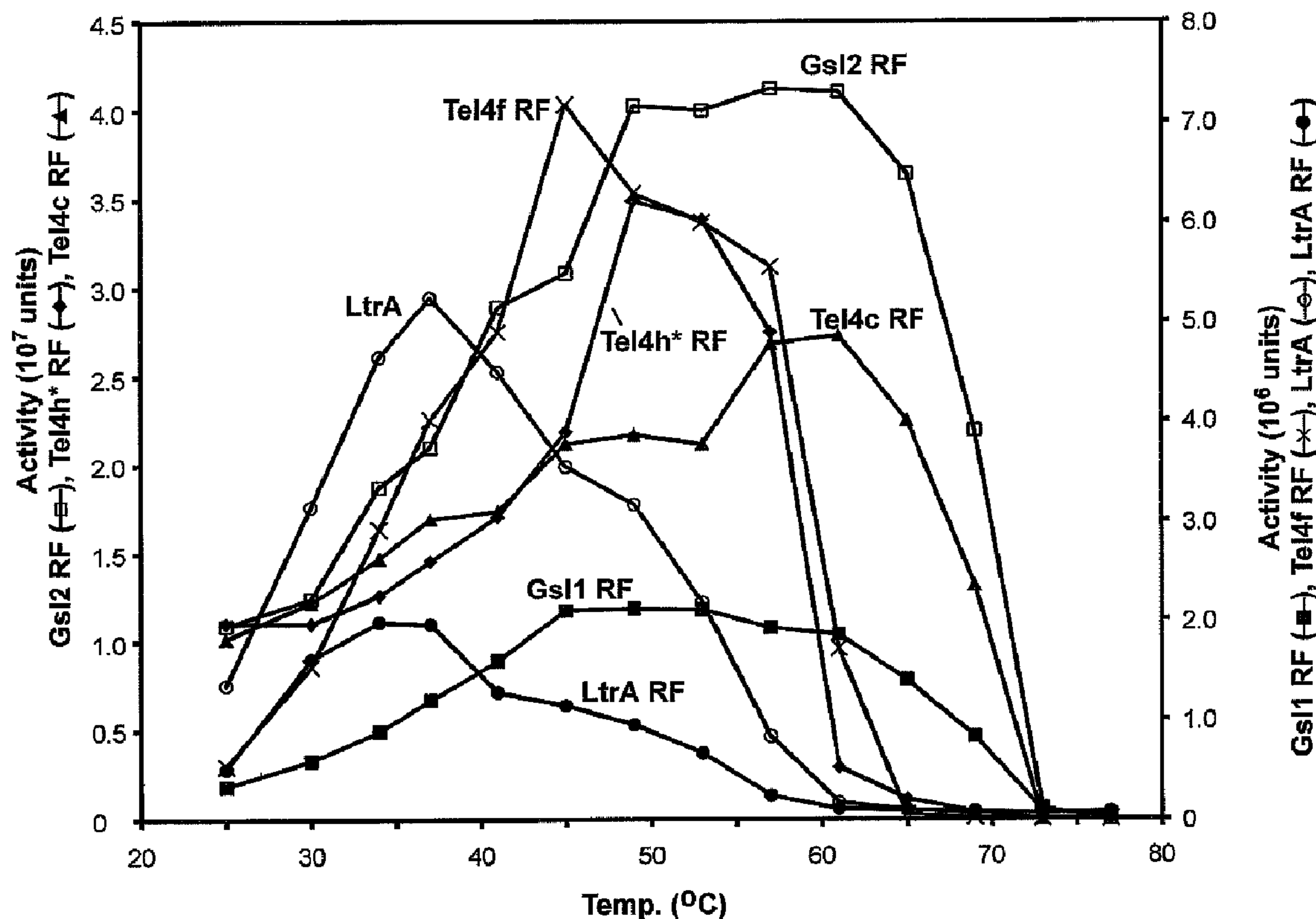


FIG. 11

(57) Abrégé/Abstract:

Stabilized reverse transcriptase fusion proteins including a thermostable reverse transcriptase connected to a stabilizer protein are described. Attaching the stabilizer protein to the thermostable reverse transcriptase stabilizes the fusion protein and can aid in its

(57) **Abrégé(suite)/Abstract(continued):**

purification, provide increased solubility, allow for longer storage, or allow the fusion protein to be used under more rigorous conditions such as higher temperature. The stabilized reverse transcriptase fusion protein can also include a linker between the stabilizer protein and the thermostable reverse transcriptase. The stabilized reverse transcriptase fusion proteins are suitable for use in nucleic acid amplification methods such as the reverse transcription polymerase chain reaction and other applications involving cDNA synthesis.

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

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
10 September 2010 (10.09.2010)

PCT

(10) International Publication Number
WO 2010/102085 A4(51) International Patent Classification:
C07K 14/195 (2006.01)

(21) International Application Number:

PCT/US2010/026165

(22) International Filing Date:

4 March 2010 (04.03.2010)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/157,332

4 March 2009 (04.03.2009)

US

(71) Applicant (for all designated States except US): **BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM** [US/US]; 201 West 7th Street, Austin, TX 78701 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **LAMBOWITZ, Alan, M.** [US/US]; 4204 Benedict Lane, Austin, TX 78746 (US). **MOHR, Sabine** [US/US]; 4514 Avenue C., Austin, TX 78751 (US). **MOHR, Georg** [US/US]; 4514 Avenue C., Austin, TX 78751 (US). **GHANEM, Eman** [EG/US]; 10926 Jollyville Road, Apt. 1408, Austin, TX 78759 (US).(74) Agent: **WIEDEMANN, John, T.**; Calfee, Halter & Griswold LLP, 800 Superior Ave., Ste. 1400, Cleveland, OH 44114 (US).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

[Continued on next page]

(54) Title: STABILIZED REVERSE TRANSCRIPTASE FUSION PROTEINS

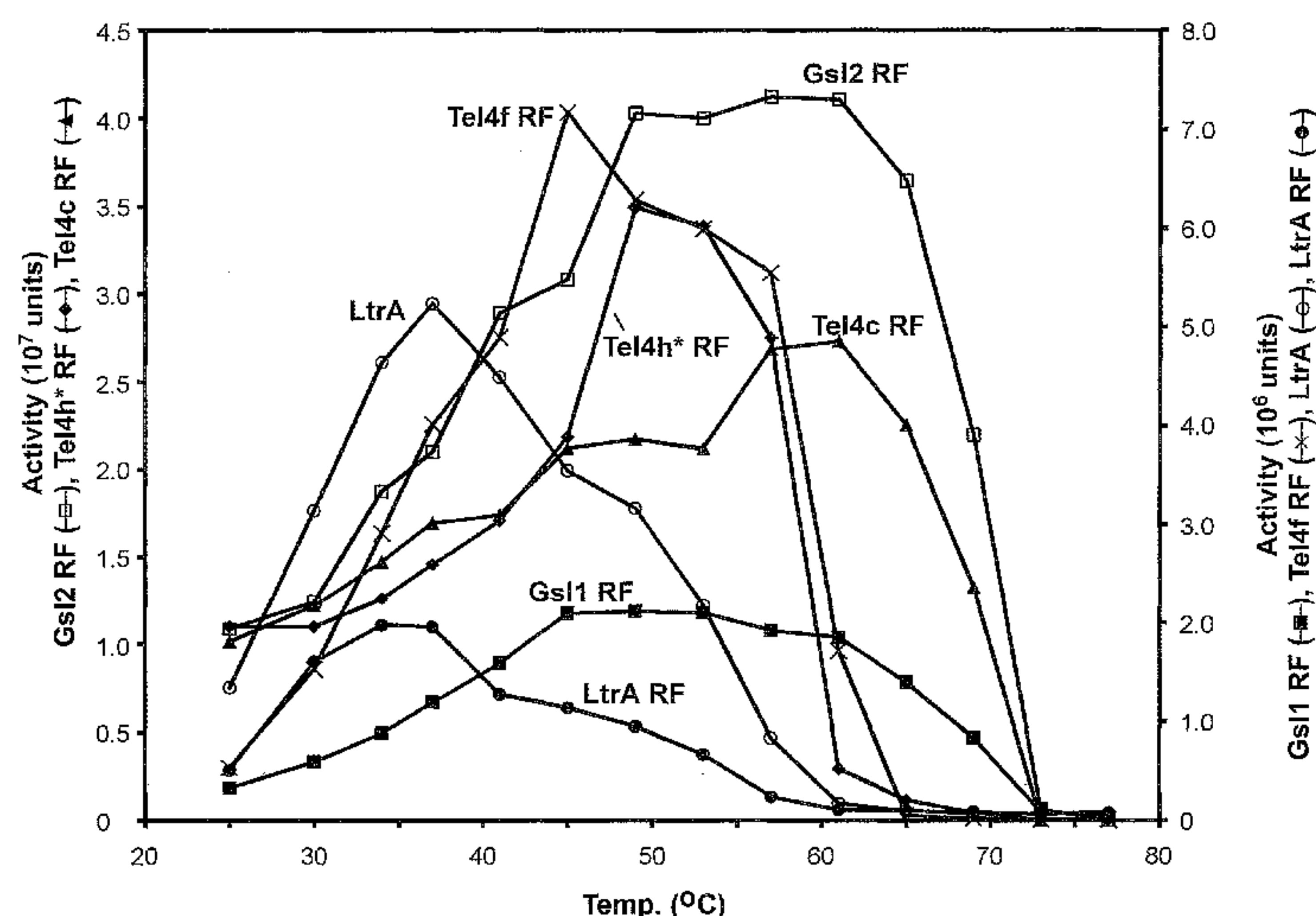


FIG. 11

(57) **Abstract:** Stabilized reverse transcriptase fusion proteins including a thermostable reverse transcriptase connected to a stabilizer protein are described. Attaching the stabilizer protein to the thermostable reverse transcriptase stabilizes the fusion protein and can aid in its purification, provide increased solubility, allow for longer storage, or allow the fusion protein to be used under more rigorous conditions such as higher temperature. The stabilized reverse transcriptase fusion protein can also include a linker between the stabilizer protein and the thermostable reverse transcriptase. The stabilized reverse transcriptase fusion proteins are suitable for use in nucleic acid amplification methods such as the reverse transcription polymerase chain reaction and other applications involving cDNA synthesis.

WO 2010/102085 A4



Published:

- *with international search report (Art. 21(3))*
- *with amended claims (Art. 19(1))*

(88) Date of publication of the international search report:
18 November 2010

Date of publication of the amended claims: 6 January 2011

STABILIZED REVERSE TRANSCRIPTASE FUSION PROTEINS

CONTINUING APPLICATION DATA

[0001] This application claims the benefit of U.S. Provisional Application Serial No. 61/157,332, filed March 4, 2009, which is incorporated by reference herein.

GOVERNMENT FUNDING

[0002] This work was supported, at least in part, by grant number GM37949-22 from the Department of Health and Human Services, National Institutes of Health. The United States government may have certain rights in this invention.

BACKGROUND OF THE INVENTION

[0003] Reverse transcription polymerase chain reaction, abbreviated as RT-PCR, is a well known technique for amplifying RNA. In RT-PCR, an RNA strand is reverse transcribed into complementary DNA (cDNA), which is then amplified using DNA polymerase in the polymerase chain reaction. In the first step of this process, cDNA is made from an RNA template using deoxyribonucleotide phosphates and reverse transcriptase together with a DNA primer.

[0004] Synthesis of cDNA from the RNA template can be hindered by RNA secondary and tertiary structures, which consist of helices and various other kinds of kinks in the RNA strand. RNA secondary and tertiary structure can be decreased by carrying out the reaction at a higher temperature (*e.g.*, above 50 °C) or by adding denaturing additives. However, the addition of denaturing additives is undesirable because it often reduces reverse transcriptase activity. Higher temperatures also provide the advantage of increasing the specificity of DNA synthesis by decreasing non-specific primer binding. Unfortunately, only a limited number of reverse transcriptases capable of operating at high temperature are currently available, and these exhibit relatively low fidelity DNA polymerization. For example,

commercially available Avian Myeloblastosis Virus reverse transcriptase includes RNase H activity and can function at 37 °C, but has a fidelity of only about 1.7×10^{-4} . RNase H activity competes with the DNA polymerase activity and the primer binding site and, therefore, cDNA yield is lower. Accordingly, there is a need for reverse transcriptase enzymes that are able to carry out reverse transcription at higher temperatures, including those that have high fidelity and processivity. Such enzymes are beneficial because higher temperatures decrease obstructing RNA secondary and tertiary structure and increase the specificity of reverse transcription by allowing the use of longer and more specific primers.

SUMMARY OF THE INVENTION

[0005] In one aspect, the invention provides a stabilized reverse transcriptase (RT) fusion protein that includes a thermostable reverse transcriptase connected to a stabilizer protein. In one embodiment of the stabilized reverse transcriptase fusion protein, the thermostable reverse transcriptase is a bacterial reverse transcriptase. In a further embodiment, the bacterial reverse transcriptase is a group II intron-derived reverse transcriptase. Examples of thermostable bacterial reverse transcriptases include *Thermosynechococcus elongatus* reverse transcriptase and *Geobacillus stearothermophilus* reverse transcriptase. In another embodiment, the thermostable reverse transcriptase exhibits high fidelity cDNA synthesis. In yet another embodiment, the thermostable reverse transcriptase includes a polypeptide with an amino acid sequence identity that is substantially similar to a sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 5.

[0006] The stabilized reverse transcriptase fusion protein includes a stabilizer protein that, when linked to the reverse transcriptase, enhances the shelf life and/or the thermal stability and/or the solubility of the thermostable reverse transcriptase. In certain embodiments, the stabilizer protein is an affinity protein or a solubility-enhancing protein (e.g., a maltose binding protein or N-utilization substance A protein). In additional embodiments, the stabilizer protein is modified by replacing certain charged amino acids with uncharged amino acids.

[0007] The stabilized reverse transcriptase fusion protein can also include a linker peptide that connects the thermostable reverse transcriptase to the stabilizer protein. In some embodiments, this linker peptide is a non-cleavable linker, while in other embodiments it is a

non-cleavable rigid linker. In some embodiments, the linker peptide consists of 1 to 20 amino acids, while in other embodiments the linker peptide consists of 1 to 5 or 3 to 5 amino acids. For example, a rigid non-cleavable linker peptide can include 5 alanine amino acids.

[0008] In additional embodiments, the stabilized reverse transcriptase fusion protein has an amino acid sequence that includes a polypeptide with an amino acid sequence identity that is substantially similar to a sequence selected from the group consisting of SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, or SEQ ID NO: 10. In some embodiments, the stabilized reverse transcriptase fusion protein is a high fidelity reverse transcriptase capable of carrying out reverse transcription with an error frequency of 2.0×10^{-5} or less at a temperature from about 45° to about 65 °C. In further embodiments, the stabilized reverse transcriptase fusion protein is capable of carrying out substantial levels of reverse transcription at temperatures up to about 81 °C.

[0009] Another aspect of the invention provides a method for preparing a cDNA from an RNA molecule that includes the steps of: (a) adding a primer nucleotide sequence to an RNA molecule and (b) incubating the RNA molecule in the presence of one or more modified or unmodified deoxy or dideoxyribonucleoside triphosphates and a stabilized reverse transcriptase fusion protein that includes a thermostable reverse transcriptase connected to a stabilizer protein under conditions sufficient to synthesize a cDNA molecule complementary to all or a portion of the RNA molecule. In particular embodiments, the thermostable reverse transcriptase is connected to the stabilizer protein by a linker peptide (*e.g.*, a non-cleavable or rigid non-cleavable linker peptide). Preferably, the reverse transcription is performed within a temperature range where RNA includes a substantially decreased amount of obstructing stable secondary or tertiary structure. Embodiments of this method include ones in which the thermostable reverse transcriptase is a group II intron-derived reverse transcriptase. In further embodiments of the method, the thermostable reverse transcriptase includes a polypeptide with an amino acid sequence identity that is substantially similar to a sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 5, a non-cleavable linker consists of 1 to 20 amino acids, and the stabilizer protein is an affinity protein or a solubility-enhancing protein. In yet further embodiments of the method, the reverse transcription is performed with an error frequency of 2.0×10^{-5} or less at a temperature from about 45° to about 65 °C.

[0010] Another aspect of the invention provides a DNA expression vector for producing a stabilized reverse transcriptase fusion protein that includes a nucleic acid that encodes a polypeptide with an amino acid sequence identity that is substantially similar to a sequence selected from the group consisting of SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, or SEQ ID NO: 10.

[0011] Another aspect of the invention provides a method of producing a stabilized reverse transcriptase fusion protein that includes the steps of: (a) culturing a host cell that includes a DNA expression vector for producing a stabilized reverse transcriptase fusion protein that includes a nucleic acid that encodes a polypeptide with an amino acid sequence identity that is substantially similar to a sequence selected from the group consisting of SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, or SEQ ID NO: 10; (b) expressing the stabilized reverse transcriptase fusion protein encoded by the DNA expression vector; and (c) isolating the stabilized reverse transcriptase fusion protein from the host cell.

[0012] The stabilized reverse transcriptase fusion protein can facilitate cDNA synthesis at higher temperature, and/or with higher processivity, and/or allow the use of longer, more stable, primers that increase the specificity (*i.e.*, fidelity) of reverse transcription. The stabilized RT fusion protein of the invention can therefore be useful for a number of applications, such as research applications.

[0013] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the invention, as claimed.

BRIEF DESCRIPTION OF THE FIGURES

[0014] Figure 1 is a listing of the amino acid sequence of a reverse transcriptase from *Thermosynechococcus elongatus* bound to a maltose binding protein by a rigid linker (SEQ ID NO: 6). Amino acid residues 1-367 represent the modified maltose binding protein (SEQ ID NO: 11); amino acid residues 368-372 represent the rigid linker (SEQ ID NO: 12); and amino acid residues 373-935 represent the TeI4c ORF (SEQ ID NO: 1).

[0015] Figure 2 is a listing of the amino acid sequence of a reverse transcriptase from *Thermosynechococcus elongatus* bound to a maltose binding protein by a rigid linker (SEQ

ID NO: 7). Amino acid residues 1-367 represent the maltose binding protein (SEQ ID NO: 11); amino acid residues 368-372 represent the rigid linker (SEQ ID NO: 12); and amino acid residues 373-935 represent the TeI4f ORF (SEQ ID NO: 2).

[0016] Figure 3 is a listing of the amino acid sequence of a reverse transcriptase from *Thermosynechococcus elongatus* bound to a maltose binding protein by a rigid linker (SEQ ID NO: 8). Amino acid residues 1-367 represent the maltose binding protein (SEQ ID NO: 11); amino acid residues 368-372 represent the rigid linker (SEQ ID NO: 12); and amino acid residues 373-935 represent the TeI4h* ORF (SEQ ID NO: 3).

[0017] Figure 4 is a listing of the amino acid sequence of a reverse transcriptase from *Geobacillus stearothermophilus* bound to a maltose binding protein by a rigid linker (SEQ ID NO: 9). Amino acid residues 1-367 represent the maltose binding protein (SEQ ID NO: 11); amino acid residues 368-372 represent the rigid linker (SEQ ID NO: 12); and amino acid residues 373-1008 represent the *Geobacillus stearothermophilus* GsI1 ORF (SEQ ID NO: 4).

[0018] Figure 5 is a listing of the amino acid sequence of a reverse transcriptase from *Geobacillus stearothermophilus* bound to a maltose binding protein by a rigid linker (SEQ ID NO: 10). Amino acid residues 1-367 represent the maltose binding protein (SEQ ID NO: 11); amino acid residues 368-372 represent the rigid linker (SEQ ID NO: 12); and amino acid residues 373-792 represent the *Geobacillus stearothermophilus* GsI2 ORF (SEQ ID NO: 5).

[0019] Figure 6 is a listing of the nucleotide sequence of the MalE-TeI4c open reading frame (ORF) rigid fusion of reverse transcriptase from *Thermosynechococcus elongatus* in the pMAL expression construct (SEQ ID NO: 13).

[0020] Figure 7 is a listing of the nucleotide sequence of the MalE-TeI4f ORF rigid fusion of a reverse transcriptase from *Thermosynechococcus elongatus* in the pMAL expression construct (SEQ ID NO: 14).

[0021] Figure 8 is a listing of the nucleotide sequence of the MalE-TeI4h* ORF rigid fusion of a reverse transcriptase from *Thermosynechococcus elongatus* in the pMAL expression construct (SEQ ID NO: 15).

[0022] Figure 9 is a listing of the nucleotide sequence of the MalE-GsI1 ORF rigid fusion of a reverse transcriptase from *Geobacillus stearothermophilus* in the pMAL expression construct (SEQ ID NO: 16).

[0023] Figure 10 is a listing of the nucleotide sequence of the MalE-GsI2 ORF rigid fusion of a reverse transcriptase from *Geobacillus stearothermophilus* in the pMAL expression construct (SEQ ID NO: 17).

[0024] Figure 11 provides a graph showing the poly(rA)/oligo(dT)₄₂ assay of reverse transcriptase (RT) activity at different temperatures. The enzymes assayed were MalE-RF-GsI1, MalE-RF-GsI2, MalE-RF-TeI4c, MalE-RF-TeI4f, MalE-RF-TeI4h*, LtrA, and MalE-RF-LtrA. Reactions were done by incubating the RT (50 nM for TeI4c and 100 nM for all other RTs) with 100 nM poly(rA)/oligo(dT)₄₂ and 5 μ l [α -³²P]-dTTP (3,000 Ci/mmol) in 75 mM KCl, 10 mM MgCl₂, 20 mM Tris-HCl, pH 7.5, and 1 mM DTT. After preincubating the RT with poly(rA)/oligo(dT)₄₂ in the reaction medium for 1 min at the indicated temperature, the reaction was initiated by adding [α -³²P]-dTTP, incubated for times verified to be within the linear range (90 sec for TeI4c RT and 5 min for all other RTs), and stopped by adding EDTA to a final concentration of 250 mM. The polymerization of [α -³²P]-dTTP into high-molecular weight material was quantified by spotting the reaction products onto Whatman DE81 chromatography paper (GE Health care Biosciences Corp), washing with 0.3 M NaCl and 0.03 M sodium citrate, and scanning with a PhosphorImager to quantify radioactivity bound to the filter, as described in Materials and Methods. The plot shows radioactivity bound to the filter (PhosphorImager units) as a function of reaction temperature.

[0025] Figure 12 shows schematic representations of Group II intron RTs and fusion proteins. Section 12(A) provides comparison of group II intron-encoded and retroviral RTs. Group II intron RTs exemplified by the LtrA protein encoded by the Ll.LtrB intron generally contains four major domains: RT, with conserved sequence blocks RT-1-7; X/thumb; DNA binding (D), and DNA endonuclease (En). The RT and thumb domains of group II intron RTs are homologous to those of retroviral RTs exemplified by HIV-1 RT, but are larger due to an N-terminal extension and insertions upstream (RT-0) and between the conserved RT sequence blocks (*e.g.*, RT-2a, 3a, 4a, and 7a and thumb domain insertion t_i in LtrA; Blocker *et al.*, RNA 11, 14-28, 2005). The positions of three α -helices characteristic of the thumb domains of retroviral RTs are shown for both LtrA and HIV-RT. The group II intron RTs

used in this work all contain the En domain, except for the GsI2 RT, which lacks the En domain. Section 12(B) shows group II intron RT fusion proteins. Group II intron RTs (IEPs) were expressed with fused N-terminal MalE or NusA solubility tags. Initial constructs contained the MalE solubility tag in expression vector pMalE-c2t fused to the N-terminus of the RT via a flexible linker with a TEV protease cleavage site (underlined). A variant of these initial constructs tested in Fig. 11 contained the pMalE-c2t linker with the TEV protease cleavage site deleted. Improved constructs used modified MalE or NusA tags fused to the N-terminus of the RT via a rigid linker containing 5 alanine residues (underlined). The modified MalE tag has charged amino acid residues changed to alanines (*italics*), and the modified NusA tag is missing the two C-terminal amino acid residues.

[0026] Figure 13 provides graphs showing the RT activity of derivatives of MalE-RF-Tel4c RT with different rigid fusion linker or solubility tag sequences. Panel 13(A) provides a bar graph showing RT activity at 60 °C. Reaction with MalE-RF-Tel4c RT (left bar) or variants containing different tag or linker sequences (right bars) were done as in Fig. 11 using 50 nM protein and 100 nM poly(rA)/oligo(dT)₄₂ and incubating for 90 sec. Values are the mean for three determinations with error bars indicating the standard deviation. Panel 13(B) provides a graph showing the temperature profile of RT activity for NusA-RF-Tel4c RT. RT activity was assayed as in Fig. 11 using 50 nM protein and 100 nM poly(rA)/oligo(dT)₄₂ and incubating for 2 min at the indicated temperature. The y-axis shows radioactivity bound to the filter (PhosphorImager units) for each protein (panel A) or for NusA-RF-Tel4c RT as a function of reaction temperature (panel B).

[0027] Figure 14 provides graphs and autoradiograms that provide a comparison of cDNA synthesis by MalE-RF-Tel4c, MalE-RF-GsI2, and SuperScript III RT activity at different temperatures. In panels (A-C), the substrate was a 531-nt RNA transcribed from AflIII-digested pBS KS(+) with an annealed 5'-labeled 37-nt primer, and in panels (D-F), the substrate was a 1.2-kb kanR RNA with an annealed 5'-labeled 44-nt DNA primer. Reactions were done by incubating 100 nM of annealed template/primer with 200 nM enzyme in 100 mM KCl, 20 mM Tris HCl pH 7.5, 10 mM MgCl₂ and 10 mM DTT for MalE-RF-Tel4c RT (panels A and D) and MalE-RF-GsI2 RT (panels B and E) and in the manufacturer's buffer for SuperScript III RT (panels C and F). Reactions were initiated by adding dNTPs to a final concentration of 1.25 mM, incubated for 30 min at the indicated temperature, and terminated by adding 0.1% SDS/250 mM EDTA (final concentrations) followed by phenol-CIA

extraction. The products were analyzed by electrophoresis in a denaturing 6% polyacrylamide gel, which was dried and quantified with a PhosphorImager. In each panel, the top and bottom autoradiograms show portions of the gel containing the full-length product (arrow) and unextended or partially extended primer, respectively, and the bar graphs show the percentage of primer that was extended to full-length cDNA based on PhosphorImager quantitation. “?” indicates unidentified bands not used in quantitation of full-length product. A 5'-labeled 10-bp ladder (Invitrogen™) was used as size markers. Schematics of two template primer substrates are shown at the bottom of the figure.

[0028] Figure 15 is a listing of the nucleotide sequence of the 1.2-kb kanR RNA template (SEQ ID NO: 21).

[0029] Figure 16 provides semi-log plots obtained from qRT-PCR to compare amounts of cDNA synthesis at different temperatures by MalE-RF-Tel4c RT and SuperScript III RT. cDNA was synthesized with MalE-RF-Tel4c RT or SuperScript III RT (SSIII RT) using the 1.2-kb kanR RNA with annealed primer P078 ($T_m = 80\text{ }^{\circ}\text{C}$) and detected with primer/probe sets at nt 188 – 257 and nt 562 – 634 (the data for detection with primer set nt 188 – 257 are shown in the figure; the data obtained with the primer set nt 562 – 634 are shown in Fig. 17). The qPCR amplification curves show a semi-log plot of fluorescence (ΔRN) versus cycle number. For each sample, duplicate wells were analyzed and are depicted in each amplification plot. The cycle threshold (C_T) values (the cycle at which the fluorescence crosses the threshold 0.4) for each cDNA synthesis reaction by MalE-RF-Tel4c or SuperScript III RT are indicated below the curves. Lower C_T values indicate a larger number of cDNAs synthesized.

[0030] Figure 17 provides semi-log plots obtained from qRT-PCR to compare processivity of cDNA synthesis by MalE-RF-Tel4c RT and SuperScript III RT. cDNA was synthesized with MalE-RF-Tel4c or SuperScript III RT using the 1.2-kb kanR RNA with annealed primer P078 ($T_m = 80\text{ }^{\circ}\text{C}$) and detected with primer/probe sets at nt 188 – 257 and nt 562 – 634. cDNA samples were obtained at $60\text{ }^{\circ}\text{C}$ (A, B) and $65\text{ }^{\circ}\text{C}$ (C, D). For each sample, triplicates were analyzed and are depicted in each amplification plot. Average copy numbers are derived from a standard curve of quantitated and diluted pET9 plasmid. Detection of similar numbers of cDNA copies with the two primer sets, as seen for MalE-RF-Tel4c RT, shows that most cDNAs extend to near the end of the RNA template, indicative of high processivity. A lower

number of cDNA copies detected with the primer set near the 5' end (nt 188-257) compared to the primer set closer to the 3' end (nt 562-634), as seen for SuperScript III RT, indicates that the RT falls off or is in some other way impeded from reaching the 5' end of the RNA template.

[0031] Figure 18 is a listing of the amino acid sequence of the NusA solubility-enhancing protein (SEQ ID NO: 38).

DETAILED DESCRIPTION OF THE INVENTION

[0032] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The terminology used in the description of the invention herein is for describing particular embodiments only and is not intended to be limiting of the invention. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety.

Definitions

[0033] As used in the description of the invention and the appended claims, the singular forms "a," "an," and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise. In addition, the recitations of numerical ranges by endpoints include all numbers subsumed within that range (e.g., 1 to 5 includes 1, 1.5, 2, 2.75, 3, 3.80, 4, 5, etc.).

[0034] As used herein, "polypeptide" refers to a polymer of amino acids and does not imply a specific length of a polymer of amino acids. Thus, for example, the terms peptide, oligopeptide, protein, antibody, and enzyme are included within the definition of polypeptide. This term also includes polypeptides with post-expression modification, such as glycosylation (e.g., the addition of a saccharide), acetylation, phosphorylation, and the like.

[0035] An "isolated" polypeptide or polynucleotide, as used herein, means a polypeptide or polynucleotide that has been either removed from its natural environment, produced using recombinant techniques, or chemically or enzymatically synthesized. Preferably, a

polypeptide or polynucleotide of this invention is purified, *i.e.*, essentially free from any other polypeptide or polynucleotide and associated cellular products or other impurities.

[0036] "Amino acid" is used herein to refer to a chemical compound with the general formula: $\text{NH}_2\text{--CRH--COOH}$, where R, the side chain, is H or an organic group. Where R is organic, R can vary and is either polar or nonpolar (*i.e.*, hydrophobic). The following abbreviations are used throughout the application: A = Ala = Alanine, T = Thr = Threonine, V = Val = Valine, C = Cys = Cysteine, L = Leu = Leucine, Y = Tyr = Tyrosine, I = Ile = Isoleucine, N = Asn = Asparagine, P = Pro = Proline, Q = Gln = Glutamine, F = Phe = Phenylalanine, D = Asp = Aspartic Acid, W = Trp = Tryptophan, E = Glu = Glutamic Acid, M = Met = Methionine, K = Lys = Lysine, G = Gly = Glycine, R = Arg = Arginine, S = Ser = Serine, H = His = Histidine. Unless otherwise indicated, the term "amino acid" as used herein also includes amino acid derivatives that nonetheless retain the general formula.

[0037] A nucleotide consists of a phosphate group linked by a phosphoester bond to a pentose (ribose in RNA, and deoxyribose in DNA) that is linked in turn to an organic base. The monomeric units of a nucleic acid are nucleotides. Naturally occurring DNA and RNA each contain four different nucleotides: nucleotides having adenine, guanine, cytosine and thymine bases are found in naturally occurring DNA, and nucleotides having adenine, guanine, cytosine and uracil bases found in naturally occurring RNA. The bases adenine, guanine, cytosine, thymine, and uracil often are abbreviated A, G, C, T and U, respectively.

[0038] Nucleotides include free mono-, di- and triphosphate forms (*i.e.*, where the phosphate group has one, two or three phosphate moieties, respectively). Thus, nucleotides include ribonucleoside triphosphates (e.g., ATP, UTP, CTG and GTP) and deoxyribonucleoside triphosphates (e.g., dATP, dCTP, dITP, dGTP and dTTP), and derivatives thereof. Nucleotides also include dideoxyribonucleoside triphosphates (ddNTPs, including ddATP, ddCTP, ddGTP, ddITP and ddTTP), and derivatives thereof.

[0039] "Substantially similar" means that a given nucleic acid or amino acid sequence shares at least 85%, more preferably at least 90%, and even more preferably at least 95% identity with a reference sequence. Furthermore, only sequences describing or encoding proteins in which only conservative substitutions are made in the conserved regions are substantially similar overall. Preferable, substantially similar sequences also retain the distinctive activity of the polypeptide. Substitutions typically seen as conservative substitutions are the

replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

[0040] A "promoter," as used herein, refers to a sequence in DNA that mediates the initiation of transcription by an RNA polymerase. Transcriptional promoters may comprise one or more of a number of different sequence elements as follows: 1) sequence elements present at the site of transcription initiation; 2) sequence elements present upstream of the transcription initiation site and; 3) sequence elements downstream of the transcription initiation site. The individual sequence elements function as sites on the DNA, where RNA polymerases and transcription factors that facilitate positioning of RNA polymerases on the DNA bind.

[0041] As used herein, the term "polymerase chain reaction" ("PCR") refers to a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. See for example Bartlett *et al.*, Methods Mol. Biol. 226:3-6 (2003), which provides an overview of PCR and its development. This process for amplifying the target sequence typically consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing and polymerase extension can be repeated many times to obtain a high concentration of an amplified segment of the desired target sequence. Unless otherwise noted, PCR, as used herein, also includes variants of PCR such as allele-specific PCR, asymmetric PCR, hot-start PCR, ligation-mediated PCR, multiplex-PCR, reverse transcription PCR, or any of the other PCR variants known to those skilled in the art.

[0042] As used in this specification, whether in a transitional phrase or in the body of the claim, the terms "comprise(s)" and "comprising" are to be interpreted as having an open-ended meaning. That is, the terms are to be interpreted synonymously with the phrases "having at least" or "including at least". When used in the context of a process, the term

"comprising" means that the process includes at least the recited steps, but may include additional steps. When used in the context of a compound or composition, the term "comprising" means that the compound or composition includes at least the recited features or components, but may also include additional features or components.

[0043] A "fusion protein," as used herein, refers to a protein having at least two heterologous polypeptides covalently linked in which one polypeptide comes from one protein sequence or domain and the other polypeptide comes from a second protein sequence or domain.

Stabilized Reverse Transcriptase Fusion Protein

[0044] The invention provides a stabilized reverse transcriptase fusion protein that includes a thermostable reverse transcriptase connected to a stabilizer protein. In many embodiments, the thermostable reverse transcriptase is connected to the stabilizer protein via a linker peptide. However, the thermostable reverse transcriptase and the stabilizer protein can also be directly fused to one another. The polypeptides that comprise the fusion protein are preferably linked N-terminus to C-terminus. However, the reverse transcriptase and the stabilizer protein can be connected together in either order. For example, the two peptide sequences can be connected from the C-terminus to N-terminus or N-terminus to the C-terminus. In some embodiments, a linker peptide is included between the connecting C-terminus and N-terminus of the reverse transcriptase and stabilizer protein.

[0045] Attaching a stabilizer protein to the thermostable reverse transcriptase can provide one or more advantages. A stabilized reverse transcriptase fusion protein can have one or more of the following advantages: (a) increased stability at elevated temperatures; (b) higher processivity, (c) increased solubility, and/or (d) higher fidelity. In some embodiments, a reverse transcriptase of the invention may have a plurality of the properties listed above. For example, a stabilized reverse transcriptase fusion protein may have increased thermostability and increased fidelity. The advantages may sometimes derive from one another. For example, by providing increased solubility, the stabilized reverse transcriptase fusion protein can provide a product able to provide increased fidelity of transcription as a result of solubilizing a previously insoluble high fidelity thermostable reverse transcriptase. The use of a stabilizer protein in the fusion protein can also provide other advantages such as increased protein expression and improved protein folding. Inclusion of a linker peptide

between the stabilizer protein and the thermostable reverse transcriptase can further enhance these advantages.

[0046] The stabilized reverse transcriptase fusion protein includes a thermostable reverse transcriptase and a stabilizer protein, as described herein. The stabilized reverse transcriptase fusion protein can also include a linker peptide. For example, the stabilized reverse transcriptase fusion protein can have an amino acid sequence as set forth in SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, or SEQ ID NO: 10, shown in Figures 1-5, respectively. Alternately, the stabilized reverse transcriptase fusion protein can have an amino acid sequence that is substantially similar to one or more of the sequences as set forth in SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, or SEQ ID NO: 10. A stabilized reverse transcriptase fusion protein amino acid sequence that is "substantially similar" to the fusion proteins provided by sequences 6-10 will share at least 85% identity, more preferably 90% identity and even more preferably 95% identity, and will include only conservative amino acid substitutions in conserved regions.

Thermostable Reverse Transcriptases

[0047] The present invention provides a reverse transcriptase fusion protein that includes a thermostable reverse transcriptase. The term "reverse transcriptases" (*i.e.*, RNA-directed DNA polymerases) refers to a group of enzymes having reverse transcriptase activity (*i.e.*, that catalyze synthesis of DNA from an RNA template). In general, such enzymes include, but are not limited to, retroviral reverse transcriptase, retrotransposon reverse transcriptase, and bacterial reverse transcriptases such as group II intron-derived reverse transcriptase, and mutants, variants or derivatives thereof. Examples of bacterial reverse transcriptase include *Lactococcus lactis* reverse transcriptase, *Thermosynechococcus elongatus* reverse transcriptase, or *Geobacillus stearothermophilus* reverse transcriptase. Further bacterial reverse transcriptases are described by Simon *et al.*, Nucleic Acids Research, 36, p. 7219-29 (2008), and Kojima and Kanehisa, Molecular Biology and Evolution, 25, p. 1395-04 (2008) which describe many classes of reverse transcriptases (*i.e.*, retrons, group II introns, and diversity-generating retroelements among others). Reverse transcriptase has been used primarily to transcribe RNA into cDNA, which can then be cloned into a vector for further manipulation or used in various amplification methods such as polymerase chain reaction, nucleic acid sequence-based amplification (NASBA), transcription mediated amplification

(TMA), self-sustained sequence replication (3SR), diverse primer extension reactions, 5'RACE, detection of chemical modifications or other techniques that require synthesis of DNA using an RNA template.

[0048] The term "thermostable" refers to the ability of an enzyme or protein (*e.g.*, reverse transcriptase) to be resistant to inactivation by heat. Typically such enzymes are obtained from a thermophilic organism (*i.e.*, a thermophile) that has evolved to grow in a high temperature environment. Thermophiles, as used herein, are organisms with an optimum growth temperature of 45 °C or more, and a typical maximum growth temperature of 70 °C or more. In general, a thermostable enzyme is more resistant to heat inactivation than a typical enzyme, such as one from a mesophilic organism. Thus, the nucleic acid synthesis activity of a thermostable reverse transcriptase may be decreased by heat treatment to some extent, but not as much as would occur for a reverse transcriptase from a mesophilic organism. "Thermostable" also refers to an enzyme which is active at temperatures greater than 38 °C, preferably between about 38-100 °C, and more preferably between about 40-81 °C. A particularly preferred temperature range is from about 45 °C to about 65 °C.

[0049] In some embodiments, a thermostable reverse transcriptase retains at least 50% (*e.g.*, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95%) of its nucleic acid synthetic activity after being heated in a nucleic acid synthesis mixture at 90 °C for 30 seconds. In contrast, typical reverse transcriptases will not work at elevated temperatures, and lose most of their nucleic acid synthetic activity after such heat treatment. Thermostable reverse transcriptases typically also have a higher optimum nucleic acid polymerization temperature.

[0050] Some reverse transcriptases are thermostable and therefore remain substantially active at temperatures commonly used in PCR-based nucleic acid synthesis. This provides the advantage of being able to carry out both reverse transcription and DNA amplification in a single reaction environment. Such temperatures vary depending upon reaction parameters, including pH, template and primer nucleotide composition, primer length, and salt concentration. Thermostable reverse transcriptases include *Thermosynechococcus elongatus* (Te) RT, *Geobacillus stearothermophilus* (Gs) RT, modified forms of these RTs, and engineered variants of Avian myoblastosis virus (AMV) RT, Moloney murine leukemia virus (M-MLV) RT, and Human immunodeficiency virus (HIV) RT. A reverse transcriptase

obtained from an organism living in an elevated temperature environment (*i.e.*, greater than 37 °C) can be expected to be stable at the living temperature of the organism, and to a reasonable degree above.

[0051] A class of reverse transcriptases that is particularly suitable for use in stabilized reverse transcriptase fusion proteins are group II intron-derived reverse transcriptases. A wide variety of group II intron-derived reverse transcriptases are known. See for example the Zimmerly Lab Website for Mobile Group II Introns that describes 105 full length group II intron-derived reverse transcriptases. The use of this website is described by Dai *et al.*, Nucleic Acids Research, 31, p. 424-26 (2003).

[0052] In certain embodiments the thermostable reverse transcriptase is one that was encoded by a group II intron. Group II intron RTs typically consist of four conserved domains: RT, which contains seven conserved sequence blocks (RT1-7) characteristic of the fingers and palm regions of retroviral RTs; X, a region required for RNA splicing activity corresponding at least in part to the thumb domain of retroviral RTs; D, a DNA-binding domain involved in DNA target site recognition; and En, a DNA endonuclease domain that cleaves the DNA target site to generate the primer for reverse transcription (Fig. 12A; Blocker *et al.*, RNA 11, 14-28, 2005). The En domain is missing in some group II intron RTs, which instead use nascent strands at DNA replication forks to prime reverse transcription (Zhong *et al.*, EMBO J. 22, 4555-4565, 2003). The RT and X/thumb domains of group II intron RTs are larger than those of retroviral RTs due to an N-terminal extension, an additional N-terminal conserved sequence block (RT-0), and insertions between the conserved sequence blocks in the RT and X/thumb domain, some of which are shared with non-LTR-retrotransposon RTs. It has been suggested that the larger-sized RT and thumb domains of group II intron and related RTs enable tighter binding of template RNAs leading to higher processivity and fidelity during reverse transcription. Unlike retroviral RTs, group II intron RTs lack an RNase H domain and typically have very low DNA-dependent DNA polymerase activity (Smith *et al.*, Genes and Development 19, 2477-2487, 2005).

[0053] Group II introns encode a class of RNAs known for their self-splicing reaction. Under certain *in vitro* conditions, group II intron-encoded RNAs can excise themselves from precursor mRNAs and ligate together their flanking exons, without the aid of a protein. The splicing reaction mechanism is similar to the splicing of nuclear pre-mRNA introns. A

number of group II introns also encode reverse transcriptase (RT) open reading frames (ORF) and are active mobile elements. The ORF is typically found in domain DIV of the group II intron encoded RNA. The group II intron RT assists RNA splicing by stabilizing the catalytically active RNA structure and then remains bound to the excised intron RNA in a ribonucleoprotein (RNP) that promotes intron mobility by a process termed "retrohoming." Retrohoming occurs by a mechanism in which the excised intron RNA in the RNPs inserts directly into a DNA target site and is reverse transcribed by the RT. During retrohoming, in which the group II intron facilitates targeting of the intron to appropriate DNA sequences, the group II intron RT must produce an accurate cDNA copy of the intron RNA, which is typically 2-2.5 kb long and folds into highly stable and compact secondary and tertiary structures. Thus, group II intron RTs must have high processivity and fidelity in order to carry out their biological function. Group II intron-derived RTs also lack RNase H activity, which can be beneficial because RNase H specifically degrades the RNA of RNA:DNA hybrids, which allows any RNA to be copied only once and can lead to reduced yields of full length cDNA.

[0054] Based on the group II intron-derived reverse transcriptases so far evaluated, these RTs typically exhibit relatively high fidelity and high processivity. The fidelity of reverse transcription refers to the reliability of nucleotide incorporation during reverse transcription of RNA to DNA, with higher fidelity describing nucleotide copying with a low number of errors (*e.g.*, misincorporations). Higher specificity can be provided by using longer and more specific primers, which requires the ability to carry out reverse transcription at higher temperatures. For example, a group II intron reverse transcriptase can provide reverse transcription with an error frequency of 2.0×10^{-5} or less, wherein the error frequency represents the proportion of nucleotide copying errors that occur relative to the number of nucleotide copying events that occur without error. Other examples of high fidelity transcription include error frequencies of 1×10^{-4} , 7.5×10^{-5} , 5×10^{-5} , 2.5×10^{-5} , 1×10^{-5} , and 5×10^{-6} . For further description of the high fidelity of group II intron-derived RTs, see Conlan *et al.*, Nucleic Acids Research, 33, p. 5262-70 (2005).

[0055] Examples of suitable group II-derived intron reverse transcriptases include the reverse transcriptases set forth in SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, and SEQ ID NO: 5, which are obtained from *Thermosynechococcus elongatus* (TeI4c, f, and h*) and *Geobacillus stearothermophilus* (GsI1 and GsI2). These sequences are shown in Figures

1-5. The invention also encompasses group II intron derived reverse transcriptases that are substantially similar to those set forth in SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, and SEQ ID NO: 5. A reverse transcriptase that is "substantially similar" to the reverse transcriptases provided by sequences 1-5 will share at least 85% identity, more preferably 90% identity and even more preferably 95% identity, and will include only conservative amino acid substitutions in conserved regions. The thermostability of a number of group II intron-derived RTs is shown in FIG. 11, which demonstrates that stabilized reverse transcriptase fusion proteins including the reverse transcriptases as set forth in SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, and SEQ ID NO: 5 have higher thermostability than mesophilic Ll.LtrB reverse transcriptase, whether or not it is part of a fusion protein, when evaluated as shown in FIG. 11. The mesophilic Ll.LtrB showed a temperature optimum of about 35 °C either alone or as part of a fusion protein.

[0056] As noted herein, modified forms of thermostable group II intron-derived RTs can also be used. For example, SEQ ID NO: 3, the TeI4h* RT, does not represent a native form of reverse transcriptase, but rather is a derivative in which the active site was modified from the amino acid sequence YAGD to the amino acid sequence YADD, to more closely resemble the active site of other active group II intron-derived RTs.

[0057] The amount by which a given amino acid sequence is "substantially similar" to a reference sequence can be determined for example, by comparing sequence information using sequence analysis software such as the Blastp program, version 2.2.10, of the BLAST 2 search algorithm, as described by Tatusova *et al.* (FEMS Microbiology Letters, 174, p. 247-50 (1999)), and available on the world wide web at the National Center for Biotechnology Information website, under BLAST in the Molecular Database section. Preferably, the default values for all BLAST 2 search parameters are used, including matrix = BLOSUM62; open gap penalty = 11, extension gap penalty = 1, gap x_dropoff = 50, expect = 10, wordsize = 3, and optionally, filter on. In the comparison of two amino acid sequences using the BLAST search algorithm, structural similarity is referred to as "similarity" and identity is referred to as "identity."

[0058] Amino acid identity is defined in the context of a comparison between a candidate polypeptides and a reference amino acid sequence, and is determined by aligning the residues of the two amino acid sequences (*i.e.*, a candidate amino acid sequence and the reference

amino acid sequence) to optimize the number of identical amino acids along the lengths of their sequences; gaps in either or both sequences are permitted in making the alignment in order to optimize the number of identical amino acids, although the amino acids in each sequence must nonetheless remain in their proper order.

[0059] Information is available to support a structure-function correlation for group II intron-derived reverse transcriptases. See for example Simon *et al.*, Nucleic Acids Research, 36, p. 7219-29 (2008), which classifies and aligns the RT domains of bacterial reverse transcriptases, and Xiong *et al.*, EMBO J., 9, p. 3353-62 (1990), which provides an alignment of 82 RT sequences showing seven conserved domains and 42 conserved positions. See also Blocker *et al.*, RNA, 11, p. 14-28 (2005), which provides a three-dimensional model of *Lactococcus lactis* Ll.LtrB intron RT (the LtrA protein), describes the proteolytic cleavage sites and conserved regions, and provides a sequence alignment analysis of LtrA relative to HIV-1 RT. Accordingly, a variety of stabilized reverse transcriptase fusion proteins that are substantially similar to those set forth in SEQ ID NO. 6-10 can readily be obtained by modification of amino acids outside of the conserved regions, and only conservative modification of amino acids within the known conserved regions.

[0060] In one embodiment, the present invention provides a stabilized reverse transcriptase fusion protein having a reverse transcriptase activity that has a half-life of greater than that of the corresponding unbound reverse transcriptase at an elevated temperature, *i.e.*, greater than 37 °C. In some embodiments, the half-life of a reverse transcriptase of the present invention may be 5 minutes or greater and preferably 10 minutes or greater at 50 °C. In some embodiments, the reverse transcriptases of the invention may have a half-life (*e.g.*, at 50 °C) equal to or greater than about 25 minutes, preferably equal to or greater than about 50 minutes, more preferably equal to or greater than about 100 minutes, and most preferably, equal to or greater than about 200 minutes.

Stabilizer Proteins

[0061] The stabilized reverse transcriptase fusion protein of the present invention also includes a stabilizer protein. A stabilizer protein, as defined herein, is a protein forming part of the fusion protein that functions to increase the overall stability of the fusion protein. Stability includes the ability of the protein to retain its conformation and activity. In addition, the stabilizer protein preferably enhances the solubility of the fusion protein, as further

described herein with regard to solubility-enhancing proteins. This can be particularly helpful with regard to group II intron RTs, which have been found to be poorly expressed and insoluble in the absence of the intron RNA to which they are ordinarily tightly bound in RNPs. (Vellore *et al.* Appl. Environ. Microbiol. 70, 7140-7147, 2004; Ng *et al.*, Gene 393, 137-144, 2007) Effective stabilizer proteins include those that include an independent folding domain and/or do not fold into long-lived misfolded intermediates that can influence the propensity of proteins to aggregate. Proteins that will provide an independent folding domain are described by Janin *et al.*, Progress in Biophysics and Molecular Biology, 42, p. 21-78 (1983), and proteins that do not fold into long-lived misfolded intermediates are described by Idicula *et al.*, Protein Science, 14, p. 582-592 (2005). For example, the stabilizer protein can be a protein that includes 50 or more amino acids. In other embodiments, the stabilizer protein can be a larger protein including 100 or more amino acids. As exemplified by the maltose binding protein and NusA proteins provided herein, the stabilizer proteins can also have a size from about 250 amino acids to about 400 amino acids. The stabilizer protein can also be a thermostable protein.

[0062] The stabilizer protein can also be or include an affinity protein. The term affinity protein, as used herein, refers to a protein for which there is a readily available ligand that exhibits a high binding constant (*i.e.*, "affinity") for the protein. Affinity proteins are often used in the role of an affinity tag. Affinity proteins, as is known to those skilled in the art, can be provided in fusion proteins to facilitate the purification of the protein connected or fused to the affinity protein by techniques such as affinity purification, in which a tag binds to a ligand within an affinity column. Suitable affinity proteins are known in the art. See for example Waugh, D., Trends in Biotechnology, 23, p. 316-320 (2005), which describes a number of suitable affinity proteins, including glutathione S-transferase, maltose-binding protein, FLAG-tag peptide, biotin acceptor peptide, streptavidin-binding peptide, and calmodulin-binding peptide. For the preparation and use of fusion proteins that include an affinity protein, see for example U.S. Patent Nos. 5,643,758, 5,654,176, and 7,001,745.

[0063] The stabilizer protein can also be a solubility-enhancing protein. Recombinantly-expressed fusion proteins can exhibit low solubility in their host cells and/or in subsequent method applications, which can be ameliorated through inclusion of a solubility-enhancing protein in the fusion protein that substantially increases the solubility of the fusion protein in aqueous environments. Some solubility-enhancing proteins used are also affinity proteins,

and can therefore be described as solubility-enhancing affinity proteins. Examples of solubility-enhancing proteins include sugar binding proteins such as arabinose binding protein, chitin binding protein, cellulose binding protein, and maltose binding protein. Other examples of solubility-enhancing proteins include the NusA and Dsb solubility tags provided by Novagen[®], and the solubility enhancing tag (SET) provided by Invitrogen[™]. Harrison has demonstrated the very high solubility provided by the NusA solubility tag, while the solubility enhancement of Dsb is described by Collins-Racie. See Harrison, R.G., *in* *Novations*, 11, p. 4-7 (2000), and Collins-Racie *et al.*, *Biotechnology*, 13, p. 982-87 (1995).

[0064] In some embodiments, stabilizer proteins such as solubility-enhancing proteins or affinity proteins can be modified to improve their performance. Modification can include providing one or more substitutions, additions or deletions of amino acids within the protein sequence of the stabilizer protein as compared to the normal, wild-type sequence of the protein. For example, a stabilizer protein such as an affinity protein or a solubility-enhancing protein can be modified by replacing the charged amino acids with uncharged amino acids in certain regions of the protein. Charged amino acids include amino acids with positively or negatively charged side chains. Examples of amino acids with positively charged side chains include arginine, histidine, lysine, and the like. Examples of amino acids with negatively charged side chains include aspartic acid and glutamic acid, and the like. Uncharged amino acids include, but are not limited to, alanine, serine, threonine, glutamine, valine, leucine, isoleucine, phenylalanine, and tyrosine. For example, a maltose binding protein can be modified by replacing one or more of the charged amino acids with alanine.

[0065] Examples of suitable affinity proteins include the maltose binding protein amino acid sequence set forth in SEQ ID NO: 11, shown in Figures 1-5, and sequences substantially similar to SEQ ID NO: 11. Note that while modification of the affinity protein is not necessary, the maltose binding protein set forth in SEQ ID NO: 11 was modified to replace three charged amino acids with alanine near the C-terminus. Another suitable protein, in this case a solubilizing protein, is the N-utilization substance A (NusA) protein, which has the amino acid sequence set forth in SEQ ID NO: 38, shown in Figure 18. In additional embodiments of the invention, fusion proteins described herein that include the maltose binding proteins can have the maltose binding protein replaced with N-utilization substance A proteins.

Linker Peptides

[0066] In some embodiments, the stabilized reverse transcriptase fusion protein also includes a linker peptide positioned between the stabilizer protein and the thermostable reverse transcriptase. Preferably, the linker peptide is a non-cleavable linker peptide. By “positioned between,” it is meant that the linker peptide is connected by a chemical linkage (*e.g.*, an amide linkage) to the N or C terminal of each of the stabilizer protein and the reverse transcriptase, as described in regard to fusion proteins herein. For example, the linker peptide can be connected through an amide linkage to the C terminal region of the stabilizer protein and the N terminal region of the thermostable reverse transcriptase. By non-cleavable, it is meant that the linker peptide is not readily susceptible to cleavage by a protease.

[0067] In additional embodiments, the linker peptide is a rigid linker peptide; *i.e.*, a relatively non-flexible peptide linker. Rigid linker peptides are not required to completely lack flexibility, but rather are significantly less flexible than flexible linker peptides such as glycine-rich peptide linkers. Rigid linker peptides, as a result of their relative lack of flexibility, decrease the movement of the two protein domains attached together by the rigid linker peptide, which in the present case are the stabilizer protein and the thermostable reverse transcriptase. Linker peptides that provide ordered chains such as alpha helical structure can provide rigid linker peptides. For example, Arginine, Leucine, Glutamate, Glutamine, and Methionine all show a relatively high propensity for helical linker formation. However, a non-helical linker including many proline residues can exhibit significant rigidity as well. Examples of rigid linkers include polylysine and poly-DL-alaninepolylysine. Further description of rigid peptide linkers is provided by Wriggers *et al.*, Biopolymers, 80, p. 736-46 (2005). In addition, rigid linker peptides are described at the linker database described by George *et al.*, Protein Engineering, 15, p. 871-79 (2003). Preferably, the rigid linker peptide is also a non-cleavable linker peptide; *i.e.*, a non-cleavable, rigid linker peptide.

[0068] Relatively short polypeptides are preferred for use as linker peptides. For example, linker peptides can include from 1 to 20 amino acids. Linker peptides can also include from 1 to 15, from 1 to 10, from 1 to 5, or from 3 to 5 amino acids. Examples of specific sequences that can be used as linker peptides include dipeptides, tripeptides, tetrapeptides, and pentapeptides formed of alanine amino acids. One suitable rigid linker peptide is AAAAA (SEQ ID NO: 12), while another suitable rigid linker peptide is AAAEF (SEQ ID

NO: 18). Use of a linker peptide (*e.g.*, a rigid linker peptide) in a fusion protein can provide one or more advantages. For example, while not intending to be bound by theory, it is believed that use of a rigid linker peptide can stabilize the fusion protein by decreasing the amount of movement of the two halves of the fusion protein relative to one another. While very short (*i.e.*, 1 or 2 amino acid) linkers can be used, it is preferable to use linkers that include from 3 to 5 amino acids.

[0069] The linker peptide can be either cleavable or non-cleavable by a protease. Affinity proteins are often associated to another protein in a fusion protein using a cleavable peptide so that the affinity protein can be removed. However, in the present invention the stabilizer protein (*e.g.*, an affinity protein) remains bound to the reverse transcriptase, for the reasons described herein. Accordingly, it is generally preferable that the linker peptide be non-cleavable. However, cleavable linkers can be used in some embodiments. For example, cleavable linkers, including rigid cleavable linker peptides, that are susceptible to protease cleavage can be used if it is desirable to remove the stabilizer protein during a subsequent step and exposure to the cleaving protease is avoided during use of the fusion protein.

Use of Stabilized Reverse Transcriptase Fusion Proteins

[0070] The invention also provides a method for preparing a cDNA from an RNA (*e.g.*, mRNA, rRNA, tRNA, and miRNA), which is required for other methods such as the reverse transcription polymerase chain reaction (RT-PCR). As used herein, the term "RT-PCR" refers to the replication and amplification of RNA sequences. In this method, reverse transcription is coupled to PCR, *e.g.*, as described in U.S. Pat. No. 5,322,770. In RT-PCR, the RNA template is converted to cDNA due to the reverse transcriptase activity of an enzyme, and then amplified using the polymerizing activity of the same or a different enzyme.

[0071] In the practice of the invention, cDNA molecules may be produced by mixing one or more nucleic acid molecules (*e.g.*, RNA) obtained from cells, tissues, or organs using methods that are well known in the art, with the composition of the invention, under conditions favoring the reverse transcription of the nucleic acid molecule by the action of the enzymes of the compositions to form a cDNA molecule (single-stranded or double-stranded). Thus, the method of the invention comprises (a) mixing one or more nucleic acid templates (preferably one or more RNA or mRNA templates, such as a population of mRNA

molecules) with stabilized RT fusion protein of the invention and (b) incubating the mixture under conditions sufficient to permit cDNA synthesis of all or a portion of the one or more nucleic acid templates.

[0072] In one aspect, the method includes the steps of (a) adding a primer to an RNA molecule and (b) incubating the RNA molecule in the presence of one or more deoxy or dideoxyribonucleoside triphosphates and a stabilized reverse transcriptase fusion protein comprising a thermostable reverse transcriptase connected to a stabilizer protein under conditions sufficient to synthesize a cDNA molecule complementary to all or a portion of the RNA molecule. Adding the primer to an RNA molecule may include hybridizing the primer to the RNA molecule. In some embodiments, the stabilized reverse transcriptase fusion protein can also include a linker peptide connecting the stabilizer protein to the thermostable reverse transcriptase. Preferably, the reverse transcription is performed within a temperature range where the RNA includes a substantially decreased amount of obstructing stable secondary or tertiary structure. This can be a temperature from about 45 °C to about 81 °C, with a more preferred temperature range being from about 45 °C to about 65 °C. This can also be described as a temperature range in which the RNA does not form a significant amount of stable secondary or tertiary structure. Due to the high fidelity and other advantages of group II intron-derived RTs, their use may be preferred. For example, the stabilized reverse transcriptase fusion protein can include a group II intron-derived reverse transcriptase with an amino acid sequence identity that is substantially similar to a sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 5, a non-cleavable linker consisting of 1 to 20 amino acids, and the stabilizer protein comprises a solubility-enhancing or affinity protein. The stabilized reverse transcriptase fusion protein can also include a linker peptide between the stabilizer peptide and the reverse transcriptase, which can have a length from 1-20 amino acids, can be a non-cleavable linker, or can be rigid linker. Embodiments of the method can perform reverse transcription with an error frequency of 2.0×10^{-5} or less. Particularly at a temperature from about 45 °C to about 65 °C.

[0073] The stabilized reverse transcriptase fusion proteins can also be used in other applications. For example, stabilized RT fusion proteins can be used for the cloning of differentially expressed 5' ends of mRNAs; a process referred to as rapid amplification of cDNA ends (RACE) and variations thereof such as RNA ligase mediated RACE (RLM-

RACE). Stabilized RT fusion proteins can also be used for the mapping of chemical footprints in RNA, differential display RT-PCR, which allows for the analysis of gene expression among cell populations, and *in-situ* PCR for medical diagnosis.

Preparation of Stabilized Reverse Transcriptase Fusion Proteins

[0074] An expression vector containing a stabilized reverse transcriptase fusion protein-encoding nucleic acid molecule may be used for high-level expression of stabilized reverse transcriptase fusion protein in a recombinant host cell. Expression vectors may include, but are not limited to, cloning vectors, modified cloning vectors, specifically designed plasmids or viruses. A variety of expression vectors may be used to express recombinant stabilized reverse transcriptase fusion sequences in appropriate cell types. For example, bacterial vectors, mammalian vectors, fungal vectors, and insect vectors may be used for expression in bacteria, mammalian cells, fungal cells, and insect cells, respectively.

[0075] Stabilized reverse transcriptase fusion proteins can be prepared by obtaining a nucleotide sequence capable of expressing a stabilized reverse transcriptase fusion protein and then expressing that nucleotide sequence in a host cell. The stabilized reverse transcriptase fusion proteins expressed by the host cell can then be purified using a variety of techniques known to those skilled in the art, depending in part on the nature of the host cell.

[0076] Nucleotide sequences capable of expressing stabilized reverse transcriptase fusion proteins of the invention can be prepared using a variety of methods known to those skilled in the art. For example, the nucleotide sequences can be prepared using recombinant plasmids in which various linkers, reverse transcriptases, and stabilizer proteins are combined, as described in Example 1 herein.

[0077] The present invention also relates to host cells transformed or transfected with vectors comprising a nucleic acid molecule capable of expressing a stabilized reverse transcriptase fusion protein. Recombinant host cells may be prokaryotic or eukaryotic, including but not limited to, bacteria such as *E. coli*, fungal cells such as yeast, mammalian cells including, but not limited to, cell lines of bovine, porcine, monkey and rodent origin; and insect cells including but not limited to *Drosophila* and silkworm derived cell lines. Such recombinant host cells can be cultured under suitable conditions to produce a stabilized reverse transcriptase fusion protein or a biologically equivalent form. As defined herein, the term

"host cell" is not intended to include a host cell in the body of a transgenic human being, human fetus, or human embryo.

[0078] As noted above, an expression vector containing DNA encoding a stabilized reverse transcriptase fusion protein may be used for expression of stabilized reverse transcriptase fusion protein in a recombinant host cell. Therefore, another aspect of this invention is a process for expressing a stabilized reverse transcriptase fusion protein in a recombinant host cell, comprising: (a) introducing a vector comprising a nucleic acid comprising a sequence of nucleotides that encodes a stabilized reverse transcriptase fusion protein into a suitable host cell, wherein the stabilized reverse transcriptase fusion protein comprises a thermostable reverse transcriptase connected to a stabilizer protein directly or via a linker and (b) culturing the host cell under conditions which allow expression of the stabilized reverse transcriptase fusion protein. The stabilized reverse transcription fusion protein can be varied to include any of the features described herein, such as the inclusion of a linker peptide connecting the thermostable reverse transcriptase and the stabilizer protein.

[0079] Following expression of a stabilized reverse transcriptase fusion protein in a host cell, the stabilized reverse transcriptase fusion protein may be recovered to provide purified stable reverse transcriptase fusion protein. Several protein purification procedures are available and suitable for use. For instance, see Example 2 provided herein. Recombinant protein may be purified from cell lysates and extracts by various combinations of, or individual application of salt fractionation, ion exchange chromatography, size exclusion chromatography, hydroxylapatite adsorption chromatography and hydrophobic interaction chromatography. The use of affinity tags in some embodiments of the invention can facilitate purification of the protein. For example, the stabilized reverse transcriptase fusion protein can be separated from other cellular proteins by use of an immunoaffinity column made with monoclonal or polyclonal antibodies specific for the reverse transcriptase or stabilizer protein portion of the fusion protein. Heating can be used to separate the stabilized reverse transcriptase fusion protein from host proteins, which are not stable at elevated temperatures and will therefore precipitate.

[0080] The nucleic acids capable of expressing a stabilized RT fusion protein may be assembled into an expression cassette which comprises sequences designed to provide for efficient expression of the fusion protein in a host cell. The cassette preferably contains a

stabilized reverse transcriptase fusion protein-encoding open reading frame, with related transcriptional and translations control sequences operatively linked to it, such as a promoter, and termination sequences. For example, the open reading frame can include a nucleic acid that encodes a polypeptide with an amino acid sequence identity that is substantially similar to a sequence selected from the group consisting of SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, or SEQ ID NO: 10, as shown in Figures 1-5, respectively. In a preferred embodiment, the promoter is a T7 or a tac promoter for expression in *E. coli*, although those skilled in the art will recognize that any of a number of other known promoters may be used. *E. coli* also has rho independent and dependent terminators and can use T7 polymerase for rapid DNA replication. In eukaryotic cells, inclusion of a polyadenylation site will be helpful for the correct processing of mRNA.

[0081] The open reading frame can also include polynucleotide sequences as set forth in SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, and SEQ ID NO: 17, as shown in Figures 6-10, respectively. Alternately, the open reading frame can include polynucleotide sequences that are substantially similar to those set forth in SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, and SEQ ID NO: 17. In this particular context, the term “substantially similar” refers to variants in the nucleotide sequence in which codons that encode the same amino acid can be used interchangeably such that the nucleotide sequence will still result in the translation of an amino acid sequence corresponding to SEQ ID NO: 6-10. The stabilized reverse transcriptase fusion protein open reading frame polynucleotide preferably has at least about 80% identity, at least about 90% identity, at least about 95% identity, or at least about 98% identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, and SEQ ID NO: 17.

[0082] Nucleotide identity is defined in the context of a comparison between a candidate stabilized reverse transcriptase fusion protein open reading frame and a polynucleotide sequence selected from the group consisting of SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, and SEQ ID NO: 17, and is determined by aligning the residues of the two polynucleotides to optimize the number of identical nucleotides along the lengths of their sequences; gaps in either or both sequences are permitted in making the alignment in order to optimize the number of shared nucleotides, although the nucleotides in each sequence must nonetheless remain in their proper order. Preferably, two nucleotide

sequences are compared using the Blastn program of the BLAST 2 search algorithm, as described by Tatusova, *et al.* (FEMS Microbiology Letters, 174, p. 247-50 (1999)), and available on the world wide web at the National Center for Biotechnology Information website, under BLAST in the Molecular Database section. Preferably, the default values for all BLAST 2 search parameters are used, including reward for match=1, penalty for mismatch=-2, open gap penalty=5, extension gap penalty=2, gap x dropoff=50, expect=10, wordsize=11, and optionally, filter on. In the comparison of two nucleotide sequences using the BLAST search algorithm, nucleotide identity is referred to as "identities."

[0083] With regard to protein preparation from nucleotide sequences, it is noted that a "triplet" codon of four possible nucleotide bases can exist in over 60 variant forms. Because these codons provide the message for only 20 different amino acids (as well as transcription initiation and termination), some amino acids can be coded for by more than one codon, a phenomenon known as codon redundancy. Accordingly, the nucleotide sequences used to prepare the particular amino acid sequences of stabilized reverse transcriptase fusion proteins can vary considerably, depending on the particular codons used. For reasons not completely understood, alternative codons are not uniformly present in the endogenous DNA of differing types of cells, and there exists a natural hierarchy or "preference" for certain codons in certain types of cells. Accordingly, in some embodiments the choice of codons used to express a stabilized reverse transcriptase fusion protein may be optimized through use of particular codons to result in higher levels of expression.

[0084] In accordance with this invention, the stabilized reverse transcriptase fusion protein expression cassette is inserted into a vector. The vector is preferably a plasmid or adenoviral vector, although linear DNA linked to a promoter, or other vectors, such as adeno-associated virus or a modified vaccinia virus, retroviral or lentiviral vector may also be used. In particular, the use of *E. coli* plasmid vectors is preferred.

[0085] A detailed description of the work conducted by the inventors to develop and evaluate stabilized reverse transcriptase fusion proteins is provided below.

Expression and purification of group II intron RTs as MalE fusion proteins

[0086] The expression and solubility of poorly behaved proteins can sometimes be improved by fusion of highly soluble proteins, like maltose-binding protein (MalE) or N utilization

substance A (NusA) (Nallamsetty *et al.*, Protein Expression and Purification 45, 175-182, 2005). The MalE tag additionally permits facile purification of the protein via amylose-affinity chromatography. The inventors therefore tested whether group II intron RTs could be expressed and purified as MalE fusions. Initially, a MalE tag was fused to the N-terminus of the RT via a TEV protease-cleavable linker in the expression vector pMal-c2t (Fig. 12B). The MalE-RT fusion proteins for several of the *T. elongatus* group II intron RTs expressed well in *E. coli* and could be purified by a procedure that involves polyethyleneimine (PEI)-precipitation to remove nucleic acids, followed by amylose-affinity and heparin-Sepharose chromatography. Further, the uncleaved MalE-RT fusion proteins assayed soon after purification had high thermostable RT activity. However, the yields of these proteins were < 0.2 mg/l for the *Thermosynechococcus* proteins. Additionally, when the MalE tag was removed by cleavage with TEV protease, the RTs immediately formed an insoluble precipitate, while if the tag was left uncleaved, the MalE-RT fusion proteins progressively lost RT activity and were degraded within days, even when stored on ice or flash frozen in 50% glycerol. The latter findings were surprising because proteins that fold properly in the presence of a solubility tag tend to remain soluble after cleavage of the tag (Nallamsetty *et al.*, Protein Expression and Purification 45, 175-182, 2005). The group II intron RTs, which were active with but not without the attached MalE tag, appear to be an exception. The finding that the stabilizer protein must remain attached to the thermostable reverse transcriptase suggests that it plays an active role in keeping the thermostable reverse transcriptase soluble and active.

[0087] To overcome these difficulties, the inventors tested whether the group II intron RTs could be stabilized in active form by attaching the MalE tag to the protein via a non-cleavable rigid linker. Such MalE-rigid fusions typically have a linker region of 3 to 5 alanine residues combined with changes at the C-terminus of the MalE tag to replace charged amino acid residues with alanines (Smyth *et al.*, Genes and Development 19, 2477-2487, 2003). These rigid fusion linkers reduce conformational heterogeneity, enabling crystallization of proteins with attached linkers for structure determination (Smyth *et al.*, *ibid*). For the MalE-RF-RT fusions tested here, the MalE/linker region of pMal-c2t TVDEALKDAQTNS₃N₁₀LENLYFQGEF (SEQ ID NO: 19) was modified to TVDAALAAAQTAAAAA (SEQ ID NO: 20) and called a MalE-RF (rigid fusion) tag (Fig. 12B).

[0088] To rapidly assess whether the MalE-RF tag affects the activity of group II intron RTs, the inventors tested whether the MalE-RF-RTs could support retrohoming *in vivo*. For initial tests, the RTs chosen were the LtrA protein encoded by the *L. lactis* Ll.LtrB intron, and TeI4h* RT, an activated derivative of the RT encoded by the thermostable *T. elongatus* TeI4h intron. In retrohoming assays at 37 °C, the MalE-RF-LtrA protein supported retrohoming at an efficiency of 20% compared to 86% for native LtrA, while in retrohoming assays at 48 °C, the MalE-RF-TeI4h* protein supported retrohoming at an efficiency of 87% compared to 100% for the unfused TeI4h* protein; see Table 1. Thus remarkably both MalE-RF-RTs retain the ability to support retrohoming with high albeit somewhat reduced efficiencies despite the presence of the attached maltose-binding protein rigid linker sequence. These findings imply that the proteins retain substantial levels of all activities required for retrohoming, including RT, RNA splicing, and DNA endonuclease activity. This mobility assay provides a convenient screen for active group II intron RTs.

Table 1: Retrohoming efficiencies for different RTs

RT	Efficiency
TeI4h* (48 °C)	100%
MalE-RF-TeI4h* (48 °C)	87%
LtrA (37 °C)	86%
MalE-RF-LtrA (37 °C)	20%

[0089] Retrohoming assays were done in *E. coli* HMS174(DE3) as described previously for the Ll.LtrB intron (LtrA protein) (Guo *et al.* Science 289, 452-457, 2000, Karberg *et al.* Nature Biotech. 19, 1162-1167, 2001) and TeI4h*. The Cap^R intron-donor plasmids use a T7lac promoter to express a ΔORF intron (I-ΔORF) with short flanking 5' and 3' exons (E1 and E2, respectively) and a T7 promoter in DIV, followed by the RT ORF downstream of E2. The Amp^R recipient plasmids contain a target site for the intron (ligated E1-E2 sequences) cloned upstream of a promoterless *tet^R* gene. Intron expression was induced with IPTG (0.1 mM for LtrA and MalE-RF-LtrA and 0.5 mM for TeI4h* and MalE-RF-TeI4h*) for 1 h at the indicated temperature. Retrohoming of the intron carrying the T7 promoter into the target site

activates the expression of the *tet*^R gene, enabling selection for Tet^R + Amp^R colonies. Retrohoming efficiencies were calculated as the ratio of (Amp^R + Tet^R)/Amp^R colonies.

[0090] Encouraged by these findings, the inventors constructed plasmids in which several group II intron RTs were expressed with a MalE tag fused to the N-terminus of the protein via a rigid linker in the vector pMal-c2t. The RTs tested included several *T. elongatus* group II intron RTs, whose ability to support retrohoming had been tested previously using the above plasmid assay and two *G. stearothermophilus* group II intron RTs related to group II intron RTs that had previously been difficult to purify with high yield and activity (Vellore *et al.*, Appl. Environ. Microbiol. 70, 7140-7147, 2004; Ng *et al.*, Gene 393, 137-144, 2007). In some constructs, the inventors added an additional C-terminal His6-tag to enrich for full-length protein in the purification. The MalE-RF-RT fusion proteins were expressed in *E. coli* and purified by a procedure that involves PEI-precipitation of nucleic acids followed by amylose-affinity and heparin-Sepharose chromatography. An additional Ni column chromatography step was included for constructs with a C-terminal His6 tag. The proteins were dialyzed against the purification buffer with 50% glycerol, flash frozen, and stored at -80°C. The final protein preparations were > 95% pure with yields of 0.5-2.2 mg/ml and their RT activity was undiminished after storage for at least six months.

RT assays

[0091] To assess their thermostability, the inventors first assayed the RT activity of fusions MalE-RF-Tel4c, Tel4h*, and Tel4f from *Thermosynechococcus elongatus* and MalE-RF-GsI1 and GsI2 from *Geobacillus stearothermophilus* at temperatures between 25 and 77°C. These initial assays were done by using poly(rA)/oligo(dT)₄₂ as the template-primer substrate and quantifying polymerization of ³²P-dTTP into high molecular weight material. The relatively long 42-nt dT primer was used so that it would remain annealed to the poly(rA) template at higher temperatures (calculated T_m = 69 °C). The LtrA protein with and without an N-terminal MalE-RF tag was assayed in parallel as a mesophilic RT control (Fig. 11). Whereas the LtrA protein had a temperature optimum of ~35 °C with or without the MalE rigid fusion tag, the other five MalE-RF-RT's had higher temperature optima ranging from 45-61 °C. The two most active and thermostable RTs, MalE-RF-GsI2 and MalE-RF-Tel4c had temperature optima of 61 °C and retained substantial activity at 70 °C (where the assay may be limited by the stability of the primer-template base pairing). Of the two RTs, MalE-RF-Tel4c had the highest activity and was assayed at lower protein concentrations (50 nM)

and for shorter times (90 sec) than the other RTs (100 nM, 5 min) in order to remain within the linear range. Tests with the MaleE-RF-Tel4c protein showed that inclusion of maltose (10 μ M to 1 mM), which can affect the conformation of the MaleE tag, had little if any effect on RT activity.

Effect of changing the tag and linker on RT activity

[0092] To determine optimal properties of the tag and linker, the inventors constructed variants of the MaleE-RF-Tel4c RT. The MaleE-RT-Tel4c RT (left bar) and variant proteins (right bars) were purified and assayed for RT activity with poly(rA)/oligo(dT)₄₂ as described above (Fig. 13A). MaleE-RT-Tel4c has a modified MaleE tag (MaleE (mod)) with 3 charged amino acid residues changed to alanines and a linker of 5 alanine residues linked to the N-terminus of the RT. Variants in which the 5 alanine-residue linker was removed or shortened to 1 or 2 alanine residues had substantial but reduced RT activity, as did a variant in which the modified MaleE tag was replaced with wild-type MaleE (MaleE (WT)) (Fig 13A). A variant of Tel4c with the MaleE (WT) tag followed by the pMal-c2t linker deleted for the TEV protease cleavage site also had substantial but reduced RT activity (Fig. 13A). A variant in which the wild-type MaleE tag was attached to the C-terminus of the Tel4c RT did not express well in *E. coli*, presumably reflecting that the nascent Tel4c RT cannot fold properly without prior expression of the MaleE tag. Finally, a variant with an N-terminal rigid fusion to NusA (N utilization substance protein) instead of MaleE had substantial thermostable RT activity (Fig. 13A and B).

Temperature profile for cDNA synthesis

[0093] Fig. 14 shows assays of cDNA synthesis at different temperatures using *in vitro* transcribed RNA templates with DNA primers annealed to their 3' ends comparing two of the thermostable group II intron RTs (MaleE-RF-Tel4c and MaleE-RF-Gsl2) with a commercially available RT, SuperScript III (Invitrogen™), which has been reported to be active at 55 °C (Potter *et al.* Focus (Invitrogen Newsletter) 25.1, 19- 24, 2003). One template was a 531-nt *in vitro* transcript synthesized from AflIII-digested pBS KS(+) with a ³²P-labeled 37-nt DNA primer annealed (Fig. 14A-C) and the other was a 1.2-kb kanR RNA (SEQ ID NO: 21; shown in Figure 15) with a ³²P-labeled 44-nt DNA primer (Fig. 14D-E). The reaction was incubated for 30 min at the indicated temperature, and the products were analyzed by electrophoresis in a denaturing 6% polyacrylamide gel. In each panel, the top and bottom autoradiograms show

portions of the gel containing the full-length product and unextended or partially extended primers, respectively, and the bar graphs show the percentage of primer that was extended to full-length cDNA.

[0094] With the 531-nt RNA template, the MaleE-RF-Tel4c RT had a temperature optimum for full-length cDNA synthesis of 61-81 °C. The MaleE-RF-GsI2 RT synthesized full-length cDNA at temperatures between 37 and 69 °C, whereas SuperScript III RT had no activity at temperatures higher than 57 °C (Fig. 14A-C). With the 1.2-kb RNA template, the MaleE-RF-Tel4c and MaleE-RF-GsI2 RT had temperature optima of 61-81 °C and 61-69 °C, respectively, while SuperScript III RT again had no activity at temperatures higher than 57 °C (Fig. 14D-E).

Analysis of cDNA synthesis by qRT-PCR

[0095] In addition to gel analysis, the inventors used qRT-PCR to compare the amounts of cDNAs synthesized by the MaleE-RF-Tel4c and SuperScript III RTs using the 1.2-kb RNA template. The inventors first compared the amounts of full-length cDNA produced at temperatures between 50 and 75 °C (Fig. 16). The cDNAs for qPCR were synthesized in reactions containing 5×10^8 copies of kanR RNA as a template, 200 nM MaleE-RF-Tel4c or 200 U of SuperScript III RT for 30 min at six different temperatures. Reactions with SuperScript III were done according to the manufacturer's specifications. The reaction mix containing all components except for dNTPs was preincubated at the desired temperatures for 2 min and started by adding the dNTPs. After 30 min, the reactions were terminated by quickly freezing on dry ice. A 5- μ l portion of each cDNA synthesis was used in qPCR reactions containing TaqMan[®] Gene Expression mix and two forward, reverse, and dual-labeled primer probe mixes located at nt 188-257 and 562-634 of the kanamycin RNA. With the primer set closest to the 5' end of the RNA (nt 188-257), the cycle threshold (C_T) values were significantly lower for the MaleE-RF-Tel4c RT than for SuperScript III RT at all temperatures tested (Fig. 16), indicating that MaleE-RF-Tel4c had synthesized larger amounts of cDNAs extending to near the 5' end of the RNA template. Notably, the difference in amounts of cDNAs synthesized was most pronounced at temperatures between 55 and 65°C, where the activity of SuperScript III falls off rapidly.

[0096] To compare the processivity of cDNA synthesis by MaleE-RF-Tel4c and SuperScript III RTs, the same cDNA samples obtained at 60 and 65 °C were analyzed with two different

amplicon primer/probe sets: 188-257, which detects cDNAs that are 920-nt long, and 562-634, which detects cDNAs that are 546 nt long (Fig. 17). In this case, cycle threshold results for cDNA samples were plotted against a standard curve obtained with Novagen[®] double-stranded DNA plasmid vector pET9a to determine copy numbers equivalents. With the 188-257 amplicon primer/probe set, 972,815 copies were detected with the MalE-RF-4c Tel4c RT versus 64,456 copies with SuperScript RT at 60°C (~15 fold difference), and that ratio increased to 732,559 versus 661 at 65 °C (~1100 fold difference). Further, at both temperatures, the MalE-RF-Tel4c RT shows little difference in the copy numbers of cDNAs detected by the two primer sets, showing that the MalE-RF-Tel4c RT synthesizes mostly full-length cDNAs, indicative of high processivity. By contrast, SuperScript III RT showed lower numbers of longer cDNAs detected by the 188-257 primer set than the 562-634 primer set at both temperatures, indicating that this RT falls off or is otherwise impeded before reaching the 5' end of the RNA, resulting in synthesis of shorter cDNAs.

Fidelity of nucleotide incorporation by Tel4c and Tel4h* RTs

[0097] The inherent fidelity of the Tel4h* and Tel4c RTs (*i.e.*, the native group II intron RT, not a stabilized RT fusion protein) was assessed initially by sequencing introns that had undergone retrohoming in *E. coli* plasmid assays (Table 2). The maximum error frequencies for the Tel4h* RNA promoting retrohoming of a Tel4h*-ΔORF intron RNA at 37 and 48 °C were 1.6×10^{-5} and 4.1×10^{-6} , respectively. The Tel4c RT is encoded by the outer intron of a “twintron”, a configuration in which one group II intron (Tel3c) has inserted into another (Tel4c), and can efficiently mobilize both introns. The maximum error frequencies for the Tel4c RT promoting retrohoming of Tel3c or Tel4c at 48 °C were 1.1×10^{-5} and 2.2×10^{-5} . These error frequencies are comparable to that estimated previously for the Ll.LtrB intron RT (LtrA) promoting retrohoming of the Ll.LtrB intron, $\sim 10^{-5}$ at 37 °C (Conlan *et al.*, Nucl. Acids Res. 33, 5262-5270, 2005).

Table 2. Fidelity of group II intron RTs as measured by frequency of nucleotide misincorporation during retrohoming

RT	TeI4h*	TeI4h*	TeI4c	TeI4c
Intron	TeI4h*-ΔORF	TeI4h*-ΔORF	TeI3c-ΔORF	TeI4c-ΔORF
Temp. (°C)	37	48	48	48
Nts sequenced	244,253	244,980	265,858	537,354
Mutations	4	1	3	12
Error Frequency	1.6×10^{-5}	4.1×10^{-6}	1.1×10^{-5}	2.2×10^{-5}

[0098] Retrohoming was done in *E. coli* HMS174(DE3) with donor plasmids expressing the indicated intron and RT and recipient plasmids containing the intron target site (ligated E1-E2) sequences cloned upstream of a promoterless *tet^R* gene. After selection of Tet^R colonies, introns that had integrated into the target site in recipient plasmid were amplified by colony PCR using the primers Rsense (5'-ACAAATAGGGGTTCCGCGCAC; SEQ ID NO: 22) and Te680rc (5'-GTTGGTGACCGCACCAGT; SEQ ID NO: 23) and Te420f (5'-AACGCGGTAAGCCCGTA; SEQ ID NO: 24) and Rev2pBRR (5'-AATGGACGATATCCCGCA; SEQ ID NO: 25) for the 5'- and 3'-integration junctions, respectively. The PCR fragments were then sequenced. Table 2 indicates the induction temperature for retrohoming, the total number of intron nucleotides sequenced, the number of mutations (errors), and the error frequency.

[0099] The following examples of methods for preparing and characterizing stabilized RT fusion proteins are included for purposes of illustration and are not intended to limit the scope of the invention.

EXAMPLES

Example 1: Recombinant plasmids

[00100] pMalE-TeI4c, pMalE-TeI4f, pMalE-TeI4h* contain the RT ORF of the indicated mobile group II intron with a fused N-terminal MalE tag cloned behind the tac promoter in the expression vector pMal-c2t. The latter is a derivative of pMal-c2x (New England Biolabs,

Ipswich MA) in which the factor Xa protease-cleavage site between MalE and the expressed protein was replaced by a TEV protease-cleavage site (Kristelly *et al.*, Acta Crystallogr D Biol Crystallogr. 59, 1859-1862, 2003). The Tel4h* RT is a derivative of the native Tel4h RT with the YAGD motif in RT-5 changed to YADD. Recombinant plasmids containing group II introns from *T. elongatus* strain BP1 cloned in pET11 (Tel4f), pUC19 (Tel4c), or pACD2X (Tel4h*) were described previously. pMalE-RT plasmids were derived from these initial constructs by PCR amplifying the RT ORF with primers that append restriction sites, and then cloning the PCR products into the corresponding sites of pMal-c2t (Tel4c RT, EcoRI and PstI sites; Tel4f RT, BamHI site; Tel4h* RT, BamHI and PstI sites). Recombinant plasmids denoted pMalE-RF-protein (*e.g.*, pMalE-RF-Tel4c) were derived from the corresponding pMalE-RT plasmids by replacing the TEV-protease cleavable linker (TVDEALKDAQTNS₃N₁₀LENLYFQG; SEQ ID NO: 19) with a rigid linker (TVDAALAAAQTAAAAA; SEQ ID NO: 20) by the QuikChange PCR procedure using the Accuprime polymerase (Invitrogen, Makarova *et al.*, BioTechniques 29, 970-972, 2000).

[00101] Derivatives of pMalE-RF-Tel4c with different linkers were constructed by PCR mutagenesis using the QuikChange procedure. The MalE tag was fused to the C-terminus of the Tel4c ORF in pMal-c2t by amplifying the MalE segment of pMal-c2t with primers that introduce a 5' EcoRI site and a 3' PstI site, and the Tel4c ORF of pMalE-Tel4c with gene specific primers that introduce a 5' NdeI site and a 3' EcoRI site, respectively, and cloning the fragments into pMal-c2t digested with NdeI and PstI.

[00102] pNusA-RF-Tel4c-His, which expresses the Tel4c RT with an N-terminal NusA tag fused to the protein via a rigid linker and a C-terminal His6 tag, was constructed by PCR amplifying the Tel4c RT ORF from pMAL-Tel4c with primers that append SacII and KpnI sites and cloning the resulting PCR product between the corresponding sites of pET-50b(+) (Novagen). PCR mutagenesis was then used to replace the last two charged residues (D and E) of NusA, the existing linker, and one of the two N-terminal His6 tags (NICWFGDEATSGSGH₆; SEQ ID NO: 26) with a rigid linker sequence (NICWFGAAAAA; SEQ ID NO: 27). The second N-terminal His6 tag was removed by PCR mutagenesis and a His6 tag was fused to the C-terminus of Tel4c RT by QuikChange PCR.

[00103] pMalE-GsI1 and pMalE-GsI2 were constructed by PCR amplifying the RT ORFs from *G. stearothermophilus* strain 10 genomic DNA (obtained from Greg Davis (Sigma-

Aldrich)) by PCR with primers that amplify the introns and appended BamHI and XbaI sites (GsI1) or BamHI sites (GsI2) and then cloning the PCR products between the corresponding sites of pMal-c2t. GsI1 is a subgroup IIB2 intron that is inserted in the *G. stearothermophilus* *recA* gene and is related to the previously described RT-encoding group II introns in the *recA* genes of *Geobacillus kaustophilus* (Chee *et al.*, Gene 363, 211-220, 2005) and *Bacillus caldolyticus* (Ng *et al.*, Gene 393, 137-144, 2007). The cloned GsI1 RT ORF was verified to correspond to the genomic sequence (CP001794). GsI2 is a group IIC intron found in multiple copies in the *G. stearothermophilus* genome. The cloned GsI2 RT ORF corresponds to the genomic sequence of one of six full-length copies of GsI2 in the *G. stearothermophilus* genome (CP001794) and has three amino acid sequence changes from the RT ORF cloned by Vellore *et al.* (Appl. Environ. Microbiol. 70, 7140-7147, 2004). The corresponding pMalE-RF-RT constructs were derived from the pMalE-RT constructs by QuikChange PCR, as described above.

[00104] pMalE-LtrA was constructed by PCR amplifying the LtrA ORF of pImp-2 (Saldanha *et al.*, Biochemistry 38, 9069-9083, 1999) using primers that append BamHI and HindIII sites and then cloning the PCR product between the corresponding sites of pMal-c2t, and pMalE-RF-LtrA was derived from pMalE-LtrA by QuikChange PCR, as described above.

Example 2: Protein purification

[00105] For expression of pMalE-RT or pMalE-RF-RT constructs, *E. coli* Rosetta 2/pRARE (Novagen, EMD Biosciences, Gibbstown NJ) or ScarabXpress/pRARE T7lac (Scarabgenomics, Madison WI) were transformed with the expression plasmid and grown at 37 °C in TB or LB medium to mid-log phase (O.D.₆₀₀ = 0.8). Expression was induced either by adding isopropyl β-D-1-thiogalactopyranoside (IPTG; 1 mM final) to mid-log phase cells (pMalE-RF-Tel4c, Tel4f, Tel4h*, GsI1, and GsI2) or by growing cells in auto-induction medium (LB containing 0.2% lactose, 0.05% glucose, 0.5% glycerol, 24 mM (NH₄)₂SO₄, 50 mM KH₂PO₄, 50 mM Na₂HPO₄) (pMalE-LtrA and pMalE-RF-LtrA). In either case, induction was for ~24 h at 18-25 °C, after which cells were pelleted by centrifugation, resuspended in buffer A (20 mM Tris-HCl, pH 7.5, 0.5 M KCl or NaCl, 1 mM EDTA, 1 mM dithiothreitol (DTT)), and frozen at -80 °C.

[00106] For purification of MalE-RF-Tel4c, Tel4f, Tel4h* and their derivatives, the cell suspension was thawed, treated with lysozyme (1 mg/ml; Sigma) for 15 min on ice, freeze-thawed three times on dry ice, sonicated (Branson 450 Sonifier, Branson Ultrasonics, Danbury CT) three or four 10 sec bursts or one 30 sec burst on ice at an amplitude of 60%, with 10 sec between bursts, and centrifuged for 30 min at 18,500 x g at 4 °C. Nucleic acids were precipitated by adding polyethyleneimine (PEI) to a final concentration of 0.1% and centrifuging for 15 min at 15,000 x g at 4 °C in a J16.25 rotor in an Avanti J-E centrifuge (Beckman Coulter, Brea CA). The resulting supernatant was applied to an amylose column (10-ml column volume; Amylose High-Flow (New England Biolabs), equilibrated in buffer A), which was washed with five column volumes each of buffer A containing 0.5 M, 1.5 M, or 0.5 M KCl, and then eluted with buffer A containing 10 mM maltose. Protein fractions were pooled and purified further via a heparin-Sepharose column (3 tandem 1-ml columns; GE Healthcare Biosciences Corp.) which had been pre-equilibrated in 20 mM Tris-HCl, pH 7.5 containing KCl (100 mM for MalE-RF-4c, 4f, 4h*, MalE-LtrA and MalE-RF-LtrA; 50 mM for MalE-RF-GsI1 or GsI2), 1 mM EDTA, 1 mM DTT, 10% glycerol. The proteins were applied to the column in the same buffer and eluted with a 40-column volume gradient from the loading concentration to 2 M KCl. The proteins eluted at ~800 mM KCl. The peak fractions were pooled and dialyzed against 20 mM Tris-HCl, pH 7.5, 0.5 M KCl, 1 mM EDTA, 1 mM DTT, and 50% glycerol for storage. The frozen proteins showed no decrease in RT activity for at least six months.

[00107] The MalE-RF-GsI1 protein, which has an N-terminal MalE tag and a C-terminal His6-tag, was purified similarly, except that nucleic acids were precipitated with 0.2% PEI, and the protein eluted from the amylose column was purified further on a nickel column prior to the final heparin-Sepharose column. The nickel column (5 ml HisTrap™ HP Nickel Sepharose; GE Healthcare Biosciences, Piscataway NJ) equilibrated with binding buffer (500 mM KCl, 20 mM Tris-HCl pH 7.5, 40 mM imidazole, and 10% glycerol) was loaded with pooled protein fractions from the amylose column, washed with 10 column volumes of binding buffer, eluted with five column volumes of elution buffer (500 mM KCl, 20 mM Tris-HCl pH 7.5, 400 mM imidazole and 10% glycerol), and the supernatant loaded directly onto the heparin-Sepharose column. The peak fractions from the heparin-Sepharose column were pooled, dialyzed against 20 mM Tris-HCl, pH 7.5, 0.5 M KCl, 50% glycerol, and stored as described above.

[00108] For the NusA fusions, *E. coli* ScarabXpress/pRARE T7lac cells were induced with 0.5 mM IPTG for 48 h at 18 °C and resuspended in nickel buffer A (20 mM Tris pH 7.5, 500 mM KCl, 30 mM imidazole, 10% glycerol). After disrupting the cells as described above, nucleic acids were precipitated from the lysate by adding a final concentration of 0.2% polyethyleneimine, followed by centrifugation at 10,000 x g for 15 min. The supernatant was applied to a 5-ml nickel-Sepharose column pre-equilibrated with nickel buffer A, and then eluted with nickel buffer A containing 500 mM imidazole. The protein fractions were pooled and loaded directly onto two connected 1-ml heparin-Sepharose columns that had been pre-equilibrated in 20 mM Tris pH 7.5, 100 mM KCl, 1 mM DTT, 1 mM EDTA, and 20% glycerol. The protein was eluted with a 20-column volume gradient of 0.1 to 1.5 M KCl, and peak fractions were pooled, dialyzed against 20 mM Tris-HCl, pH 7.5, 0.5 M KCl, 1 mM EDTA, 1 mM DTT, 50% glycerol, and stored as described above.

Example 3: Reverse transcriptase assays

[00109] RT activity at different temperatures was assayed by quantifying incorporation of ³²P-dTTP using poly(rA)/oligo(dT)₄₂ as the template-primer. The RT (50 nM MalE-RF-Tel4c RT or 100 nM of all other RTs) was pre-incubated with 100 nM poly(rA)/oligo(dT)₄₂ in 1x RT buffer (75 mM KCl, 10 mM MgCl₂, 20 mM Tris-HCl, pH 7.5, and 1 mM DTT) at different temperatures (ranging from 25-77 °C), and reactions were initiated by adding 5 µCi [α-³²P]-dTTP (3,000 Ci/mmol; Perkin Elmer, Waltham MA). The reactions were incubated for times within the linear range and stopped by adding EDTA to a final concentration of 250 mM. Reaction products were spotted onto Whatman DE81 chromatography paper (10 x 7.5-cm sheets; GE Healthcare), washed 3 times in 0.3 M NaCl and 0.03 M sodium citrate, and scanned with a PhosphorImager (Typhoon Trio Variable Mode Imager; GE Healthcare) to quantify bound radioactivity.

[00110] Other RT assays used RNA templates with annealed DNA oligonucleotide primers. The RNA template was either a 531-nt *in vitro* transcript synthesized from pBluescript KS (+) digested with AflIII transcribed using T7 Megascript kits (Ambion, Applied Biosystems, Austin, TX) or a 1.2-kb kanR RNA purchased from Promega (Promega, Madison WI). *In vitro* transcription was done according to the manufacturer's instructions for 4 h at 37 °C. After digesting the DNA template with Turbo DNase I (5 min, 37 °C), RNAs were extracted with phenol:chloroform:isoamyl alcohol (25:24:1; phenol-CIA) and purified by two cycles of

gel filtration through Sephadex G-50 (Sigma, St Louis, MO) spin columns. The RNA concentration was determined by using a Nanodrop (Thermo Scientific, Wilmington, DE). RNAs were stored in Milli-Q-grade H₂O and stored at -20 °C.

[00111] DNA oligonucleotide primers complementary to the 3' ends of the RNAs were synthesized by IDT (Coralville, IA; AflIII primer: 5'-CCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCG; SEQ ID NO: 28; P078 Kanamycin Rev 5'-GGTGGACCAGTTGGTGATTTTGAAC TTTTGCTTTGCCACGGAAC; SEQ ID NO: 29). Primer concentrations were determined by A₂₆₀. The primers were 5' ³²P- labeled with T4 polynucleotide kinase (New England Biolabs) according to the manufacturer's instructions, and free nucleotides were removed by gel filtration through a Sephadex G-25 column. The primers were mixed with the template at a molar ratio of 1.0:1.1 and annealed by heating to 82°C for 2 min and then cooling to room temperature in a GeneAmp 9700 PCR cycler with the ramp setting of 10%.

[00112] For gel analysis of cDNA synthesis, 100 nM of annealed template/primer was incubated with 200 nM enzyme in 100 mM KCl, 20 mM Tris HCl pH 7.5, 10 mM MgCl₂ and 1 mM DTT for MaleE-RF-Tel4c RT and in 10 mM NaCl, 20 mM Tris HCl pH 7.5, 10 mM MgCl₂ and 1 mM DTT for MaleE-RF-GsI2 RT. Reactions were initiated by adding dNTPs and MgCl₂ to final concentrations of 1.25 mM and 10 mM, respectively, incubated for 30 min at the indicated temperature, and terminated by adding 0.1% SDS/250 mM EDTA (final concentrations) followed by phenol-CIA extraction. The products were analyzed by electrophoresis in a denaturing 6% polyacrylamide gel, which was dried and quantified with a PhosphorImager. A 5'-labeled 10-bp ladder (Invitrogen™) was used as size markers.

Example 4: Quantitative Real-Time Polymerase Chain Reaction (qPCR)

[00113] cDNAs for qPCR analysis were generated in 20 µl reactions containing 1X RT buffer (75 mM KCl, 10 mM MgCl₂, 20 mM Tris-HCl, pH 7.5), 1 mM DTT, 5 x 10⁸ copies of kanR RNA, 200 nM MaleE-RF-Tel4c RT and 1 mM dNTPs for 30 min at temperatures specified for individual experiments. Parallel reactions with SuperScript III (Invitrogen) were done according to the manufacturers specifications. Reactions were incubated at the different temperatures for 2 min and started by adding dNTPs. After incubating for 30 min, the

reactions were quickly frozen on dry ice to stop the reactions. 5 µl of cDNA reaction were used for the qPCR.

[00114] qPCR analysis was done in 96-well plates with optical caps with each well containing 25 µl of reaction mix consisting of 12.5 µl of 2X TaqMan[®] Gene Expression Master Mix (Applied Biosystems, Foster City, CA), 7.5 µl of forward, reverse, and dual-labeled probe mix (oligonucleotides purchased individually from Integrated DNA Technologies, Coralville, IA), and 5 µl cDNA template. The mixture was incubated in the 7900HT Fast Real-Time PCR System (Applied Biosystems), using the 9600 emulation mode protocol (50 °C for 2 min, 95 °C for 10 min, then cycled for a total of 45 cycles at 95 °C for 15 sec and 60 °C for 60 sec). Data were collected and analyzed using the Applied Biosystems Sequence Detection System Software, Versions 2.2 or 2.3.

[00115] The Novagen[®] double-stranded DNA plasmid vector pET9a (EMD Chemicals) was used to quantitate kanR cDNA levels. The pET9a vector contains the kanR coding sequence (bases 3523-4335) and has 100% sequence homology at each primer/probe binding site with the Promega 1.2-kb kanR RNA. Purified and quantitated pET9a DNA vector was initially diluted to 1 x 10⁹ copies/µl stock aliquots and stored at -20 °C. For each run, fresh stocks were thawed and then serially diluted to generate a quantitative standard curve used in qPCR. Cycle threshold results for cDNA samples were then plotted against the standard curve to determine copy numbers equivalents.

[00116] Primers used were:

P078 Kanamycin RT-1107R 5'-

GGTGGACCAGTTGGTGATTTTGAAC TTTTGCTTTGCCACGGAAC-3'; SEQ ID NO:
29 (T_m= 80°C)

primer sets nt 188 – 257 :

Forward -- P029 kan-188F: 5'-GGGTATAAATGGGCTCGCG-3'; SEQ ID NO: 30

Reverse -- P030 kan-257R: 5'-CGGGCTTCCCATAACAATCG-3'; SEQ ID NO: 31

Taqman Probe -- P031 kan -213T: 5' (6-carboxyfluorescein (6FAM))-TCGGGCAATCAGGTGCGACAATC-3'; (Iowa Black FQ; a dark non-fluorescent quencher); SEQ ID NO: 32

Amplicon 70 bp:

5'GGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCG-3'; SEQ ID NO: 33

Primer Set (nt 562 – 634) :

Forward -- P001 kan-562F: 5'-CGCTCAGGCGCAATCAC-3'; SEQ ID NO: 34

Reverse -- P002 kan-634R: 5'-CCAGCCATTACGCTCGTCAT-3'; SEQ ID NO: 35

Taqman Probe -- P003 kan-581T: 5' (6-FAM)-ATGAATAACGGTTTGGTTGATGCGAGTGA-3'-(TAMRA); SEQ ID NO: 36

Amplicon 73 bp

5'CGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGTGACGAGCGTAATGGCTGG-3'; SEQ ID NO: 37

Example 5: Retrohoming assays.

[00117] Retrohoming assays were done in *E. coli* HMS174(DE3) (Novagen™) grown on LB medium, with antibiotics added at the following concentrations: ampicillin, 100 µg/ml; chloramphenicol, 25 µg/ml; tetracycline, 25 µg/ml. The intron-donor plasmids, derivatives of pACD2X (San Filippo *et al.*, Journal of Molecular Biology, 324, 933-951, 2002), carry a *cap^R* marker and use a T7lac promoter to express a ΔORF intron (I-ΔORF) with short flanking 5' and 3' exons (E1 and E2, respectively) and a T7 promoter in DIV, followed by the RT ORF downstream of E2. The recipient plasmids, derivatives of pBRR-tet (Guo *et al.*, Science 289, 452-457, 2000; Karberg *et al.*, Nature Biotech. 19, 1162-1167, 2001), carry an *amp^R* marker and contain a target site for the intron (ligated E1-E2 sequences) cloned upstream of a promoterless *tet^R* gene. The latter is activated by insertion of the intron carrying the T7 promoter, enabling selection for Tet^R + Amp^R colonies. For the assays, cells were co-transformed with the Cap^R donor and Amp^R recipient plasmids, inoculated into 5 ml of LB medium containing chloramphenicol and ampicillin, and grown with shaking (200 rpm)

overnight at 37 °C. A small portion (50 µl) of the overnight culture was inoculated into 5 ml of fresh LB medium containing the same antibiotics and grown for 1 h as above. The cells were then induced with IPTG for 1 h under conditions specified in the legend of Table 1 for individual experiments. The cultures were then placed on ice, diluted with ice-cold LB, and plated at different dilutions onto LB agar containing ampicillin or ampicillin + tetracycline. After incubating the plates overnight at 37 °C, the mobility efficiency was calculated as the ratio of (Tet^R + Amp^R)/Amp^R colonies.

AMENDED CLAIMS

received by the International Bureau on 19 November 2010 (19.11.2010)

WHAT IS CLAIMED IS:

1. A stabilized reverse transcriptase fusion protein comprising a thermostable reverse transcriptase connected to a stabilizer protein by a linker peptide.
2. The stabilized reverse transcriptase fusion protein of claim 1, wherein the thermostable reverse transcriptase comprises a bacterial reverse transcriptase.
3. The stabilized reverse transcriptase fusion protein of claim 2, wherein the bacterial reverse transcriptase comprises a group-II intron-derived reverse transcriptase.
4. The stabilized reverse transcriptase fusion protein of claim 3, wherein the bacterial reverse transcriptase is a *Thermosynechococcus* reverse transcriptase or *Geobacillus* reverse transcriptase.
5. The stabilized reverse transcriptase fusion protein of claim 1, wherein the thermostable reverse transcriptase comprises a polypeptide with at least 85% amino acid sequence identity to a sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 5.
6. The stabilized reverse transcriptase fusion protein of claim 1, wherein the stabilizer protein comprises an affinity protein or a solubility-enhancing protein.
7. The stabilized reverse transcriptase fusion protein of claim 6, wherein the stabilizer protein comprises a maltose binding protein or an N-utilization substance A protein.
8. The stabilized reverse transcriptase fusion protein of claim 6, wherein the stabilizer protein has been modified by replacing charged amino acids with uncharged amino acids.
9. The stabilized reverse transcriptase fusion protein of claim 1, wherein the linker peptide is a non-cleavable linker peptide.
10. The stabilized reverse transcriptase fusion protein of claim 9, wherein the non-cleavable linker peptide is a rigid linker peptide.

11. The stabilized reverse transcriptase fusion protein of claim 9, wherein the linker peptide consists of 1 to 20 amino acids.
12. The stabilized reverse transcriptase fusion protein of claim 9, wherein the linker peptide consists of 1 to 5 amino acids.
13. The stabilized reverse transcriptase fusion protein of claim 9, wherein the linker peptide consists of 3 to 5 amino acids.
14. The stabilized reverse transcriptase fusion protein of claim 10, wherein the rigid linker peptide consists of SEQ ID NO: 12.
15. The stabilized reverse transcriptase fusion protein of claim 1, wherein the fusion protein has an amino acid sequence comprising a polypeptide with at least 85% amino acid sequence identity to a sequence selected from the group consisting of SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, or SEQ ID NO: 10.
16. The stabilized reverse transcriptase fusion protein of claim 1, wherein the stabilized reverse transcriptase fusion protein is capable of carrying out reverse transcription with an error frequency of 2.0×10^{-5} or less at a temperature from about 45° to about 65 °C.
17. A method for preparing a cDNA from an RNA molecule, comprising the steps of:
 - (a) adding a primer nucleotide sequence to the RNA molecule
 - (b) incubating the RNA molecule in the presence of one or more deoxy or dideoxyribonucleoside triphosphates and a stabilized reverse transcriptase fusion protein comprising a thermostable reverse transcriptase connected to a stabilizer protein under conditions sufficient to synthesize a cDNA molecule complementary to all or a portion of the RNA molecule.
18. The method of claim 17, wherein the thermostable reverse transcriptase is connected to the stabilizer protein by a non-cleavable linker peptide.
19. The method of claim 17, wherein reverse transcription is performed within a temperature range where the RNA has a substantially decreased amount of obstructing stable secondary or tertiary structure.

20. The method of claim 17, wherein the thermostable reverse transcriptase is a group-II intron-derived reverse transcriptase.
21. The method of claim 18, wherein the thermostable reverse transcriptase comprises a polypeptide with at least 85% amino acid sequence identity to a sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 5, the non-cleavable linker consists of 1 to 20 amino acids, and the stabilizer protein comprises an affinity protein or a solubility-enhancing protein.
22. The method of claim 17, wherein the reverse transcription is performed with an error frequency of 2.0×10^{-5} or less at a temperature from about 45 °C to about 65 °C.
23. A DNA expression vector for producing a stabilized reverse transcriptase fusion protein comprising an isolated nucleic acid that encodes a polypeptide with at least 85% amino acid sequence identity to a sequence selected from the group consisting of SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, or SEQ ID NO: 10.
24. A method of producing a stabilized reverse transcriptase fusion protein, comprising the steps of
- (a) culturing a host cell comprising the DNA expression vector of claim 23;
 - (b) expressing the stabilized reverse transcriptase fusion protein encoded by the DNA expression vector; and
 - (c) isolating the stabilized reverse transcriptase fusion protein from the host cell.
25. A stabilized reverse transcriptase fusion protein comprising a thermostable reverse transcriptase connected to a stabilizer protein; wherein the thermostable reverse transcriptase comprises a polypeptide with at least 85% amino acid sequence identity to a sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4.
26. The stabilized reverse transcriptase fusion protein of claim 25, wherein the stabilizer protein comprises an affinity protein or a solubility-enhancing protein.
27. The stabilized reverse transcriptase fusion protein of claim 25, wherein the stabilizer protein has been modified by replacing charged amino acids with uncharged amino acids.

28. The stabilized reverse transcriptase fusion protein of claim 25, wherein the thermostable reverse transcriptase is connected to the stabilizer protein by a linker peptide.

29. The stabilized reverse transcriptase fusion protein of claim 25, wherein the fusion protein has an amino acid sequence comprising a polypeptide with at least 85% amino acid sequence identity to a sequence selected from the group consisting of SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, or SEQ ID NO: 9.

30. The stabilized reverse transcriptase fusion protein of claim 25, wherein the stabilized reverse transcriptase fusion protein is capable of carrying out reverse transcription with an error frequency of 2.0×10^{-5} or less at a temperature from about 45° to about 65 °C.

SEQ ID NO: 6

```
1 MKIEEGKLVI WINGDKGYNG LAEVGKKFEK DTGIKVTVEH PDKLEEKFPQ VAATGDGPDI
61 IFWAHDRFGG YAQSGLLAEI TPDKAFQDKL YPFTWDAVRY NGKLIAYPIA VEALSIIYNK
121 DLLPNPPKTW EEIPALDKEL KAKGKSALMF NLQEPYFTWP LIAADGGYAF KYENGKYDIK
181 DVGVDNAGAK AGLTFLVDLI KNKHMNADTD YSIAEAAFNK GETAMTINGP WAWSNIDTSK
241 VNYGVTVLPT FKGQPSKPFV GVLSAGINAA SPNKELAKEF LENYLLTDEG LEAVNKDKPL
301 GAVALKSYEE ELAKDPRIAA TMENAQKGEI MPNIPQMSAF WYAVRTAVIN AASGRQTVDA
361 ALAAAQTAAA AAMETRQMTV DQTTGAVTNQ TETSWHSINW TKANREVKRL QVRIAKAVKE
421 GRWGKVKALQ WLLTHSFY GK ALAVKRVTDN SGSRTPGVDG ITWSTQEQKT QAIKSLRRRG
481 YKPQPLRRVY IPKANGKQRP LGIPTMKDRA MQALYALALE PVAETTADRN SYGFRRGRCT
541 ADAAGQCFLA LAKAKSAEHV LDADISGCFD NISHEWLLAN TPLDKGILRK WLKSGFVWKQ
601 QLFPTHAGTP QGGVISPVLA NITLDGMEEL LAKHLRGQKV NLIRYADDFV VTGKDEETLE
661 KARNLIQEFL KERGLTLSPE KTKIVHIEEG FDFLGWNIRK YNGVLLIKPA KKNVKAFLKK
721 IRDTLRELRT ATQEIVIDL NPIIRGWANY HKGQVSKETF NRVD FATWHK LWRWARRRHP
781 NKPAQWVKDK YFIKNGSRDW VFGMVMKDKN GELRTKRLIK TSDTRIQRHV KIKADANPFL
841 PEWAEIFEKR KKLKKAPAY RRIRRELWKK QGGICPVC GG EIEQDMLTDI HHILPKHKGG
901 SDDLNLVLI HANCHKQVHS RDGQHSRSL L KEGL*
```

FIG. 1

SEQ ID NO: 7

```
1 MKIEEGKLVI WINGDKGYNG LAEVGKKFEK DTGIKVTVEH PDKLEEKFPQ VAATGDGPDI
61 IFWAHDRFGG YAQSGLLAEI TPDKAFQDKL YPFTWDAVRY NGKLIAYPIA VEALSIIYNK
121 DLLPNPPKTW EEIPALDKEL KAKGKSALMF NLQEPYFTWP LIAADGGYAF KYENGKYDIK
181 DVGVDNAGAK AGLTFLVDLI KNKHMNADTD YSIAEAAFNK GETAMTINGP WAWSNIDTSK
241 VNYGVTVLPT FKGQPSKPFV GVLSAGINAA SPNKELAKEF LENYLLTDEG LEAVNKDKPL
301 GAVALKSYEE ELAKDPRIAA TMENAQKGEI MPNIPQMSAF WYAVRTAVIN AASGRQTVDA
361 ALAAAQTAAA AAMETRQMAV EQTTGAVTNQ TETSWHSIDW AKANREVKRL QVRIAKAVKE
421 GRWGKVKALQ WLLTHSFY GK ALAVKRVTDN SGSKTPGVDG ITWSTQEQKA QAIKSLRRRG
481 YKPQPLRRVY IPKANGKQRP LGIPTMKDRA MQALYALALE PVAETTADRN SYGFRRGRCI
541 ADAATQCHIT LAKTDRAQYV LDADIAGCFD NISHEWLLAN IPLDKRILRK WLKSGFVWKQ
601 QLFPIHAGTP QGGVISPMLA NMTLDGMEEL LNKFPRAHKV KLIRYADDFV VTGETKEVLY
661 IAGAVIQAFI KERGLTLSKE KTKIVHIEEG FDFLGWNIRK YDGKLLIKPA KKNVKAFLKK
721 IRDTLRELRT APQEIVIDL NPIIRGWTNY HKNQASKETF VGV DHLIWQK LWRWARRRHP
781 SKSVRWVSKS YFIQIGNRKW MFGIWT KDKN GDPWAKHLIK ASEIRIQRRG KIKADANPFL
841 PEWAEIFEQR KKLKEAPAY RRTRRELWKK QGGICPVC GG EIEQDMLTEI HHILPKHKGG
901 TDDLNLVLI HTNCHKQVHN RDGQHSRFL L KEGL*
```

FIG. 2

2/24

SEQ ID NO: 8

```

1  MKIEEGKLVI  WINGDKGYNG  LAEVGKKFEK  DTGIKVTVEH  PDKLEEKFPQ  VAATGDGPDI
61  IFWAHDRFGG  YAQSGLLAEI  TPDKAFQDKL  YPFTWDAVRY  NGKLIAYPIA  VEALSIIYNK
121  DLLPNPPKTW  EEIPALDKEL  KAKGKSALMF  NLQEPYFTWP  LIAADGGYAF  KYENGKYDIK
181  DVGVDNAGAK  AGLTFLVDLI  KNKHMNADTD  YSIAEAAFNK  GETAMTINGP  WAWSNIDTSK
241  VNYGVTVLPT  FKGQPSKPFV  GVLSAGINAA  SPNKELAKEF  LENYLLTDEG  LEAVNKDKPL
301  GAVALKSYEE  ELAKDPRIAA  TMENAQKGEI  MPNIPQMSAF  WYAVRTAVIN  AASGRQTVDA
361  ALAAQAQAAA  AAMETRQMAV  EQTTGAVTNQ  TETSWHSIDW  AKANREVKRL  QVRIAKAVKE
421  GRWGKVKALQ  WLLTHSFY GK  ALAVKRVTDN  SGSKTPGVDG  ITWSTQEQKA  QAIKSLRRRG
481  YKPQPLRRVY  IPKASGKQRP  LGIPTTKDRA  MQALYALALE  PVAETTADRN  SYGFRQGRCT
541  ADAAGQCFTV  LGRSDCAKYI  LDADITGCFD  NISHEWLLDN  IPLDKEVLRK  WLKSGFVWKQ
601  QLFPTHAGTP  QGGVISPLMA  NMTLDGMEEL  LKKHLRKQKV  NLIRYADDFV  VTGESKETLE
661  KVTTVIQEFL  KERGLTLSEE  KTKVVHIEEG  FDFLGWNIRK  YGEKLLIKPA  KKNIKAFHKK
721  IRDALKELRT  ATQEAVIDTL  NPIIKGWANY  HRNQVSKRIF  NRADDNIWHK  LWRWAKRRHP
781  NKPARWTKNK  YFIKIGNRHW  VFGTWKKDKE  GRLRSRYLIK  AGDTRIQRHV  KIKADANPFL
841  PEWAIFYEER  KKLKEAPAQY  RRIRRELWKK  QGGICPVC GG  EIEQDMLTEI  HHILPKHKGG
901  SDDLNLVLI  HANCHKQVHS  RDGQHSRFL  L  KEGL*

```

FIG. 3

SEQ ID NO: 9

```

1  MKIEEGKLVI  WINGDKGYNG  LAEVGKKFEK  DTGIKVTVEH  PDKLEEKFPQ  VAATGDGPDI
61  IFWAHDRFGG  YAQSGLLAEI  TPDKAFQDKL  YPFTWDAVRY  NGKLIAYPIA  VEALSIIYNK
121  DLLPNPPKTW  EEIPALDKEL  KAKGKSALMF  NLQEPYFTWP  LIAADGGYAF  KYENGKYDIK
181  DVGVDNAGAK  AGLTFLVDLI  KNKHMNADTD  YSIAEAAFNK  GETAMTINGP  WAWSNIDTSK
241  VNYGVTVLPT  FKGQPSKPFV  GVLSAGINAA  SPNKELAKEF  LENYLLTDEG  LEAVNKDKPL
301  GAVALKSYEE  ELAKDPRIAA  TMENAQKGEI  MPNIPQMSAF  WYAVRTAVIN  AASGRQTVDA
361  ALAAQAQAAA  AAMKVNKL VV  KSEQDLRNCL  DLLYQEAKKG  KHFGMLELL  QNDVVILEAI
421  RNIKSNKGSK  TAGIDQKIVD  DYLLMPTEKV  FGMIAKALND  YKPIPVRRCN  KPKGNAKSSK
481  RKGNSPNEEG  ETRPLGISAV  TDRIIQEMLR  IVLEPIFEAQ  FYPHSYGFRP  YRSTEHALAW
541  MLKIINGSKL  YWVVKGDIES  YFDHINHKKL  LNIMWNMGVR  DKRVL CIVKK  MLKAGQVIQG
601  KFYPTAKGIP  QGGIISPLLA  NVYLNSFDWM  VGQEYEHHPN  NANYREKKNA  LAALRNKGHH
661  PVFYIRYADD  WVILTDTKEY  AEKIREQCKQ  YLACELHLTL  SDEKTFIADI  REQ RVKFLGF
721  CIEAGKRRFH  KKGFAARMIP  DMEKVNKAVK  EIKRDIRLLR  TRKSELEKAL  DIENINTKII
781  GLANHLKIGI  SKYIMGKVDR  VIEETAYRTW  VKMYGKEKAA  QYKRPVSEFH  NRIDRHKG YQ
841  MKHFSVVTED  GIRVGITHAK  ITPIQYATVF  KQEMTPYTAD  GRKMYEEKHR  KIRLPDKMSL
901  FDHDSIFIYI  LSEHNDGKYN  LEYFLNRVNV  FHRDKGKCKI  CAVYLSPGNF  HCHHIDPSKP
961  LSEINKTVNL  ISLCNQCHRL  VHSNQEPFFT  ERKMFDKLT K  YRNKLKI*

```

FIG. 4

SEQ ID NO: 10

```

1  MKIEEGKLVI  WINGDKGYNG  LAEVGKKFEK  DTGIKVTVEH  PDKLEEKFPQ  VAATGDGPDI
61  IFWAHDRFGG  YAQSGLLAEI  TPDKAFQDKL  YPFTWDAVRY  NGKLIAYPIA  VEALSIIYNK
121  DLLPNPPKTW  EEIPALDKEL  KAKGKSALMF  NLQEPYFTWP  LIAADGGYAF  KYENGKYDIK
181  DVGVDNAGAK  AGLTFLVDLI  KNKHMNADTD  YSIAEAAFNK  GETAMTINGP  WAWSNIDTSK
241  VNYGVTVLPT  FKGQPSKPFV  GVLSAGINAA  SPNKELAKEF  LENYLLTDEG  LEAVNKDKPL
301  GAVALKSYEE  ELAKDPRIAA  TMENAQKGEI  MPNIPQMSAF  WYAVRTAVIN  AASGRQTVDA
361  ALAAQAQAAA  AAMALLERIL  ARDNLITALK  RVEANQGAPG  IDGVSTDQLR  DYIRAHWSTI
421  HAQLLAGTYR  PAPVRRVEIP  KPGGGTRQLG  IPTVVDR LIQ  QAILQELTPI  FDPDFSSSS F
481  GFRPGRNAHD  AVRQAQGYIQ  EGYRYVVDMD  LEKFFDRVNH  DILMSRVARK  VKDKRVLKLI
541  RAYLQAGVMI  EGVKVQTEEG  TPQGGPLSPL  LANILLDDLD  KELEKRGLKF  CRYADDCNIY
601  VKSLRAGQRV  KQSIQRFLEK  TLKLKVNEEK  SAVDRPWKRA  FLGFSFTPER  KARIRLAPRS
661  IQRLKQRI RQ  LTNP NWSISM  PERIHRVNQY  VMGWIGYFRL  VETPSVLQTI  EGWIRRLRL
721  CQWLQWKVR  TRIREL RALG  LKETAVMEIA  NTRKGAWRTT  KTPQLHQALG  KTYWTAQGLK
781  SLTQRYFELR  QG*

```

FIG. 5

SEQ ID NO: 13

```
1  CCGACACCAT CGAATGGTGC AAAACCTTTC GCGGTATGGC ATGATAGCGC CCGGAAGAGA
61  GTCAATTCAG GGTGGTGAAT GTGAAACCAG TAACGTTATA CGATGTGCGA GAGTATGCCG
121 GTGTCTCTTA TCAGACCGTT TCCCGCGTGG TGAACCAGGC CAGCCACGTT TCTGCGAAAA
181 CGCGGGAAAA AGTGGAAGCG GCGATGGCGG AGCTGAATTA CATTCCCAAC CGCGTGGCAC
241 AACAACTGGC GGGCAAACAG TCGTTGCTGA TTGGCGTTGC CACCTCCAGT CTGGCCCTGC
301 ACGCGCCGTC GCAAATTGTC GCGGCGATTA AATCTCGCGC CGATCAACTG GGTGCCAGCG
361 TGGTGGTGTG GATGGTAGAA CGAAGCGGCG TCGAAGCCTG TAAAGCGGCG GTGCACAATC
421 TTCTCGCGCA ACGCGTCAGT GGGCTGATCA TTAACATATC GCTGGATGAC CAGGATGCCA
481 TTGCTGTGGA AGCTGCCTGC ACTAATGTTC CGGCGTTATT TCTTGATGTC TCTGACCAGA
541 CACCCATCAA CAGTATTATT TTCTCCCATG AAGACGGTAC GCGACTGGGC GTGGAGCATC
601 TGGTCGCATT GGGTCACCAG CAAATCGCGC TGTTAGCGGG CCCATTAAGT TCTGTCTCGG
661 CGCGTCTGCG TCTGGCTGGC TGGCATAAAT ATCTCACTCG CAATCAAATT CAGCCGATAG
721 CGGAACGGGA AGGCGACTGG AGTGCCATGT CCGGTTTTTC ACAAACCATG CAAATGCTGA
781 ATGAGGGCAT CGTTCCCACT GCGATGCTGG TTGCCAACGA TCAGATGGCG CTGGGCGCAA
841 TCGCGCCCAT TACCGAGTCC GGGCTGCGCG TTGGTGC GGA TATCTCGGTA GTGGGATACG
901 ACGATACCGA AGACAGCTCA TGTATATATC CGCCGTTAAC CACCATCAAA CAGGATTTTC
961 GCCTGCTGGG GCAAACCAGC GTGGACCGCT TGCTGCAACT CTCTCAGGGC CAGGCGGTGA
1021 AGGGCAATCA GCTGTTGCCC GTCTCACTGG TGAAAAGAAA AACCACCCTG GCGCCCAATA
1081 CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT GCAGCTGGCA CGACAGGTTT
1141 CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG TAAGTTAGCT CACTCATTAG
1201 GCACAATTCT CATGTTTGAC AGCTTATCAT CGACTGCACG GTGCACCAAT GCTTCTGGCG
1261 TCAGGCAGCC ATCGGAAGCT GTGGTATGGC TGTGCAGGTC GTAAATCACT GCATAATTCTG
1321 TGTGCTCAA GGCGCACTCC CGTTCTGGAT AATGTTTTTT GCGCCGACAT CATAACGGTT
1381 CTGGCAAATA TTCTGAAATG AGCTGTTGAC AATTAATCAT CGGCTCGTAT AATGTGTGGA
1441 ATTGTGAGCG GATAACAATT TCACACAGGA AACAGCCAGT CCGTTTAGGT GTTTTCACGA
1501 GCACTTCACC AACAAGGACC ATAGCATATG AAAATCGAAG AAGGTAAACT GGTAATCTGG
1561 ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT
1621 ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT
1681 GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC
1741 GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCAGGA CAAGCTGTAT
1801 CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT
1861 GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA
1921 GAGATCCCGG CGCTGGATAA AGAACTGAAA GCGAAAGGTA AGAGCGCGCT GATGTTCAAC
1981 CTGCAAGAAC CGTACTTCAC CTGGCCGCTG ATTGCTGCTG ACGGGGGTTA TCGTTC AAG
2041 TATGAAAACG GCAAGTACGA CATTAAAGAC GTGGGCGTGG ATAACGCTGG CGCGAAAGCG
2101 GGTCTGACCT TCCTGGTTGA CCTGATTAAA AACAAACACA TGAATGCAGA CACCGATTAC
2161 TCCATCGCAG AAGCTGCCTT TAATAAAGGC GAAACAGCGA TGACCATCAA CGGCCCGTGG
2221 GCATGGTCCA ACATCGACAC CAGCAAAGTG AATTATGGTG TAACGGTACT GCCGACCTTC
2281 AAGGGTCAAC CATCCAAACC GTTCGTTGGC GTGCTGAGCG CAGGTATTAA CGCCGCCAGT
2341 CCGAACAAAG AGCTGGCAAA AGAGTTCCTC GAAAACTATC TGCTGACTGA TGAAGGTCTG
2401 GAAGCGGTTA ATAAAGACAA ACCGCTGGGT GCCGTAGCGC TGAAGTCTTA CGAGGAAGAG
2461 TTGGCGAAAG ATCCACGTAT TGCCGCCACT ATGGAAAACG CCCAGAAAGG TGAAATCATG
2521 CCGAACATCC CGCAGATGTC CGCTTTCTGG TATGCCGTGC GTACTGCGGT GATCAACGCC
2581 GCCAGCGGTC GTCAGACTGT CGATGCCGCC CTGGCCGCCG CGCAGACTGC CGCCGCCGCC
2641 GCCATGGAGA CAAGGCAAAT GACGGTGGAC CAAACCACTG GTGCGGTCAC CAACCAAACG
2701 GAAACAAGCT GGCACAGCAT AAACCTGGACC AAAGCCAACC GTGAGGTAAA GAGGCTGCAA
2761 GTGCGTATCG CAAAGGCTGT GAAGGAAGGA CGCTGGGGCA AAGTGAAAGC TTTGCAATGG
2821 CTCCTGACCC ACTCGTTCTA CGGCAAAGCC CTCGCCGTGA AACGGGTAAC TGACAACTCA
2881 GGCAGTAGAA CACCTGGTGT GGACGGGATA ACCTGGTCCA CACAAGAGCA GAAAACCCAA
2941 GCCATAAAGT CCCTCAGGAG AAGAGGCTAT AAACCCCAAC CCCTGAGGCG GGTATACATC
3001 CCGAAAGCAA ACGGCAAACA GCGCCCGCTA GGAATCCCGA CAATGAAGGA CAGGGCAATG
3061 CAGGCACTAT ATGCCCTAGC CCTAGAACCA GTCGCGGAAA CCACAGCGGA CCGGAACTCC
3121 TATGGGTTCC GCCGAGGGCG ATGTACGGCA GATGCGGCAG GACAATGCTT CCTTGCTCTG
3181 GCAAAAGCCA AGTCGGCTGA ACACGTCCTT GACGCTGACA TATCCGGATG CTTTGATAAC
3241 ATCAGCCATG AGTGGCTACT AGCCAACACT CCACTGGACA AAGGGATCTT ACGGAAATGG
3301 CTTAAATCTG GGTTCGTCTG GAAACAGCAA CTCTTCCCCA CCCATGCTGG GACACCTCAG
3361 GGAGGGGTAA TCTCCCCAGT TCTTGCCAAT ATAACCCTAG ATGGGATGGA AGAACTGTTG
3421 GCCAAACACC TCAGAGGTCA AAAAGTCAAC CTCATCCGAT ATGCTGACGA TTTTGTCTGTG
3481 ACGGGAAAAG ATGAGGAAAC CCTGGAGAAA GCCAGAAACC TAATCCAGGA GTTCCTAAAA
```

FIG. 6

```

3541 GAACGGGGCT TGACCCTGTC CCCCAGAGAAG AAAAAAATCG TCCATATTGA GGAAGGCTTC
3601 GACTTTCTCG GATGGAACAT TCGCAAGTAC AACGGGGTTC TTCTCATCAA ACCCGCGAAG
3661 AAGAACGTGA AAGCGTTCCT CAAGAAAATC CGAGACACTC TAAGGGAAC TTAGGACAGCA
3721 ACCCAGGAAA TCGTGATAGA CACACTCAAC CCAATCATTA GAGGTTGGGC CAACTATCAC
3781 AAAGGACAAG TCTCTAAGGA AACCTTCAAC CGAGTGGACT TCGCCACCTG GCACAAATTG
3841 TGGCGATGGG CAAGGCGCCG GCACCCAAAC AAACCTGCCC AATGGGTGAA GGACAAATAC
3901 TTCATCAAAA ACGGAAGCAG AGACTGGGTG TTCGGTATGG TGATGAAAGA CAAGAACGGG
3961 GAACTGAGGA CCAAACGCCT AATCAAAACC TCTGACACCC GAATCCAACG CCACGTCAAA
4021 ATCAAGGCAG ACGCCAATCC GTTTCTCCCA GAGTGGGCAG AATACTTTGA GAAACGCAAG
4081 AAACCTCAAAA AAGCCCCTGC TCAATATCGG CGCATCCGCC GAGAACTATG GAAGAAACAG
4141 GGTGGTATCT GTCCAGTATG CGGGGGTGAA ATTGAGCAAG ACATGCTCAC TGACATCCAC
4201 CACATATTGC CCAAACACAA GGGTGGTTCT GACGACCTGG ATAATCTTGT CTTAATCCAC
4261 GCCAACTGCC ACAAACAGGT GCACAGCCGA GATGGTCAGC ACAGCCGGTC CCTCTTGAAA
4321 GAGGGGCTTT GACTGCAGGC AAGCTTGGCA CTGGCCGTCG TTTTACAACG TCGTGACTGG
4381 GAAAACCTTG GCGTTACCCA ACTTAATCGC CTTGCAGCAC ATCCCCCTTT CGCCAGCTGG
4441 CGTAATAGCG AAGAGGCCCG CACCGATCGC CCTTCCCAAC AGTTGCGCAG CCTGAATGGC
4501 GAATGGCAGC TTGGCTGTTT TGGCGGATGA GATAAGATTT TCAGCCTGAT ACAGATTAAA
4561 TCAGAACGCA GAAGCGGTCT GATAAAACAG AATTGCGCTG GCGGCAGTAG CGCGGTGGTC
4621 CCACCTGACC CCATGCCGAA CTCAGAAGTG AAACGCCGTA GCGCCGATGG TAGTGTGGGG
4681 TCTCCCCATG CGAGAGTAGG GAACTGCCAG GCATCAAATA AAACGAAAGG CTCAGTCGAA
4741 AGACTGGGCC TTTCTGTTTTA TCTGTTGTTT GTCGGTGAAC GCTCTCCTGA GTAGGACAAA
4801 TCCGCCGGGA GCGGATTTGA ACGTTGCGAA GCAACGGCCC GGAGGGTGGC GGGCAGGACG
4861 CCCGCCATAA ACTGCCAGGC ATCAAATTAA GCAGAAGGCC ATCTGACGG ATGGCCTTTT
4921 TGCGTTTCTA CAAACTCTTT TTGTTTATTT TTCTAAATAC ATTCAAATAT GTATCCGCTC
4981 ATGAGACAAT AACCTTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTATT
5041 CAACATTTCC GTGTCGCCCT TATTCCCTTT TTTGCGGCAT TTTGCCTTCC TGTFTTTGCT
5101 CACCCAGAAA CGCTGGTGAA AGTAAAAGAT GCTGAAGATC AGTTGGGTGC ACGAGTGGGT
5161 TACATCGAAC TGGATCTCAA CAGCGGTAAG ATCCTTGAGA GTTTTCGCCC CGAAGAACGT
5221 TCTCCAATGA TGAGCACTTT TAAAGTTCTG CTATGTGGCG CGGTATTATC CCGTGTGAC
5281 GCCGGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACTT GGTGAGTAC
5341 TCACCAAGTCA CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGAGAATT ATGCAGTGCT
5401 GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT CGGAGGACCG
5461 AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG
5521 GAACCGGAGC TGAATGAAGC CATAACAAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA
5581 ATGGCAACAA CGTTGCGCAA ACTATTAAC TGGCGAACTAC TTACTCTAGC TTCCCGGCAA
5641 CAATTAATAG ACTGGATGGA GGCGGATAAA GTTGCGAGGAC CACTTCTGCG CTCGGCCCTT
5701 CCGGCTGGCT GGTATTATTG TGATAAATCT GGAGCCGGTG AGCGTGGGTC TCGCGGTATC
5761 ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC TCCCGTATCG TAGTTATCTA CACGACGGGG
5821 AGTCAGGCAA CTATGGATGA ACGAAATAGA CAGATCGCTG AGATAGGTGC CTCACTGATT
5881 AAGCATTTGGT AACTGTCAGA CCAAGTTTAC TCATATATAC TTTAGATTGA TTTACCCCGG
5941 TTGATAATCA GAAAAGCCCC AAAAACAGGA AGATTGTATA AGCAAATATT TAAATTGTAA
6001 ACGTTAATAT TTTGTTAAAA TTCGCGTTAA ATTTTGTGTA AATCAGCTCA TTTTTTAACC
6061 AATAGGCCGA AATCGGCAAA ATCCCTTATA AATCAAAAGA ATAGACCGAG ATAGGGTTGA
6121 GTGTTGTTCC AGTTTGGAAC AAGAGTCCAC TATTAAAGAA CGTGGACTCC AACGTCAAAG
6181 GGCGAAAAAC CGTCTATCAG GGCGATGGCC CACTACGTGA ACCATCACCC AAATCAAGTT
6241 TTTTGGGGTC GAGGTGCCGT AAAGCACTAA ATCGGAACCC TAAAGGGAGC CCCCATTATA
6301 GAGCTTGACG GGGAAAGCCG GCGAACGTGG CGAGAAAGGA AGGGAAGAAA GCGAAGGAG
6361 CGGGCGCTAG GCGCTGGCA AGTGTAGCGG TCACGCTGCG CGTAACCACC ACACCCCGCG
6421 CGCTTAATGC GCCGCTACAG GGCGCGTAAA AGGATCTAGG TGAAGATCCT TTTTGATAAT
6481 CTCATGACCA AAATCCCTTA ACGTGAGTTT TCGTTCCACT GAGCGTCAGA CCCCCTAGAA
6541 AAGATCAAAG GATCTTCTTG AGATCCTTTT TTTCTGCGCG TAATCTGCTG CTTGCAAACA
6601 AAAAAACCAC CGCTACCAGC GGTGGTTTGT TTGCCGGATC AAGAGCTACC AACTCTTTTT
6661 CCGAAGGTAA CTGGCTTCAG CAGAGCGCAG ATACCAAATA CTGTCCTTCT AGTGTAGCCG
6721 TAGTTAGGCC ACCACTTCAA GAACTCTGTA GCACCGCCTA CATACCTCGC TCTGCTAATC
6781 CTGTTACCAG TGGCTGCTGC CAGTGGCGAT AAGTCGTGTC TTACCGGGTT GGACTCAAGA
6841 CGATAGTTAC CGGATAAGGC GCAGCGGTCT GGCTGAACGG GGGGTTCTGT CACACAGCCC
6901 AGCTTGGAGC GAACGACCTA CACCGAAGT AGATACCTAC AGCGTGAGCT ATGAGAAAGC
6961 GCCACGCTTC CCGAAGGGAG AAAGGCGGAC AGGTATCCGG TAAGCGGCAG GGTGGAACA
7021 GGAGAGCGCA CGAGGGAGCT TCCAGGGGGA AACGCCTGGT ATCTTTATAG TCCTGTCGGG
7081 TTTCGCCACC TCTGACTTGA GCGTCGATTT TTGTGATGCT CGTCAGGGGG GCGGAGCCTA
7141 TGGAAAAACG CCAGCAACGC GGCCTTTTTT CGGTTCCCTG CCTTTTGCTG GCCTTTTGCT

```

FIG. 6 (cont.)

5/24

7201 CACATGTTCT TTCCTGCGTT ATCCCCTGAT TCTGTGGATA ACCGTATTAC CGCCTTTGAG
7261 TGAGCTGATA CCGCTCGCCG CAGCCGAACG ACCGAGCGCA GCGAGTCAGT GAGCGAGGAA
7321 GCGGAAGAGC GCCTGATGCG GTATTTTCTC CTTACGCATC TGTGCGGTAT TTCACACCGC
7381 ATATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC ATAGTTAAGC CAGTATACAC
7441 TCCGCTATCG CTACGTGACT GGGTCATGGC TGCGCCCCGA CACCCGCCAA CACCCGCTGA
7501 CGCGCCCTGA CGGGCTTGTC TGCTCCCGGC ATCCGCTTAC AGACAAGCTG TGACCGTCTC
7561 CGGGAGCTGC ATGTGTCAGA GGTTTTCACC GTCATCACCG AAACGCGCGA GGCAGCTGCG
7621 GTAAAGCTCA TCAGCGTGGT CGTGCAGCGA TTCACAGATG TCTGCCTGTT CATCCGCGTC
7681 CAGCTCGTTG AGTTTCTCCA GAAGCGTTAA TGTCTGGCTT CTGATAAAGC GGGCCATGTT
7741 AAGGGCGGTT TTTTCCTGTT TGGTCACTGA TGCCTCCGTG TAAGGGGGAT TTCTGTTTCA
7801 GGGGGTAATG ATACCGATGA AACGAGAGAG GATGCTCACG ATACGGGTTA CTGATGATGA
7861 ACATGCCCGG TTA CTGGAAC GTTGTGAGGG TAAACAAC TG GCGGTATGGA TGCGGCGGGA
7921 CCAGAGAAAA ATCACTCAGG GTCAATGCCA GCGCTTCGTT AATACAGATG TAGGTGTTCC
7981 ACAGGGTAGC CAGCAGCATC CTGCGATGCA GATCCGGAAC ATAATGGTGC AGGGCGCTGA
8041 CTTCCGCGTT TCCAGACTTT ACGAAACACG GAAACCGAAG ACCATTTCATG TTGTTGCTCA
8101 GGTCGCAGAC GTTTTGCAGC AGCAGTCGCT TCACGTTTCG TCGCGTATCG GTGATTTCAT
8161 CTGCTAACCA GTAAGGCAAC CCCGCCAGCC TAGCCGGGTC CTCAACGACA GGAGCACGAT
8221 CATGCGCACC CGTGGCCAGG ACCCAACGCT GCCCGAAATT

FIG. 6 (cont.)

SEQ ID NO: 14

```
1 CCGACACCAT CGAATGGTGC AAAACCTTTC GCGGTATGGC ATGATAGCGC CCGGAAGAGA
61 GTCAATTCAG GGTGGTGAAT GTGAAACCAG TAACGTTATA CGATGTCGCA GAGTATGCCG
121 GTGTCTCTTA TCAGACCGTT TCCCGCGTGG TGAACCAGGC CAGCCACGTT TCTGCGAAAA
181 CGCGGGAAAA AGTGAAGCG GCGATGGCGG AGCTGAATTA CATTCCCAAC CGCGTGGCAC
241 AACAACTGGC GGGCAAACAG TCGTTGCTGA TTGGCGTTGC CACCTCCAGT CTGGCCCTGC
301 ACGCGCCGTC GCAAATTGTC GCGGCGATTA AATCTCGCGC CGATCAACTG GGTGCCAGCG
361 TGGTGGTGTG GATGGTAGAA CGAAGCGGCG TCGAAGCCTG TAAAGCGGCG GTGCACAATC
421 TTCTCGCGCA ACGCGTCAGT GGGCTGATCA TTAACATATC GCTGGATGAC CAGGATGCCA
481 TTGCTGTGGA AGCTGCCTGC ACTAATGTTT CGGCGTTATT TCTTGATGTC TCTGACCAGA
541 CACCCATCAA CAGTATTATT TTCTCCCATG AAGACGGTAC GCGACTGGGC GTGGAGCATC
601 TGGTCGCATT GGGTCACCAG CAAATCGCGC TGTTAGCGGG CCCATTAAGT TCTGTCTCGG
661 CGCGTCTGCG TCTGGCTGGC TGGCATAAAT ATCTCACTCG CAATCAAATT CAGCCGATAG
721 CGGAACGGGA AGGCGACTGG AGTGCCATGT CCGGTTTTCA ACAAACCATG CAAATGCTGA
781 ATGAGGGCAT CGTTCCCACT GCGATGCTGG TTGCCAACGA TCAGATGGCG CTGGGCGCAA
841 TGCGCGCCAT TACCGAGTCC GGGCTGCGCG TTGGTGC GGA TATCTCGGTA GTGGGATACG
901 ACGATACCGA AGACAGCTCA TGTATATATC CGCCGTTAAC CACCATCAA CAGGATTTTC
961 GCCTGCTGGG GCAAACCAGC GTGGACCGCT TGCTGCAACT CTCTCAGGGC CAGGCGGTGA
1021 AGGGCAATCA GCTGTTGCCC GTCTCACTGG TGAAAAGAAA AACCACCCTG GCGCCCAATA
1081 CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT GCAGCTGGCA CGACAGGTTT
1141 CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG TAAGTTAGCT CACTCATTAG
1201 GCACAATTCT CATGTTTGAC AGCTTATCAT CGACTGCACG GTGCACCAAT GCTTCTGGCG
1261 TCAGGCAGCC ATCGGAAGCT GTGGTATGGC TGTGCAGGTC GTAAATCACT GCATAATTCTG
1321 TGTCGCTCAA GGCGCACTCC CGTTCTGGAT AATGTTTTTT GCGCCGACAT CATAACGGTT
1381 CTGGCAAATA TTCTGAAATG AGCTGTTGAC AATTAATCAT CGGCTCGTAT AATGTGTGGA
1441 ATTGTGAGCG GATAACAATT TCACACAGGA AACAGCCAGT CCGTTTAGGT GTTTTCACGA
1501 GCACTTCACC AACAAAGGACC ATAGCATATG AAAATCGAAG AAGGTAAACT GGTAATCTGG
1561 ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT
1621 ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT
1681 GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC
1741 GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT
1801 CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT
1861 GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA
1921 GAGATCCCGG CGCTGGATAA AGAACTGAAA GCGAAAGGTA AGAGCGCGCT GATGTTCAAC
1981 CTGCAAGAAC CGTACTTCAC CTGGCCGCTG ATTGCTGCTG ACGGGGGTTA TCGTTCAAG
2041 TATGAAAACG GCAAGTACGA CATTAAAGAC GTGGGCGTGG ATAACGCTGG CGCGAAAGCG
2101 GGTCTGACCT TCCTGGTTGA CCTGATTAAA AACAAACACA TGAATGCAGA CACCGATTAC
2161 TCCATCGCAG AAGCTGCCTT TAATAAAGGC GAAACAGCGA TGACCATCAA CGGCCCGTGG
2221 GCATGGTCCA ACATCGACAC CAGCAAAGTG AATTATGGTG TAACGGTACT GCCGACCTTC
2281 AAGGGTCAAC CATCCAAACC GTTCGTTGGC GTGCTGAGCG CAGGTATTAA CGCCGCCAGT
2341 CCGAACAAAG AGCTGGCAAA AGAGTTCCTC GAAAACATAT TGCTGACTGA TGAAGGTCTG
2401 GAAGCGGTTA ATAAAGACAA ACCGCTGGGT GCCGTAGCGC TGAAGTCTTA CGAGGAAGAG
2461 TTGGCGAAAG ATCCACGTAT TGCCGCCACT ATGGAAAACG CCCAGAAAGG TGAAATCATG
2521 CCGAACATCC CGCAGATGTC CGCTTTCTGG TATGCCGTGC GTACTGCGGT GATCAACGCC
2581 GCCAGCGGTC GTCAGACTGT CGATGCCGCC CTGGCCGCCG CGCAGACTGC CGCCGCCGCC
2641 GCCATGGAGA CAAGGCAAAT GGCAGTGGA CAAACCACTG GTGCGGTCAC CAACCAAACG
2701 GAAACAAGCT GGCACAGCAT AGACTGGGCC AAAGCCAACC GTGAGGTAAA GAGGCTGCAA
2761 GTGCGTATCG CAAAGGCTGT GAAGGAAGGA CGTGGGGCA AAGTGAAAGC TTTGCAATGG
2821 CTCCTGACCC ACTCGTTCTA CGGCAAAGCC CTCGCCGTGA AACGGGTAAC TGACAACCTC
2881 GGCAGCAAAA CACCTGGTGT GGACGGGATA ACCTGGTCCA CACAAGAGCA GAAAGCCCAA
2941 GCCATAAAGT CCCTCAGGAG AAGAGGCTAT AAACCCCAAC CCCTGAGGCG GGTATACATC
3001 CCGAAAGCAA ACGGCAAACA GCGCCCGCTA GGAATCCCGA CAATGAAGGA CAGGGCAATG
3061 CAGGCACTAT ATGCCCTAGC CCTAGAACCA GTCGCGGAAA CCACAGCAGA CCGGAACTCC
3121 TATGGGTTCC GCGGAGGACG ATGCATAGCC GATGCAGCGA CGCAGTGTC CATCACGCTA
3181 GCCAAAACAG ACCGTGCACA ATACGTTCTC GACGCCGATA TTGCTGGGTG CTTTGACAAC
3241 ATCAGCCATG AGTGGCTACT AGCTAACATT CCACTAGACA AAAGAATTCT ACGGAAATGG
3301 CTTAAATCTG GGTTTGTCTG GAAGCAGCAA CTCTTCCCCA TCCATGCTGG AACACCTCAG
3361 GGAGGGGTAA TCTCCCCGAT GCTTGCCAAC ATGACACTGG ATGGGATGGA AGAATTGTTA
3421 AACAAAGTTT CCAGGGCGCA CAAGGTCAA CTCATCCGAT ATGCCGACGA CTTGCTCGTA
3481 ACCGGTGAAA CGAAGGAAGT GCTCTATATT GCCGGTGCGG TAATACAAGC ATTCCTCAAG
```

FIG. 7

```

3541 GAAAGGGGCC TTACCCTATC AAAGGAAAAG ACGAAGATCG TACACATTGA AGAAGGGTTT
3601 GACTTTCTCG GATGGAACAT TCGCAAATAT GATGGGAAAC TGCTCATCAA ACCTGCGAAG
3661 AAGAACGTTA AAGCGTTCCT CAAGAAAATC CGAGACACCT TAAGAGAACT TAGGACAGCA
3721 CCCCAGGAGA TTGTGATAGA CACACTCAAC CCAATCATCA GAGGTTGGAC TAACTATCAC
3781 AAAAATCAGG CATCCAAAGA AACCTTCGTC GGAGTGGACC ACCTCATATG GCAAAAATTA
3841 TGGCGATGGG CAAGGCGCCG ACACCCAAGC AAATCTGTCC GATGGGTGAA GAGTAAGTAC
3901 TTCATCCAAA TCGGGAACAG AAAATGGATG TTCGGAATAT GGACGAAAGA CAAAAACGGA
3961 GACCCGTGGG CCAAGCATTT AATCAAAGCC TCGGAAATCC GAATCCAACG TCGCGGTAAA
4021 ATCAAGGCAG ACGCCAACCC GTTTCTCCCA GAATGGGCAG AATACTTTGA GCAGCGCAAG
4081 AAATCAAAG AGGCCCTGTC CCAATACCGG CGCACCCGTC GGAATTTGTG GAAGAAACAA
4141 GCGGCGCATCT GTCCAGTATG TGGGGGAGAA ATTGAGCAAG ACATGCTCAC CGAAATCCAC
4201 CACATACTGC CCAAACACAA GGGTGGTACT GACGACCTGG ACAATCTTGT CCTAATCCAC
4261 ACTAACTGCC ACAAACAGGT GCACAACCGA GATGGTCAGC ACAGCCGGTT CCTCTTGAAG
4321 GAGGGGCTTT GACTGCAGGC AAGCTTGGCA CTGGCCGTCG TTTTACAACG TCGTGACTGG
4381 GAAAACCCCTG GCGTTACCCA ACTTAATCGC CTTGCAGCAC ATCCCCCTTT CGCCAGCTGG
4441 CGTAATAGCG AAGAGGCCCG CACCGATCGC CCTTCCCAAC AGTTGCGCAG CCTGAATGGC
4501 GAATGGCAGC TTGGCTGTTT TGGCGGATGA GATAAGATTT TCAGCCTGAT ACAGATTAAA
4561 TCAGAACGCA GAAGCGGTCT GATAAAACAG AATTTGCCTG GCGGCAGTAG CGCGGTGGTC
4621 CCACCTGACC CCATGCCGAA CTCAGAAGTG AAACGCCGTA GCGCCGATGG TAGTGTGGGG
4681 TCTCCCCATG CGAGAGTAGG GAACTGCCAG GCATCAAATA AAACGAAAGG CTCAGTCGAA
4741 AGACTGGGCC TTTCGTTTTA TCTGTTGTTT GTCGGTGAAC GCTCTCCTGA GTAGGACAAA
4801 TCCGCCGGGA GCGGATTTGA ACGTTGCGAA GCAACGGCCC GGAGGGTGGC GGGCAGGACG
4861 CCCGCCATAA ACTGCCAGGC ATCAAATTA GCAGAAGGCC ATCCTGACGG ATGGCCTTTT
4921 TCGGTTTCTA CAAACTCTTT TTGTTTATTT TTCTAAATAC ATTCAAATAT GTATCCGCTC
4981 ATGAGACAAT AACCCTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTATT
5041 CAACATTTCC GTGTCGCCCT TATTCCCTTT TTTGCGGCAT TTTGCCTTCC TGTTTTTGCT
5101 CACCCAGAAA CGCTGGTGAA AGTAAAAGAT GCTGAAGATC AGTTGGGTGC ACGAGTGGGT
5161 TACATCGAAC TGGATCTCAA CAGCGGTAAG ATCCTTGAGA GTTTTCGCCC CGAAGAACGT
5221 TCTCCAATGA TGAGCACTTT TAAAGTTCTG CTATGTGGCG CGGTATTATC CCGTGTGAC
5281 GCCGGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACTT GGTGAGTAC
5341 TCACCAGTCA CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGAGAATT ATGCAGTGCT
5401 GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT CGGAGGACCG
5461 AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG
5521 GAACCGGAGC TGAATGAAGC CATAACAAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA
5581 ATGGCAACAA CGTTGCGCAA ACTATTAAC TGGCGAACTAC TTACTCTAGC TTCCCGGCAA
5641 CAATTAATAG ACTGGATGGA GGCGGATAAA GTTGCAAGAC CACTTCTGCG CTCGGCCCTT
5701 CCGGCTGGCT GGTTTATTGC TGATAAATCT GGAGCCGGTG AGCGTGGGTC TCGCGGTATC
5761 ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC TCCCGTATCG TAGTTATCTA CACGACGGGG
5821 AGTCAGGCAA CTATGGATGA ACGAAATAGA CAGATCGCTG AGATAGGTGC CTCACTGATT
5881 AAGCATTTGGT AACTGTCAGA CCAAGTTTAC TCATATATAC TTTAGATTGA TTTACCCCGG
5941 TTGATAATCA GAAAAGCCCC AAAAACAGGA AGATTGTATA AGCAAATATT TAAATTGTAA
6001 ACGTTAATAT TTTGTTAAAA TTCGCGTTAA ATTTTGTGTA AATCAGCTCA TTTTTTAACC
6061 AATAGGCCGA AATCGGCAAA ATCCCTTATA AATCAAAGA ATAGACCGAG ATAGGGTTGA
6121 GTGTTGTTCC AGTTTGGAAC AAGAGTCCAC TATTAAAGAA CGTGGACTCC AACGTCAAAG
6181 GCGGAAAAAC CGTCTATCAG GGCGATGGCC CACTACGTGA ACCATCACCC AAATCAAGTT
6241 TTTTGGGGTC GAGGTGCCGT AAAGCACTAA ATCGGAACCC TAAAGGGAGC CCCCATTATA
6301 GAGCTTGACG GGGAAAGCCG GCGAACGTGG CGAGAAAGGA AGGGAAGAAA GCGAAAGGAG
6361 CGGGCGCTAG GCGCTGGCA AGTGTAGCGG TCACGCTGCG CGTAACCACC ACACCCCGCG
6421 CGCTTAATGC GCCGCTACAG GGCGCGTAAA AGGATCTAGG TGAAGATCCT TTTTGATAAT
6481 CTCATGACCA AAATCCCTTA ACGTGAGTTT TCGTTCCACT GAGCGTCAGA CCCCCTAGAA
6541 AAGATCAAAG GATCTTCTTG AGATCCTTTT TTTCTGCGCG TAATCTGCTG CTTGCAAACA
6601 AAAAAACCAC CGCTACCAGC GGTGGTTTGT TTGCCGGATC AAGAGCTACC AACTCTTTTT
6661 CCGAAGGTAA CTGGCTTCAG CAGAGCGCAG ATACCAAATA CTGTCCTTCT AGTGTAGCCG
6721 TAGTTAGGCC ACCACTTCAA GAACTCTGTA GCACCGCCTA CATACCTCGC TCTGCTAATC
6781 CTGTTACCAG TGGCTGCTGC CAGTGGCGAT AAGTCGTGTC TTACCGGGTT GGAATCAAGA
6841 CGATAGTTAC CGGATAAGGC GCAGCGGTCG GGCTGAACGG GGGGTTCTGT CACACAGCCC
6901 AGCTTGGAGC GAACGACCTA CACCGAACTG AGATACCTAC AGCGTGAGCT ATGAGAAAGC
6961 GCCACGCTTC CCGAAGGGAG AAAGGCGGAC AGGTATCCGG TAAGCGGCAG GGTCCGAACA
7021 GGAGAGCGCA CGAGGGAGCT TCCAGGGGGA AACGCCTGGT ATCTTTATAG TCCTGTCGGG
7081 TTTCGCCACC TCTGACTTGA GCGTCGATTT TTGTGATGCT CGTCAGGGGG GCGGAGCCTA
7141 TGGAAAAACG CCAGCAACGC GGCCTTTTTA CGGTTCCCTG CCTTTTGCTG GCCTTTTGCT

```

FIG. 7 (cont.)

8/24

7201 CACATGTTCT TTCCTGCGTT ATCCCCTGAT TCTGTGGATA ACCGTATTAC CGCCTTTGAG
7261 TGAGCTGATA CCGCTCGCCG CAGCCGAACG ACCGAGCGCA GCGAGTCAGT GAGCGAGGAA
7321 GCGGAAGAGC GCCTGATGCG GTATTTTCTC CTTACGCATC TGTGCGGTAT TTCACACCGC
7381 ATATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC ATAGTTAAGC CAGTATACAC
7441 TCCGCTATCG CTACGTGACT GGGTCATGGC TGCGCCCCGA CACCCGCCAA CACCCGCTGA
7501 CGCGCCCTGA CGGGCTTGTC TGCTCCCGGC ATCCGCTTAC AGACAAGCTG TGACCGTCTC
7561 CGGGAGCTGC ATGTGTCAGA GGTTTTCACC GTCATCACCG AAACGCGCGA GGCAGCTGCG
7621 GTAAAGCTCA TCAGCGTGGT CGTGCAGCGA TTCACAGATG TCTGCCTGTT CATCCGCGTC
7681 CAGCTCGTTG AGTTTCTCCA GAAGCGTTAA TGTCTGGCTT CTGATAAAGC GGGCCATGTT
7741 AAGGGCGGTT TTTTCCTGTT TGGTCACTGA TGCCTCCGTG TAAGGGGGAT TTCTGTTTCA
7801 GGGGGTAATG ATACCGATGA AACGAGAGAG GATGCTCACG ATACGGGTTA CTGATGATGA
7861 ACATGCCCCG TTA CTGGAAC GTTGTGAGGG TAAACAAC TG GCGGTATGGA TGC GCGCGGGA
7921 CCAGAGAAAA ATCACTCAGG GTCAATGCCA GCGCTTCGTT AATACAGATG TAGGTGTTCC
7981 ACAGGGTAGC CAGCAGCATC CTGCGATGCA GATCCGGAAC ATAATGGTGC AGGGCGCTGA
8041 CTTCCGCGTT TCCAGACTTT ACGAAACACG GAAACCGAAG ACCATTTCATG TTGTTGCTCA
8101 GGTCGCAGAC GTTTTGCAGC AGCAGTCGCT TCACGTTTCG TCGCGTATCG GTGATTTCATT
8161 CTGCTAACCA GTAAGGCAAC CCCGCCAGCC TAGCCGGGTC CTCAACGACA GGAGCACGAT
8221 CATGCGCACC CGTGGCCAGG ACCCAACGCT GCCCGAAATT

FIG. 7 (cont.)

9/24

SEQ ID NO: 15

```
1 CCGACACCAT CGAATGGTGC AAAACCTTTC GCGGTATGGC ATGATAGCGC CCGGAAGAGA
61 GTC AATTCAG GGTGGTGAAT GTGAAACCAG TAACGTTATA CGATGTCGCA GAGTATGCCG
121 GTGTCTCTTA TCAGACCGTT TCCCGCGTGG TGAACCAGGC CAGCCACGTT TCTGCGAAAA
181 CGCGGGAAAA AGTGGAAGCG GCGATGGCGG AGCTGAATTA CATTCCCAAC CGCGTGGCAC
241 AACAACTGGC GGGCAAACAG TCGTTGCTGA TTGGCGTTGC CACCTCCAGT CTGGCCCTGC
301 ACGCGCCGTC GCAAATTGTC GCGGCGATTA AATCTCGCGC CGATCAACTG GGTGCCAGCG
361 TGGTGGTGTG GATGGTAGAA CGAAGCGGCG TCGAAGCCTG TAAAGCGGCG GTGCACAATC
421 TTCTCGCGCA ACGGTCAGT GGGCTGATCA TTAAC TATCC GCTGGATGAC CAGGATGCCA
481 TTGCTGTGGA AGCTGCCTGC ACTAATGTTT CCGCGTTATT TCTTGATGTC TCTGACCAGA
541 CACCCATCAA CAGTATTATT TTCTCCCATG AAGACGGTAC GCGACTGGGC GTGGAGCATC
601 TGGTCGCATT GGGTCACCAG CAAATCGCGC TGTTAGCGGG CCCATTAAGT TCTGTCTCGG
661 CGCGTCTGCG TCTGGCTGGC TGGCATAAAT ATCTCACTCG CAATCAAATT CAGCCGATAG
721 CGGAACGGGA AGGCGACTGG AGTGCCATGT CCGGTTTTCA ACAAACCATG CAAATGCTGA
781 ATGAGGGCAT CGTTCCCACT GCGATGCTGG TTGCCAACGA TCAGATGGCG CTGGGCGCAA
841 TGCGCGCCAT TACCGAGTCC GGGCTGCGCG TTGGTGCGGA TATCTCGGTA GTGGGATACG
901 ACGATACCGA AGACAGCTCA TGTTATATCC CGCCGTTAAC CACCATCAA CAGGATTTTC
961 GCCTGCTGGG GCAAACCAGC GTGGACCGCT TGCTGCAACT CTCTCAGGGC CAGGCGGTGA
1021 AGGGCAATCA GCTGTTGCCC GTCTCACTGG TGAAAAGAAA AACCACCCTG GCGCCCAATA
1081 CGAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT GCAGCTGGCA CGACAGGTTT
1141 CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG TAAGTTAGCT CACTCATTAG
1201 GCACAATTCT CATGTTTGAC AGCTTATCAT CGACTGCACG GTGCACCAAT GCTTCTGGCG
1261 TCAGGCAGCC ATCGGAAGCT GTGGTATGGC TGTGCAGGTC GTAAATCACT GCATAATTCTG
1321 TGTCGCTCAA GGCGCACTCC CGTTCTGGAT AATGTTTTTT GCGCCGACAT CATAACGGTT
1381 CTGGCAAATA TTCTGAAATG AGCTGTTGAC AATTAATCAT CGGCTCGTAT AATGTGTGGA
1441 ATTGTGAGCG GATAACAATT TCACACAGGA AACAGCCAGT CCGTTTAGGT GTTTTCACGA
1501 GCACTTCACC AACAAGGACC ATAGCATATG AAAATCGAAG AAGGTAAACT GGTAATCTGG
1561 ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT
1621 ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT
1681 GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC
1741 GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT
1801 CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT
1861 GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA
1921 GAGATCCCGG CGCTGGATAA AGAACTGAAA GCGAAAGGTA AGAGCGCGCT GATGTTCAAC
1981 CTGCAAGAAC CGTACTTCAC CTGGCCGCTG ATTGCTGCTG ACGGGGGTTA TCGTTCAAG
2041 TATGAAAACG GCAAGTACGA CATTAAAGAC GTGGGCGTGG ATAACGCTGG CGCGAAAGCG
2101 GGTCTGACCT TCCTGGTTGA CCTGATTAAA AACAAACACA TGAATGCAGA CACCGATTAC
2161 TCCATCGCAG AAGCTGCCTT TAATAAAGGC GAAACAGCGA TGACCATCAA CGGCCCGTGG
2221 GCATGGTCCA ACATCGACAC CAGCAAAGTG AATTATGGTG TAACGGTACT GCCGACCTTC
2281 AAGGGTCAAC CATCCAAACC GTTCGTTGGC GTGCTGAGCG CAGGTATTAA CGCCGCCAGT
2341 CCGAACAAAG AGCTGGCAAA AGAGTTCCTC GAAAAC TATC TGCTGACTGA TGAAGGTCTG
2401 GAAGCGGTTA ATAAAGACAA ACCGCTGGGT GCCGTAGCGC TGAAGTCTTA CGAGGAAGAG
2461 TTGGCGAAAG ATCCACGTAT TGCCGCCACT ATGGAAAACG CCCAGAAAGG TGAAATCATG
2521 CCGAACATCC CGCAGATGTC CGCTTTCTGG TATGCCGTGC GTACTGCGGT GATCAACGCC
2581 GCCAGCGGTC GTCAGACTGT CGATGCCGCC CTGGCCGCCG CGCAGACTGC CGCCGCCGCC
2641 GCCATGGAGA CAAGGCAAAT GGCAGTGGA CAAACCACTG GTGCGGTCAC CAACCAAACG
2701 GAAACAAGCT GGCACAGCAT AGACTGGGCC AAAGCCAACC GTGAGGTAAA GAGGCTGCAA
2761 GTGCGTATCG CAAAGGCTGT GAAGGAAGGA CGCTGGGGCA AAGTGAAAGC TTTGCAATGG
2821 CTCCTGACCC ACTCGTTCTA CGGCAAAGCC CTCGCCGTGA AACGGGTAAC TGACAAC TCG
2881 GGCAGCAAAA CACCTGGTGT GGACGGGATA ACCTGGTCCA CACAAGAGCA GAAAGCCCAA
2941 GCCATAAAGT CCCTCAGGAG AAGAGGCTAC AAACCCCAAC CCCTGAGGCG GGTATACATC
3001 CCGAAAGCAA GCGGCAAGCA GCGCCCGCTA GGAATCCCGA CAACGAAGGA CAGGGCAATG
3061 CAGGCATTAT ATGCCCTAGC TCTAGAACCT GTCGCGGAAA CCACAGCGGA TCGGAACTCA
3121 TACGGGTTCC GTCAAGGACG GTGCACGGCA GATGCTGCCG GGCAGTGCTT CACTGTGCTA
3181 GGCCGATCTG ACTGTGCAAA ATATATCCTT GATGCTGACA TCACCGGATG CTTTGACAAC
3241 ATTAGCCACG AATGGCTACT AGACAACATC CCGCTGGACA AAGAGGTTCT GCGGAAGTGG
3301 CTTAAATCTG GGTTCGTCTG GAAACAGCAA CTCTTCCCAA CCCATGCTGG GACACCTCAG
3361 GGAGGGGTAA TCTCCCCAAT GCTGGCCAAT ATGACCCTAG ATGGGATGGA AGAATTGCTG
3421 AAGAAACACC TCAGAAAACA AAAAGTCAAC CTCATACGAT ATGCAGACGA CTTTGTCGTA
3481 ACTGGTGAAT CAAAGGAAAC CTTGGAAAAG GTTACAAC TG TAATCCAAGA ATTCCTCAAG
```

FIG. 8

10/24

3541 GAAAGGGGCC TTACCCTATC AGAAGAAAAG ACAAAGGTCG TTCATATCGA AGAAGGATTT
3601 GACTTTCTTG GATGGAACAT TCGCAAATAT GGTGAGAAGC TTCTCATCAA ACCTGCGAAG
3661 AAGAACATCA AGGCGTTCCA CAAGAAAATC CGAGACGCAC TGAAGGAACT CAGAACAGCC
3721 ACCCAGGAAG CTGTGATAGA CACACTCAAC CCAATTATCA AAGGCTGGGC TAACTATCAC
3781 AGAAACCAGG TTTCCAAAAG AATCTTCAAC AGAGCGGATG ACAATATCTG GCATAAATTA
3841 TGGCGATGGG CAAAACGTCG GCACCCAAAC AAACCAGCCC GATGGACAAA GAACAAATAC
3901 TTCATCAAAA TCGGGAATAG GCACTGGGTG TTTGGCACAT GGAAAAAGGA CAAAGAGGGA
3961 AGGTTACGGT CCAGATACCT AATTAAAGCC GGAGATACTC GAATCCAACG TCATGTCAAA
4021 ATCAAGGCAG ACGCCAATCC GTTTCTCCCA GAGTGGGCAG AATACTTTGA GGAACGCAAG
4081 AAACCTCAAAG AAGCCCCGTC TCAATATCGG CGCATCCGCC GAGAACTATG GAAGAAACAG
4141 GGTGGTATCT GTCCAGTATG CGGGGGTGAA ATTGAGCAAG ACATGCTCAC TGAAATCCAC
4201 CACATATTGC CCAAACACAA GGGTGGTTCT GACGACCTGG ATAATCTTGT CTTAATCCAC
4261 GCCAACTGTC ACAAACAGGT GCACAGCCGA GACGGTCAGC ACAGCCGGTT CCTCTTGAAA
4321 GAGGGGCTTT GACTGCAGGC AAGCTTGGA CTGGCCGTCG TTTTACAACG TCGTGACTGG
4381 GAAAACCCTG GCGTTACCCA ACTTAATCGC CTTGCAGCAC ATCCCCCTTT CGCCAGCTGG
4441 CGTAATAGCG AAGAGGCCCG CACCGATCGC CCTTCCCAAC AGTTGCGCAG CCTGAATGGC
4501 GAATGGCAGC TTGGCTGTTT TGGCGGATGA GATAAGATTT TCAGCCTGAT ACAGATTAAA
4561 TCAGAACGCA GAAGCGGTCT GATAAAACAG AATTTGCCTG GCGGCAGTAG CGCGGTGGTC
4621 CCACCTGACC CCATGCCGAA CTCAGAAGTG AAACGCCGTA GCGCCGATGG TAGTGTGGGG
4681 TCTCCCCATG CGAGAGTAGG GAACTGCCAG GCATCAAATA AAACGAAAGG CTCAGTCGAA
4741 AGACTGGGCC TTTCGTTTTA TCTGTTGTTT GTCGGTGAAC GCTCTCCTGA GTAGGACAAA
4801 TCCGCCGGGA GCGGATTTGA ACGTTGCGAA GCAACGGCCC GGAGGGTGGC GGGCAGGACG
4861 CCCGCCATAA ACTGCCAGGC ATCAAATTAA GCAGAAGGCC ATCCTGACGG ATGGCCTTTT
4921 TCGGTTTCTA CAAACTCTTT TTGTTTATTT TTCTAAATAC ATTCAAATAT GTATCCGCTC
4981 ATGAGACAAT AACCCTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTATT
5041 CAACATTTCC GTGTCGCCCT TATTCCCTTT TTTGCGGCAT TTTGCCTTCC TGTTTTTGCT
5101 CACCCAGAAA CGCTGGTGAA AGTAAAAGAT GCTGAAGATC AGTTGGGTGC ACGAGTGGGT
5161 TACATCGAAC TGGATCTCAA CAGCGGTAAG ATCCTTGAGA GTTTTCGCCC CGAAGAACGT
5221 TCTCCAATGA TGAGCACTTT TAAAGTTCTG CTATGTGGCG CGGTATTATC CCGTGTGAC
5281 GCCGGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACTT GGTGAGTAC
5341 TCACCAGTCA CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGAGAATT ATGCAGTGCT
5401 GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT CGGAGGACCG
5461 AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG
5521 GAACCGGAGC TGAATGAAGC CATAACAAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA
5581 ATGGCAACAA CGTTGCGCAA ACTATTAACT GGCGAACTAC TTACTCTAGC TTCCCGGCAA
5641 CAATTAATAG ACTGGATGGA GGCGGATAAA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT
5701 CCGGCTGGCT GGTTTATTGC TGATAAATCT GGAGCCGGTG AGCGTGGGTC TCGCGGTATC
5761 ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC TCCCGTATCG TAGTTATCTA CACGACGGGG
5821 AGTCAGGCAA CTATGGATGA ACGAAATAGA CAGATCGCTG AGATAGGTGC CTCACTGATT
5881 AAGCATTGGT AACTGTCAGA CCAAGTTTAC TCATATATAC TTTAGATTGA TTTACCCCGG
5941 TTGATAATCA GAAAAGCCCC AAAAACAGGA AGATTGTATA AGCAAATATT TAAATTGTAA
6001 ACGTTAATAT TTTGTTAAAA TTCGCGTTAA ATTTTTGTTA AATCAGCTCA TTTTTTAACC
6061 AATAGGCCGA AATCGGCAAA ATCCCTTATA AATCAAAAGA ATAGACCGAG ATAGGGTTGA
6121 GTGTTGTTCC AGTTTGGAAC AAGAGTCCAC TATTAAAGAA CGTGGACTCC AACGTCAAAG
6181 GGCGAAAAAC CGTCTATCAG GCGGATGGCC CACTACGTGA ACCATCACCC AAATCAAGTT
6241 TTTTGGGGTC GAGGTGCCGT AAAGCACTAA ATCGGAACCC TAAAGGGAGC CCCCATTATA
6301 GAGCTTGACG GGGAAAGCCG GCGAACGTGG CGAGAAAGGA AGGGAAGAAA GCGAAAGGAG
6361 CGGGCGCTAG GCGCTGGCA AGTGTAGCGG TCACGCTGCG CGTAACCACC ACACCCGCCG
6421 CGCTTAATGC GCCGCTACAG GGCGCGTAAA AGGATCTAGG TGAAGATCCT TTTTGATAAT
6481 CTCATGACCA AAATCCCTTA ACGTGAGTTT TCGTTCCACT GAGCGTCAGA CCCCCTAGAA
6541 AAGATCAAAG GATCTTCTTG AGATCCTTTT TTTCTGCGCG TAATCTGCTG CTTGCAAACA
6601 AAAAAACCAC CGCTACCAGC GGTGGTTTGT TTGCCGGATC AAGAGCTACC AACTCTTTTT
6661 CCGAAGGTAA CTGGCTTCAG CAGAGCGCAG ATACCAAATA CTGTCCTTCT AGTGAGCCG
6721 TAGTTAGGCC ACCACTTCAA GAACTCTGTA GCACCGCCTA CATACCTCGC TCTGCTAATC
6781 CTGTTACCAG TGGCTGCTGC CAGTGGCGAT AAGTCGTGTC TTACCGGGTT GGAATCAAGA
6841 CGATAGTTAC CGGATAAGGC GCAGCGGTCG GGCTGAACGG GGGGTTCGTG CACACAGCCC
6901 AGCTTGAGC GAACGACCTA CACCGAAGT AGATACCTAC AGCGTGAGCT ATGAGAAAGC
6961 GCCACGCTTC CCGAAGGGAG AAAGGCGGAC AGGTATCCGG TAAGCGGCAG GGTGGAACA
7021 GGAGAGCGCA CGAGGGAGCT TCCAGGGGGA AACGCTGGT ATCTTTATAG TCCTGTCGGG
7081 TTTCGCCACC TCTGACTTGA GCGTCGATTT TTGTGATGCT CGTCAGGGGG GCGGAGCCTA
7141 TGGAAAAACG CCAGCAACGC GGCCTTTTTA CGGTTCTTGG CTTTTTGCTG GCCTTTTGCT

FIG. 8 (cont.)

7201 CACATGTTCT TTCCTGCGTT ATCCCCTGAT TCTGTGGATA ACCGTATTAC CGCCTTTGAG
7261 TGAGCTGATA CCGCTCGCCG CAGCCGAACG ACCGAGCGCA GCGAGTCAGT GAGCGAGGAA
7321 GCGGAAGAGC GCCTGATGCG GTATTTTCTC CTTACGCATC TGTGCGGTAT TTCACACCGC
7381 ATATATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC ATAGTTAAGC CAGTATACAC
7441 TCCGCTATCG CTACGTGACT GGGTCATGGC TGCGCCCCGA CACCCGCCAA CACCCGCTGA
7501 CGCGCCCTGA CGGGCTTGTC TGCTCCCGGC ATCCGCTTAC AGACAAGCTG TGACCGTCTC
7561 CGGGAGCTGC ATGTGTCAGA GGTTTTTACC GTCATCACCG AAACGCGCGA GGCAGCTGCG
7621 GTAAAGCTCA TCAGCGTGGT CGTGCAGCGA TTCACAGATG TCTGCCTGTT CATCCGCGTC
7681 CAGCTCGTTG AGTTTCTCCA GAAGCGTTAA TGTCTGGCTT CTGATAAAGC GGGCCATGTT
7741 AAGGGCGGTT TTTTCCTGTT TGGTCACTGA TGCCTCCGTG TAAGGGGGAT TTCTGTTTCA
7801 GGGGGTAATG ATACCGATGA AACGAGAGAG GATGCTCACG ATACGGGTTA CTGATGATGA
7861 ACATGCCCCG TTA CTGGAAC GTTGTGAGGG TAAACAAC TG GCGGTATGGA TCGGCGGGGA
7921 CCAGAGAAAA ATCACTCAGG GTCAATGCCA GCGCTTCGTT AATACAGATG TAGGTGTTCC
7981 ACAGGGTAGC CAGCAGCATC CTGCGATGCA GATCCGGAAC ATAATGGTGC AGGGCGCTGA
8041 CTTCCGCGTT TCCAGACTTT ACGAAACACG GAAACCGAAG ACCATTTCATG TTGTTGCTCA
8101 GGTCGCAGAC GTTTTGCAGC AGCAGTCGCT TCACGTTTCG TCGCGTATCG GTGATTTCATT
8161 CTGCTAACCA GTAAGGCAAC CCCGCCAGCC TAGCCGGGTC CTCAACGACA GGAGCACGAT
8221 CATGCGCACC CGTGGCCAGG ACCCAACGCT GCCCGAAATT

FIG. 8 (cont.)

12/24

SEQ ID NO: 16

```
1 CCGACACCAT CGAATGGTGC AAAACCTTTC GCGGTATGGC ATGATAGCGC CCGGAAGAGA
61 GTCAATTCAG GGTGGTGAAT GTGAAACCAG TAACGTTATA CGATGTCGCA GAGTATGCCG
121 GTGTCTCTTA TCAGACCGTT TCCCGCGTGG TGAACCAGGC CAGCCACGTT TCTGCGAAAA
181 CGCGGGAAAA AGTGGAAGCG GCGATGGCGG AGCTGAATTA CATTCCCAAC CGCGTGGCAC
241 AACAACTGGC GGGCAAACAG TCGTTGCTGA TTGGCGTTGC CACCTCCAGT CTGGCCCTGC
301 ACGCGCCGTC GCAAATTGTC GCGGCGATTA AATCTCGCGC CGATCAACTG GGTGCCAGCG
361 TGGTGGTGTG GATGGTAGAA CGAAGCGGCG TCGAAGCCTG TAAAGCGGCG GTGCACAATC
421 TTCTCGCGCA ACGCGTCAGT GGGCTGATCA TTAACATATC GCTGGATGAC CAGGATGCCA
481 TTGCTGTGGA AGCTGCCTGC ACTAATGTTC CGGCGTTATT TCTTGATGTC TCTGACCAGA
541 CACCCATCAA CAGTATTATT TTCTCCCATG AAGACGGTAC GCGACTGGGC GTGGAGCATC
601 TGGTCGCATT GGGTCACCAG CAAATCGCGC TGTTAGCGGG CCCATTAAGT TCTGTCTCGG
661 CGCGTCTGCG TCTGGCTGGC TGGCATAAAT ATCTCACTCG CAATCAAATT CAGCCGATAG
721 CGGAACGGGA AGGCGACTGG AGTGCCATGT CCGGTTTTCA ACAAACCATG CAAATGCTGA
781 ATGAGGGCAT CGTTCCCACT GCGATGCTGG TTGCCAACGA TCAGATGGCG CTGGGCGCAA
841 TGCGCGCCAT TACCGAGTCC GGGCTGCGCG TTGGTGCGGA TATCTCGGTA GTGGGATACG
901 ACGATACCGA AGACAGCTCA TGTATATATC CGCCGTTAAC CACCATCAAA CAGGATTTTC
961 GCCTGCTGGG GCAAACCAGC GTGGACCGCT TGCTGCAACT CTCTCAGGGC CAGGCGGTGA
1021 AGGGCAATCA GCTGTTGCCC GTCTCACTGG TGAAAAGAAA AACCACCCTG GCGCCCAATA
1081 CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT GCAGCTGGCA CGACAGGTTT
1141 CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG TAAGTTAGCT CACTCATTAG
1201 GCACAATTCT CATGTTTGAC AGCTTATCAT CGACTGCACG GTGCACCAAT GCTTCTGGCG
1261 TCAGGCAGCC ATCGGAAGCT GTGGTATGGC TGTGCAGGTC GTAAATCACT GCATAATTCTG
1321 TGTGCTCAA GGCGCACTCC CGTTCTGGAT AATGTTTTTT GCGCCGACAT CATAACGGTT
1381 CTGGCAAATA TTCTGAAATG AGCTGTTGAC AATTAATCAT CGGCTCGTAT AATGTGTGGA
1441 ATTGTGAGCG GATAACAATT TCACACAGGA AACAGCCAGT CCGTTTAGGT GTTTTCACGA
1501 GCACTTCACC AACAAAGGACC ATAGCATATG AAAATCGAAG AAGGTAAACT GGTAATCTGG
1561 ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT
1621 ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT
1681 GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC
1741 GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT
1801 CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT
1861 GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA
1921 GAGATCCCGG CGCTGGATAA AGAACTGAAA GCGAAAGGTA AGAGCGCGCT GATGTTCAAC
1981 CTGCAAGAAC CGTACTTCAC CTGGCCGCTG ATTGCTGCTG ACGGGGGTTA TCGTTCAAG
2041 TATGAAAACG GCAAGTACGA CATTAAAGAC GTGGGCGTGG ATAACGCTGG CGCGAAAGCG
2101 GGTCTGACCT TCCTGGTTGA CCTGATTAAA AACAAACACA TGAATGCAGA CACCGATTAC
2161 TCCATCGCAG AAGCTGCCTT TAATAAAGGC GAAACAGCGA TGACCATCAA CGGCCCGTGG
2221 GCATGGTCCA ACATCGACAC CAGCAAAGTG AATTATGGTG TAACGGTACT GCCGACCTTC
2281 AAGGGTCAAC CATCCAAACC GTTCGTTGGC GTGCTGAGCG CAGGTATTAA CGCCGCCAGT
2341 CCGAACAAAG AGCTGGCAAA AGAGTTCCTC GAAAACTATC TGCTGACTGA TGAAGGTCTG
2401 GAAGCGGTTA ATAAAGACAA ACCGCTGGGT GCCGTAGCGC TGAAGTCTTA CGAGGAAGAG
2461 TTGGCGAAAG ATCCACGTAT TGCCGCCACT ATGGAAAACG CCCAGAAAGG TGAAATCATG
2521 CCGAACATCC CGCAGATGTC CGCTTCTGG TATGCCGTGC GTACTGCGGT GATCAACGCC
2581 GCCAGCGGTC GTCAGACTGT CGATGCCGCC CTGGCCGCCG CGCAGACTGC CGCCGCCGCC
2641 GCCATGAAGG TAAACAAACT TGTCGTAAAA AGCGAACAGG ACTTGAGAAA CTGCTTGAT
2701 CTTCTTTTATC AAGAAGCTAA AAAGGGAAAA CATTTTTACG GCATGCTTGA GTTGCTTCAA
2761 AATGATGTTG TCATTTTAGA AGCTATTCGC AATATTAAAA GCAATAAAGG TAGCAAAACG
2821 GCGGGGATTG ATCAGAAAAT AGTAGATGAT TATTTGCTTA TGCCAACGGA AAAGGTTTTTC
2881 GGGATGATAA AAGCCAAACT CAATGACTAT AAGCCTATAC CAGTGAGAAG GTGCAACAAG
2941 CCCAAAGGAA ATGCCAAAAG CTCAAAAAGA AAAGGCAATA GTCCGAATGA GGAAGGGGAA
3001 ACGAGGCCCT TAGGAATATC CGCAGTGACG GATAGAATCA TCCAAGAGAT GCTACGGATA
3061 GTGCTCGAGC CGATTTTCGA AGCCCAATTC TATCCGCACA GTTATGGGTT CAGACCGTAT
3121 CGCTCCACCG AACATGCCTT AGCCTGGATG CTGAAAATCA TCAACGGAAG CAAACTGTAT
3181 TGGGTTGTAA AAGGTGACAT TGAAAGTTAT TTTGATCACA TCAATCATAA GAAGCTTCTG
3241 AACATCATGT GGAATATGGG CGTTAGGGAT AAACGGGTAC TATGCATCGT TAAGAAAATG
3301 CTGAAGGCGG GGCAAGTGAT ACAAGGTAAA TTCTATCCAA CCGCTAAGGG GATTCCCTCAG
3361 GGAGGAATTA TTAGCCCGTT GTTGGCTAAT GTATATCTCA ACAGCTTTGA CTGGATGGTT
3421 GGCCAAGAAT ATGAGTATCA CCCTAATAAC GCAAATATC GGGAAAAGAA AAACGCATTA
3481 GCGGCGTTAA GGAACAAGGG ACATCATCCC GTCTTTTACA TTCGTTATGC TGATGATTGG
```

FIG. 9

13/24

3541 GTTATTCTTA CGGATACGAA AGAATATGCG GAAAAAATAA GGGAGCAATG TAAGCAGTAT
3601 TTAGCCTGTG AGTTGCACTT AACTCTATCG GATGAGAAAA CGTTCATTGC AGATATCCGC
3661 GAACAACGGG TTAAGTTTCT AGGCTTTTGT ATTGAGGCAG GAAAGCGGCG TTTTCATAAA
3721 AAAGGATTCT CCGCTAGAAT GATTCCCGAT ATGGAAAAAG TCAATGCCAA GGTCAAAGAA
3781 ATTAAGCGCG ATATTGATT GTTAAGAACG AGAAAATCGG AATTAGAGAA AGCCCTTGAT
3841 ATTGAAAACA TTAACACCAA AATTATAGGA TTAGCCAATC ATCTAAAAAT AGGCATTTCC
3901 AAGTACATTA TGGGCAAAGT AGATCGCGTC ATTGAAGAGA CAGCCTACCG CACCTGGGTT
3961 AAAATGTATG GGAAAGAAAA AGCGGCGCAA TATAAAAGGC CTGTGTCAGA GTTTCACAAT
4021 CGGATTGACA GACATAAAGG CTATCAAATG AAACATTTTT CTGTGTCAC AGAGGATGGC
4081 ATAAGAGTAG GGATTACCCA TGCAAAAATA ACGCCTATAC AGTATGCAAC AGTATTTCAA
4141 CAAGAAATGA CCCCATACAC TGCAGACGGC AGAAAAATGT ATGAAGAAAA GCATAGAAAA
4201 ATACGATTGC CGGATAAAAT GAGTCTGTTC GATCACGATT CGATATTCAT CTACATTTTA
4261 TCTGAGCATA ATGATGGGAA ATATAATCTT GAATATTTCT TAAATAGGGT GAATGTATTT
4321 CACAGAGATA AAGGAAAATG CAAAATATGT GCCGTATACT TAAGTCCCGG TAACTTCCAC
4381 TGCCATCATA TTGACCCGAG TAAACCTTTA AGTGAGATCA ATAAGACCGT TAATCTAATT
4441 AGCTTATGCA ACCAATGCCA TAGGCTTGTC CATAGCAACC AAGAACCGCC GTTTACAGAA
4501 CGAAAAATGT TTGACAAACT AACGAAATAT AGGAACAAGC TGAAAATATA AGGATCCTCT
4561 AGCTGCAGGC AAGCTTGGCA CTGGCCGTCG TTTTACAACG TCGTGAAGTG GAAAACCTG
4621 GCGTTACCCA ACTTAATCGC CTTGCAGCAC ATCCCCCTTT CGCCAGCTGG CGTAATAGCG
4681 AAGAGGCCCC CACCGATCGC CCTTCCCAAC AGTTGCGCAG CCTGAATGGC GAATGGCAGC
4741 TTGGCTGTTT TGGCGGATGA GATAAGATTT TCAGCCTGAT ACAGATTAAA TCAGAACGCA
4801 GAAGCGGTCT GATAAAACAG AATTGCTCTG GCGGCAGTAG CGCGGTGGTC CCACCTGACC
4861 CCATGCCGAA CTCAGAAAGT AAACGCCGTA GCGCCGATGG TAGTGTGGGG TCTCCCCATG
4921 CGAGAGTAGG GAACTGCCAG GCATCAAATA AAACGAAAGG CTCAGTCGAA AGACTGGGCC
4981 TTTCTGTTTT TCTGTTGTTT GTCGGTGAAC GCTCTCCTGA GTAGGACAAA TCCGCCGGGA
5041 GCGGATTTGA ACGTTGCGAA GCAACGGCCC GGAGGGTGGC GGGCAGGACG CCCGCCATAA
5101 ACTGCCAGGC ATCAAATTAA GCAGAAGGCC ATCCTGACGG ATGGCCTTTT TCGGTTTCTA
5161 CAAACTCTTT TTGTTTATTT TTCTAAATAC ATTCAAATAT GTATCCGCTC ATGAGACAAT
5221 AACCTTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTATT CAACATTTCC
5281 GTGTCGCCCT TATTCCCTTT TTTGCGGCAT TTTGCCTTCC TGTTTTTGCT CACCCAGAAA
5341 CGCTGGTGAA AGTAAAAGAT GCTGAAGATC AGTTGGGTGC ACGAGTGGGT TACATCGAAC
5401 TGGATCTCAA CAGCGGTAAG ATCCTTGAGA GTTTTCGCCC CGAAGAACGT TCTCCAATGA
5461 TGAGCACTTT TAAAGTTCTG CTATGTGGCG CGGTATTATC CCGTGTGAC GCCGGGCAAG
5521 AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACTT GGTGAGTAC TCACCAGTCA
5581 CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGAGAATT ATGCAGTGCT GCCATAACCA
5641 TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT CGGAGGACCG AAGGAGCTAA
5701 CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG GAACCGGAGC
5761 TGAATGAAGC CATAACAAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA ATGGCAACAA
5821 CGTTGCGCAA ACTATTAAT TGGGAACTAC TTACTCTAGC TTCCCGGCAA CAATTAATAG
5881 ACTGGATGGA GGCGGATAAA GTTGCAAGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT
5941 GGTTTATTGC TGATAAATCT GGAGCCGGTG AGCGTGGGTC TCGCGGTATC ATTGCAGCAC
6001 TGGGGCCAGA TGGTAAGCCC TCCCGTATCG TAGTTATCTA CACGACGGGG AGTCAGGCAA
6061 CTATGGATGA ACGAAATAGA CAGATCGCTG AGATAGGTGC CTCACTGATT AAGCATTGGT
6121 AACTGTCAGA CCAAGTTTAC TCATATATAC TTTAGATTGA TTTACCCCGG TTGATAATCA
6181 GAAAAGCCCC AAAAACAGGA AGATTGTATA AGCAAATATT TAAATTGTAA ACGTTAATAT
6241 TTTGTTAAAA TTCGCGTTAA ATTTTGTGTA AATCAGCTCA TTTTTTAACC AATAGGCCGA
6301 AATCGGCAAA ATCCCTTATA AATCAAAAGA ATAGACCGAG ATAGGGTTGA GTGTGTGTTCC
6361 AGTTTGGAAC AAGAGTCCAC TATTAAAGAA CGTGGACTCC AACGTCAAAG GCGGAAAAAC
6421 CGTCTATCAG GGCGATGGCC CACTACGTGA ACCATCACCC AAATCAAGTT TTTTGGGGTC
6481 GAGGTGCCGT AAAGCACTAA ATCGGAACCC TAAAGGGAGC CCCCATTATA GAGCTTGACG
6541 GGGAAAGCCG GCGAACGTGG CGAGAAAGGA AGGGAAGAAA GCGAAAGGAG CGGGCGCTAG
6601 GGCGCTGGCA AGTGTAGCGG TCACGCTGCG CGTAACCACC ACACCCGCCG CGCTTAATGC
6661 GCCGCTACAG GGCGCGTAAA AGGATCTAGG TGAAGATCCT TTTTGATAAT CTCATGACCA
6721 AAATCCCTTA ACGTGAGTTT TCGTTCCACT GAGCGTCAGA CCCCCTAGAA AAGATCAAAG
6781 GATCTTCTTG AGATCCTTTT TTTCTGCGCG TAATCTGCTG CTTGCAAACA AAAAACCCAC
6841 CGCTACCAGC GGTGGTTTGT TTGCCGGATC AAGAGCTACC AACTCTTTTT CCGAAGGTAA
6901 CTGGCTTCAG CAGAGCGCAG ATACCAAATA CTGTCTTCTT AGTGTAGCCG TAGTTAGGCC
6961 ACCACTTCAA GAACTCTGTA GCACCGCCTA CATACCTCGC TCTGCTAATC CTGTTACCAG
7021 TGGCTGCTGC CAGTGGCGAT AAGTCGTGTC TTACCGGGTT GGAATCAAGA CGATAGTTAC
7081 CGGATAAGGC GCAGCGGTGC GGCTGAACGG GGGGTTCTGT CACACAGCCC AGCTTGGAGC
7141 GAACGACCTA CACCGAACTG AGATACCTAC AGCGTGAGCT ATGAGAAAGC GCCACGCTTC

FIG. 9 (cont.)

7201 CCGAAGGGAG AAAGGCGGAC AGGTATCCGG TAAGCGGCAG GGTCGGAACA GGAGAGCGCA
7261 CGAGGGGAGCT TCCAGGGGGA AACGCCTGGT ATCTTTATAG TCCTGTCGGG TTTCGCCACC
7321 TCTGACTTGA GCGTCGATTT TTGTGATGCT CGTCAGGGGG GCGGAGCCTA TGGAAAAACG
7381 CCAGCAACGC GGCCTTTTTTA CGGTTCCCTGG CCTTTTGCTG GCCTTTTGCT CACATGTTCT
7441 TTCCTGCGTT ATCCCCTGAT TCTGTGGATA ACCGTATTAC CGCCTTTGAG TGAGCTGATA
7501 CCGCTCGCCG CAGCCGAACG ACCGAGCGCA GCGAGTCAGT GAGCGAGGAA GCGGAAGAGC
7561 GCCTGATGCG GTATTTTCTC CTTACGCATC TGTGCGGTAT TTCACACCGC ATATATGGTG
7621 CACTCTCAGT ACAATCTGCT CTGATGCCGC ATAGTTAAGC CAGTATACAC TCCGCTATCG
7681 CTACGTGACT GGGTCATGGC TGCGCCCGGA CACCCGCCAA CACCCGCTGA CGCGCCCTGA
7741 CGGGCTTGTC TGCTCCCGGC ATCCGCTTAC AGACAAGCTG TGACCGTCTC CGGGAGCTGC
7801 ATGTGTCAGA GGTTTTCACC GTCATCACCG AAACGCGCGA GGCAGCTGCG GTAAAGCTCA
7861 TCAGCGTGGT CGTGCAGCGA TTCACAGATG TCTGCCTGTT CATCCGCGTC CAGCTCGTTG
7921 AGTTTCTCCA GAAGCGTTAA TGTCTGGCTT CTGATAAAGC GGGCCATGTT AAGGGCGGTT
7981 TTTTCCTGTT TGGTCACTGA TGCCTCCGTG TAAGGGGGAT TTCTGTTTCAAT GGGGGTAATG
8041 ATACCGATGA AACGAGAGAG GATGCTCACG ATACGGGTTA CTGATGATGA ACATGCCCGG
8101 TTA CTGGAAC GTTGTGAGGG TAAACAAC TG GCGGTATGGA TGCGGCGGGA CCAGAGAAAA
8161 ATCACTCAGG GTCAATGCCA GCGCTTCGTT AATACAGATG TAGGTGTTCC ACAGGGTAGC
8221 CAGCAGCATC CTGCGATGCA GATCCGGAAC ATAATGGTGC AGGGCGCTGA CTTCCGCGTT
8281 TCCAGACTTT ACGAAACACG GAAACCGAAG ACCATT CATG TTGTTGCTCA GGTGCGCAGAC
8341 GTTTTGCAGC AGCAGTCGCT TCACGTTTCG TCGCGTATCG GTGATT CAT CTGCTAACCA
8401 GTAAGGCAAC CCCGCCAGCC TAGCCGGGTC CTCAACGACA GGAGCACGAT CATGCGCACC
8461 CGTGGCCAGG ACCCAACGCT GCCCGAAATT

FIG. 9 (cont.)

SEQ ID NO: 17

```
1 CCGACACCAT CGAATGGTGC AAAACCTTTC GCGGTATGGC ATGATAGCGC CCGGAAGAGA
61 GTCAATTCAG GGTGGTGAAT GTGAAACCAG TAACGTTATA CGATGTCGCA GAGTATGCCG
121 GTGTCTCTTA TCAGACCGTT TCCCGCGTGG TGAACCAGGC CAGCCACGTT TCTGCGAAAA
181 CGCGGGAAAA AGTGGAAGCG GCGATGGCGG AGCTGAATTA CATTCCCAAC CGCGTGGCAC
241 AACAACTGGC GGGCAAACAG TCGTTGCTGA TTGGCGTTGC CACCTCCAGT CTGGCCCTGC
301 ACGCGCCGTC GCAAATTGTC GCGGCGATTA AATCTCGCGC CGATCAACTG GGTGCCAGCG
361 TGGTGGTGTC GATGGTAGAA CGAAGCGGCG TCGAAGCCTG TAAAGCGGCG GTGCACAATC
421 TTCTCGCGCA ACGCGTCAGT GGGCTGATCA TTAATATCC GCTGGATGAC CAGGATGCCA
481 TTGCTGTGGA AGCTGCCTGC ACTAATGTTT CCGCGTTATT TCTTGATGTC TCTGACCAGA
541 CACCCATCAA CAGTATTATT TTCTCCCATG AAGACGGTAC GCGACTGGGC GTGGAGCATC
601 TGGTCGCATT GGGTCACCAG CAAATCGCGC TGTTAGCGGG CCCATTAAGT TCTGTCTCGG
661 CGCGTCTGCG TCTGGCTGGC TGGCATAAAT ATCTCACTCG CAATCAAATT CAGCCGATAG
721 CGGAACGGGA AGGCGACTGG AGTGCCATGT CCGGTTTTCA ACAAACCATG CAAATGCTGA
781 ATGAGGGCAT CGTTCCCACT GCGATGCTGG TTGCCAACGA TCAGATGGCG CTGGGCGCAA
841 TGC GCGCCAT TACCGAGTCC GGGCTGCGCG TTGGTGCGGA TATCTCGGTA GTGGGATACG
901 ACGATACCGA AGACAGCTCA TGTATATATC CGCCGTTAAC CACCATCAAA CAGGATTTTC
961 GCCTGCTGGG GCAAACCAGC GTGGACCGCT TGCTGCAACT CTCTCAGGGC CAGGCGGTGA
1021 AGGGCAATCA GCTGTTGCCC GTCTCACTGG TGAAAAGAAA AACCACCCTG GCGCCCAATA
1081 CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT GCAGCTGGCA CGACAGGTTT
1141 CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG TAAGTTAGCT CACTCATTAG
1201 GCACAATTCT CATGTTTGAC AGCTTATCAT CGACTGCACG GTGCACCAAT GCTTCTGGCG
1261 TCAGGCAGCC ATCGGAAGCT GTGGTATGGC TGTGCAGGTC GTAAATCACT GCATAATTCG
1321 TGTCGCTCAA GGCGCACTCC CGTCTGGAT AATGTTTTTT GCGCCGACAT CATAACGGTT
1381 CTGGCAAATA TTCTGAAATG AGCTGTTGAC AATTAATCAT CGGCTCGTAT AATGTGTGGA
1441 ATTGTGAGCG GATAACAATT TCACACAGGA AACAGCCAGT CCGTTTAGGT GTTTTCACGA
1501 GCACTTCACC AACAAGGACC ATAGCATATG AAAATCGAAG AAGGTAAACT GGTAATCTGG
1561 ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT
1621 ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT
1681 GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC
1741 GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCAGGA CAAGCTGTAT
1801 CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT
1861 GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA
1921 GAGATCCCGG CGCTGGATAA AGAACTGAAA GCGAAAGGTA AGAGCGCGCT GATGTTCAAC
1981 CTGCAAGAAC CGTACTTCAC CTGGCCGCTG ATTGCTGCTG ACGGGGGTTA TCGTTC AAG
2041 TATGAAAACG GCAAGTACGA CATTAAAGAC GTGGGCGTGG ATAACGCTGG CCGGAAAGCG
2101 GGTCTGACCT TCCTGGTTGA CCTGATTAAA AACAAACACA TGAATGCAGA CACCGATTAC
2161 TCCATCGCAG AAGCTGCCTT TAATAAAGGC GAAACAGCGA TGACCATCAA CGGCCCGTGG
2221 GCATGGTCCA ACATCGACAC CAGCAAAGTG AATTATGGTG TAACGGTACT GCCGACCTTC
2281 AAGGGTCAAC CATCCAAACC GTTCGTTGGC GTGCTGAGCG CAGGTATTAA CGCCGCCAGT
2341 CCGAACAAAG AGCTGGCAAA AGAGTTCCTC GAAAACTATC TGCTGACTGA TGAAGGTCTG
2401 GAAGCGGTTA ATAAAGACAA ACCGCTGGGT GCCGTAGCGC TGAAGTCTTA CGAGGAAGAG
2461 TTGGCGAAAG ATCCACGTAT TGCCGCCACT ATGGAAAACG CCCAGAAAGG TGAAATCATG
2521 CCGAACATCC CGCAGATGTC CGTTTTCTGG TATGCCGTGC GTACTGCGGT GATCAACGCC
2581 GCCAGCGGTC GTCAGACTGT CGATGCCGCC CTGGCCGCCG CGCAGACTGC CGCCGCCGCC
2641 GCCATGGCTT TGTGGAACG CATCTTAGCG AGAGACAACC TCATCACGGC GCTCAAACGG
2701 GTCGAAGCCA ACCAAGGAGC ACCGGGAATC GACGGAGTAT CAACCGATCA ACTCCGTGAT
2761 TACATCCGCG CTCCTGGAG CACGATCCAC GCCCAACTCT TGGCGGGAAC CTACCGGCCG
2821 GCGCCTGTCC GCAGGGTCGA AATCCCGAAA CCGGGCGGCG GCACACGGCA GCTAGGCATT
2881 CCCACCGTGG TGGACCGGCT GATCCAACAA GCCATTCTTC AAGAACTCAC ACCCATTTTC
2941 GATCCAGACT TCTCCTCTTC CAGCTTCGGA TTCCGTCCGG GCCGCAACGC CCACGATGCC
3001 GTGCGGCAAG CGCAAGGCTA CATCCAGGAA GGGTATCGGT ACGTGGTCTGA CATGGACCTG
3061 GAAAAGTTCT TTGATCGGGT CAACCATGAC ATCTTGATGA GTCGGGTGGC CCGAAAAGTC
3121 AAGGATAAAC GCGTGCTGAA ACTGATCCGT GCCTACCTGC AAGCCGGCGT TATGATCGAA
3181 GGGGTGAAGG TGCAGACGGA GGAAGGGACG CCGCAAGGCG GCCCCCTCAG CCCCCTGCTG
3241 GCGAACATCC TTCTCGACGA TTTAGACAAG GAATTGGAGA AGCGAGGATT GAAATTCTGC
3301 CGTTACGCAG ATGACTGCAA CATCTATGTG AAAAGTCTGC GGGCAGGACA ACGGGTGAAA
3361 CAAAGCATCC AACGGTTCTT GGAGAAAACG CTCAAACCTA AAGTAAACGA GGAGAAAAGT
3421 GCGGTGGACC GCCCGTGGAA ACGGGCCTTT CTGGGGTTTA GCTTCACACC GGAACGAAAA
3481 GCGCGAATCC GGCTCGCCCC AAGGTCGATT CAACGTCTGA AACAGCGGAT TCGACAGCTG
```

FIG. 10

3541 ACCAACCCTAA ACTGGAGCAT ATCGATGCCA GAACGAATTC ATCGCGTCAA TCAATACGTC
3601 ATGGGATGGA TCGGGTATTT TCGGCTCGTC GAAACCCCGT CTGTCCCTTCA GACCATCGAA
3661 GGATGGATTTC GGAGGAGGCT TCGACTCTGT CAATGGCTTC AATGGAAACG GGTCAGAACC
3721 AGAATCCGTG AGTTAAGAGC GCTGGGGCTG AAAGAGACAG CGGTGATGGA GATCGCCAAT
3781 ACCCGAAAAG GAGCTTGGCG AACAACGAAA ACGCCGCAAC TCCACCAGGC CCTGGGCAAG
3841 ACCTACTGGA CCGCTCAAGG GCTCAAGAGT TTGACGCAAC GATATTTTCA ACTCCGTCAA
3901 GGTGACTGTC AGGCAAGCTT GGCCTGGGCC GTCGTTTTAC AACGTCGTGA CTGGGAAAAC
3961 CCTGGCGTTA CCCAACTTAA TCGCCTTGCA GCACATCCCC CTTTCGCCAG CTGGCGTAAT
4021 AGCGAAGAGG CCCGCACCGA TCGCCCTTCC CAACAGTTGC GCAGCCTGAA TGGCGAATGG
4081 CAGCTTGGCT GTTTTGGCGG ATGAGATAAG ATTTTCAGCC TGATACAGAT TAAATCAGAA
4141 CGCAGAAGCG GTCTGATAAA ACAGAATTTG CCTGGCGGCA GTAGCGCGGT GGTCCCACCT
4201 GACCCCATGC CGAACTCAGA AGTGAAACGC CGTAGCGCCG ATGGTAGTGT GGGGTCTCCC
4261 CATGCGAGAG TAGGGAAGT CCAGGCATCA AATAAAACGA AAGGCTCAGT CGAAAGACTG
4321 GGCCTTTCGT TTTATCTGTT GTTTGTGCGT GAACGCTCTC CTGAGTAGGA CAAATCCGCC
4381 GGGAGCGGAT TTGAACGTTG CGAAGCAACG GCCCGGAGGG TGGCGGGCAG GACGCCCGCC
4441 ATAAACTGCC AGGCATCAAA TTAAGCAGAA GGCCATCCTG ACGGATGGCC TTTTGTGCTT
4501 TCTACAACT CTTTTTGT TTATTTCTAA ATACATTCAA ATATGTATCC GCTCATGAGA
4561 CAATAACCTT GATAAATGCT TCAATAATAT TGAAAAAGGA AGAGTATGAG TATTCACAT
4621 TTCCGTGTCT CCCTTATTCC CTTTTTTGCG GCATTTTGCC TTCTGTGTTT TGCTCAGCCA
4681 GAAACGCTGG TGAAAGTAAA AGATGCTGAA GATCAGTTGG GTGCACGAGT GGGTTACATC
4741 GAACTGGATC TCAACAGCGG TAAGATCCTT GAGAGTTTTT GCCCGAAGA ACGTTCTCCA
4801 ATGATGAGCA CTTTTAAAGT TCTGCTATGT GGCGCGGTAT TATCCCGTGT TGACGCCGGG
4861 CAAGAGCAAC TCGGTCGCCG CATACTAT TCTCAGAATG ACTTGTTGA GTACTACCA
4921 GTCACAGAAA AGCATCTTAC GGATGGCATG ACAGTAAGAG AATTATGCAG TGCTGCCATA
4981 ACCATGAGTG ATAACACTGC GGCCAACTTA CTTCTGACAA CGATCGGAGG ACCGAAGGAG
5041 CTAACCGCTT TTTTGCACAA CATGGGGGAT CATGTAATC GCCTTGATCG TTGGGAACCG
5101 GAGCTGAATG AAGCCATACC AAACGACGAG CGTGACACCA CGATGCCTGT AGCAATGGCA
5161 ACAACGTTGC GCAAACTATT AACTGGCGAA CTACTTACTC TAGCTTCCCG GCAACAATTA
5221 ATAGACTGGA TGGAGGCGGA TAAAGTTGCA GGACCACTT TCGCTCGGC CCTTCCGGCT
5281 GGCTGGTTTA TTGCTGATAA ATCTGGAGCC GGTGAGCGTG GGTCTCGCG TATCATTGCA
5341 GCACTGGGGC CAGATGGTAA GCCCTCCCGT ATCGTAGTTA TCTACACGAC GGGGAGTCAG
5401 GCAACTATGG ATGAACGAAA TAGACAGATC GCTGAGATAG GTGCCTCACT GATTAAGCAT
5461 TGGTAACTGT CAGACCAAGT TTACTCATAT ATACTTTAGA TTGATTTACC CCGGTTGATA
5521 ATCAGAAAAG CCCCCAAAAC AGGAAGATTG TATAAGCAAA TATTTAAATT GTAAACGTTA
5581 ATATTTTGTT AAAATTCGCG TTAAATTTTT GTTAAATCAG CTCATTTTTT AACCAATAGG
5641 CCGAAATCGG CAAAATCCCT TATAAATCAA AAGAATAGAC CGAGATAGGG TTGAGTGTG
5701 TTCCAGTTTG GAACAAGAGT CCACTATTAA AGAACGTGGA CTCCAACGTC AAAGGGCGAA
5761 AAACCGTCTA TCAGGGCGAT GGCCCACTAC GTGAACCATC ACCCAAATCA AGTTTTTTGG
5821 GGTGAGGCTG CCGTAAAGCA CTAAATCGGA ACCCTAAAGG GAGCCCCGA TTTAGAGCTT
5881 GACGGGGAAA GCCGGCGAAC GTGGCGAGAA AGGAAGGGAA GAAAGCGAAA GGAGCGGGCG
5941 CTAGGGCGCT GGCAAGTGTA GCGGTCACGC TCGCGGTAAC CACCACACCC GCCGCGCTTA
6001 ATGCGCCGCT ACAGGGCGCG TAAAAGGATC TAGGTGAAGA TCCTTTTTGA TAATCTCATG
6061 ACCAAAATCC CTTAACGTGA GTTTTCGTTT CACTGAGCGT CAGACCCCGT AGAAAAGATC
6121 AAAGGATCTT CTTGAGATCC TTTTTTTCTG CGCGTAATCT GCTGCTTGCA AACAAAAAA
6181 CCACCGCTAC CAGCGGTGGT TTGTTTGCCG GATCAAGAGC TACCAACTCT TTTTCCGAAG
6241 GTAACCTGGT TCAGCAGAGC GCAGATACCA AATACTGTCC TTCTAGTGTA GCCGTAGTTA
6301 GGCCACCACT TCAAGAACTC TGTAGCACCG CCTACATACC TCGCTCTGCT AATCCTGTTA
6361 CCAGTGGCTG CTGCCAGTGG CGATAAGTCG TGTCTTACCG GGTTGGACTC AAGACGATAG
6421 TTACCGGATA AGGCGCAGCG GTCGGGCTGA ACGGGGGGTT CGTGCACACA GCCCAGCTTG
6481 GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG AGCTATGAGA AAGCGCCACG
6541 CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGGTCGG AACAGGAGAG
6601 CGCACGAGGG AGCTTCCAGG GGGAAACGCC TGGTATCTTT ATAGTCTGTG CGGGTTTCGC
6661 CACCTCTGAC TTGAGCGTCG ATTTTTGTGA TGCTCGTCAG GGGGGCGGAG CCTATGGAAA
6721 AACGCCAGCA ACGCGGCCTT TTTACGGTTC CTGGCCTTTT GCTGGCCTTT TGCTCACATG
6781 TTCTTTCCTG CGTTATCCCC TGATTCTGTG GATAACCGTA TTACCGCCTT TGAGTGAGCT
6841 GATACCGCTC GCCGCAGCCG AACGACCGAG CGCAGCGAGT CAGTGAGCGA GGAAGCGGAA
6901 GAGCGCCTGA TCGGCTATTT TCTCCTTACG CATCTGTGCG GTATTTTACA CCGCATATAT
6961 GGTGCACTCT CAGTACAATC TGCTCTGATG CCGCATAGTT AAGCCAGTAT AACTCCGCT
7021 ATCGCTACGT GACTGGGTCA TGGCTGCGCC CCGACACCCG CCAACACCCG CTGACGCGCC
7081 CTGACGGGCT TGTCTGCTCC CGGCATCCGC TTACAGACAA GCTGTGACCG TCTCCGGGAG
7141 CTGCATGTGT CAGAGGTTTT CACCGTCATC ACCGAAACGC GCGAGGCAGC TGCGGTAAAG

FIG. 10 (cont.)

17/24

7201 CTCATCAGCG TGGTCGTGCA GCGATTCACA GATGTCTGCC TGTTTCATCCG CGTCCAGCTC
7261 GTTGAGTTTC TCCAGAAGCG TTAATGTCTG GCTTCTGATA AAGCGGGCCA TGTTAAGGGC
7321 GGTTTTTTCC TGTTTGGTCA CTGATGCCTC CGTGTAAGGG GGATTTCTGT TCATGGGGGT
7381 AATGATACCG ATGAAACGAG AGAGGATGCT CACGATACGG GTTACTGATG ATGAACATGC
7441 CCGGTTACTG GAACGTTGTG AGGGTAAACA ACTGGCGGTA TGGATGCGGC GGGACCAGAG
7501 AAAAATCACT CAGGGTCAAT GCCAGCGCTT CGTTAATACA GATGTAGGTG TTCCACAGGG
7561 TAGCCAGCAG CATCCTGCGA TGCAGATCCG GAACATAATG GTGCAGGGCG CTGACTTCCG
7621 CGTTTCCAGA CTTTACGAAA CACGGAAACC GAAGACCATT CATGTTGTTG CTCAGGTCGC
7681 AGACGTTTTG CAGCAGCAGT CGCTTCACGT TCGCTCGCGT ATCGGTGATT CATTCTGCTA
7741 ACCAGTAAGG CAACCCCGCC AGCCTAGCCG GGTCCTCAAC GACAGGAGCA CGATCATGCG
7801 CACCCGTGGC CAGGACCCAA CGCTGCCCCA AATT

FIG. 10 (cont.)

18/24

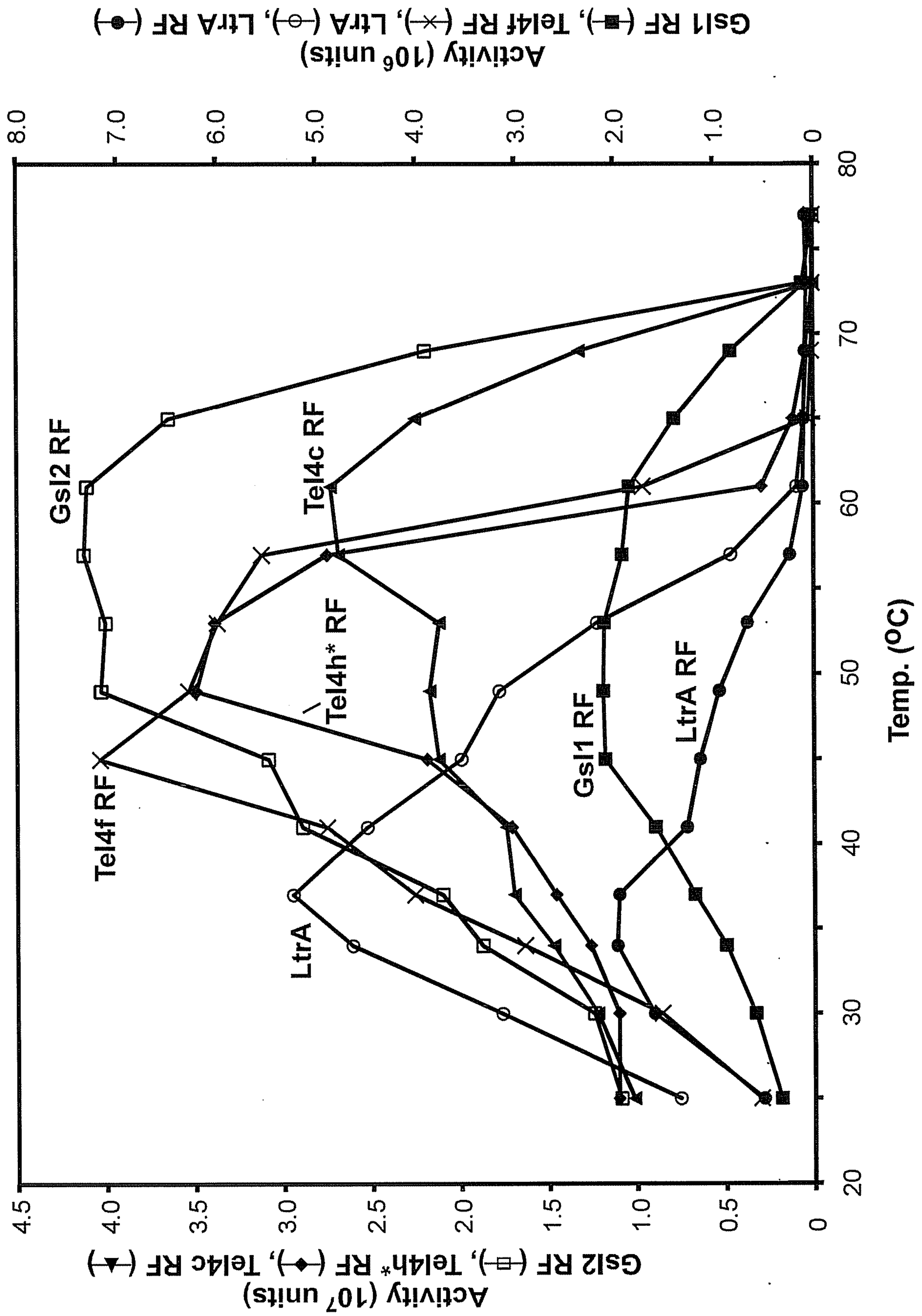
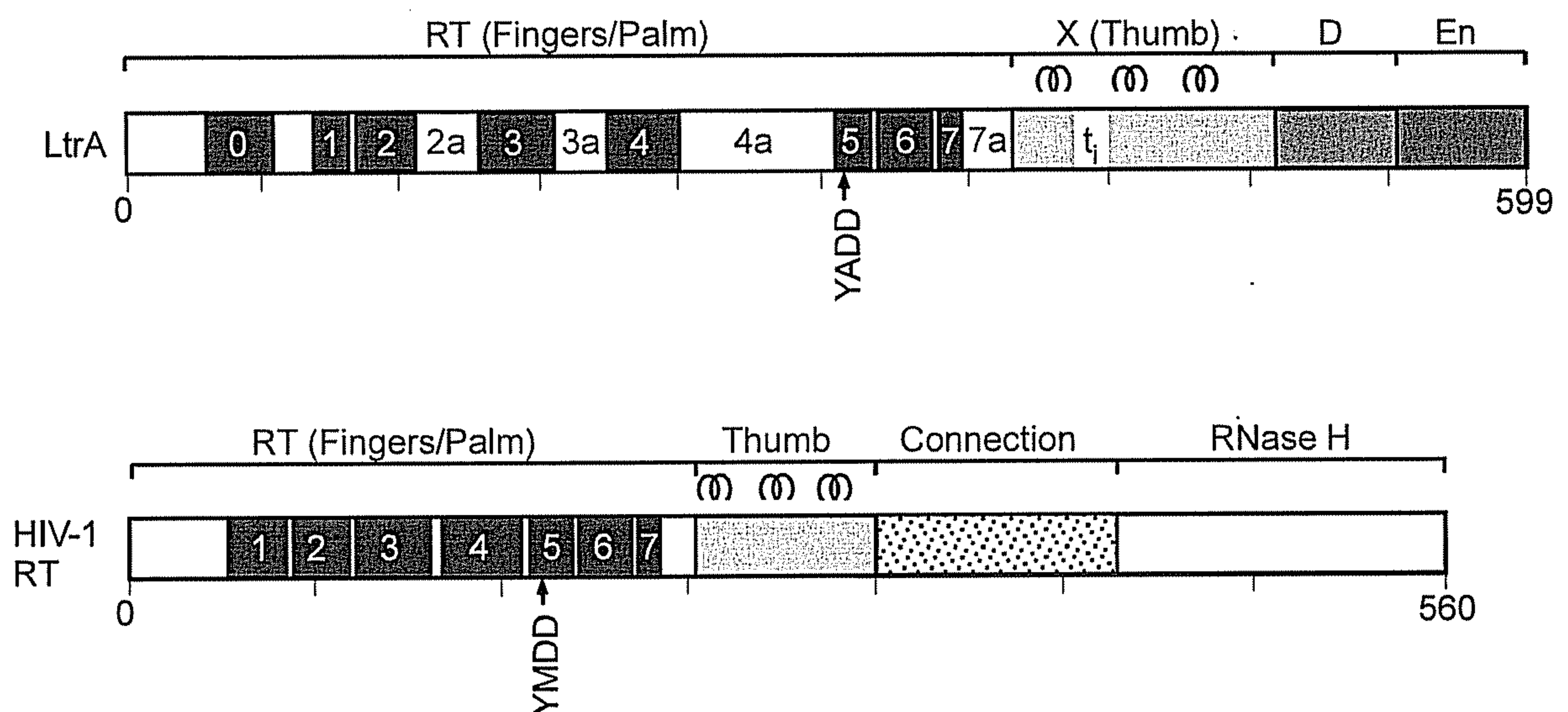


FIG. 11

A



B

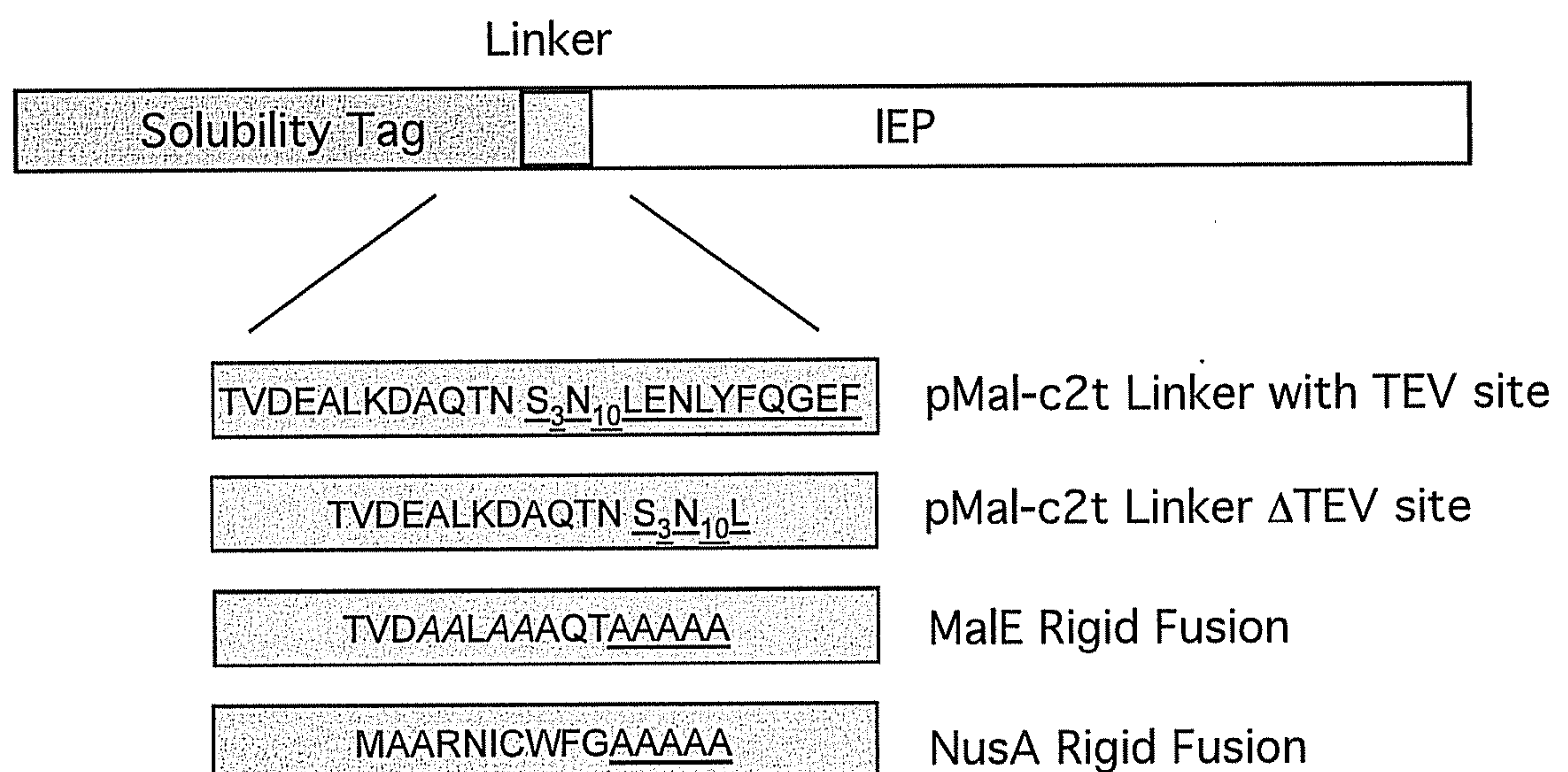


FIG. 12

20/24

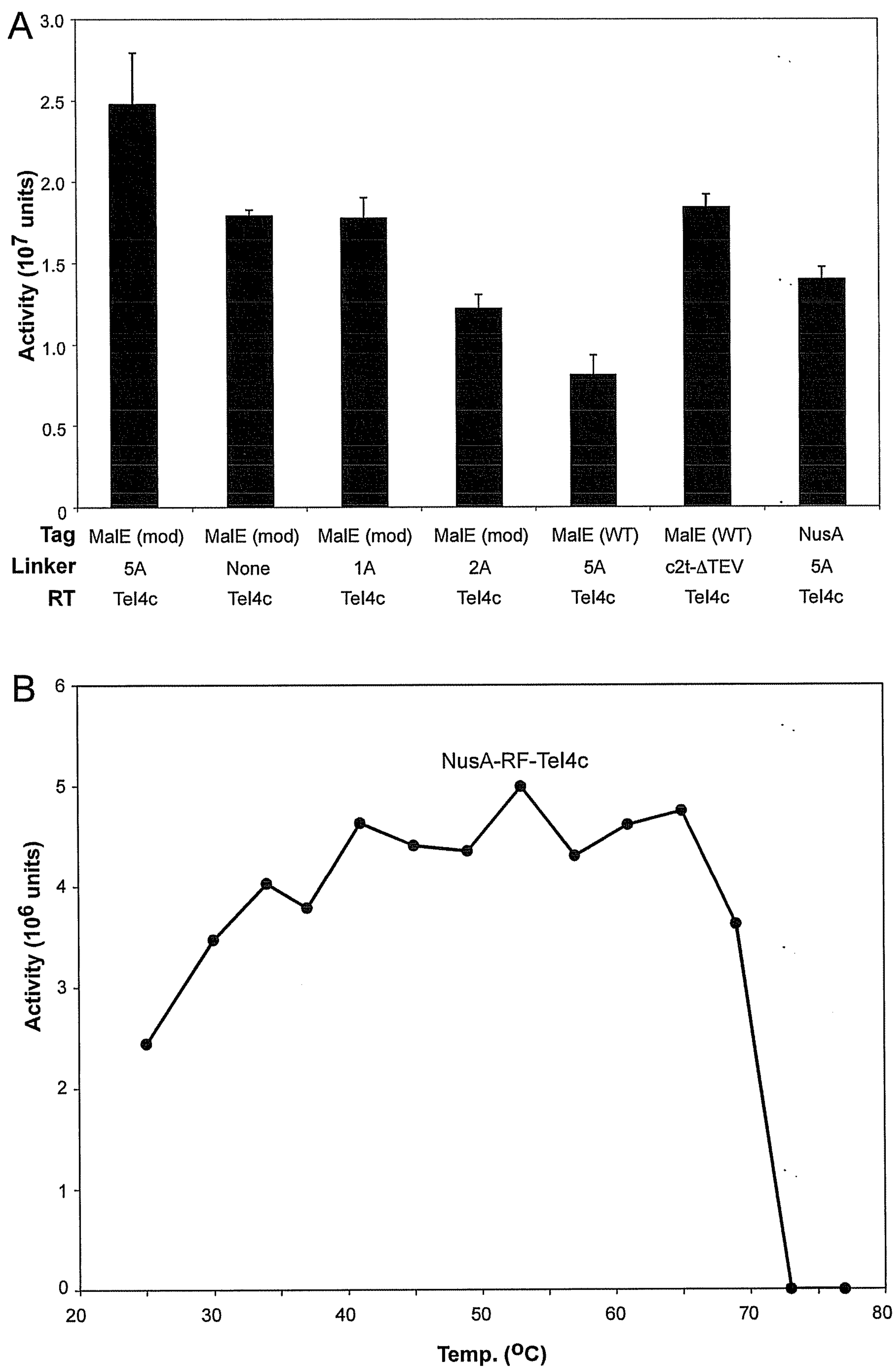


FIG. 13

21/24

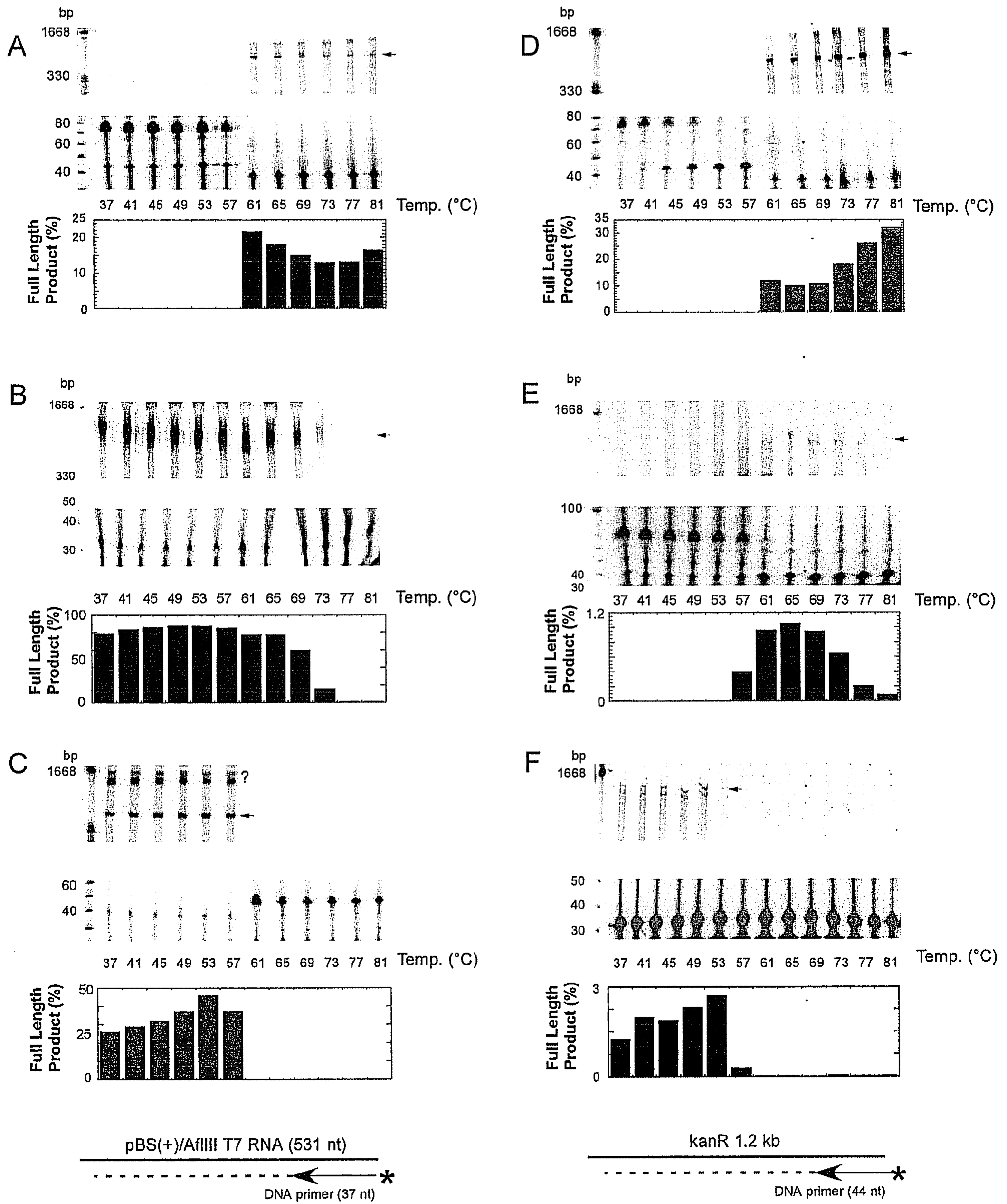


FIG. 14

SEQ ID NO: 21

5'GAATACAAGCTTGGGCGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAA
AAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAG
GGGTGTTATGAGCCATATTCAACGGGAAACGTCTTGCTCGAGGCCGCGATTAAAT
TCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGC
AATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCCGATGCGCCAGAGTTGTT
TCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAG
ACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGT
ACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGGAAAACAGCATTCC
AGGTATTAGAAGAATATCCTGAGTCAGGTGAAAATATTGTTGATGCGCTGGCAGT
GTTCTGCGCCGGTTGCATTTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATC
GCGTATTTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGC
GAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGA
AATGCATAAGCTTTTGCCATTCTCACCGGATTCAGTCGTCACATCATGGTGATTCT
CACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGG
ACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACCTGCCTC
GGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATA
ATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCA
GAATTGGTTAATTGGTTGTAACACTGGCAGAGCATTACGCTGACTTGACGGGACG
GCGGCTTTGTTGAATAAATCGAACTTTTGCTGAGTTGAAGGATCAGATCACGCAT
CTTCCCGACAACGCAGACCGTTCCGTGGCAAAGCAAAAGTTCAAAATCACCAAC
TGGTCCACCTACAACAAAGCTCTCATCAACCGTGGCGACTCTAGAGGATCCCCGG
GCGAGCTCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACCGAATT-3'

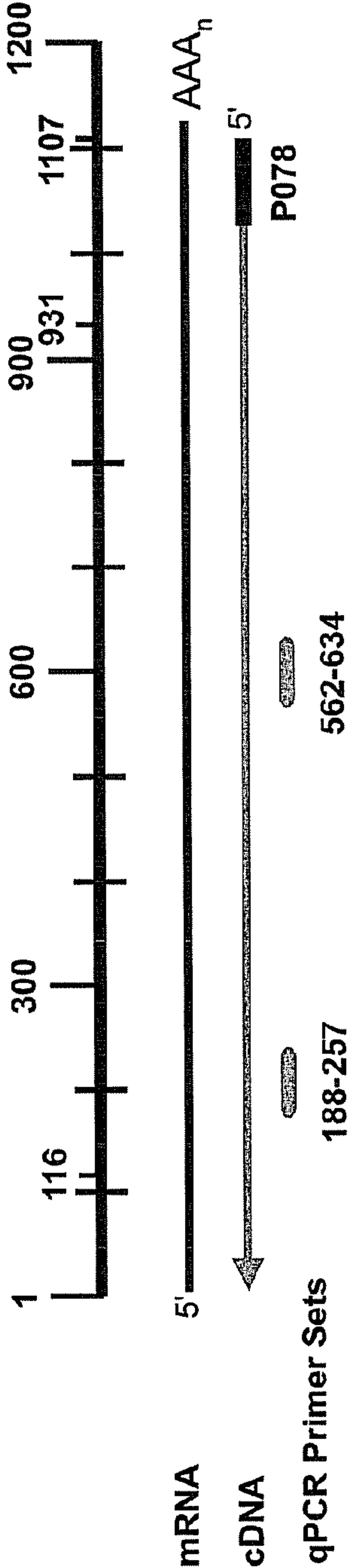
FIG. 15

SEQ ID NO: 38

MNKEILAVVEAVSNEKALPREKIFEALESALATATKKKYEQEIDVRVQIDRKSGDFD
TFRRWLVVDEVTQPTKEITLEAARYEDES LNLGDYVEDQIESVTFDRITTQTAKQVIV
QKVREAERAMVVDQFREHEGEIITGVVKKVNRDNISLDLGNNAEAVILREDMLPRE
NFRPGDRVRGVLYSVRPEARQAQLFVTRSKPEMLIELFRIEVPEIGEEVIEIKAAARDP
GSRAKIAVKTNDRIDPVGACVGMRGARVQAVSTELGGERIDIVLWDDNPAQFVIN
AMAPADVASIVVDEDKHTMDIAVEAGNLAQAIGRNGQNVRLASQLSGWELNVMTV
DDLQAKHQAEAHAAIDTFTKYLDIDEDFATVLVEEGFSTLEELAYVPMKELLEIEGL
DEPTVEALRERAKNALATIAQAQEE SLGDNKPADDLLNLEGVDRDLAFKLAARGVC
TLEDLAEQGIDDLADIEGLTDEKAGALIMAARNICWFG

FIG. 18

kanR RNA Coding Region nt 116 - 931



qPCR Amplification Plots, $\Delta RN_{(Log)}$ vs Cycle

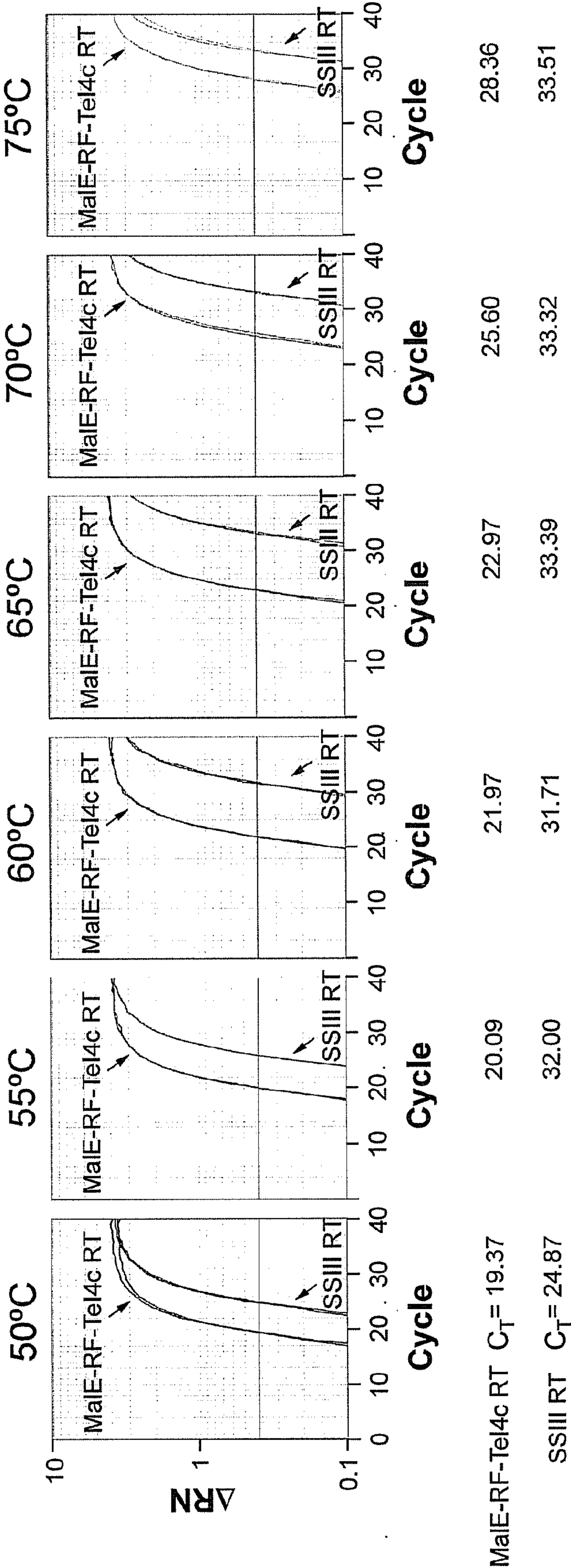


FIG. 16

kanR RNA Coding Region nt 116 - 931

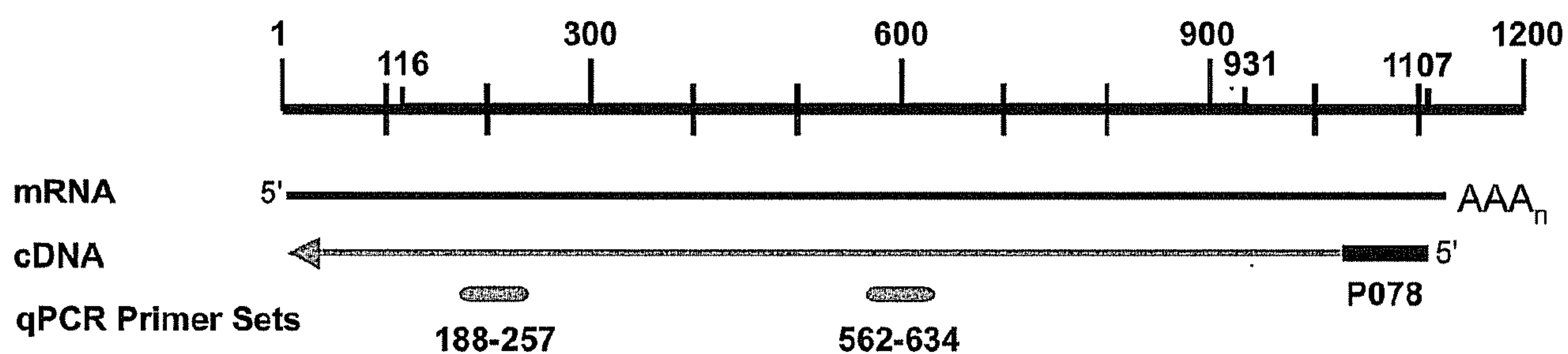
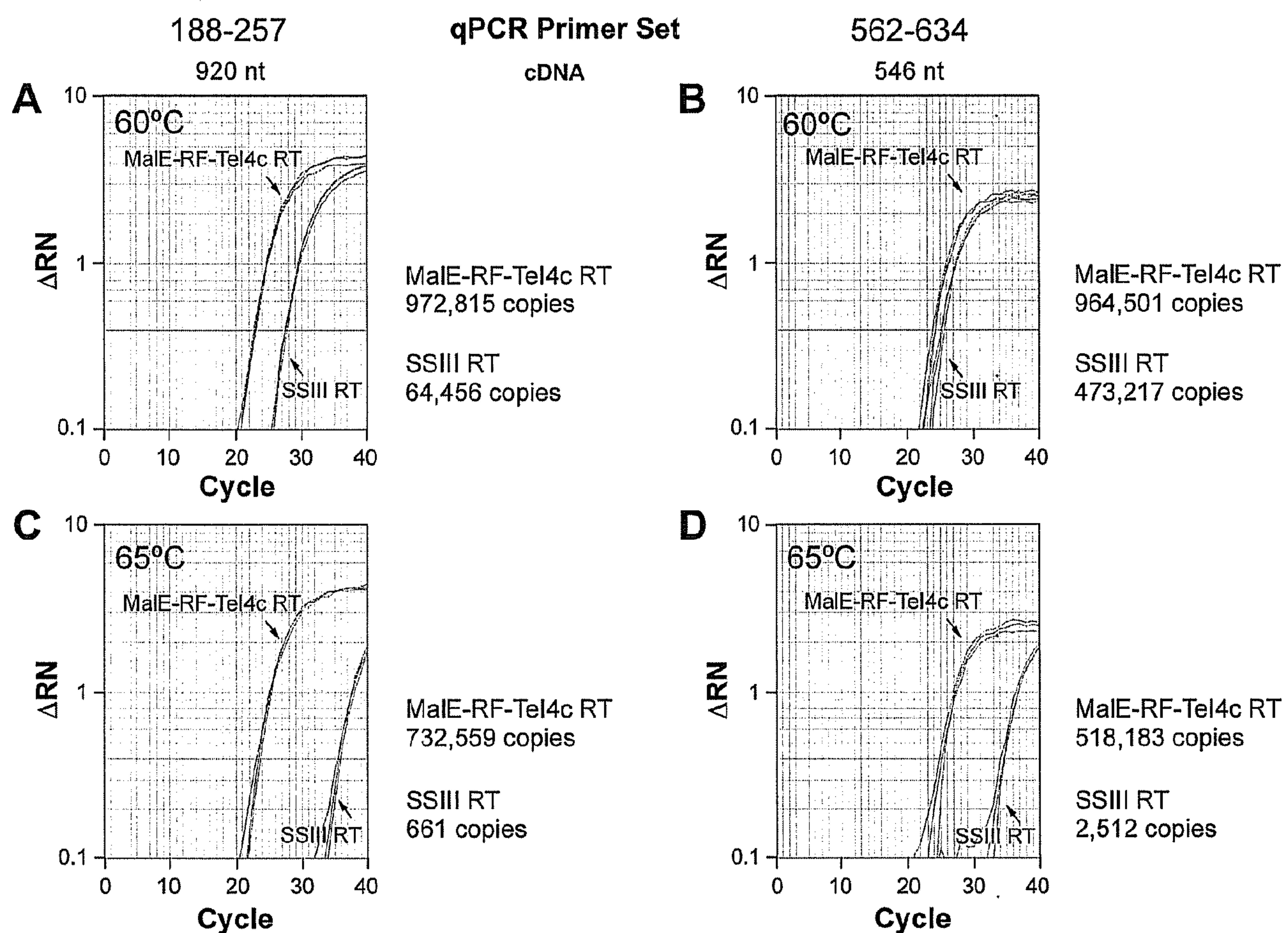
qPCR Amplification Plots, $\Delta RN_{(Log)}$ vs Cycle

FIG. 17

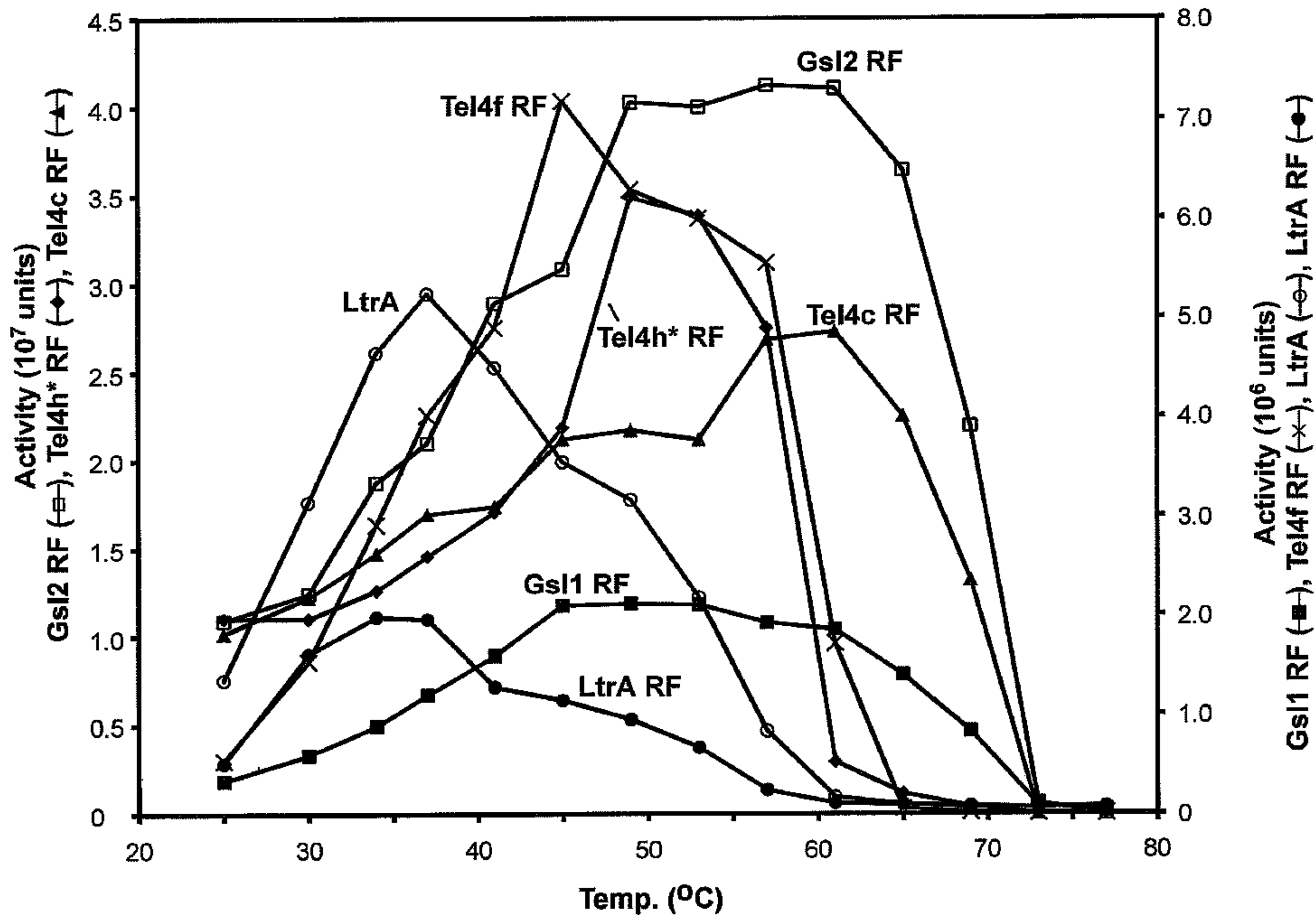


FIG. 11