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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
04 March 2021 (04.03.2021)

WIPO | PCT

(10) International Publication Number
WO 2021/041924 A2

(51) International Patent Classification:
C12N 15/11 (2006.01)

(21) International Application Number:
PCT/US2020/048561

(22) International Filing Date:
28 August 2020 (28.08.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/894,611	30 August 2019 (30.08.2019)	US
62/904,635	23 September 2019 (23.09.2019)	US
63/043,504	24 June 2020 (24.06.2020)	US

(71) Applicant: **BAYLOR COLLEGE OF MEDICINE** [US/US]; One Baylor Plaza, S234, Houston, Texas 77030 (US).

(72) Inventors: **YEN, Laising**; 2507 Rose Bay Drive, Pearland, Texas 77584 (US). **LUO, Liming**; 2922 Brookview Drive, Pearland, Texas 77584 (US). **JEA, Jocelyn Duen-Ya**; c/o Baylor College of Medicine, One Baylor Plaza, S234, Houston, Texas 77030 (US).

(74) Agent: **VETTER, Michael L.** et al.; Choate, Hall & Stewart LLP, Two International Place, Boston, Massachusetts 02110 (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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) Title: SYSTEM FOR REGULATING GENE EXPRESSION

(57) Abstract: Compositions and methods relating to regulation of gene expression are described.

WO 2021/041924 A2

SYSTEM FOR REGULATING GENE EXPRESSION

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Application No. 62/894,611, filed on August 30, 2019, U.S. Provisional Application No. 62/904,635, filed on September 23, 2019, and U.S. Provisional Application No. 63/043,504, filed June 24, 2020, the contents of each of which are incorporated herein by reference in their entirety.

Government License Rights

[0002] This invention was made with government support under EB013584 awarded by the National Institutes of Health. The government has certain rights in the invention.

Background

[0003] Nucleic acid based constructs for modulating expression of genes can be improved by increasing sensitivity and reducing leakiness.

Summary

[0004] The present disclosure recognizes a discovery of nucleic acid constructs related to regulatable gene product expression. In some embodiments, the present disclosure provides compositions and methods for the regulation of gene expression using nucleic acid constructs. In some embodiments, the present disclosure recognizes the utility of alternative splicing in regulation of gene expression in a nucleic acid construct. In some embodiments, the present disclosure recognizes the utility of regulating gene expression utilizing ligand-binding aptamers.

[0005] In some embodiments, the present disclosure provides a system for modulating gene expression, comprising a polyA aptamer polynucleotide that comprises in a 5' to 3' direction: a 5' splice donor site; an engineered intron; a first 3' splice acceptor site; a polyA switch comprising two or more ligand-binding aptamers with one or more ligand binding pockets, and at least one polyA cleavage signal therein; a second 3' splice acceptor site; and a nucleic acid sequence encoding an expressible polypeptide.

[0006] In some embodiments, a polyA aptamer polynucleotide of the present disclosure comprises two ligand-binding aptamers. In some embodiments, a polyA aptamer

polynucleotide comprises three ligand-binding aptamers. In some embodiments, a polyA aptamer polynucleotide comprises a polyA switch comprising a three way junction. In some embodiments, a three way junction comprises a junction of one or more RNA double stranded stems. In some embodiments, portions of a three way junction are single stranded. In some embodiments, a RNA double stranded stem comprises a ligand-binding aptamer. In some embodiments, a nucleic acid sequence encoding an expressible polypeptide comprises a 5'UTR.

[0007] In some embodiments, the present disclosure provides a method for modulating expression of a gene product in a cell. The method comprises the steps of: introducing into the cell a system comprising in a 5' to 3' direction: a 5' splice donor site; an engineered intron; a first 3' splice acceptor site; a polyA switch comprising two or more ligand-binding aptamers with one or more ligand binding pockets, and at least one polyA cleavage signal therein; a second 3' splice acceptor site. In some embodiments a gene product expressed by the methods described herein is exogenous to the cell. In some embodiments, a gene product expressed by the methods described herein is endogenous to the cell. In some embodiments, a method provided by the present disclosure occurs in one or more cells of an individual, the ligand is glucose, the individual has diabetes, pre-diabetes, or complications from diabetes, and/or the expressible polynucleotide is insulin. In some embodiments, a method provided by the present disclosure occurs in one or more cells of an individual, the expressible polynucleotide is a therapeutic gene product such as human growth hormone, coagulation factor X, or dystrophin. In some embodiments, a method provided by the present disclosure occurs in one or more cells of an individual, the ligand is the gene product of a cancer biomarker, and the expressible polynucleotide is a suicide gene. In some embodiments, a method provided by the present disclosure occurs in an individual, the expressible polynucleotide is a reporter gene, and the location and/or intensity of the expression of the reporter gene provides information about spatial distribution, temporal fluctuation, or both, of a ligand in one or more cells of the individual. In some embodiments, a method provided by the present disclosure occurs in an individual, tissue, or cell, wherein the expressible polynucleotide encodes a detectable gene product, and wherein the respective individual, tissue, or cell is imaged.

Brief Description of the Drawing

[0008] Figures 1A-1C provide schematics of aspects of a polyA aptamer polynucleotide described herein. Figure 1A depicts mechanism of the 'hybrid' switch based on ligand-inducible alternative splicing and polyA signal cleavage. Figure 1B depicts configuration of Y-shape polyA switch. The name of different parts of Y-shape structure is labeled. Figure C shows the configuration of a representative Y-shaped polyA switch Y196CAA.

[0009] Figures 2A-2C demonstrate results of additional Y-shape structures that are configured differently and with the polyA cleavage signal positioned differently. polyA signal is indicated by a red line. 3-way junction is indicated by a box. Figure 2A and 2B shows alternative Y-shape configurations with three aptamers (aptamer A, B, and C) arranged differently around the 3-way junction. Figure 2C shows three aptamers stacked on each other without 3-way junction.

[0010] Figures 3A-3C demonstrate results of modification of the number of polyA cleavage signals in a polyA aptamer polynucleotide described herein. Figure 3A shows 2 polyA signal (red box) located on two different stems. Figure 3B shows only one polyA signal partially buried in arm 1-2. Figure 3C shows 2 polyA signals (red box) are embedded in arm1-2.

[0011] Figures 4A- 4L demonstrate results of modification of a 3-way junction of a polyA aptamer polynucleotide described herein. Figure 4L shows the best 3-way junction sequences.

[0012] Figures 5 demonstrate results of modification of a polyA signal relative to the location of a 3-way junction of a polyA aptamer polynucleotide described herein.

[0013] Figures 6A-6B demonstrate results of modification of the third double strand stems (refer to as arm 3-1 and 3-2 in Figure 1B) of a polyA aptamer polynucleotide described herein. Figures 6A demonstrates results of modification of arm 3-1. Figures 6B demonstrates results of modification of arm 3-2.

[0014] Figures 7A-7B demonstrate results of modification of the second double strand stems (refer to as arm2-1 and 2-2 in Figure 1B) of a polyA aptamer polynucleotide

described herein. Figures 7A demonstrates results of modification of arm 2-2. Figures 7B demonstrates results of modification of arm 2-1.

[0015] Figure 8 demonstrates results of modification of the upper part of the first double strand stem (refer to as arm1-2 in Figure 1B) of a polyA aptamer polynucleotide described herein.

[0016] Figure 9 demonstrates results of modification of the lower part of the first double strand stem (refer to as arm 1-1 in Figure 1B) of a polyA aptamer polynucleotide described herein.

[0017] Figures 10A-10B demonstrate results of modification of aptamer orientation of a polyA aptamer polynucleotide described herein. Figure 10A shows the results with the orientation of aptamer B reversed. Figure 10B shows the results with the orientation of aptamer A orientation reversed.

[0018] Figures 11A-11B demonstrate the contribution of each aptamer in a polyA aptamer polynucleotide described herein. Figures 11A shows the effect of inactivating each aptamer by an A to C point mutation (indicated by the arrow). Figure 11B shows the effect of deleting aptamer A on induction.

[0019] Figures 12A-12D demonstrate results of modification of a 5'UTR of the expressible polynucleotide following a polyA aptamer polynucleotide described herein. Figure 12A shows results of inserting CAA repeats (underlined) in the 5'UTR of the expressible polynucleotide using different parental constructs. Figure 12B shows results of testing new 5'UTR sequence with strong 3' splice site using S56 as the parental construct. Figure 12C shows the results of inserting unstructured spacer sequence into 5'UTR of Y305 and Y300. Figure 12D shows inserting CAA repeats before the 3' splice site in 5'UTR.

[0020] Figures 13A-13B show the importance of G quad sequences of a polyA aptamer polynucleotide described herein. Figure 13A shows the effects of G-quad sequence on induction using Y196CAA as the parental construct. Figure 13B shows results of testing different G-quad sequences to replace 4MAZ G-quad using S56 as the parental construct.

[0021] Figure 14 demonstrates confirmation of tetracycline-induced alternative splicing of a polyA aptamer polynucleotide described herein. In the absence of Tc, IVS2-spliced RNA is degraded by polyA cleavage (lane 1 and 3). The presence of Tc induces

alternative splicing in both Y196CAA-2MAZ and Y196CAA-4MAZ (lane 2 and 4). Ligand-induced alternative splicing is much more pronounced with the presence of 4MAZ.

[0022] Figures 15A- 15G demonstrate results of modification of a first 3'splice acceptor site of a polyA aptamer polynucleotide described herein. Figure 15A shows results of moving IVS2 3' splice site into arm1-1 of Y196CAA-4MAZ. Figure 15B shows that the first 3' splice site is strongly inhibited when completely embedded into the arm1-1 near aptamer A (red arrow), resulting in very low induction. Diminishing the clamping effect of aptamer A by deleting part of its sequence restores the induction. Figure 15C shows results of moving the IVS 3' splice site (blue box) along the arm 1 of S9m, and Figure 15D shows results of placing the IVS 3' splice site in the bulge of arm1-2. Figure 15E shows results of changing the predicted strength of splicing by mutating the base after IVS2 3' splice site. Figure 15F shows results of moving mini-IVS2 3' splice site further into or away from aptamer A in arm 1-1. Figure 15G shows randomization of the three bases after the first 3' splice site (CAGNNN).

[0023] Figures 16A-16C demonstrate results of modification of a second 3'splice acceptor site of a polyA aptamer polynucleotide described herein. Figure 16A shows results of modifications of 5'UTR to alter the strength of the alternative 3' splice site. Figure 12B shows results of randomization of the three bases after 'TAG' in 5'UTR (TAGNNN) to modulate the strength of the alternative 3' splice site in order to improve the induction. Figure 12C shows the results of incorporating the best TAGNNN sequences selected from randomization into Y329 5'UTR.

[0024] Figures 17A and B demonstrate results of modification of the size of an engineered intron of a polyA aptamer polynucleotide described herein. Figure 17A shows results of varying the size and splicing elements of the IVS2 intron. Figure 17B shows results of removing CAA repeats from the constructs (S159, S164 and S169) with the shorter engineered intron.

[0025] Figures 18A-18C demonstrate results of inclusion of an upstream open reading frame (μORF) in a polyA aptamer polynucleotide described herein. Figure 18A shows the schematics of inclusion of an upstream open reading frame in a polyA aptamer. The inserted upstream ATG start codon is boxed. Figure 18B shows results of fine-tuning

the 5'UTR sequence of constructs with an upstream open reading frame. Figure 18C shows one representative hybrid switch with the inclusion of an upstream open reading frame.

[0026] Figures 19A-19E demonstrate the ability of a polyA aptamer polynucleotide described herein to control the gene expression of an expressible polypeptide in the presence of a ligand. Figure19A show the performance of representative S series constructs vs. Y196CAA-4MAZ. Figure 19B shows dose response of representative S series constructs vs. Y196CAA-4MAZ visualized by microscopy. Figure 19C shows the performance of Y300 and Y301. Figure 19D shows the dose response of Y362 and Y367 determined by luciferase reporter assays. Figure 19E shows the response to 1ug/ml tetracycline of Y362 and Y367 as determined by fluorescence activated cell sorting (FACS) using eGFP reporter signal. 'Induction in fold' in all results is calculated as the ratio of transgene expression in the presence vs. absence of tetracycline.

[0027] Figure 20 demonstrates the ability of a polyA aptamer polynucleotide described herein to function as an endogenous switch to control the expression of an endogenous gene in the genome.

[0028] Figure 21 depicts configuration of a Y-shape polyA switch combining single base changes at three locations. The Y387 construct shown here contains all the three changes.

[0029] Figure 22 demonstrates that the combination of three single base changes significantly increase the induced expression of an expressible polypeptide at low drug concentration. Four different parental constructs (Y359, Y360, Y361, Y362C) were used to demonstrate the effects of single base changes on induction. The effects on induction by these single base changes are similar across all four different parental constructs. Upper panel shows the induction in fold with standard variation. Lower panel plots the induction in fold for each construct.

[0030] Figures 23A and 23B demonstrate a dose response analysis of induction of expression from constructs Y362 and Y386 comprising a Y-shape polyA switch combining single base changes at three locations. Figures 23A shows that the induction by tetracycline reaches 50% of the maximal level (EC_{50}) at as low as 0.5 to 1 μg/ml Tc using the maximum induction in fold as the EC_{100} reference. Figures 23B shows a similar calculation using the

maximum expression level of parental construct (HDM-Luc, which has similar sequence but without the Y-shape structure) as the EC_{100} reference. In this case, EC_{50} is reached by tetracycline as low as 0.5 to 1.2 μg/ml.

Detailed Description of Certain Embodiments

[0031] In some embodiments, the present disclosure provides compositions and methods for regulatable gene product expression. In some embodiments, compositions and methods for regulatable gene product expression comprise a polyA aptamer polynucleotide. In some embodiments, a polyA aptamer polynucleotide comprises, amongst other things, one or more splice donor sites, one or more splice acceptor sites, an engineered intron; a polyA switch; and a nucleic acid sequence encoding an expressible polypeptide. In some embodiments, a polyA switch comprises at least one ligand-binding aptamer. In some embodiments, a polyA switch comprises at least one polyA cleavage signal. In some embodiments, a polyA aptamer polynucleotide comprises RNA double strand stems.

Aptamer

[0032] Aptamers are short RNA sequences that fold like receptors and bind to specific ligands. Efficient in vitro evolution methods for generating aptamers with high affinity to specific ligands are well established. The binding affinity of aptamers can often reach nanomolar range, comparable to that of antibodies. In this regard, aptamers can be viewed as antibodies made of RNA. What distinguishes an aptamer from an antibody are its small size (often smaller than 50 bases) and its modular nature. These features enable aptamers to integrate with and control other RNA structures without losing its binding function. It has been demonstrated that aptamers can transform the self-cleaving RNA ribozymes to operate in a ligand-dependent manner, and function like a molecular switch in test tubes and in cells.

[0033] In some embodiments, a polyA aptamer polynucleotide comprises one or more RNA double stranded stems. In some embodiments, a RNA double stranded stem is a

nucleic acid structure formed by intramolecular base pairing of complementary nucleic acids contained within a single polyA aptamer polynucleotide. In some embodiments, a RNA double stranded stem may also be referred to as an arm. In some embodiments, a polyA aptamer polynucleotide comprises one or more RNA double strand stems. In some embodiments, a polyA aptamer polynucleotide comprises two RNA double strand stems. In some embodiments, a polyA aptamer polynucleotide comprises three RNA double strand stems. In some embodiments, a RNA double stranded stem comprises ligand binding aptamer. In some embodiments, a polyA aptamer polynucleotide comprises two ligand binding aptamers. In some embodiments, a polyA aptamer polynucleotide comprises three ligand binding aptamers.

[0034] In some embodiments, at least two RNA double stranded stems are joined to form a junction. In some embodiments, a junction of RNA double stranded stems comprises a single stranded region. In some embodiments, three RNA stems meet to form a three way junction. In some embodiments, a three way junction comprises at least one single stranded region. In some embodiments, a three way junction comprises one, two, or three single stranded regions.

[0035] In some embodiments the sequence of a double stranded RNA stem is selected from one of the following:

SEQ ID NO.:	**SEQUENCE (5' to 3')**
2	GGGUGUUUGUGGC
3	CACGAGAUCUGG
4	GCGUUUUAUACUU
5	CUCUGCAGAUGUU

[0036] In some embodiments, a single stranded region formed by a junction of RNA double stranded stems comprises at least one nucleic acid. In some embodiments, a single stranded region formed by a junction of RNA double stranded stems comprises one, two,

three, four, five, six, seven, eight, nine, ten, eleven, twelve, or more nucleic acids. In some embodiments, a three way junction comprises a first, second, and third single stranded regions. In some embodiments, a first single stranded region comprises at least one base selected from C and A. In some embodiments, a second single stranded region comprises at least one base selected from C and A.

[0037] In some embodiments, a RNA double stranded stem is 30, 20, 10, or 5 base pairs in length. In some embodiments, a RNA double stranded stem is 5 to 30, 10 to 30, 20 to 30, 5 to 10, 5 to 20, 5 to 30, or 10 to 20 base pairs in length. In some embodiments, a RNA double stranded stem is up to 30 base pairs in length. In some embodiments, a RNA double stranded stem is less than 30, 20, or 10 base pairs in length.

[0038] In some embodiments, a polyA aptamer polynucleotide comprises one or more aptamers. In some embodiments, a polyA aptamer polynucleotide comprises two aptamers. In some embodiments, a polyA aptamer polynucleotide comprises three aptamers.

In some embodiments, an aptamer included in a polyA aptamer polynucleotide described herein comprises at least one single stranded region and at least one aptamer RNA double stranded stem. In some embodiments, an aptamer RNA double stranded stem comprises a single stranded region. In some embodiments, an aptamer RNA has an RNA double stranded stem with a sequence of AATAAGATTACCGAAAGGCAATCTTATT (e.g., arm2-2). In some embodiments, an aptamer RNA has an RNA double stranded stem with a sequence of CCAGATCGAATTCGATCTGG (e.g., are 3-2). In some embodiments, an aptamer RNA has an RNA double stranded stem with a length ranging from 6-10; 7-11; 8-12; 9-13; 10-14 base pairs in length.

PolyA cleavage signal

[0039] In accordance with various embodiments, any of a variety of polyA signals (e.g., encoded by a polyA signal sequence) may be used. By way of non-limiting example, a polyA signal sequences used in mammalian cells include: AAUAAA, AUUAAA, AGUAAA, ACUAAA, UAUAAA, CAUAAA, GAUAAA, AAUAUA, AAUACA, and AAUAGA. In some embodiments, a polyA switch may include two or more polyA signal sequences (e.g., 3, 4, 5, 6, 7, 8, 9, 10 or more).

[0040] Polyadenylation is a foundational mRNA processing mechanism that is present in all mammalian cells. Typically, mammalian polyA signals are found in the 3' untranslated region (UTR). In contrast, the present disclosure provides compositions and methods that comprise a polyA cleavage signal present in an expression construct at a location other than at the 3' untranslated region (UTR) of an expressible polynucleotide, such as a gene. When a polyA signal is artificially created in the 5' UTR, where it is not normally found in cells, efficient cleavage of the polyA signal leads to the addition of polyA tail at the site. This results in the removal and degradation of the second half of the associated mRNA with transgene sequence, and therefore a loss of gene expression. In some embodiments, the polyA signal is present upstream of the translation start site of a nucleic acid sequence encoding an expressible polynucleotide (mRNA) encoding an expressed polypeptide. In some embodiments, the polyA signal is located in the 5' UTR of the mRNA. In some embodiments, a single stranded region of a 3-way junction comprises all or a portion of the polyA cleavage signal. In some embodiments, the third single stranded region of a 3-way junction comprise all or a portion of the polyA cleavage signal. In some embodiments, a RNA double stranded stem comprises all or a portion of the polyA cleavage signal. In some embodiments, the third RNA double stranded stem comprises all or a portion of the polyA cleavage signal. In some embodiments, a portion of the polyA cleavage signal, as used herein, includes one, two, three, or four nucleotides. In some embodiments, a polyA cleavage signal has a sequence of AAUAAA. In some embodiments, a polyA cleavage signal has a sequence of AUUAAA, AGUAAA, ACUAAA, UAUAAA, CAUAAA, GAUAAA, AAUAUA, AAUACA, AAUAGA, AAAAAG, or ACUAAA. In embodiments wherein two or more polyA signals are utilized in the construct, the polyA signals may be the same or may be different. In particular embodiments, the expressible polynucleotide is able to be transcribed by RNA polymerase II.

[0041] In some embodiments, the presence of the polyA cleavage signal in the 5' UTR targets the second half of mRNA after the polyA signal for degradation, and this ability is exploited in the various compositions and methods of the present disclosure. In some embodiments, the presence of the polyA cleavage signal in the 5' UTR results in cleavage of a pre-mRNA/mRNA encoded by a polyA aptamer polynucleotide. In some embodiments, cleavage of a pre-mRNA/mRNA encoded by a polyA aptamer polynucleotide

results in degradation of the second half of pre-mRNA/mRNA. In some embodiments, cleavage of a pre-mRNA/mRNA encoded by a polyA aptamer polynucleotide results in no expression of a polypeptide.

[0042] In particular embodiments, the polyA cleavage signal is within a polyA aptamer polynucleotide comprising at least one ligand-binding aptamer to which one or more ligands can bind. In some embodiments, binding of the ligand to the ligand-binding aptamer determines whether or not the polyA cleavage signal is present in the pre-mRNA/mRNA after alternative splicing. In some embodiments, binding of the ligand to the ligand-binding aptamer determines whether or not the pre-mRNA/mRNA is cleaved after alternative splicing. In some embodiments, binding of the ligand to the ligand-binding aptamer determines whether or not an expressible polypeptide is expressed after alternative splicing.

Engineered Intron

[0043] In some embodiments, a polyA aptamer polynucleotide comprises an engineered intron. In some embodiments, an engineered intron comprises one or more splice sites. In some embodiments, a splice site is or comprises a splice donor site (e.g, comprising a GU sequence). In some embodiments a splice site is or comprises a splice acceptor site (e.g., comprising an AG sequence). In some embodiments, splice sites in an engineered intron function (e.g., in conjunction with each other and/or in conjunction with one or more endogenous splice site(s)) to excise an engineered intron from a polyA aptamer polynucleotide.

[0044] In some embodiments, an engineered intron is preceded by a 5' splice donor site. In some embodiments, a polyA aptamer polynucleotide comprises a 5' splice donor site in the region 5' of an engineered intron. In some embodiments, a polyA aptamer polynucleotide comprises a first 3' splice acceptor site 3' of an engineered intron. In some embodiments, an engineered intron of a polyA aptamer polynucleotide described herein comprises a 5' splice donor site and a first 3' splice acceptor site. In some embodiments, a polyA aptamer polynucleotide comprises a nucleic acid sequence encoding an expressible polypeptide. In some embodiments, a polyA aptamer polynucleotide comprises a second

3'splice acceptor site immediately 5' of a nucleic acid sequence encoding an expressible polypeptide.

[0045] In some embodiments, a polyA aptamer polynucleotide comprises a promoter 5' of the splice donor site. Exemplary promoters include, e.g., CMV, E1F, VAV, TCRvbeta, MCSV, an SV40 promoter, an RSV promoter, and PGK promoter.

[0046] In some embodiments, in the absence of a ligand bound to a ligand-binding aptamer, splicing of the pre-mRNA encoded by a polyA aptamer polynucleotide described herein occurs between the 5' splice donor site and the first 3' splice acceptor site. In some embodiments, splicing between the 5' splice donor site and the first 3' splice acceptor site of a pre-mRNA encoded by a polyA aptamer polynucleotide described herein results in an mRNA comprising a polyA cleavage signal preceding a 5'UTR of a nucleic acid sequence encoding an expressible polypeptide. In some embodiments, presence of a polyA cleavage signal preceding a 5'UTR of a nucleic acid sequence encoding an expressible polypeptide results in cleavage at the polyA cleavage site and degradation of the sequence encoding an expressible polypeptide.

[0047] In some embodiments, in the presence of a ligand bound to a ligand-binding aptamer, splicing of the pre-mRNA encoded by a polyA aptamer polynucleotide described herein occurs between the 5' splice donor site and the second 3' splice acceptor site. In some embodiments, splicing of the pre-mRNA encoded by a polyA aptamer polynucleotide described herein between the 5' splice donor site and the second 3' splice acceptor site results in an mRNA comprising a nucleic acid sequence encoding an expressible polypeptide. In some embodiments, splicing of the pre-mRNA encoded by a polyA aptamer polynucleotide described between the 5' splice donor site and the second 3' splice acceptor site results in removal of polyA cleavage signal by splicing it out. In some embodiments, splicing between the 5' splice donor site and the second 3' splice acceptor site of the pre-mRNA encoded by a polyA aptamer polynucleotide described herein results in the expression of an expressible polypeptide.

[0048] In some embodiments, a polyA aptamer polynucleotide comprises two or more ligand-binding aptamers. In some embodiments, each of two or more ligand binding aptamers binds a different ligand. In some embodiments, a polyA aptamer polynucleotide

comprises two or more separate polyA switches. In some embodiments, a first polyA switch comprises a first aptamer that binds a first ligand, and a second polyA switch comprises a second aptamer that binds a second ligand. In some embodiments the first and second aptamers are non-identical and the first and second ligands are non-identical. In some embodiments, the first and second aptamers are non-identical and the first and second ligands are identical.

[0049] In some embodiments, an engineered intron is any sequence. In some embodiments, an engineered intron is approximately 100, 200, 300, 400, or 500 nucleotides in length. In some embodiments, an engineered intron is in the range of 100-200; 110-200; 120-200; 130-200; 140-200; 150-200; 160-200; 170-200; or 180-200 bases in length. In some embodiments, an engineered intron is at most 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220 bases in length. In some embodiments, an engineered intron has the following sequence:

GTGAGTCTTAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTG
GATTATTCTGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCT
TCCTCTGCAG (SEQ ID NO.: 1)

[0050] In some embodiments, an engineered intron has the following sequence:

GTGAGTCTATGGGACCCTTGATGTTTTCTTTCCCCTTCTTTTCTATGGTTAAGTTC
ATGTCATAGGAAGGGGAGAAGTAACAGGGTACACATATTGACCAAATCAGGGT
AATTTTGCATTTGTAATTTTAAAAAATGCTTTCTTCTTTTAATATACTTTTTTGTT
TATCTTATTTCTAATACTTTCCCTAATCTCTTTCTTTCAGGGCAATAATGATACAA
TGTATCATGCCTCTTTGCACCATTCTAAAGAATAACAGTGATAATTTCTGGGTTA
AGGCAATAGCAATATTTCTGCATATAAATATTTCTGCATATAAATTGTAACTGAT
GTAAGAGGTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATTCTGCTTTT
ATTTTATGGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTTTGC
TAATCATGTTCATACCTCTTATCTTCCTCCCACAG (SEQ ID NO.: 49)

[0051] As used herein, an intron can refer to either a DNA sequence or its corresponding RNA sequence.

[0052] In some embodiments a polyA aptamer polynucleotide comprises additional sequences to facilitate, regulate or assist polyA signal cleavage within a polyA aptamer

polynucleotide. In some embodiments, a polyA aptamer polynucleotide comprises a G-U rich region 5' of the nucleic acid sequence encoding the expressible polypeptide and 3' of the polyA cleavage signal. In some embodiments a polyA aptamer polynucleotide comprises additional sequences to facilitate, regulate or assist splicing within a polyA aptamer polynucleotide. In some embodiments, a polyA aptamer polynucleotide comprises a nucleic acid triplet sequence capable of modulating the strength of alternative splicing. In some embodiments, a nucleic acid triplet sequence is 3' relative to the second 3'acceptor site in the 5'UTR. In some embodiments, a nucleic acid triplet sequence is 3' of an engineered intron. In some embodiments, a sequence of a nucleic acid triplet sequence comprises any three nucleotides. In some embodiments, a sequence of a nucleic acid triplet sequence comprises TAG, TCT, TTC, TTG, TGA, TGC, TCC, ACA, AAC, ACC, AGC, AGG, CCT, CCC, TTT, TGA, TCT, TAC, CAC, or CAT.

[0053] In some embodiments, a polyA aptamer polynucleotide comprises a G-U rich region 5' of the nucleic acid sequence encoding the expressible polypeptide and 3' of the polyA cleavage signal. In some embodiments, a polyA aptamer polynucleotide comprises a G rich region 5' of the nucleic acid sequence encoding the expressible polypeptide and 3' of the G-U rich region. In some embodiments, a G rich region is understood in the art to be a MAZ sequence. In some embodiments, a polyA aptamer polynucleotide comprises one or more G rich regions. In some embodiments, a polyA aptamer polynucleotide comprises one or more consecutive G rich regions. In some embodiments, a polyA aptamer polynucleotide comprises one or more MAZ sequences. In some embodiments, a polyA aptamer polynucleotide comprises one or more consecutive MAZ sequences. In some embodiments, a polyA aptamer polynucleotide comprises one, two, three, four, five, six MAZ sequences. The consecutive MAZ may be separated by one or more spacer sequences. In some embodiments the sequence of a G rich region is AACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGG AAAGGGGGAGGGGGA (SEQ ID NO.: 47).

[0054] In some embodiments, a polyA aptamer polynucleotide comprises one or more start codons. In some embodiments, a polyA aptamer polynucleotide comprises one or more out of frame start codons. In some embodiments, an out of frame start codon is out of frame relative to the coding sequence of a nucleic acid sequence encoding an expressible

polypeptide. In some embodiments, a polyA aptamer polynucleotide comprises at least one out of frame start codon. In some embodiments, a polyA aptamer polynucleotide comprises at least one out of frame start codon 3' of a first 3' splice acceptor site 3' of an engineered intron.

Expressible polypeptide

[0055] In some embodiments, a polyA aptamer polynucleotide comprises a nucleic acid sequence encoding an expressible polypeptide. In some embodiments, a nucleic acid sequence encoding an expressible polypeptide comprises a 5'UTR. In some embodiments, a 5' UTR of a nucleic acid sequence encoding an expressible polypeptide comprises a 3'splice acceptor site. In some embodiments, a 5' UTR of a nucleic acid sequence encoding an expressible polypeptide comprises a branch point and a 3'splice acceptor site. A branch point is understood in the art to comprise a nucleotide or nucleotides involved in initiating a nucleophilic attack on the 5′ donor splice site. In some embodiments, a 5' UTR of a nucleic acid sequence encoding an expressible polypeptide does not comprise a branch point. In some embodiments, a 5' UTR of a nucleic acid sequence encoding an expressible polypeptide comprises a spacer sequence. In some embodiments, a spacer sequence comprises at least one CAA repeat. In some embodiments a 5'UTR of a nucleic acid sequence encoding an expressible polypeptide has a sequence of GCGGCCGCCTTAATTAACAGTGTTCACTAGAGCCAACAACAACAACAACAACA ACAACAACAACGACACC (SEQ ID NO.: 48)

[0056] In some embodiments, a nucleic acid sequence encoding an expressible polypeptide contemplated in the present disclosure can be any nucleic acid sequence or any gene encoding any polypeptide. In some embodiments, a nucleic acid sequence encoding a non-coding RNA. In some embodiments, a nucleic acid sequence encoding an expressible polypeptide contemplated in the present disclosure can be an exogenous nucleic acid. In some embodiments, a nucleic acid sequence encoding an expressible polypeptide contemplated in the present disclosure can be a gene endogenous to a subject to which a polyA aptamer polynucleotide has been introduced. In some embodiments, a polyA aptamer polynucleotide of the present disclosure is introduced into a region of an individual's genome that regulates expression of a gene of interest. Accordingly, in some embodiments, a polyA aptamer polynucleotide of the present disclosure can be used to regulate expression

of genes endogenous to an individual. In some embodiments, a nucleic acid sequence encoding an expressible polypeptide of a polyA aptamer polynucleotide of the present disclosure is an endogenous nucleic acid sequence.

[0057] In some embodiments, an expressible polypeptide is insulin. In some embodiments, an expressible polypeptide is human growth hormone. In some embodiments, an expressible polypeptide is coagulation factor X. In some embodiments, an expressible polypeptide is dystrophin. In some embodiments, an expressible polypeptide is a suicide protein. In some embodiments, a suicide protein is a protein that induces cell death. Exemplary suicide proteins include Mixed Lineage Kinase Domain Like Pseudokinase (MLKL), Receptor-interacting serine/threonine-protein kinase 3 (RIPK3), Receptor-interacting serine/threonine-protein kinase 1 (RIPK1), Fas-associated protein with death domain (FADD), or gasdermin D (GSDMD), cysteine-aspartic proteases, cysteine aspartases or cysteine-dependent aspartate-directed proteases (CASPASE-1 or CASP-1), CASPASE-4, CASPASE-5, CASPASE-12, PYCARD/ASC (PYD and CARD domain containing / Fas-associated protein with death domain) or variants thereof.

[0058] In some embodiments, an expressible polypeptide is a detectable gene product. In some embodiments a detectable gene product is a reporter. In some embodiments a reporter is a protein capable of providing a detectable signal and/or comprise the ability to generate a detectable signal (e.g. by catalyzing reaction converting a compound to a detectable product). Detectable signals can comprise, for example, fluorescence or luminescence. Detectable signals, methods of detecting them, and methods of incorporating them into reagents (e.g. polypeptides comprising a reporter protein) are well known in the art. In some embodiments of any of the aspects, detectable signals can include signals that can be detected by spectroscopic, photochemical, biochemical, immunochemical, electromagnetic, radiochemical, or chemical means, such as fluorescence, chemifluoresence, or chemiluminescence, or any other appropriate means. In some embodiments of any of the aspects, the reporter protein is selected from the group consisting of luciferase, nanoluciferase, beta-lactamase, beta-galactosidase, horseradish peroxidase, alkaline phosphatase, catalase, carbonic anhydrase, green fluorescent protein, red fluorescent protein, cyan fluorescent protein, yellow fluorescent protein, trypsin, a protease, a peptide that complements and activates a truncated reporter protein, a kinase.

[0059] In some embodiments, activity or function of a polyA aptamer polynucleotide of the present disclosure is measured by expression of an expressible polypeptide. In some embodiments, activity or function of a polyA aptamer polynucleotide of the present disclosure is measured by fold induction. In some embodiments, fold induction is calculated as the ratio of expressible polypeptide in the presence of a ligand and expressible polypeptide in the absence of a ligand. In some embodiments, fold induction is calculated as the ratio of expressible polypeptide in the presence of an aptamer and expressible polypeptide in the presence of a different aptamer. In some embodiments, fold induction is calculated as the ratio of expressible polypeptide in the presence of an aptamer comprising at least one splice acceptor site and one splice donor site and expressible polypeptide in the presence of a different aptamer with no splice sites. In some embodiments, fold induction is calculated as the ratio of expression of an endogenous gene before introduction of a polyA aptamer polynucleotide and expression of an endogenous gene after introduction of a polyA aptamer polynucleotide regulating expression of the same endogenous gene.

Ligand

[0060] In accordance with various embodiments, a ligand may be selected so as to facilitate a desired end purpose of a provided system. Accordingly, a ligand may be or comprise a polypeptide, nucleic acid, small molecule, drug, metabolite, or combination thereof. In some embodiments, a ligand may be or comprise a cellular metabolite, aberrant cellular protein, or a protein expressed by a pathogenic organisms (e.g., a virus, bacteria, or fungus). For example, in some embodiments, a ligand may be an exogenously administered small molecule so that dosing and function of the system can be modulated easily as desired in a particular therapeutic context. For example, in some embodiments, a ligand is tetracycline or its derivatives. In some embodiments, a ligand may be selected such that expression of an expressible polypeptide occurs in response to a particular biological condition (e.g., infection, tumorigenesis, high or low glucose), for example, as a biosensor system that can detect one or more intracellular “signatures” in a cell, tissue, or subject. Accordingly, in some embodiments, a ligand is endogenous to a subject (e.g., an endogenous protein) In some embodiments, a ligand is neomycin or its derivatives. In some

embodiments, a ligand is theophylline or its derivatives. In some embodiments, a ligand is glucose. In some embodiments a ligand is a cancer biomarker.

Vectors

[0061] In some embodiments, a polyA aptamer polynucleotide of the present disclosure can be introduced by a vector. In some embodiments, a vector can be a viral vector. Suitable viral vectors include, without limitation, lentiviral vectors, retroviral vectors, alphaviral, picornal (e.g., polio) vaccinial, adenoviral, adeno-associated viral, herpes viral, and fowl pox viral vectors.

Exemplary Uses Including Treatment

[0062] In accordance with the present disclosure, polyA aptamer polynucleotides and/or systems including one or more polyA aptamer polynucleotides, may be used in any of a variety of applications. For example, in some embodiments, a polyA aptamer polynucleotide of the present disclosure is used for treatment of an individual suffering from a disease, for example, by providing controllable expression of a therapeutic protein encoded by an expressible polynucleotide. In some embodiments, a disease is the lack of certain protein(s) caused by a genetic disorder. In some embodiments, a disease is diabetes, pre-diabetes, or complications from diabetes. In some embodiments, a disease is cancer. In some embodiments, a disease is muscular dystrophy. In some embodiments, a disease is hereditary Factor X deficiency. In some embodiments, a polyA aptamer polynucleotide of the present disclosure is provided in combination with other treatments for a disease. In some embodiments, a polyA aptamer polynucleotide of the present disclosure is used for inducing reprogramming of cells into pluripotent stem cells (induced pluripotent stem cells or iPSCs). In some embodiments, a polyA aptamer polynucleotide of the present disclosure is introduced or administered prior to, during, or subsequent to other treatments for a disease. In some embodiments, a therapeutic protein maybe or comprise insulin, growth hormones, dystrophin, albumin, factor IX, Oct4, Sox2, Klf4, cMyc, and any combination thereof.

[0063] In some embodiments, a system comprising a polyA aptamer polynucleotide may be used to provide information regarding whether or not a therapy is effective in a particular subject. In some embodiments wherein it is desirable to determine whether one or more therapies are effective in a subject, a system may be employed in the subject before the therapy is provided, such as to detect the presence or absence of a specific indicative compound for the therapy, and then after the therapy is provided one or more times the system may be employed in the subject to detect the presence or absence of the specific indicative compound. In other embodiments, the system is not employed for monitoring therapy until after the therapy is provided one or more times to the subject, such as to identify the presence or absence of a specific compound that is indicative of the efficacy of the therapy.

[0064] In some embodiments, polyA aptamer polynucleotides and/or systems including one or more polyA aptamer polynucleotides may be used as a biosensor. In accordance with various embodiments, provided systems may provide spatial and/or temporal information regarding a particular environment (e.g., an intracellular, extracellular, and/or environmental environment). For example, in some embodiments, a system comprising at least one polyA aptamer polynucleotide may be used to detect one or more specific molecular signatures in a subject and to allow for production of a desired expressible polypeptide in order to achieve a desired biological state in response to the presence of the molecular signature(s). In some embodiments, a molecular signature may be or comprise: the presence of a particular endogenous gene product (e.g., a disease-associated gene product/protein), the presence of a toxin, the presence of an exogenous gene product, the presence of a metabolite (e.g., a metabolite from an environmental contaminant), and any combination thereof.

[0065] In some embodiments, a polyA aptamer polynucleotide may comprise one or more reporter moieties (e.g., a reporter gene product, for example, an imaging reporter). In some embodiments, an expressible polynucleotide comprised in a polyA aptamer polynucleotide encodes a reporter gene product (e.g., protein). In some embodiments, a reporter gene product may be or comprise luciferase, green fluorescent protein, red fluorescent protein, β-galactosidase, infrared fluorescent proteins, near-infrared fluorescent proteins, opsin, and any combination thereof.

[0066] In some embodiments, a system comprising a polyA aptamer polynucleotide may encode both a reporter gene product and a therapeutic gene product. In some such embodiments, expression of the reporter gene product and the therapeutic gene product may be controlled by the same aptamer. In some embodiments, expression of the reporter gene product and the therapeutic gene product may be controlled by different aptamers.

Exemplification

[0067] The present examples describe a highly responsive gene regulation mechanism that harnesses the power of drug-inducible alternative splicing to control polyA cleavage. Figure 1 provides a representation of some embodiments of the present disclosure. As demonstrated in Figure 1A, when an engineered short intron (mini-IVS2) and a new polyA signal (in red) are artificially created at the 5' UTR of a transgene, efficient splicing of the intron and the cleavage of polyA signal lead to destruction of the second half of mRNA and therefore loss of gene expression. Binding of a specific ligand to the aptamers engineered as part of the Y-shape switch (in green) efficiently induces an alternative splicing. The ligand-induced alternative splicing results in the removal of the Y-shape structure and the artificial 5' UTR polyA signal. This in turn leads to the preservation of the intact mRNA and therefore the induced gene expression. Note, a second 3' splice site (3'ss) is built in the 5'UTR sequence. This 3' splice site is only activated after ligand (e.g., tetracycline, "Tc") binding to the aptamers. The 4MAZ sequence next to the Y structure is to reinforce the alternative splicing upon ligand binding.

[0068] Figure 1B provides a demonstration of a polyA switch comprising three aptamers as described herein. Each aptamer is located on one arm of the Y shape RNA structure. This Y-shape design has several important advantages: It incorporates 3 aptamers to control the polyA signal (pA) which is strategically placed at the central 3-way junction. By doing so, it harnesses the combined power of Tetracycline-binding effects generated from three different aptamers; The Y-shape structure is compact and requires overall shorter sequences to incorporate 3 aptamers; The Y-shape structure is designed to fold intrinsically during RNA biosynthesis. The three aptamers are arranged in a forward-forward-reverse orientation to minimize the chance of alternative folding between the aptamers. Further,

double-stranded RNA stems longer than 35bp are known to evoke innate immune response in cells. Therefore, all stems in the Y structure are made to be significantly shorter than 35pb to eliminate innate immune response.

[0069] Figure 1C provides an example (Y196CAA) of the nucleic acid sequence of a polyA switch as described herein. More than 370 constructs were designed and tested to extensively probe the effect of every component of the Y shape structure. These include (1) the length of each arm, (2) the sequence of each arm, (3) the loop of each arm, and (4) the sequence and size of the central 3-way junction where polyA signal is placed. The effect of modifications of those components are described further in these non-limiting examples.

Example 1: Modulation of PolyA Cleavage Signal

Location

[0070] Constructs were made to test additional Y-shape structures that are configured differently and with the polyA cleavage signal positioned differently. Four different constructs were made: B1-B4 where the polyA signal (in red) is placed near aptamer C and clamped by the 3-way junction (Figure 2A; B1 construct is shown). These showed no or minimal induction. An additional four constructs with polyA signal near the 3-way junction were made: T1-T4 (Figure 2B). These also showed minimal or moderate induction. Figure 2C exemplifies a polyA switch in which the 3 aptamers are stacked on each other without 3-way junction. Minimal induction was observed for this configuration. The particular Y-shape configuration shown in Figure 1B, in which polyA signal is placed close to the three way junction, is used for additional testing. In this configuration the three way junction bends with different orientation to provide a unique geometry for clamping the polyA signal. The stability of each arm is determined by two factors: the number of base pair and the composition of base pair (for example, G-C is more stable than A-U or G-U pair).

Number of PolyA Cleavage Signals

[0071] Tests were performed to evaluate the optimal number of polyA signal(s) in Y-shape structure. Figure 3A demonstrates testing of three structures from the Y series with 2 polyA signals indicated by the red boxes. Y1 shows ~12 fold induction, the highest in these three constructs. In this group, the majority of arm 3-1 is A-U or G-U pair, so it requires a longer stem to reach certain stability. As demonstrated in the figures, arms of the constructs exemplified herein comprise double stranded nucleic acid stems. Shorter arm 3-1 gives lower induction. Figure 3B further demonstrates effect of length of arms. Y5 to Y9 have only one polyA signal (red box) with variable length of arm3-1 (blue box) and arm2-1 (green box). The length of arm 3-1 and arm2-1 are shortened by 1 bp stepwise from Y5 to Y9. This one polyA configuration leads to better induction. Figure 3C demonstrates that when there are 2 polyA signals (Y6mut) in a row in arm1-2, the induction is reduced by approximately half. Y6mut: is identical to Y6 except that 2 polyA signals (red box) are embedded in arm1-2. Based on these results, the optimal number and position of polyA signal are determined: a single polyA signal partially embedded in arm1-2 and in 3-way junction. The configuration is used as the basis for further optimization.

Example 2: Optimization of Three Way Junction

[0072] Modifying the environment of a 3-way junction directly affects the clamping of polyA signal. Therefore, the performance of Y-shape switch is very sensitive to any change in the 3-way junction. Extensive mutation/insertion/deletion studies around the 3-way junction were performed to identify the best sequences. Figure 4A shows that an U to G mutation in Y22 doubles the induction, presumably because this mutation generates a new G-U base pair on arm3-1 that tightens the clamping of polyA signal. Figure 4B provides examples showing the effects of different 3-way junction sequences on induction. Figure 4C compares constructs having 3 bases vs. 1 base in box-1 of the three way junctions. Y107 to Y110 are derivatives of Y79 which has 3 bases in box 1. Y107 to Y110 have only one base in box1. Y107 performs similarly to Y79, indicating one unpaired base in box1 is sufficient. Figure 4D shows results of inserting one base into box 2 of the 3-way junction, which leads to subtle changes of folding in the 3-way junction. The results suggest that the best

configuration is one unpaired base in box2. For the constructs in Figure 4E the single base in box 1 and box2 were randomized. 16 combinations were tested and the results showed that Y127, Y130 and Y134 are the best among them when compared to the parental Y79 tested on the same day. Figure 4F shows further optimization of the constructs using Y130 as the basis. None of the modifications tested lead to significant improvement. Figure 4G shows additional modifications made relative to Y143 that resulted in little change in induction. Figure 4H shows additional modifications made relative to Y147. Y163 slightly improves induction while Y162 slightly decreases the induction as compared to Y147. Figure 4I shows additional modifications made relative to Y163. Y177 improves induction while Y178 decreases the induction compared to Y163. Figure 4J shows modifications made relative to Y152. These modifications lead to significant improvement compared to Y152. In particular, Y166 nearly doubles the induction. Y166 serves as the new basis for further optimization. Figure 4K shows additional modifications made relative to Y166. These modifications lead to significant improvement as compared to Y166. They also serve as the new bases for optimization.

[0073] Y174, Y175, Y176, and Y177 (See Figure 4L) are among the best 3-way junction sequences. All these constructs have a single base C or A in Box1 and Box2. In these constructs, the first 3 bases of polyA signal AAUAAA (red box) are open in the pocket of 3-way junction. The last 2 bases of polyA signal are embedded in arm 1-2.

[0074] Changing the polyA signal position relative to the pocket of the 3-way junction can alter induction capability (Figure 5). In Y135-Y140, changes made relative to Y101, the pocket of the 3-way junction is moved along the polyA signal. As a result, the polyA signal is embedded deeper in arm1-2. These modifications lead to lower induction. Y101mut, a derivative of Y101, contains a flipped C-G pair in arm2-1 (indicated by the red arrow) that removes a potential 3' splice site. Constructs Y141-Y159 are based on Y101mut. The 3-way junction pocket is moved along the polyA signal. The induction results of moving the 3-way junction pocket along the polyA signal are shown in the last part of Figure 5.

Example 3: Double Strand Stems

[0075] PolyA aptamer polynucleotide constructs as described herein comprise nucleic acid (e.g., RNA) double strand stems. Such double stranded regions are also referred to in the present disclosure as arms. Modifications of the length, stability, and nucleotide composition can affect the strength and effectiveness of the polyA aptamer polynucleotide.

[0076] Earlier results (using constructs Y1 to Y9, Figure 3) indicated that the stability of arm 3-1 needed to be within certain range. Arm3 is a very sensitive area because it is very close to the polyA signal. Minor changes in stability of arm3 can result in significant change in polyA signal clamping therefore the induction. Using Y35 as the basis, we made many modifications to optimize arm3. Figures 6A to 6B demonstrate the induction variation based on changes in arm 3. In these figures, the parental construct is on the right side, and the results of modification shown on the left side. Figure 6A shows results of modification of arm 3-1. Constructs Y43 to Y45 with decreasing strength of arm 3-2 are based on Y35; constructs Y188C and Y189C with decreasing strength of arm 3-2 are based on Y175; constructs Y188D and Y189D with decreasing strength of arm 3-2 are based on Y176. Constructs Y219A-224A with weaker strength of arm3-2 by changing a G-C pair to G-U pair at various locations are based on Y197. Figure 6B shows results of modification of arm 3-2. Constructs Y201 –Y203 are based on Y175. Constructs Y216B-217B with weaker arm 3-2 are based on Y208. The results demonstrate that increasing the length of arm3-2 and changing the loop sequence greatly reduce induction.

[0077] The majority of these modifications significantly reduce induction, and none surpasses Y35. Therefore, the arm3 of Y35 represents the optimal arm3 sequence for the Y shape structure of those tested. Some other parental constructs used for arm3 modification, such as Y175, Y197, and Y210, all share the same arm3 sequence of Y35.

[0078] Modifications to the double strand stems that are arm 2 (i.e., arm2-1 and arm2-2) alter the stability of arm 2. The modifications include variations in length, sequences, as well as point mutations that create mismatches in the stem (Figure 7).

Figure 7A shows the results of modification of arm2-2. Constructs Y48 to Y53 are based on Y35. Figure 7B shows the results of arm2-1 modifications. The results of these modifications indicate that induction is less sensitive to changes in the stability of arm2 as

compared to that of arm3. Presumably this is because that arm2 is not directly connected to polyA signal. Nonetheless, arm2 requires certain levels of stability to achieve good induction. Unstable arm2 leads to very low induction. The sequences of arm2 shown in these results are empirically determined. Some of the arm2 sequences are already within the optimal range of stability, and represent near optimal sequences that lead to very efficient induction. Further increase in stability either increases or decreases induction.

[0079] Figure 8 shows results of various modifications arm 1-2. Figure 9 shows results of various modifications of arm 1-1.

Example 4: Orientation of Aptamers

[0080] Orientation of each of the aptamers relative to the other aptamers may have an effect of the function of polyA aptamer polynucleotide. Figure 10A shows the results of constructs Y54 to Y57 which are based on Y35, with aptamer B orientation reversed. Reversing the orientation of aptamer B largely eliminates the induction. Figure 10B shows induction results of constructs Y240 to Y252 which are based on Y196CAA, with aptamer A orientation reversed. Reversing the orientation of aptamer A completely eliminates the induction regardless of the length of arm1-2.

Example 5: Contribution of each aptamer to induction

[0081] Figure 11A demonstrates the contribution of each aptamer of the Y-shape structure to induction. Each aptamer of the Y-shape structure can be disabled by an A to C mutation (arrows) in the binding pocket which eliminates the binding to its ligand tetracycline. NA: Aptamer A is disabled; NB: Aptamer B is disabled; NC: Aptamer C is disabled; NAB: Aptamers A and B are disabled; NBC: Aptamers B and C are disabled; NAC: Aptamers A and C are disabled. These results indicate that aptamer C contributes most significantly to the final induction. This is followed by aptamer B, then by aptamer A.

[0082] Figure 11B demonstrates the effect of removing aptamer A from the Y-shape structure. The boxes indicate the sequence removed for each construct. Removing aptamer

A retains moderate induction, although the level is significantly reduced compared to the parental Y196CAA.

Example 6: Modifications of 5'UTR

[0083] Figure 12A demonstrates that inserting CAA repeats (underlined) in the 5'UTR can alter induction levels. Here inserting CAA repeats in Y196, Y208, Y209, and Y211 all lead to higher induction. Inserting spacer sequences that contain CAA repeats into 5'UTR of Y301 results in variable effect on induction. These spacer sequences are only slightly different from each other, yet resulting in large difference in induction, indicating that this area is very sensitive to changes. Figure 12B shows some examples of testing a new 5'UTR sequence with a strong 3' splice site using S56 as parental construct. Figure 12C shows the results of adding intrinsically unstructured RNA sequences to the 5'UTR near the translational start ATG without using CAA repeats. These constructs are based on Y300 and Y305. Of the Y300-based constructs, Y329 is the best. While it does not surpass the performance of Y305, it has the advantage of not using the CAA repeats. Figure 12D shows that the insertion location of CAA repeats also significantly affects induction.

Example 7: Importance of G Quad Sequence

[0084] We tested the effects of G-quad sequence on induction. Figure 13A shows 3MAZ or CD44 G-quad reaches a similar induction level as compared to 2MAZ using Y196CAA as the parental. However, 4MAZ dramatically doubled the induction due to its ability to effectively induce alternative splicing. Figure 13B shows induction results when different G-quad sequences were tested to replace 4MAZ G-quad using the S56 construct as the parental. In these constructs, 4MAZ is replaced by the following: one CD44 G-quad 'TGGTGGTGGAATGGT' (S177), two CD44 G-quad 'TGGTGGTGGAATGGTAAATGGTGGTGGAATGGT' (S178), or four CD44 G-quad 'TGGTGGTGGAATGGTAAATGGTGGTGGAATGGTAAATGGTGGTGGAATGGTAAATGGTGGTGGAATGGT' (S179). The results indicate that the effect of 4MAZ is unique and cannot be replaced by other G-quad sequences. The 4MAZ sequence possesses unique properties and is a key element of the hybrid switch that requires both efficient polyA signal

cleavage and Tc-induced alternative splicing. Figure 14 further demonstrates the importance of the 4MAZ sequence. RT-PCR revealed the mechanism of drug-induced alternative splicing. In the absence of Tc, IVS2-spliced RNA is degraded by polyA cleavage (lane 1 and 3). The presence of Tc induces alternative splicing in both Y196CAA-2MAZ and Y196CAA-4MAZ (lane 2 and 4). Sanger sequencing confirmed that the Tc-induced band (lower band) contains the expected alternative splices RNA junction. Tc-induced alternative splicing is far more pronounced in Y196CAA-4MAZ as compared to Y196CAA-2MAZ (lane 4 vs. 2). With this induced alternative splicing, both the polyA signal and the Y-shape structure are removed in the presence of Tc, and the induction of protein expression is significantly increased.

Example 8: Modulating First 3' Splice Acceptor Site

[0085] To further optimize the mechanism of Tc-induced alternative splicing, we have extensively probed the effects of IVS2 3' splice site location and surrounding sequence/structure. The modifications include: embed IVS23' splice site into the arm1; move IVS2 3' splice site closer or further away from the aptamers binding site; put IVS2 3' splice site in a loose bulge in the arm1; change the length or stability of the arm1 that hosts IVS2 3' splice site; change splicing strength of IVS2 3' splice site. Figure 15 A shows the results of gradually moving IVS2 3' splice site into arm1-1 of Y196CAA-4MAZ (S1-S4). It shows also that when the IVS2 3' splice site is mutated from CAG to CCC (S5), the induction is nearly eliminated. Figure 15B demonstrates that when IVS 3' splice site is completely embedded into the arm1-1 near the Tc binding pocket of aptamer A (red arrow; S9), this splice site is strongly inhibited, resulting in very low induction. This indicates that clamping of IVS2 3' splice site by aptamer cannot be too strong. Further, diminishing the clamping effect of aptamer A by deleting part of its sequence (S9m) restores the induction. Moving the IVS 3' splice site along the arm 1 of S9m leads to S19 which is shorter and has similar induction levels compared to the parental S9m (Figure 15C). Figure 15D demonstrate the effect on induction when the IVS2 3' splice site CAG is placed in the bulge of arm1-2. S47 to S50 are based on S19. At 1 ug/mL Tc, most of them yield lower induction. At 5 ug/mL Tc, they give similar or higher induction compared to S19 with the exception of S50. Figure 15E shows results of changing the predicted strength of splicing by

mutating the base after IVS2 3' splice site. Changing the strength of IVS2 3' splice site does not significantly alter the induction in the S9m-based and Y196CAA-4MAZ based configurations. Figure 15F shows results of moving mini-IVS2 3' splice site further into or away from stem, which all lead to lower induction. Figure 15G shows effects of randomization of the three bases after the cag of the 3' splice site of mini-IVS2 to select the sequences with the highest performance. This group of constructs (in particular Y362, Y366, and Y367) exhibited superb switching efficiency, surpassing the performance of Y300 and Y301. Best NNN sequences identified by testing: Y344-based: Y359 (CAT) ,Y360 (TTT), Y361 (TGA), Y362 (TCT); Y358-based: Y363 (CAT), Y366 (TAC), Y367 (TTT)

Example 9: Modulating a Second 3' Splice Acceptor Site in 5'UTR

[0086] Assays were performed to test the effect of modulating the strength second 3' splice acceptor site in the 5'UTR. The 5'UTR sequence of Y196CAA-4MAZ located after 4MAZ and before the start codon ATG has the following sequence: gcggccgccaacaacaacaacaacaacaacaacaacaacaacaacaacataa**cag**tgttcactagcaacctcaaacagacaccA TG. Adding an additional branch point (S10), ppt (S11), or mutating CAG to CCC (S12) or AAG (S13) all lead to reduced induction (Figure 16A). To activate the correct 3' splice site (IVS2 3' splice site) in the absence of Tc, and in the presence of Tc (the second alternative 3' splice site), we used the constructs with short introns as the starting point and used a randomization approach to select the best three bases after the TAG in 5'UTR (TAGNNN) to improve the induction (Figure 16B). We also inserted these best three bases (NNN) intothe 5'UTR of Y329 to assess the performance (Figure 16C). Among these, Y344 performed best.

Example 10: Intron Size

[0087] We tested the effect of shortening the overall size of the hybrid switch by reducing the size of IVS2 intron. Figure 17 A shows exemplary intron sequences. Constructs S164 to S168 are similar to S159-S163 but have a branch point TACTAAC inserted at the same location before IVS2 3' splice site. The intron sequence of S164 is shown as an example:

Gtgagtctatgccagctaccattctgcttttatttatggttgggataaggctggattattctgagtccaagctaggccctttgctaatcat CttcaTACTAACctcttatcttcctctg**CAG.** Constructs S169 to S173 are similar to S159-S163 but have a branch point TACTAAC and one more 3' splice site CAG inserted at the same location before IVS2 3' splice site. The intron sequence of S169 is shown as an example: **GT**gagtctatgccagctaccattctgcttttatttatggttgggataaggctggattattctgagtccaagcTACTAACttttcctg tgcttcttcagacctcttatcttcctctg**CAG.** Reducing the IVS2 intron size from 476 bases to 120 -200 bases reduced the induction significantly (Figure 16B). The results from Y164 to Y173, which have different splicing elements added to enforce IVS2 intron splicing, lead to even lower induction compared to the ones without those added elements. This indicates that shortening or adding elements to IVS2 intron alter the choice of 3' splice site activation in the presence of Tc. Previously we have shown that CAA repeats alter the splicing strength of the 3' splice site in 5'UTR. Here the CAA repeats (in red) are to be removed from S159, S164, and S169. As compared to S56, S192 (with 120 bases intron) gave better induction at 1 ug/mL Tc, and similar induction at 5 ug/mL Tc. S192, which is more compact due to shorter intron, is used as a new basis for further modification.

Example 11: Addition of an Upstream out-of-frame AUG (μORF)

[0088] An upstream out-of-frame AUG was introduce to construct S192 to test the effect on reporter gene translation from IVS2-spliced transcript. The modifications include: (1) changing TAC to ATG immediately after IVS2 3' splice site to create a new start codon (red box), (2) changing the corresponding base on the other side of arm1 to maintain the base paring in the stem, and (3) mutating an in-frame stop codon tga into aga in arm2-1 (red arrow), so the translation from this new ATG can produce fairly long protein. See Figure 18A.

[0089] The sequence after IVS2 3' splice site **CAG** is shown. The new μORF is underlined:

ctg**CAG**ATGttcctcgagatctggggaggtgaagaatacgaccacctaataagattaccgaaaggcaatcttattaaaacatac cagatcttgagagggtgtttgtggcaaaacataccagatcgaattcgatctggggaggtgaagaatacgaccacctgctacaagtac ctaataaaCATtagCGGaGaaacataccactgtgtgttggttttttgtgtgttaacgggggaggggggaggaaaggggggagggg gaggaaaggggggaggggggaggaaaggggggaggggggagcggccgcca**taa**cagtgttcactagcaaccTcaaacagacacc

ATG. This approach significantly lowers the leakage expression from IVS2-spliced transcript, therefore significantly increases the induction as demonstrated by the result of S206.

[0090] This construct is further optimized by fine-tuning the 5'UTR sequence based on S206 (Figure 18B). All of these constructs demonstrate very good induction. These constructs are more compact due to shorter intron and partially deleted aptamer A. They perform very well at Tc concentration as low as 1 ug/mL, and reach as high as ~700 fold induction at 5 ug/mL.

[0091] In summary, in the process of optimizing Tc effects on splicing choice between IVS2 3' splice site and the alternative 3' splice site, we found that the best location for placing IVS2 3' splice site is to embed it inside the arm1 of Y structure. In order to place IVS2 3' splice site in that location, the aptamer A is deleted from the Y structure. Creating an upstream out-of-frame AUG (μORF) which eliminates reporter gene translation from IVS2-spliced transcript decreases leakage expression. Compared to Y196CAA-4MAZ, S222 (Figure 17C) shows higher induction in fold at lower drug concentration, higher gene expression levels, and perhaps more important, S222 is highly sensitive to Tc and performs well at low Tc concentrations.

Construct Performance

[0092] Figure 19A demonstrates comparison of performance of representative S series constructs relative to Y196CAA-4MAZ. Figure 18B shows a dose response of expression from the hybrid switch constructs visualized by microscopy.

[0093] To avoid potential immunogenicity generated by the protein translation of upstream open reading frames (μORF), we built another hybrid switch without the μORF aimed at surpassing the performance of S222. To build this new hybrid switch, we returned to the Y196CAA-4MAZ design as it has 3 aptamers as compared to 2 aptamers in S222. To further improve Y196CAA-4MAZ, we (1) use the mini-IVS2 intron with 120 bases, (2) optimizing the 3' splice site of mini-IVS2 sequence, (3) optimizing the 5'UTR sequence containing the downstream alternative 3' splice site. These efforts led to a group of constructs surpassing S222 in performance. The induction by tetracycline is so efficient that

they induce gene expression to 50% of the maximal level (EC_{50}) at a drug concentration as low as 0.5 to 1 μg/ml. This concentration of tetracycline can be routinely achieved in human serum using FDA-approved dosage, and is an order of magnitude lower than what has been previously achieved using any RNA-based gene regulation technology. Figure 19C demonstrates a comparison of the performance of these new constructs to that of S222. 5'UTR sequence of Y300: **gcggccgcCataacagtgttcactagcaTccCcaaacagacaccATG.** Y301: based on Y300 with modified 5'UTR **gcggccTTaATtaacagtgttcactaggacaccATG.**

Figure 19D demonstrates the performance of Y362 and Y367 determined by luciferase assays. Figure 19E shows the response to 1ug/ml tetracycline of Y362 and Y367 as determined by fluorescence activated cell sorting (FACS) using eGFP reporter signal. 'Induction in fold' in all results is calculated as the ratio of transgene expression in the presence vs. absence of tetracycline.

Example 12: Insertion of Riboswitch at Endogenous Location

[0094] The Y-shape polyA switch, when combined with CRISPR, creates a powerful technology platform to control the expression of any endogenous gene in mammalian genome. Figure 20 provides a schematic of using CD133, a stem cell membrane protein, to demonstrate the principle. The conditional gene expression of endogenous CD133 is achieved by inserting Y196 riboswitch at the 5'UTR of CD133 using CRISPR-Cas9 and a repair matrix. Figure 20A Top: three gRNAs (g1, g2, and g3) are used to specify the locations for CRISPR-Cas9 cleavage near the translational start of CD133. Figure 20A Bottom: repair matrix containing mini-CMV promoter, IVS2 intron, and Y196 riboswitch flanked by upstream and downstream homologous sequences to CD133 is used for repair. Figure 20B provides schematics of experimental procedures. Y196 riboswitch was first inserted into parental $CD133^-$ cells by CRISPR-Cas9. The successfully engineered cells then respond to Tc in a dose-dependent manner to turn on CD133 expression. FITC-conjugated antibody against CD133 protein was used to label and isolate the cells responding to Tc. Figure 19C shows that conditional expression of endogenous CD133 was regulated by Tc. CD133 expression in engineered cell clone (293T cell in this case) showed little or no background leakage. The CD133 expression is specifically induced by Tc, but not its analog

Doxy. ND: no drug treatment, Tc: Tetracycline, Doxy: Doxycycline. Cell clone was treated with or without drug for 2 days and then harvested for flow analysis. X-axis showed the intensity of antibody staining of individual cells. Figure 20D shows as expected, the CD133 protein induced by Tc (as revealed by FITC-anti CD133 antibody) was localized to cell membrane as normal endogenous CD133 protein would. The stable cell clone was treated with or without drug at 2 μg/ ml for 2 days and then harvested for Image flow analysis (Amnis). Again, the induction is clearly specific to Tc but not Doxy.

[0095] The data described represent a highly responsive gene regulation mechanism that harnesses the power of drug-inducible alternative splicing to control polyA cleavage. The combination engineered creates a sensitive RNA-based switch that can be controlled by small molecule drugs and enables tight regulation of gene expression in mammalian cells. In contrast to other reported methods, this hybrid switch technology described herein exhibits very low leaky expression, and effectively turns on the transgene expression close to 700-folds in human cells. Furthermore, the induction by tetracycline is so efficient that it induces gene expression to 50% of the maximal level (EC_{50}) at a drug concentration as low as 0.5 to 1 μg/ml. This concentration of tetracycline can be routinely achieved in human serum using FDA-approved dosage, and is an order of magnitude lower than what has been previously achieved using other RNA-based gene regulation technology.

[0096] This hybrid switch technology therefore is advantageously safe to use in human patients for controlling the expression of a therapeutic gene or transgene. The present disclosure thus satisfies a long-felt need in the art to provide a highly efficient and non-immunogenic technology to regulate genes of interest in cells at a drug concentration that is safe for human consumption.

Example 13: Combination of Single Base Changes at Three Locations

[0097] A combination of three base changes to the sequence of the Y-shape structure was tested to determine the cumulative effects on induction performance of the poly A aptamer. The three mutations, as noted in Figure 21, consist of an 'A' deletion in Arm1-1; an 'A' to 'G' change to close the unpaired break in Arm2-2; and an "A" insertion in the 3-way junction preceding the polyA signal. These mutations were implemented using four

different parental constructs that have different bases posterior to mini-IVS2 intron. In all, 12 constructs, described in Table 1, were designed to probe the cumulative effects.

Table 1	**Y359**	**Y392**	**Y395**	**Y360**	**Y393**	**Y396**	**Y361**	**Y394**	**Y397**	**Y362C**	**Y362**	**Y387**
'A' deletion in Arm1-1	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
'A' to 'G' change in Arm2-2	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
"A" insertion in 3 way junction	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Bases after mini-IVS2	CAT	CAT	CAT	TTT	TTT	TTT	TGA	TGA	TGA	TCT	TCT	TCT

[0098] Figure 22 demonstrates that the combination of the three single base changes significantly increase induction at lower drug concentration. Additionally, Figures 23A and 23B demonstrate dose response analysis for constructs Y362 and Y387. Y362 and Y387 effectively turn on the transgene expression up to 650~700-folds in 293T cells using only 1ug/ml of tetracycline. For both constructs, the induction by tetracycline reaches 50% of the maximal level (EC_{50}) at as low as 0.5 to 1 μg/ml Tc using the maximum induction in fold as the EC_{100} reference (Fig. 23A). Calculations using the maximum expression level of parental construct (HDM-Luc, which has similar sequence but without the Y-shape structure) as the EC_{100} reference also show similar EC_{50} values as low as 0.5 to 1.2 μg/ml (Fig. 23B). Y387 is a particularly effective design as it exhibits an EC_{50} value of 0.5 μg/ml regardless of the EC_{100} references used.

Example 14: Methods

[0099] Assays described in the figures filed herewith were performed as follows:

Luciferase assay

[0100] Cells were seeded in 96-well plates at a density of 25000-30000 cells/well. After 24 hours of incubation, each well was transfected with 50 ng of DNA vectors and were

incubated with culture medium containing none or various concentration of tetracycline for an additional 18 hours. Luciferase activity was measured in relative light units (RLU) with a Polarstar Omega plate reader (BMG Labtech, USA). To make 36 mL of assay buffer, 144 µL 1M DTT, 108 µL 0.1 M ATP, 252 µL 0.1M luciferin and 360 µl 0.05M CoA were added to 35 mL of basic buffer (25mM Tricine, 0.5mM EDTA-Na_2, 0.54mM Na-triphosphate, 16.3 mM MgSO4.7H_2O,and 0.8% Triton X-100). After the cell medium was removed, 40 µL of assay buffer was added to each well, and luciferase activity was read twice with the Polarstar Omega plate reader. Induction in fold is calculated as the ratio of transgene expression in the presence vs absence of tetracycline.

RT-PCR

[0101] Cells transfected with the respective constructs were grown 18 hours at 37 °C in medium in the absence or presence of tetracycline. Total RNA was isolated according to the protocol supplied with RiboPure™ RNA Purification Kit (Ambion, Austin, TX). For RT-PCR, RT was performed using SuperScript III (invitrogen, Carlsbad, CA) according to manufacturer's protocol and PCR was performed using the primers targeting the beginning of the transcript and reporter gene.

Fluorescence Microscopy

[0102] Cells were seeded in 12-well plates at a density of 1.2×10^5 cells/well. After 24 hours of incubation, each well was transfected with 500 ng of DNA vectors and were incubated with culture medium containing none or various concentration of tetracycline for an additional 18 hours. Images were taken on a fluorescence microscope (Zeiss Axiovert 40CFL) at a magnification of 200x.

Example 15: Exemplary Construct Sequences

[0103] The following sequences are additional examples of embodiments of components of the system described herein. The sequences are provided as DNA sequences that when transcribed components of form RNA aptamers :

+1: Transcriptional start
Black: 5' leading RNA sequence
Underline: IVS2 intron or mini-IVS2 intron
Bold: Y-shape polyA switch (with 4MAZ underlined)
Italic: 5'UTR
ATG: Translational start in bold

Y196CAA-4MAZ

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
ATGGGACCCTTGATGTTTTCTTTCCCCTTCTTTTCTATGGTTAAGTTCATGTCATA
GGAAGGGGAGAAGTAACAGGGTACACATATTGACCAAATCAGGGTAATTTTGC
ATTTGTAATTTTAAAAAATGCTTTCTTCTTTTAATATACTTTTTTGTTTATCTTATT
TCTAATACTTTCCCTAATCTCTTTCTTTCAGGGCAATAATGATACAATGTATCAT
GCCTCTTTGCACCATTCTAAAGAATAACAGTGATAATTTCTGGGTTAAGGCAAT
AGCAATATTTCTGCATATAAATATTTCTGCATATAAATTGTAACTGATGTAAGAG
GTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATTCTGCTTTTATTTTAT
GGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTTTGCTAATCAT
GTTCATACCTCTTATCTTCCTCCCACAGCTCCTGGGCAACGTGCTGGTCTGTGTG
CTGGCCCATCACTTTGGCAAAGAATTG**GGCTAGCCACACACACAAATCTGGGG**
AGGTGAAGAATACGACCACCTGCGTTTTATACTTCCACGAGATCTGGGGAG
GTGAAGAATACGACCACCTAATAAGATTACCGAAAGGCAATCTTATTAAAA
CATACCAGATCTTGTGAGGGTGTTTGTGGCAAAACATACCAGATCGAATTC
GATCTGGGGAGGTGAAGAATACGACCACCTGCTACAAGTACCTAATAAAGT
ATAAAGTGCAAAACATACCAGATCTGTGTGTTGGTTTTTTGTGTGTTAACG
GGGGAGGGGGAGGAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGGA
AAGGGGGAGGGGGA*GCGGCCGCCAACAACAACAACAACAACAACAACAACAACAACAA*
*CAACAACATAACAGTGTTCACTAGCAACCTCAAACAGACACC***ATG** (SEQ ID NO.:6)

Y208

+1TGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGAT
CCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCTATGGGACCC
TTGATGTTTTCTTTCCCCTTCTTTTCTATGGTTAAGTTCATGTCATAGGAAGGGGA
GAAGTAACAGGGTACACATATTGACCAAATCAGGGTAATTTTGCATTTGTAATT
TTAAAAAATGCTTTCTTCTTTTAATATACTTTTTTGTTTATCTTATTTCTAATACTT

TCCCTAATCTCTTTCTTTCAGGGCAATAATGATACAATGTATCATGCCTCTTTGC
ACCATTCTAAAGAATAACAGTGATAATTTCTGGGTTAAGGCAATAGCAATATTT
CTGCATATAAATATTTCTGCATATAAATTGTAACTGATGTAAGAGGTTTCATATT
GCTAATAGCAGCTACAATCCAGCTACCATTCTGCTTTTATTTTATGGTTGGGATA
AGGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTTTGCTAATCATGTTCATACCT
CTTATCTTCCTCCCACAGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATC
ACTTTGGCAAAGAATTG**GCTAGCCACACACACAAATCTGGGGAGGTGAAGA**
ATACGACCACCTGCGTTTTATACTTCCGcGAGATCTGGGGAGGTGAAGAAT
ACGACCACCTAATAAGATTACCGAAAGGCAATCTTATTAAAACATACCAGA
TCTTGCGAGGGTGTTTGTGGCAAAACATACCAGATCGAATTCGATCTGGGG
AGGTGAAGAATACGACCACCTGCTACAAGTACCTAATAAAGTATAAAGTGC
AAAACATACCAGATCTGTGTGTTGGTTTTTTGTGTGTTAACGGGGGAGGGG
GAGGAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGGAAAGGGGGAG
GGGGA*GCGGCCGCCAACAACAACAACAACAACAACAACAACAACAACAACATAA*
*CAGTGTTCACTAGCAACCTCAAACAGACACC***ATG** (SEQ ID NO.:7)

Y209

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
ATGGGACCCTTGATGTTTTCTTTCCCCTTCTTTTCTATGGTTAAGTTCATGTCATA
GGAAGGGGAGAAGTAACAGGGTACACATATTGACCAAATCAGGGTAATTTTGC
ATTTGTAATTTTAAAAAATGCTTTCTTCTTTTAATATACTTTTTTGTTTATCTTATT
TCTAATACTTTCCCTAATCTCTTTCTTTCAGGGCAATAATGATACAATGTATCAT
GCCTCTTTGCACCATTCTAAAGAATAACAGTGATAATTTCTGGGTTAAGGCAAT
AGCAATATTTCTGCATATAAATATTTCTGCATATAAATTGTAACTGATGTAAGAG
GTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATTCTGCTTTTATTTTAT
GGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTTTGCTAATCAT
GTTCATACCTCTTATCTTCCTCCCACAGCTCCTGGGCAACGTGCTGGTCTGTGTG
CTGGCCCATCACTTTGGCAAAGAATTG**GCTAGCCACACACACAAATCTGGGG**
AGGTGAAGAATACGACCACCTGCGTTTTATACTTCCGcGAGATCTGGGGAG
GTGAAGAATACGACCACCTAATAAGATTACCGAAAGGCAATCTTATTAAAA
CATACCAGATCTTGCGAGGGTGTTTGTGGCAAAACATACCAGATCGAATTC
GATCTGGGGAGGTGAAGAATACGACCACCTGCTACAAGTACCTAAATAAAG
TATAAAGTGCAAAACATACCAGATCTGTGTGTTGGTTTTTTGTGTGTTAACG
GGGGAGGGGGAGGAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGGA
AAGGGGGAGGGGGA*GCGGCCGCCAACAACAACAACAACAACAACAACAACAACAACAA*
*CAACAACATAACAGTGTTCACTAGCAACCTCAAACAGACACC***ATG** (SEQ ID NO.:8)

Y211

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
ATGGGACCCTTGATGTTTTCTTTCCCCTTCTTTTCTATGGTTAAGTTCATGTCATA
GGAAGGGGAGAAGTAACAGGGTACACATATTGACCAAATCAGGGTAATTTTGC

ATTTGTAATTTTAAAAAATGCTTTCTTCTTTTAATATACTTTTTTGTTTATCTTATT
TCTAATACTTTCCCTAATCTCTTTCTTTCAGGGCAATAATGATACAATGTATCAT
GCCTCTTTGCACCATTCTAAAGAATAACAGTGATAATTTCTGGGTTAAGGCAAT
AGCAATATTTCTGCATATAAATATTTCTGCATATAAATTGTAACTGATGTAAGAG
GTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATTCTGCTTTTATTTTAT
GGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTTTGCTAATCAT
GTTCATACCTCTTATCTTCCTCCCACAGCTCCTGGGCAACGTGCTGGTCTGTGTG
CTGGCCCATCACTTTGGCAAAGAATTG**GCTAGCCACACACACAAATCTGGGG**
AGGTGAAGAATACGACCACCTGCGTTTTATACTTCCAcGAGATCTGGGGAG
GTGAAGAATACGACCACCTAATAAGATTACCGAAAGGCAATCTTATTAAAA
CATACCAGATCTTgTGAGGGTGTTTGTGGCAAAACATACCAGATCGAATTC
GATCTGGGGAGGTGAAGAATACGACCACCTGCTACAAGTACCTAAATAAAG
TATAAAGTGCAAAACATACCAGATCTGTGTGTTGGTTTTTTGTGTGTTAACG
GGGGAGGGGGAGGAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGGA
AAGGGGGAGGGGGA*GCGGCCGCCAACAACAACAACAACAACAACAACAACAACAA*
*CAACAACATAACAGTGTTCACTAGCAACCTCAAACAGACACC***ATG** (SEQ ID NO.:9)

Y226

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
ATGGGACCCTTGATGTTTTCTTTCCCCTTCTTTTCTATGGTTAAGTTCATGTCATA
GGAAGGGGAGAAGTAACAGGGTACACATATTGACCAAATCAGGGTAATTTTGC
ATTTGTAATTTTAAAAAATGCTTTCTTCTTTTAATATACTTTTTTGTTTATCTTATT
TCTAATACTTTCCCTAATCTCTTTCTTTCAGGGCAATAATGATACAATGTATCAT
GCCTCTTTGCACCATTCTAAAGAATAACAGTGATAATTTCTGGGTTAAGGCAAT
AGCAATATTTCTGCATATAAATATTTCTGCATATAAATTGTAACTGATGTAAGAG
GTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATTCTGCTTTTATTTTAT
GGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTTTGCTAATCAT
GTTCATACCTCTTATCTTCCTCCCACAGCTCCTGGGCAACGTGCTGGTCTGTGTG
CTGGCCCATCACTTTGGCAAAGAATTG**GCTAGCCACACACACAAACCTGGGG**
AGGTGAAGAATACGACCACCTGCGTTTTATACTTCCACGAGATCTGGGGAG
GTGAAGAATACGACCACCTAATAAGATTACCGAAAGGCAATCTTATTAAAA
CATACCAGATCTTGTGAGGGTGTTTGTGGCAAAACATACCAGATCGAATTC
GATCTGGGGAGGTGAAGAATACGACCACCTGCTACAAGTACCTAATAAAGT
ATAAAGTGCAAAACATACCAGGTCTGTGTGTTGGTTTTTTGTGTGTTAACG
GGGGAGGGGGAGGAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGGA
AAGGGGGAGGGGGA*GCGGCCGCCAACAACAACAACAACAACAACAACAACAACAACAA*
*CAACAACATAACAGTGTTCACTAGCAACCTCAAACAGACACC***ATG** (SEQ ID NO.:10)

Y227

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
ATGGGACCCTTGATGTTTTCTTTCCCCTTCTTTTCTATGGTTAAGTTCATGTCATA
GGAAGGGGAGAAGTAACAGGGTACACATATTGACCAAATCAGGGTAATTTTGC
ATTTGTAATTTTAAAAAATGCTTTCTTCTTTTAATATACTTTTTTGTTTATCTTATT

TCTAATACTTTCCCTAATCTCTTTCTTTCAGGGCAATAATGATACAATGTATCAT
GCCTCTTTGCACCATTCTAAAGAATAACAGTGATAATTTCTGGGTTAAGGCAAT
AGCAATATTTCTGCATATAAATATTTCTGCATATAAATTGTAACTGATGTAAGAG
GTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATTCTGCTTTTATTTTAT
GGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTTTGCTAATCAT
GTTCATACCTCTTATCTTCCTCCCACAGCTCCTGGGCAACGTGCTGGTCTGTGTG
CTGGCCCATCACTTTGGCAAAGAATTG**GCTAGCCACACACACAAACCTGGGG**
AGGTGAAGAATACGACCACCTGCGTTTTATACTTCCAcGAGATCTGGGGAG
GTGAAGAATACGACCACCTAATAAGATTACCGAAAGGCAATCTTATTAAAA
CATACCAGATCTTgTGAGGGTGTTTGTGGCAAAACATACCAGATCGAATTC
GATCTGGGGAGGTGAAGAATACGACCACCTGCTACAAGTACCTAAATAAAG
TATAAAGTGCAAAACATACCAGGTCTGTGTGTTGGTTTTTTGTGTGTTAAC
GGGGGAGGGGGAGGAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGG
AAAGGGGGAGGGGGA*GCGGCCGCCAACAACAACAACAACAACAACAACAACAACA*
*ACAACAACATAACAGTGTTCACTAGCAACCTCAAACAGACACC***ATG** (SEQ ID NO.:11)

Y300

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**CACACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCATAACAGTG*
*TTCACTAGCATCCCCAAACAGACACC***ATG** (SEQ ID NO.:12)

Y329

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**CACACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG

GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCATAACAGTG TTCACTAGCATCCCCCAGACCATCTACCACCGACACC***ATG** (SEQ ID NO.:13)

Y305

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC AG**CACACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA***GCGGCCGCCTTAATTAAC AGTGTTCACTAG*[*CAT*]*CAACAACAACAACAACAACAACAACAACAACAACGACACC***ATG** (SEQ ID NO.:14)

Y305D1

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC AG**CACACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA***GCGGCCGCCTTAATTAAC AGTGTTCACTAG*[*TAG*]*CAACAACAACAACAACAACAACAACAACAACAACGACACC***ATG** (SEQ ID NO.:15)

Y305D2

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC AG**CACACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG**

TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCTTAATTAAC*
AGTGTTCACTAG[*ACA*]*CAACAACAACAACAACAACAACAACAACAACGACACC***ATG**
(SEQ ID NO.:16)

Y305D3

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**CACACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCTTAATTAAC*
AGTGTTCACTAG[*AAC*]*CAACAACAACAACAACAACAACAACAACAACGACACC***ATG**
(SEQ ID NO.:17)

Y305D4

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**CACACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCTTAATTAAC*
AGTGTTCACTAG[*TGC*]*CAACAACAACAACAACAACAACAACAACAACGACACC***ATG**
(SEQ ID NO.:18)

Y305D5

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**CACACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG

TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCTTAATTAAC*
AGTGTTCACTAG[*TTG*]*CAACAACAACAACAACAACAACAACAACAACGACACC***ATG**
(SEQ ID NO.:19)

Y305D6

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**CACACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCTTAATTAAC*
AGTGTTCACTAG[*ACC*]*CAACAACAACAACAACAACAACAACAACAACGACACC***ATG**
(SEQ ID NO.:20)

Y305D7

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**CACACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCTTAATTAAC*
AGTGTTCACTAG[*CCC*]*CAACAACAACAACAACAACAACAACAACAACGACACC***ATG**
(SEQ ID NO.:21)

Y301

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**CACACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**

ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGCGGCC*TTAATTAACAGT*
*GTTCACTAGGACACC***ATG** (SEQ ID NO.:22)

Y305D9

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**CACACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCTTAATTAAC*
AGTGTTCACTAG[*AGG*]*CAACAACAACAACAACAACAACAACAACAACGACACC***ATG**
(SEQ ID NO.:23)

Y305D10

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**CACACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCTTAATTAAC*
AGTGTTCACTAG[*TGA*]*CAACAACAACAACAACAACAACAACAACAACGACACC***ATG**
(SEQ ID NO.:24)

Y305D11

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT

TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**CACACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCTTAATTAAC*
AGTGTTCACTAG[*TCC*]*CAACAACAACAACAACAACAACAACAACAACGACACC***ATG**
(SEQ ID NO.:25)

Y305D12

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**CACACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCTTAATTAAC*
AGTGTTCACTAG[*CCT*]*CAACAACAACAACAACAACAACAACAACAACGACACC***ATG**
(SEQ ID NO.:26)

Y305D13

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**CACACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCTTAATTAAC*
AGTGTTCACTAG[*TCT*]*CAACAACAACAACAACAACAACAACAACAACGACACC***ATG**
(SEQ ID NO.:27)

Y344

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AGCAC**ACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCATAACAGTG*
*TTCACTAGCCCCCCCCAGACCATCTACCACCGACACC***ATG** (SEQ ID NO.:28)

Y359

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AGCAT**ACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCATAACAGTG*
*TTCACTAGCCCCCCCCAGACCATCTACCACCGACACC***ATG** (SEQ ID NO.:29)

Y360

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AGTTT**ACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCATAACAGTG*
*TTCACTAGCCCCCCCCAGACCATCTACCACCGACACC***ATG** (SEQ ID NO.:30)

Y361

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AGTGA**ACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCATAACAGTG*
*TTCACTAGCCCCCCCCAGACCATCTACCACCGACACC***ATG** (SEQ ID NO.:31)

Y362

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AGTCT**ACACAC**AA**TCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTATA**
CTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTACC
GAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGCA
AAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCTG
CTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTGT
TGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGG
AAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCATAACAGTGT*
*TCACTAGCCCCCCCCAGACCATCTACCACCGACACC***ATG** (SEQ ID NO.:32)

Y358

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AGCAC**ACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT**
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCTTAATTAAC*
*AGTGTTCACTAGAGCCAACAACAACAACAACAACAACAACAACAACGACACC***ATG**
(SEQ ID NO.:33)

Y363

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**CAT**ACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCTTAATTAAC*
*AGTGTTCACTAGAGCCAACAACAACAACAACAACAACAACAACAACAACGACACC***ATG**
(SEQ ID NO.:34)

Y366

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**TAC**ACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCTTAATTAAC*
*AGTGTTCACTAGAGCCAACAACAACAACAACAACAACAACAACAACAACGACACC***ATG**
(SEQ ID NO.:35)

Y367

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**TTT**ACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCTTAATTAAC*
*AGTGTTCACTAGAGCCAACAACAACAACAACAACAACAACAACAACAACGACACC***ATG**
(SEQ ID NO.:36)

Y375

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**TCT**ACACACAAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTAT
ACTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTAC
CGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGC
AAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCT
GCTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTG
TTGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAG
GAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCATAACAGTG*
*TTCACTAGCCCCCCCCAGACCATCTACCACCGACACC***ATG**

Y376

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**TTT**ACACAC**AA**TCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTATA
CTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTACC
GAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGCA
AAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCTG
CTACAAGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTGT
TGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGG
AAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCTTAATTAACA*
*GTGTTCACTAGAGCCAACAACAACAACAACAACAACAACAACAACGACACC***ATG**

S206

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
ATGCCAGCTACCATTCTGCTTTTATTTTATGGTTGGGATAAGGCTGGATTATTCT
GAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**ATGTTCCTCGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGAT**
TACCGAAAGGCAATCTTATTAAAACATACCAGATCTTGAGAGGGTGTTTGT
GGCAAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCA
CCTGCTACAAGTACCTAATAAACATTAGCGGAGAAACATACCACTGTGTGT
TGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGG

AAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCATAACAGTGT TCACTAGCAACCTCAAACAGACACC***ATG** (SEQ ID NO.:37)

S210

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT ATGCCAGCTACCATTCTGCTTTTATTTTATGGTTGGGATAAGGCTGGATTATTCT GAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC AG**ATGTTCCTCGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGAT TACCGAAAGGCAATCTTATTAAAACATACCAGATCTTGAGAGGGTGTTTGT GGCAAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCA CCTGCTACAAGTACCTAATAAACATTAGCGGAGAAACATACCACTGTGTGT TGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGG AAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA***GCGGCCGCCATAACAGTGT TCACTAGCAACCCCAAACAGACACC***ATG** (SEQ ID NO.:38)

S211

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT ATGCCAGCTACCATTCTGCTTTTATTTTATGGTTGGGATAAGGCTGGATTATTCT GAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC AG**ATGTTCCTCGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGAT TACCGAAAGGCAATCTTATTAAAACATACCAGATCTTGAGAGGGTGTTTGT GGCAAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCA CCTGCTACAAGTACCTAATAAACATTAGCGGAGAAACATACCACTGTGTGT TGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGG AAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA***GCGGCCGCCACCCATAACAG TGTTCACTAGCAACCCCAAACAGACACC***ATG** (SEQ ID NO.:39)

S212

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT ATGCCAGCTACCATTCTGCTTTTATTTTATGGTTGGGATAAGGCTGGATTATTCT GAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC AG**ATGTTCCTCGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGAT TACCGAAAGGCAATCTTATTAAAACATACCAGATCTTGAGAGGGTGTTTGT GGCAAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCA CCTGCTACAAGTACCTAATAAACATTAGCGGAGAAACATACCACTGTGTGT TGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGG AAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA***GCGGCCGCCACCATGATAA CAGTGTTCACTAGCAACCCCAAACAGACACC***ATG** (SEQ ID NO.:40)

S213

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
ATGCCAGCTACCATTCTGCTTTTATTTTATGGTTGGGATAAGGCTGGATTATTCT
GAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**ATGTTCCTCGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGAT**
TACCGAAAGGCAATCTTATTAAAACATACCAGATCTTGAGAGGGTGTTTGT
GGCAAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCA
CCTGCTACAAGTACCTAATAAACATTAGCGGAGAAACATACCACTGTGTGT
TGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGG
AAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCACCACG**ATAA*
*CAGTGTTCACTAGCAACCCCAAACAGACACC***ATG** (SEQ ID NO.:41)

S214

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
ATGCCAGCTACCATTCTGCTTTTATTTTATGGTTGGGATAAGGCTGGATTATTCT
GAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**ATGTTCCTCGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGAT**
TACCGAAAGGCAATCTTATTAAAACATACCAGATCTTGAGAGGGTGTTTGT
GGCAAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCA
CCTGCTACAAGTACCTAATAAACATTAGCGGAGAAACATACCACTGTGTGT
TGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGG
AAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCACCATAACAG*
*TGTTCACTAGCA7CCCCAAACAGACACC***ATG** (SEQ ID NO.:42)

S215

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
ATGCCAGCTACCATTCTGCTTTTATTTTATGGTTGGGATAAGGCTGGATTATTCT
GAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**ATGTTCCTCGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGAT**
TACCGAAAGGCAATCTTATTAAAACATACCAGATCTTGAGAGGGTGTTTGT
GGCAAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCA
CCTGCTACAAGTACCTAATAAACATTAGCGGAGAAACATACCACTGTGTGT
TGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGG
AAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCACCATAACAG*
*TGTTCACCAGCA7CCCCAAACAGACACC***ATG** (SEQ ID NO.:43)

S222

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
ATGCCAGCTACCATTCTGCTTTTATTTTATGGTTGGGATAAGGCTGGATTATTCT
GAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC

AG**ATGTTCCTCGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGAT**
TACCGAAAGGCAATCTTATTAAAACATACCAGATCTTGAGAGGGTGTTTGT
GGCAAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCA
CCTGCTACAAGTACCTAATAAACATTAGCGGAGAAACATACCACTGTGTGT
TGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGG
AAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCATAACAGTGT*
*TCACTAGCATCCCCCAAACAGACACC***ATG** (SEQ ID NO.:44)

S223

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
ATGCCAGCTACCATTCTGCTTTTATTTTATGGTTGGGATAAGGCTGGATTATTCT
GAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**ATGTTCCTCGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGAT**
TACCGAAAGGCAATCTTATTAAAACATACCAGATCTTGAGAGGGTGTTTGT
GGCAAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCA
CCTGCTACAAGTACCTAATAAACATTAGCGGAGAAACATACCACTGTGTGT
TGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGG
AAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCATAACAGTGT*
*TCACCAGCATCCCCCAAACAGACACC***ATG** (SEQ ID NO.:45)

S272

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCT
TAAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGC
AG**ATTTTCCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGAT**
TACCGAAAGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGT
GGCAAAACATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCA
CCTGCTACAAGTACCTAATAAAAATTAGCGGAGAAACATACCACTGTGTGT
TGGTTTTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGG
AAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCATAACAGTGT*
*TCACTAGCATCCCCCAAACAGACACC***ATG** (SEQ ID NO.:46)

Y387

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCTT
AAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTCTGA
GTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGCAGT
CTACACACAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTATACTTC
CACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTGCCGAAA
GGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGCAAAACA
TACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCTGCTACA

**AGTACCTAAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTGTTGGTT
TTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGGAAAG
GGGGAGGGGGAGGAAAGGGGGAGGGGGA***GCGGCCGC****CCATAACAGTGTTCACT**
*AGCCCCCCCCCAGACCATCTACCACCGACACC***ATG** (SEQ ID NO.:50)

Y392

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCTT
AAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTCTGA
GTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGCAGC
**ATACACACAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTATACTTC
CACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTACCGAAA
GGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGCAAAACA
TACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCTGCTACA
AGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTGTTGGTTTT
TTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGGAAAGGG
GGAGGGGGAGGAAAGGGGGAGGGGGA***GCGGCCGCCATAACAGTGTTCACTAGC
*CCCCCCCCAGACCATCTACCACCGACACC***ATG** (SEQ ID NO.:51)

Y393

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCTT
AAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTCTGA
GTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGCAGTT
**TACACACAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTATACTTCC
ACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTACCGAAAG
GCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGCAAAACAT
ACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCTGCTACAA
GTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTGTTGGTTTTT
TGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGGAAAGGGG
GAGGGGGAGGAAAGGGGGAGGGGGA***GCGGCCGCCATAACAGTGTTCACTAGCC
*CCCCCCCAGACCATCTACCACCGACACC***ATG** (SEQ ID NO.:52)

Y394

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCTT
AAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTCTGA

GTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGCAGT
GAACACACAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTATACTT
CCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTACCGAA
AGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGCAAAAC
ATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCTGCTACA
AGTACCTAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTGTTGGTTTT
TTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGGAAAGGG
GGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCATAACAGTGTTCACTAGC*
*CCCCCCCAGACCATCTACCACCGACACC***ATG** (SEQ ID NO.:53)

Y395

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCTT
AAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTCTGA
GTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGCAGCA
TACACACAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTATACTTCC
ACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTGCCGAAAG
GCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGCAAAACAT
ACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCTGCTACAA
GTACCTAAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTGTTGGTTT
TTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGGAAAGG
GGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCATAACAGTGTTCACTA*
*GCCCCCCCCCAGACCATCTACCACCGACACC***ATG** (SEQ ID NO.:54)

Y396

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCTT
AAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTCTGA
GTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGCAGTT
TACACACAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTATACTTCC
ACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTGCCGAAAG
GCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGCAAAACAT
ACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCTGCTACAA
GTACCTAAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTGTTGGTTT
TTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGGAAAGG
GGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCATAACAGTGTTCACTA*
*GCCCCCCCCCAGACCATCTACCACCGACACC***ATG** (SEQ ID NO.:55)

Y397

+1TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCCCTCGAAGCTGATCCTGAGAACTTCAGGGTGAGTCTT
AAGCCAGCTACCATTCTGCTTTTATTTTATCGTTGGGATAAGGCTGGATTATTCTGA
GTCCAAGCTAGGCCCTTTTGCTAATCATCTTCATACCTCTTATCTTCCTCTGCAGT
GAACACACCAATCTGGGGAGGTGAAGAATACGACCACCTGCGTTTTATACTT
CCACGAGATCTGGGGAGGTGAAGAATACGACCACCTAATAAGATTGCCGAA
AGGCAATCTTATTAAAACATACCAGATCTTGTGAGGGTGTTTGTGGCAAAAC
ATACCAGATCGAATTCGATCTGGGGAGGTGAAGAATACGACCACCTGCTACA
AGTACCTAAATAAAGTATAAAGTGCAAAACATACCAGATCTGTGTGTTGGTT
TTTTGTGTGTTAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGGAAAG
GGGGAGGGGGAGGAAAGGGGGAGGGGGA*GCGGCCGCCATAACAGTGTTCACT*
*AGCCCCCCCCCAGACCATCTACCACCGACACC***ATG** (SEQ ID NO.:56)

Equivalents

[0104] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. The scope of the present invention is not intended to be limited to the above Description, but rather is as set forth in the following claims:

Claims

We claim:

1. A system for modulating gene expression, comprising a polyA aptamer polynucleotide that comprises in a 5' to 3' direction:

a) a 5’ splice donor site;

b) an engineered intron;

c) a first 3’ splice acceptor site;

d) a polyA switch comprising two or more ligand-binding aptamers with one or more ligand binding pockets, and at least one polyA cleavage signal therein;

e) a second 3’ splice acceptor site; and

f) a nucleic acid sequence encoding an expressible polypeptide.

2. The system of claim 1, wherein the polyA switch comprises two ligand binding aptamers.

3. The system of claim 1, wherein the polyA switch comprises three ligand binding aptamers.

4. The system of claim 1, wherein the polyA switch comprises a three way junction.

5. The system of claim 4, wherein the three way junction comprises a junction of a first, a second, and a third double stranded RNA stem.

6. The system of claim 5, wherein the first double stranded RNA stem does not comprise a ligand binding aptamer.

7. The system of claim 5, wherein each of the first, second, and third double stranded RNA stems comprise a ligand binding aptamer.

8. The system of claim 5, wherein the three way junction comprises at least one single stranded region.

9. The system of claim 8, wherein the three way junction comprises a first, a second, and a third single stranded region.

10. The system of claim 9, wherein the first single stranded region is located between the first double stranded RNA stem and the second double stranded RNA stem.

11. The system of claim 9, wherein the second single stranded region is located between the second double stranded RNA stem and the third double stranded RNA stem.

12. The system of claim 9, wherein the third single stranded region is located between the third double stranded RNA stem and the first double stranded RNA stem of the first aptamer.

13. The system of any one of the preceding claims, wherein the first aptamer and the second aptamer, in a 5' to 3' orientation, are in the same orientation.

14. The system of any one of the preceding claims, wherein the third aptamer, in a 5' to 3' orientation, is in the opposite orientation relative to the first and second aptamers.

15. The system of claim 1, wherein one or more nucleotides of the the polyA cleavage signal are within the 3 way junction, the third double stranded RNA stem, the third single stranded region, or the first double stranded RNA stem.

16. The system of claim 15, wherein the third single stranded region comprises the first four bases of the polyA cleavage signal.

17. The system of claim 15, wherein the first double stranded RNA stem comprises the last two bases of the polyA cleavage signal.

18. The system of claim 15, wherein the first double stranded RNA stem comprises the entirety of the polyA cleavage signal.

19. The system of claim 3, wherein the double stranded RNA stem between the binding pocket of the third aptamer and the three way junction is between 10 and 15 base pairs in length.

20. The system of claim 10, wherein the first single stranded region comprises at least one base selected from C and A.

21. The system of claim 11, wherein the second single stranded region comprises at least one base selected from C and A.

22. The system of claim 5, wherein the sequence of the second double stranded RNA stem is SEQ ID NO.: 3.

23. The system of claim 5, wherein the sequence of the third double stranded RNA stem is SEQ ID NO.: 2.

24. The system of claim 5, wherein the sequence of the first double stranded RNA stem is SEQ ID NO.: 4.

25. The system of claim 5, wherein the sequence of the first double stranded RNA stem is SEQ ID NO.: 5.

26. The system of claim 1, wherein the nucleic acid sequence encoding the expressible polypeptide further comprises a 5'UTR.

27. The system of claim 26, wherein the 5'UTR further comprises a CAA repeat.

28. The system of claim 26, wherein the 5'UTR further comprises one or more 3' splice acceptor sites.

29. The system of claim 26, wherein the engineered 5'UTR has sequence SEQ ID NO.: 48.

30. The system of claim 1, further comprising a G-U rich region 5' of the nucleic acid sequence encoding the expressible polypeptide and 3' of the polyA cleavage signal.

31. The system of claim 29, where the 3'acceptor site is followed by a nucleic acid triplet sequence that modulates the strength of the alternative splicing.

32. The system of claim 31, wherein the nucleic acid triplet is 3' relative to the second 3'acceptor site in the 5'UTR and has a sequence selected from the following: TAG, TCT, TTC, TTG, TGA, TGC, TCC, ACA, AAC, ACC, AGC, AGG, CCT, and CCC.

33. The system of claim 1, further comprising a G rich region 5' of the nucleic acid sequence encoding the expressible polypeptide and 3' of the G-U rich region.

34. The system of claim 33, wherein the G rich-region comprises 4 MAZ sequence.

35. The system of claim 1, wherein the engineered intron has a sequence of between 100 and 200 bases in length.

36. The system of claim 1, wherein the engineered intron has sequence SEQ ID NO 1.

37. The system of claim 1, where the engineered intron is followed by a nucleic acid triplet sequence that modulates the strength of the intron splicing.

38. The system of claim 37, wherein the nucleic acid triplet sequence is a sequence selected from: TTT, TGA, TCT, TAC, CAC, and CAT.

39. The system of claim 1, wherein the system comprises a sequence selected from the group SEQ ID NO.:6 to SEQ ID NO.: 56.

40. The system of claim 39, wherein the system comprises a sequence selected from the group SEQ ID NO.:6 SEQ ID NO.:13; SEQ ID NO.:14; SEQ ID NO.:28; SEQ ID NO.:32; SEQ ID NO.:33; SEQ ID NO.:36; SEQ ID NO.:38; SEQ ID NO.:44; SEQ ID NO.:46; SEQ ID NO.: 50; NO.: 51; NO.: 52; NO.: 53; NO.: 54; NO.: 55; NO.: 56.

41. A vector for delivery of the system of claim 1.

42. The vector of claim 41, wherein the vector is a viral vector.

43. The vector of claim 42, wherein the vector is selected from an adenoviral vector, a lentiviral vector; an adeno-associated viral vector, a poliovirus vector, and a retrovirus vector.

44. A method for modulating expression of a gene product in a cell the method comprising the steps of:

introducing into the cell a system comprising in a 5' to 3' direction:

a) a 5’ splice donor site

b) an engineered intron

c) a first 3’ splice acceptor site

d) a polyA switch comprising two or more ligand-binding aptamers with one or more ligand binding pockets, and at least one polyA cleavage signal therein; and

e) a second 3’ splice acceptor site.

45. The method of claim 44, wherein the gene product is exogenous to the cell.

46. The method of claim 45, wherein the system further comprises a nucleic acid sequence encoding the gene product immediately 3’ of the splice site of e).

47. The method of claim 44, wherein the gene product is endogenous to the cell.

48. The method of claim 47, wherein the method does not comprise administering the ligand to inhibit expression of the endogenous gene product.

49. The method of claim 44, wherein the system further comprises a promoter 5’ of the splice site of a).

50. The method of claim 49, wherein the promoter is a CMV promoter.

51. The method of any one of the preceding claims, wherein the method occurs in one or more cells of an individual, the ligand is glucose, the individual has diabetes, pre-diabetes, or complications from diabetes, and/or the expressible polynucleotide is insulin.

52. The method of any one of the preceding claims, wherein the method occurs in one or more cells of an individual, the ligand is the gene product of a cancer biomarker, and the expressible polynucleotide is a suicide gene.

53. The method of any one of the preceding claims, wherein the method occurs in an individual, the expressible polynucleotide is a reporter gene, and the location and/or intensity of the expression of the reporter gene provides information about spatial distribution, temporal fluctuation, or both, of a ligand in one or more cells of the individual.

54. The method of any one of the preceding claims, wherein the method occurs in an individual, tissue, or cell, wherein the expressible polynucleotide encodes a detectable gene product, and wherein the respective individual, tissue, or cell is imaged.

55. The method of claim 50, wherein the vector of a) and/or the cells of b) are provided to the individual before the therapy, during the therapy, and/or after the therapy.

56. A nucleic acid molecule encoding the poly A aptamer polynucleotide comprising in a 5' to 3' direction:

a) a 5’ splice donor site;

b) an engineered intron;

c) a first 3' splice acceptor site;

d) a polyA switch comprising two or more ligand-binding aptamers with one or more ligand binding pockets, and at least one polyA cleavage signal therein;

e) a second 3' splice acceptor site; and

f) a nucleic acid sequence encoding an expressible polypeptide.

57. The nucleic acid molecule of claim 56, wherein the nucleic acid is DNA.

58. The nucleic acid molecule of claim 56, wherein the nucleic acid is RNA.

59. A vector for delivery of the nucleic acid of claim 56.

60. The vector of claim 59, wherein the vector is a viral vector.

61. The vector of claim 59, wherein the vector is selected from an adenoviral vector, a lentiviral vector; an adeno-associated viral vector, a poliovirus vector, and a retrovirus vector.

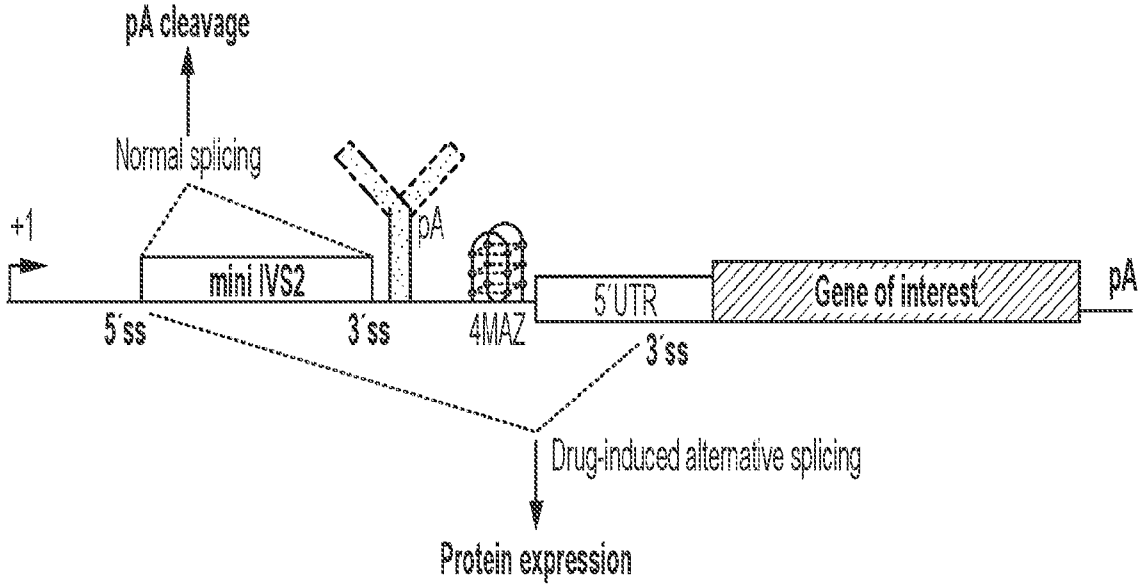


Figure 1A

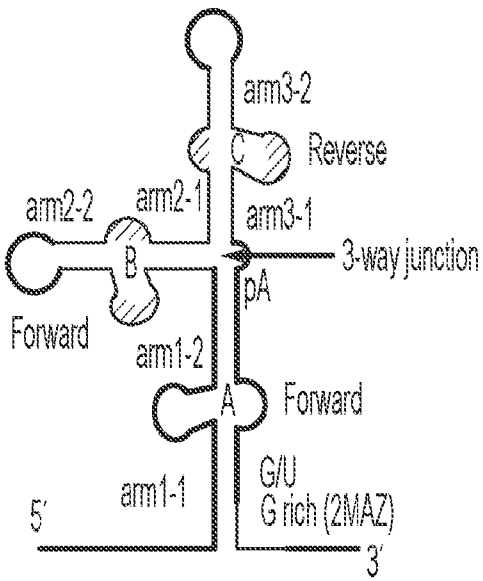


Figure 1B

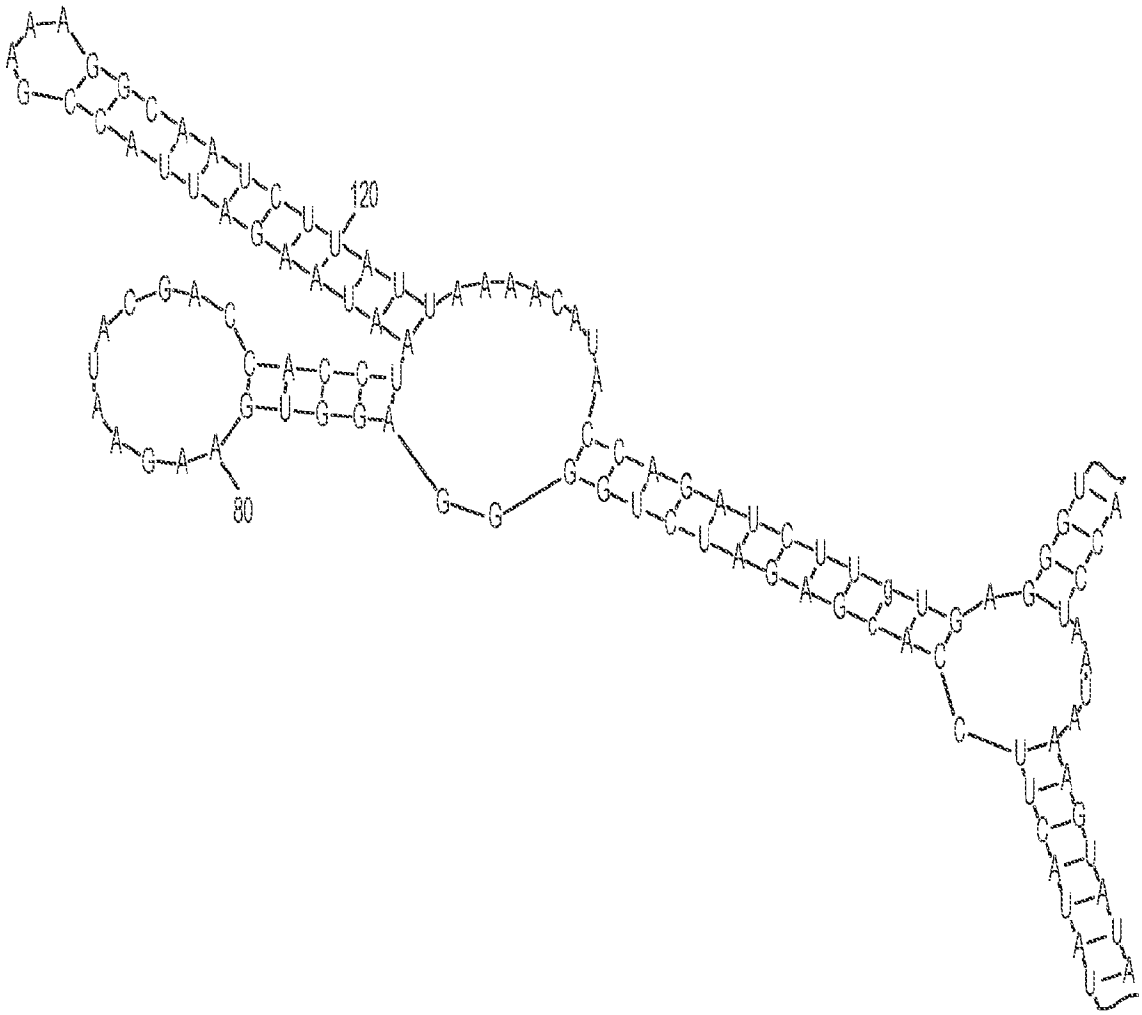


Figure 1C

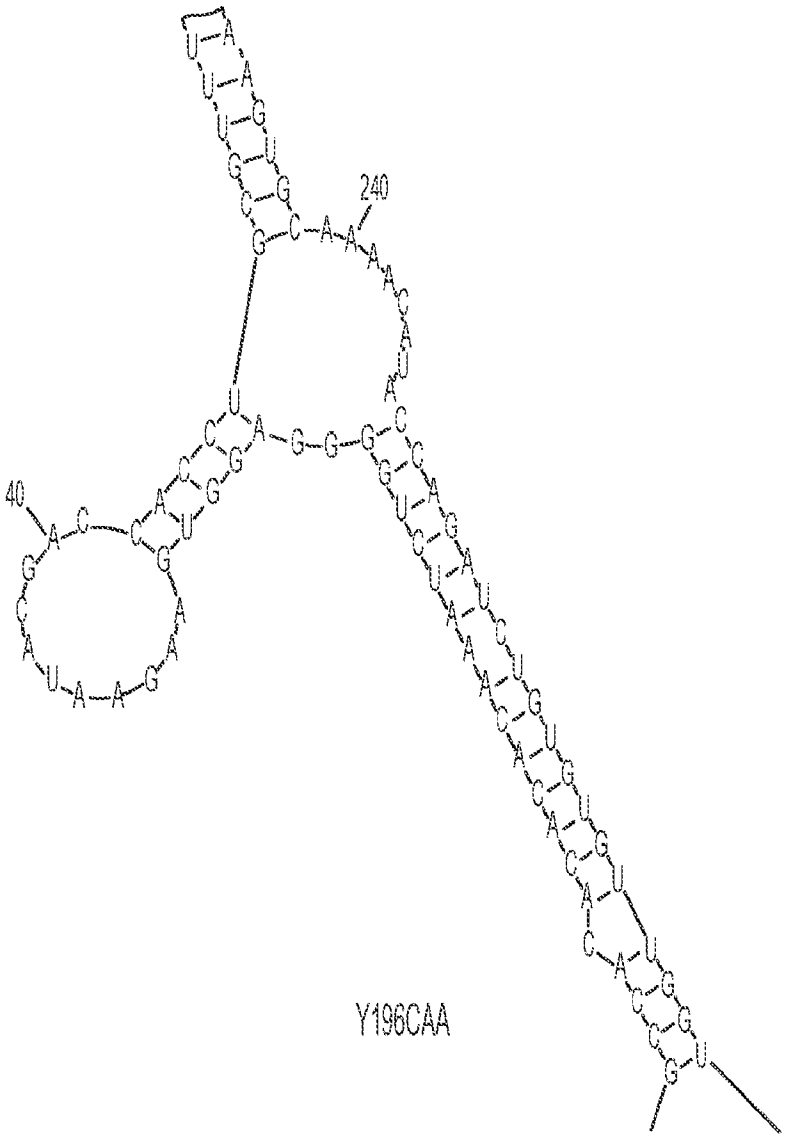


Figure 1C
CONTINUED

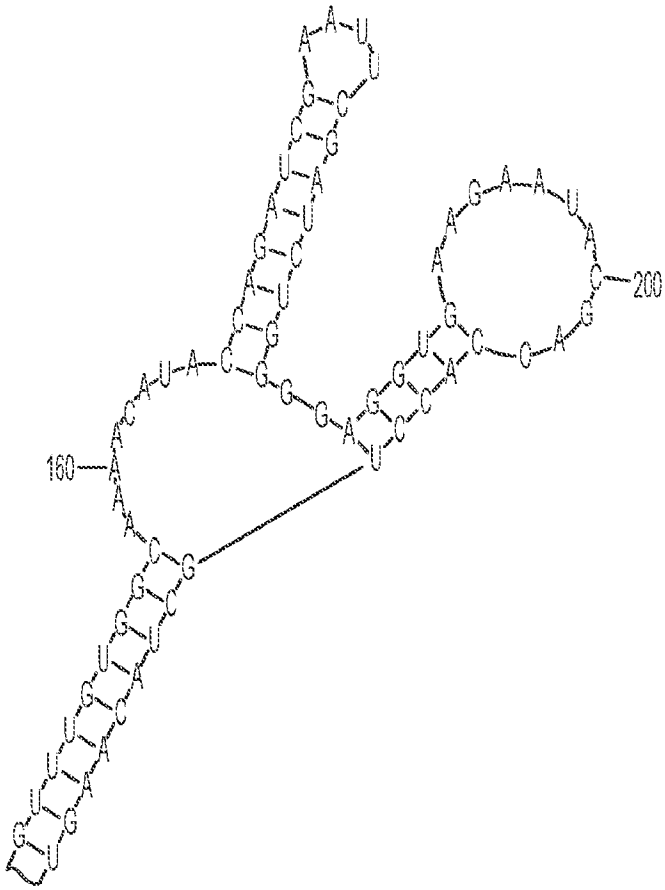


Figure 1C
CONTINUED

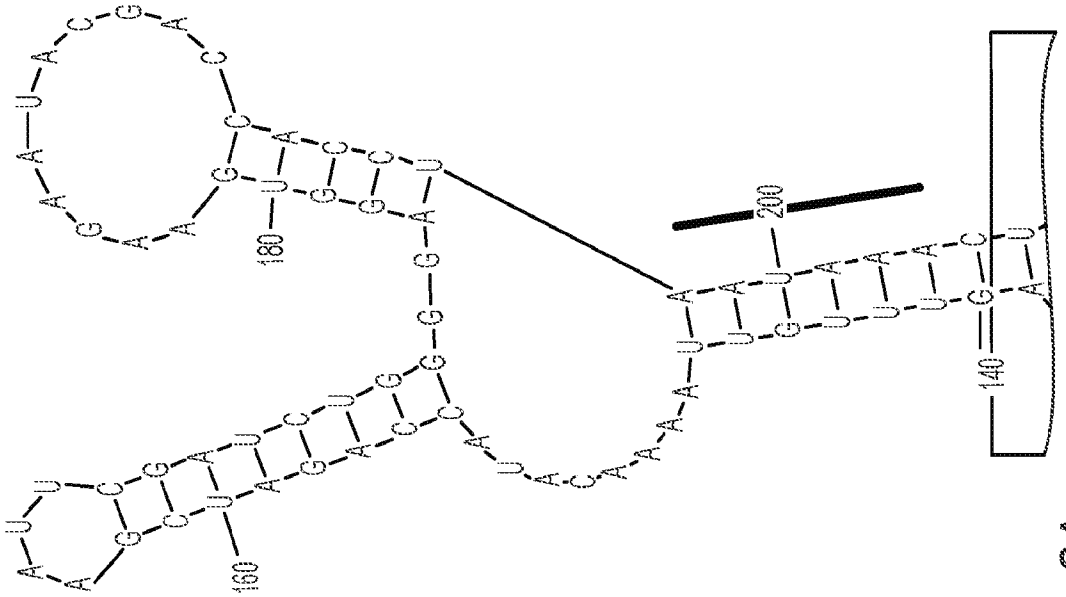
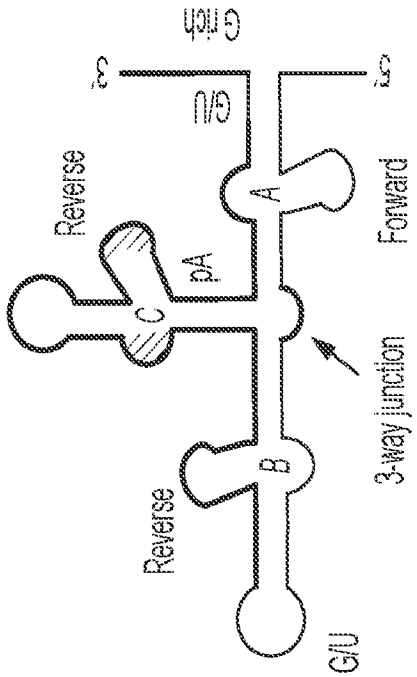
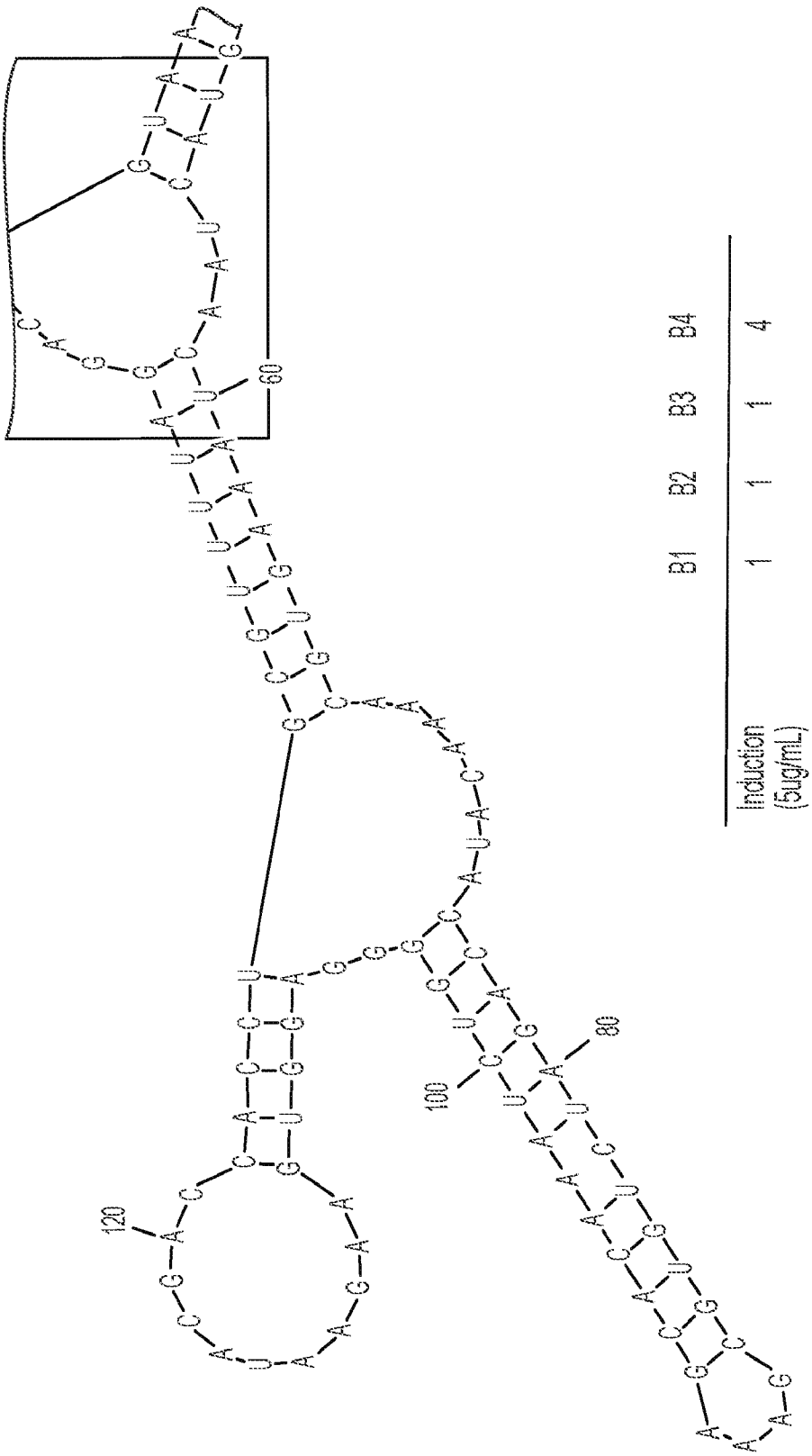

180
200
160
140

Figure 2A


Gluc
Reverse
pA
Forward
A
C
B
3-way junction
Reverse
G/U



Induction (5ug/mL)	B1	B2	B3	B4
	1	1	1	4

Figure 2A
CONTINUED

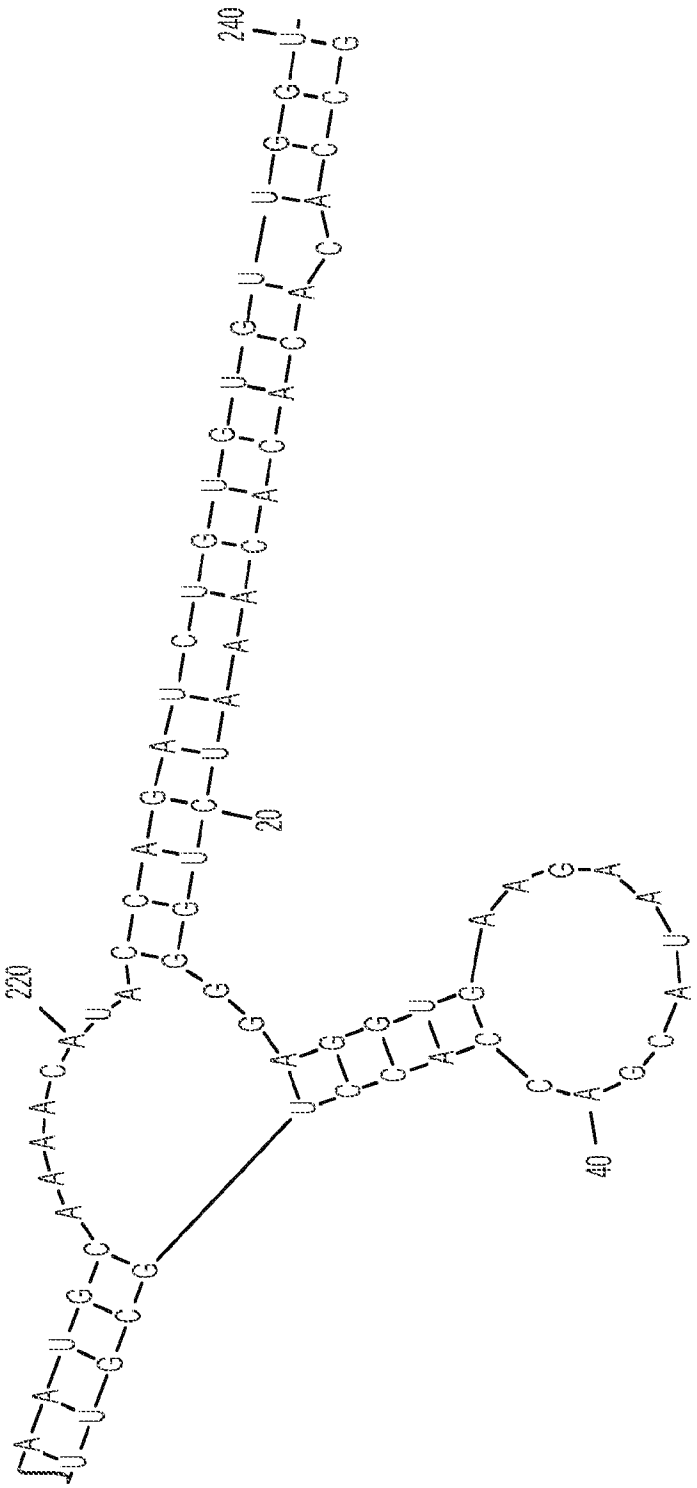


Figure 2A
CONTINUED

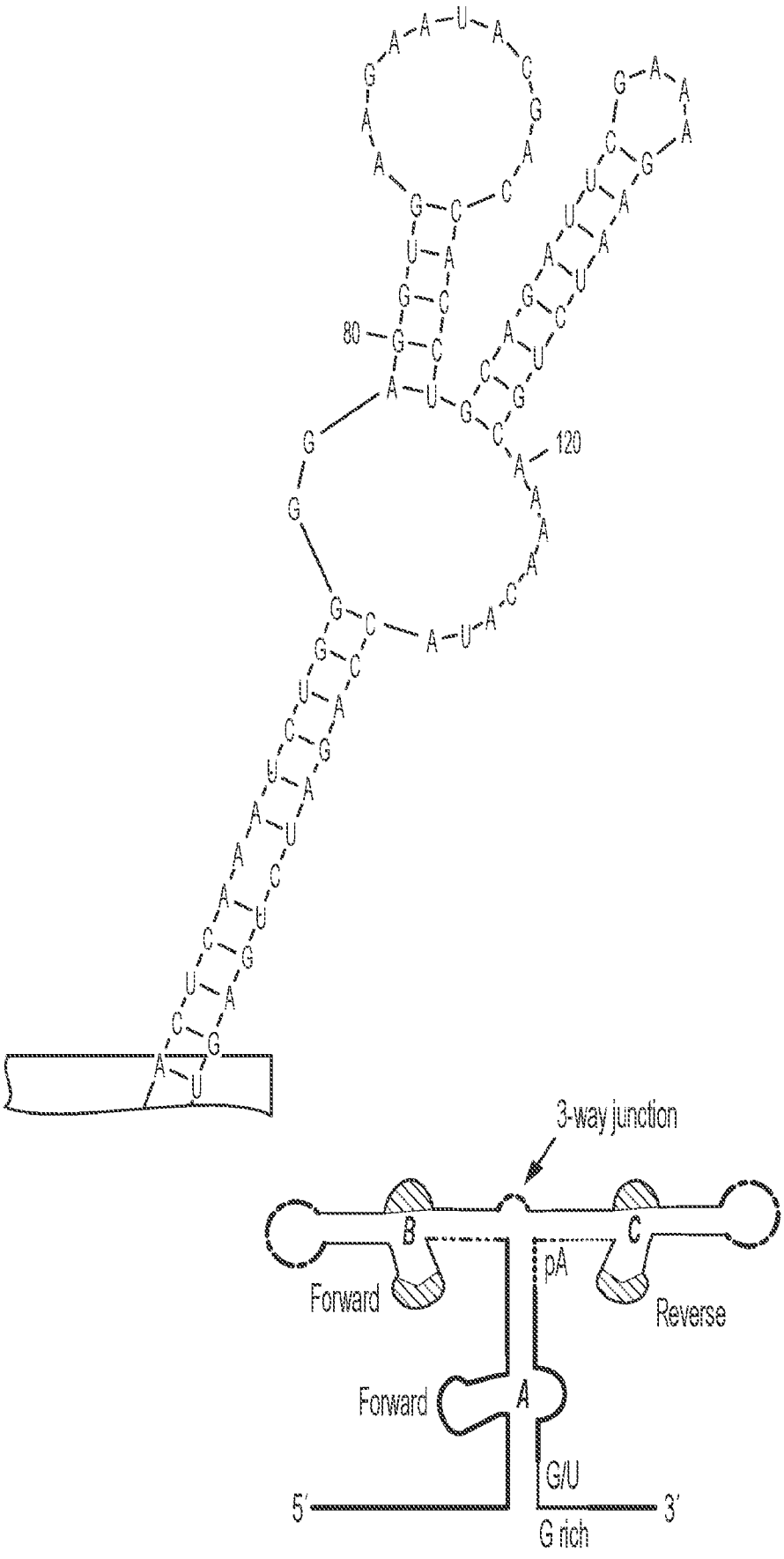

80
120
3-way junction
B
C
pA
Forward
Reverse
Forward
A
G/U
5'
3'
G rich

Figure 2B

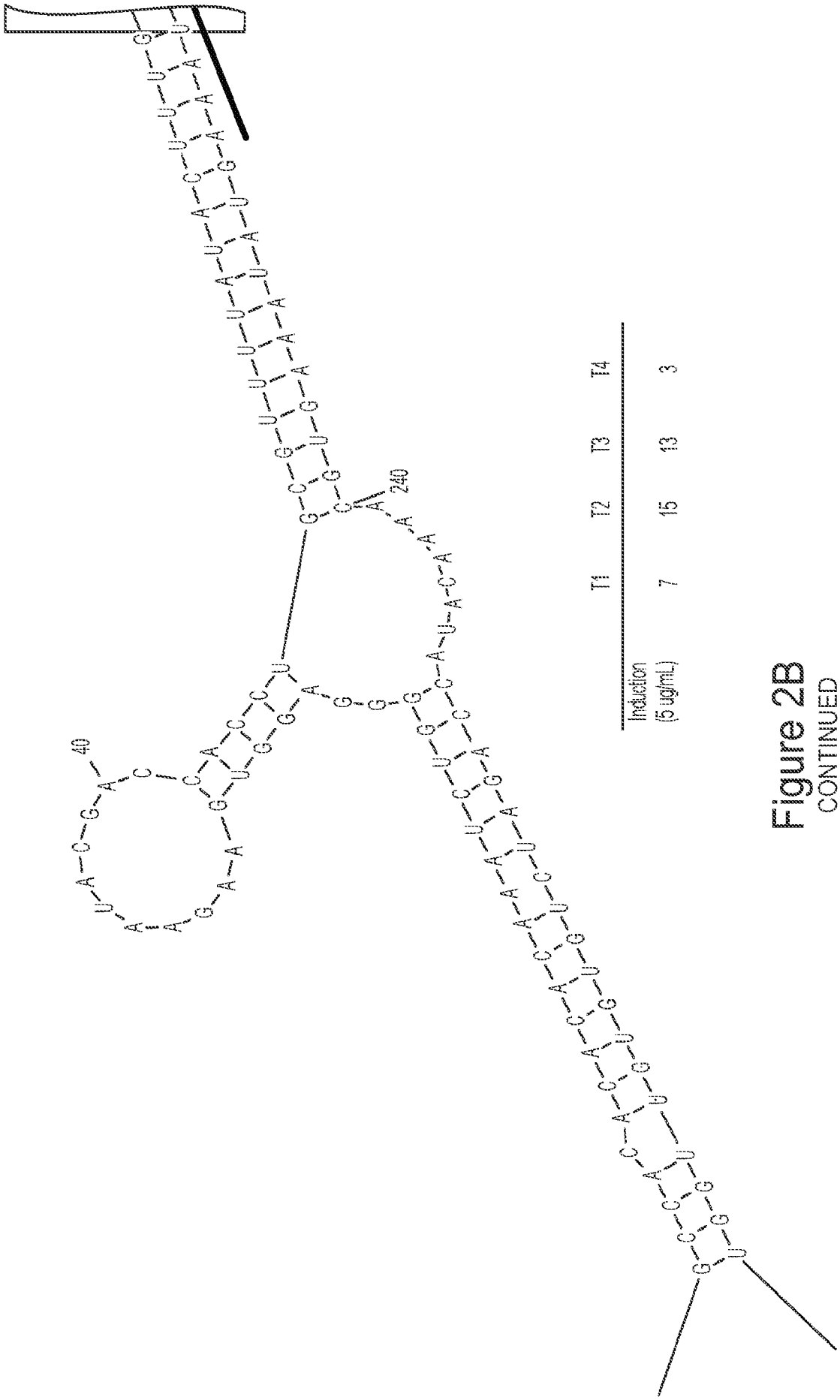


Figure 2B
CONTINUED

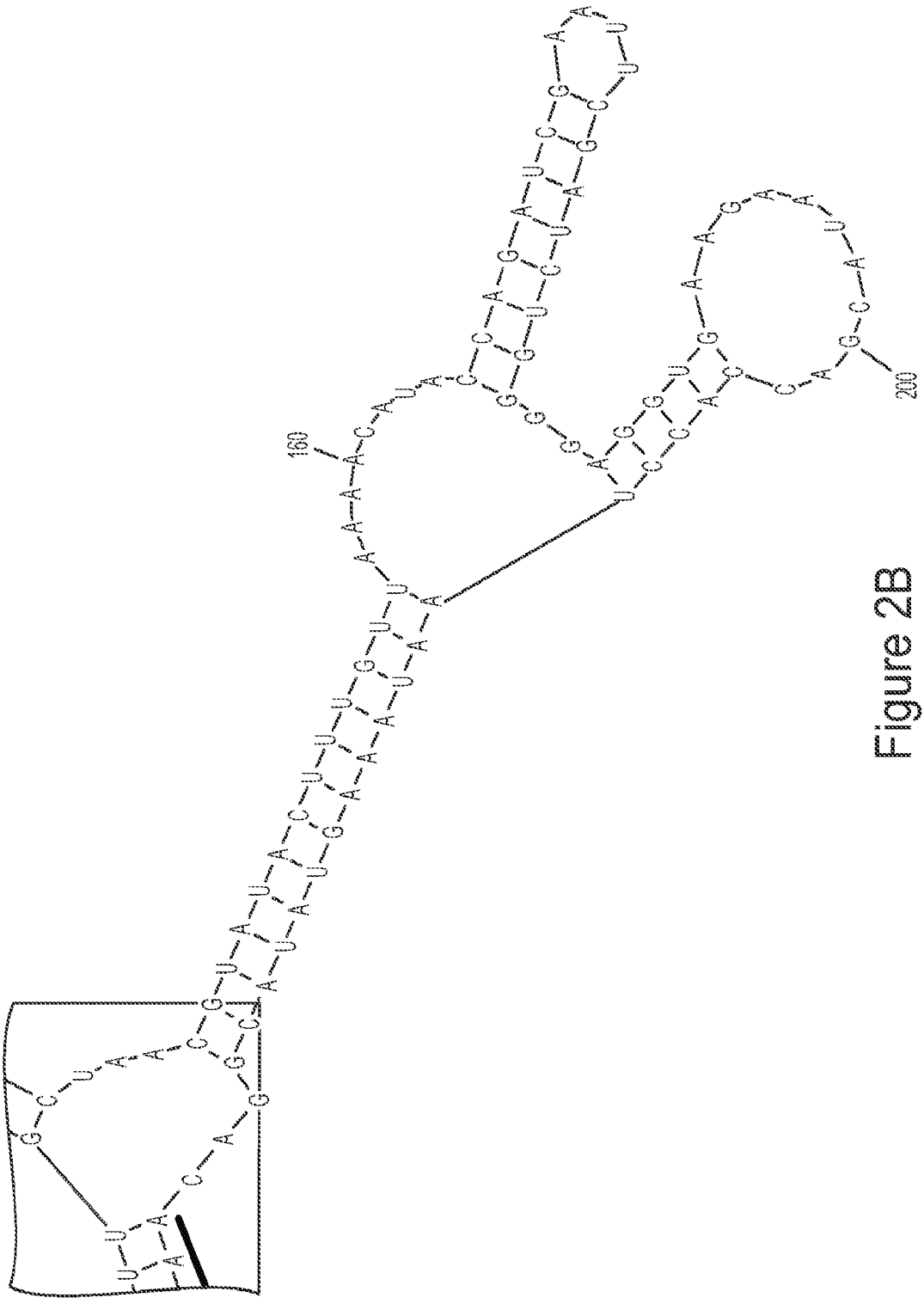


Figure 2B
CONTINUED

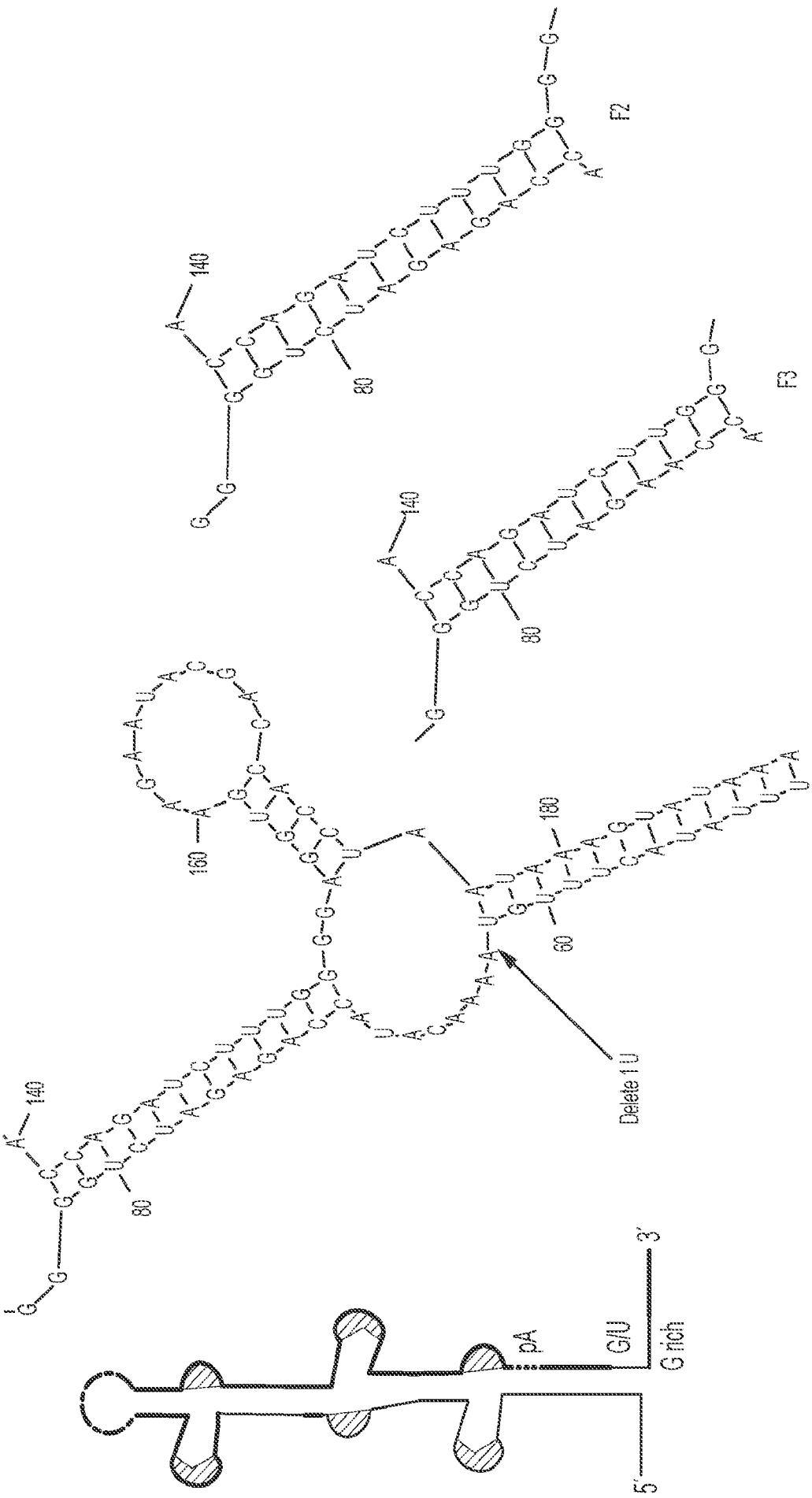


Figure 2C

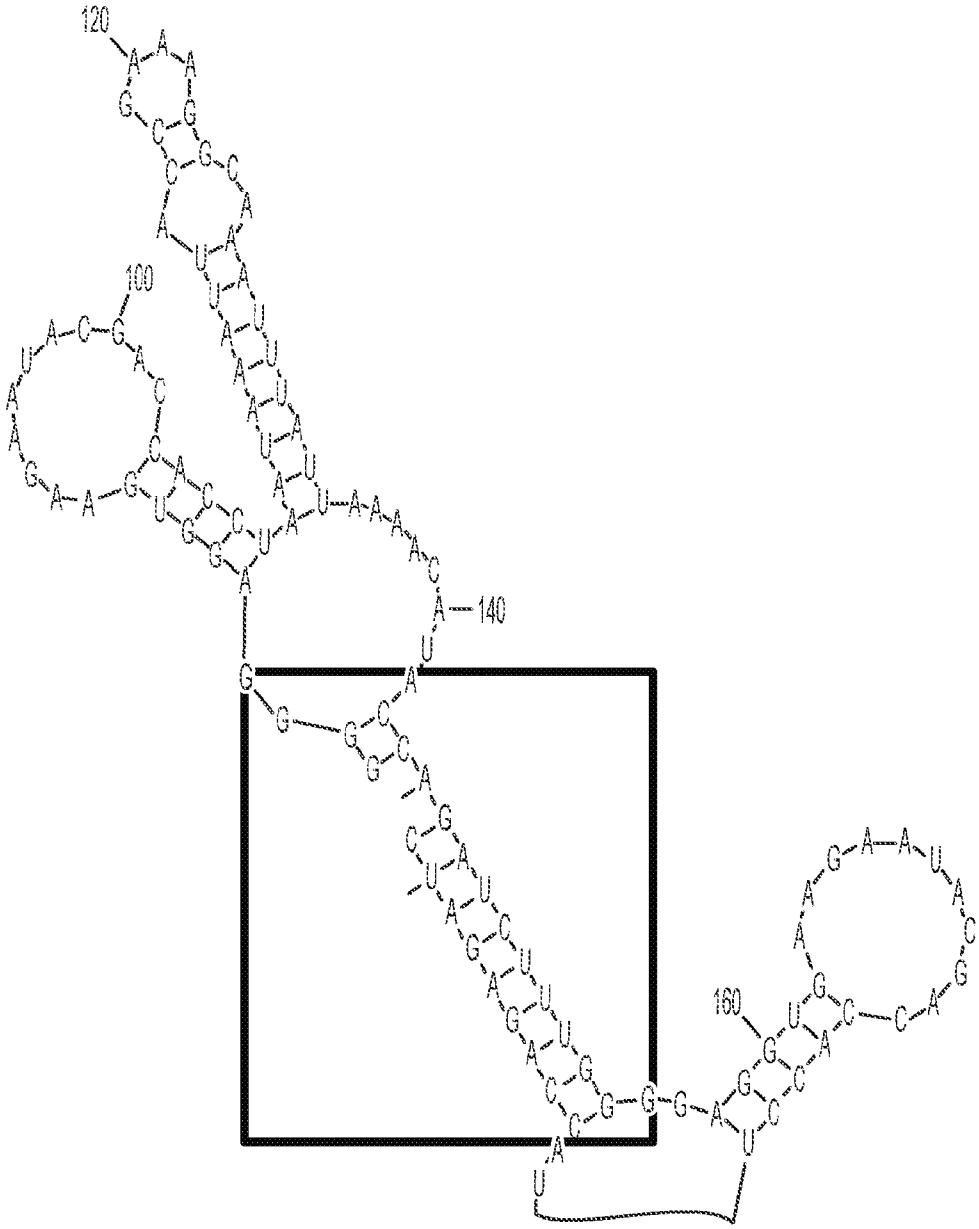


Figure 2C
CONTINUED

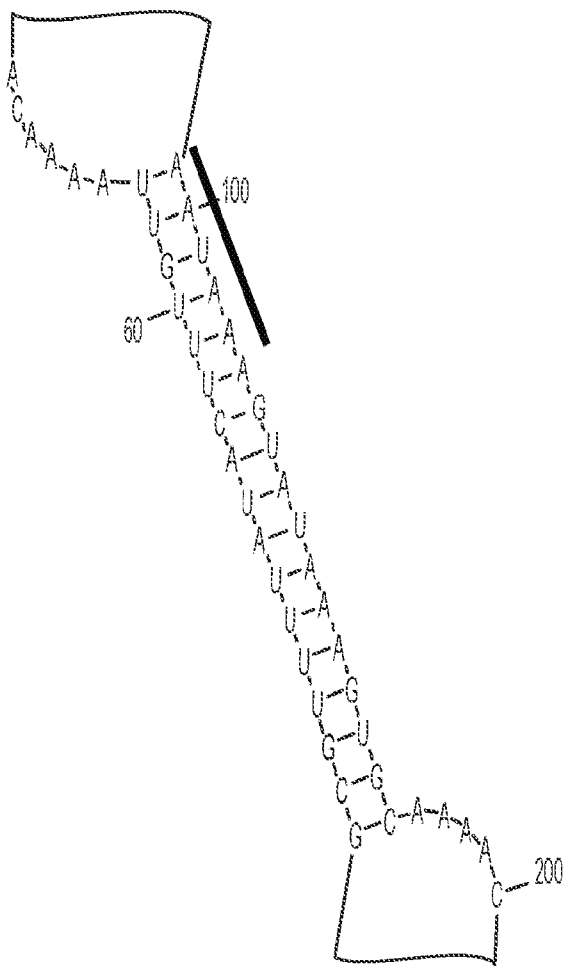
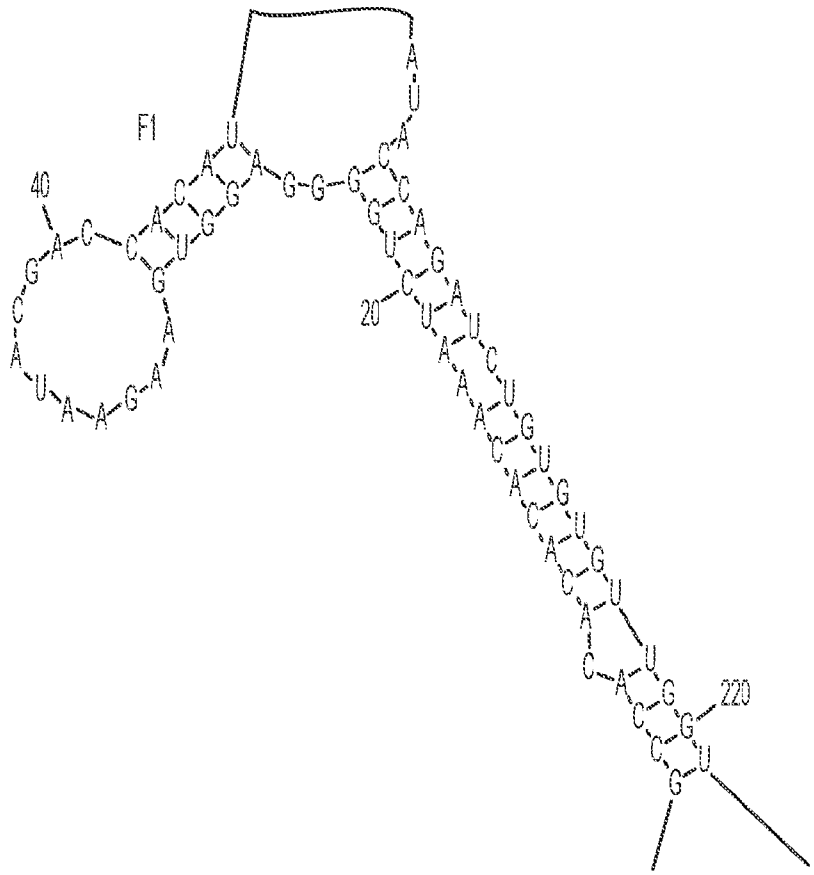
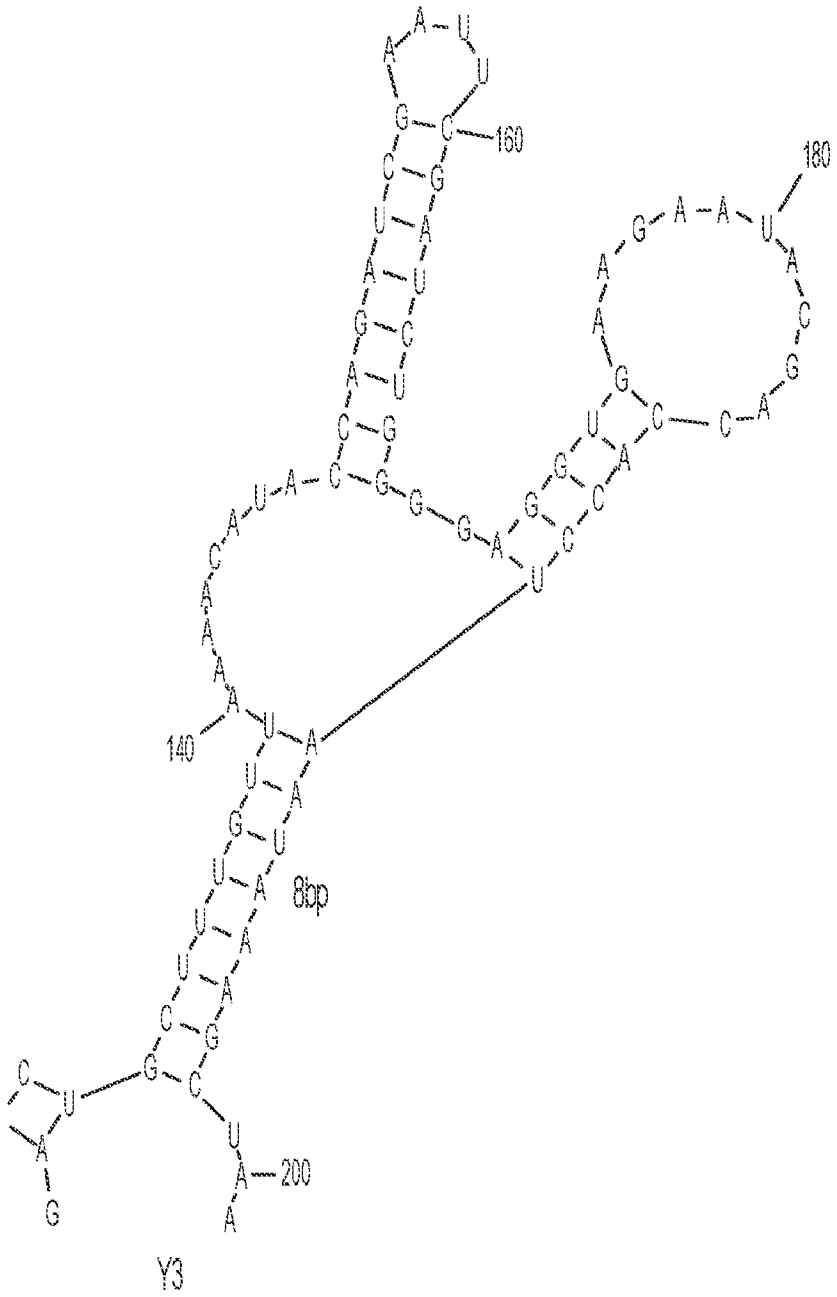


Figure 2C
CONTINUED



	F2	F3	F4	F5
Induction (5 ug/mL)	2	3	1	1

Figure 2C
CONTINUED



	Y1	Y2	Y3
Induction (5 ug/mL)	12	5	7

Figure 3A

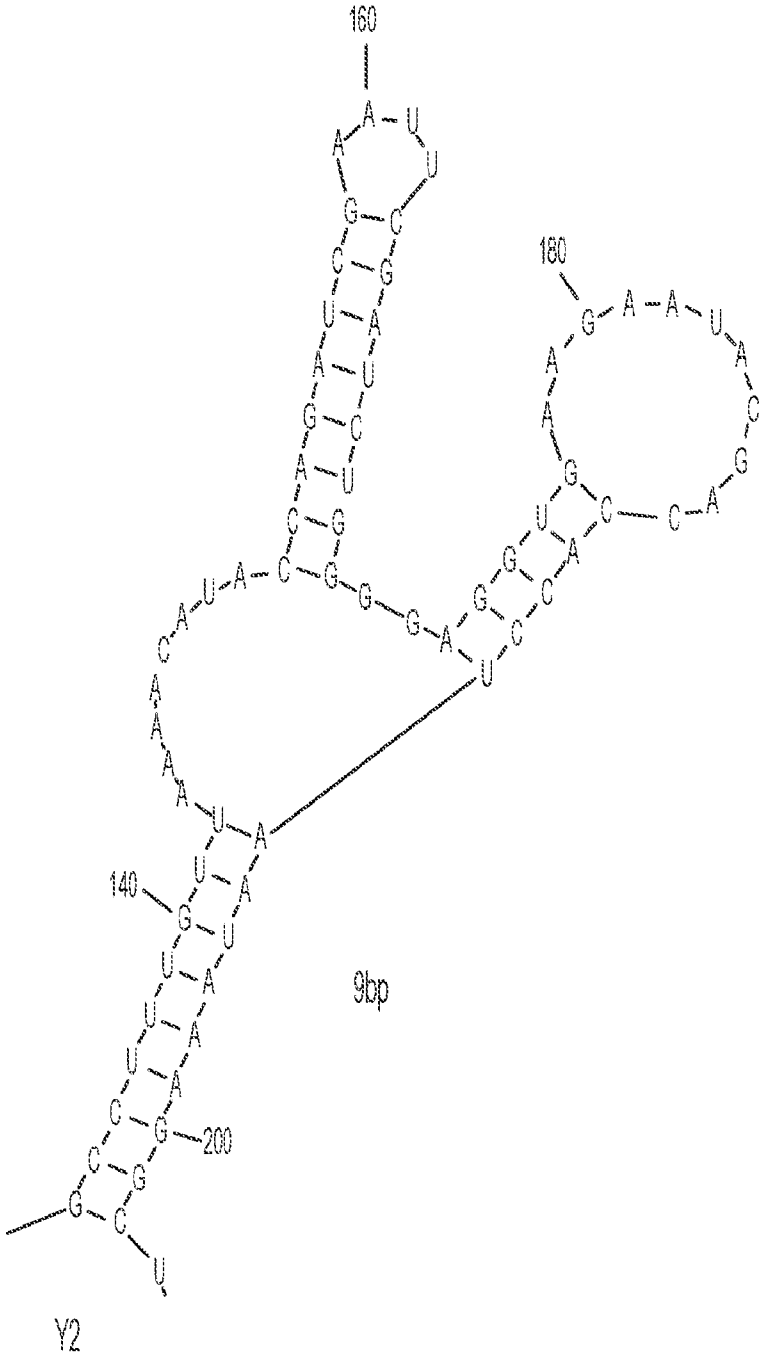


Figure 3A
CONTINUED

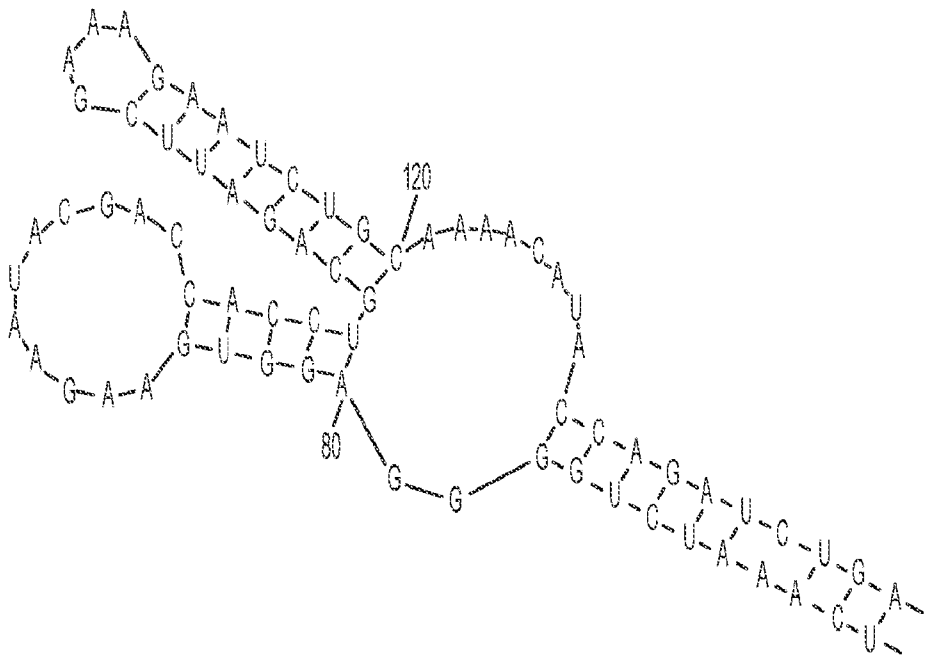


Figure 3A
CONTINUED

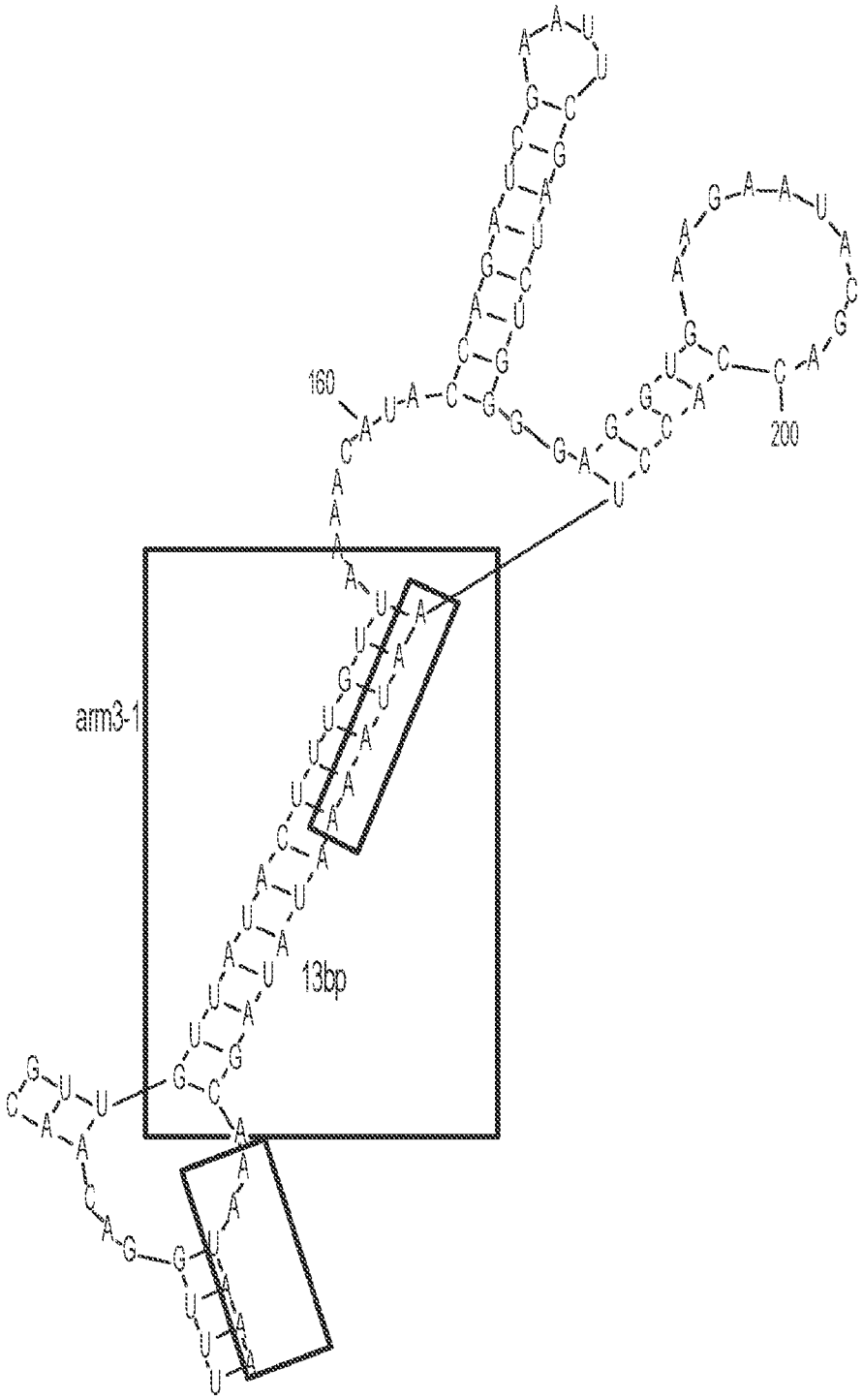


Figure 3A
CONTINUED

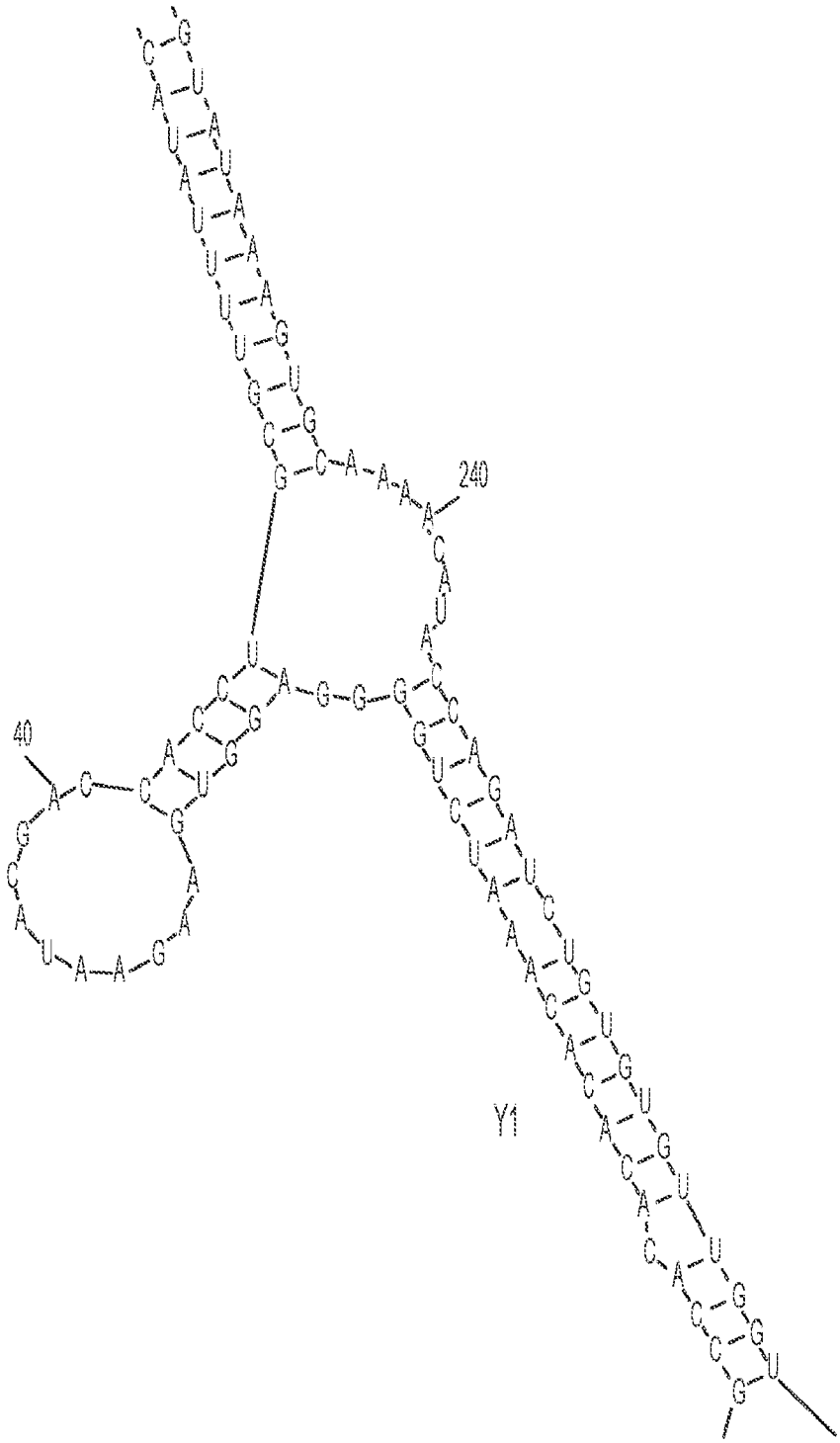


Figure 3A
CONTINUED

	Y5	Y6	Y7	Y8	Y9
Induction (5 ug/mL)	12	20	27	26	12
The length of arm 3-1	12bp	11bp	10bp	9bp	8bp
The length of arm 2-1	12bp	11bp	10bp	9bp	8bp

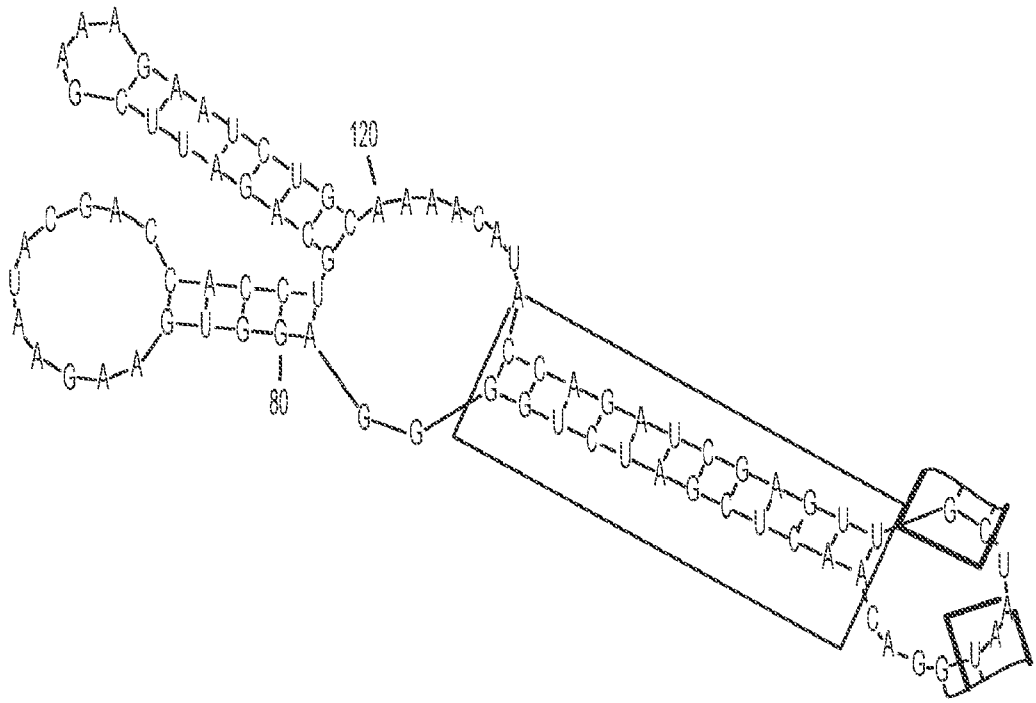


Figure 3B

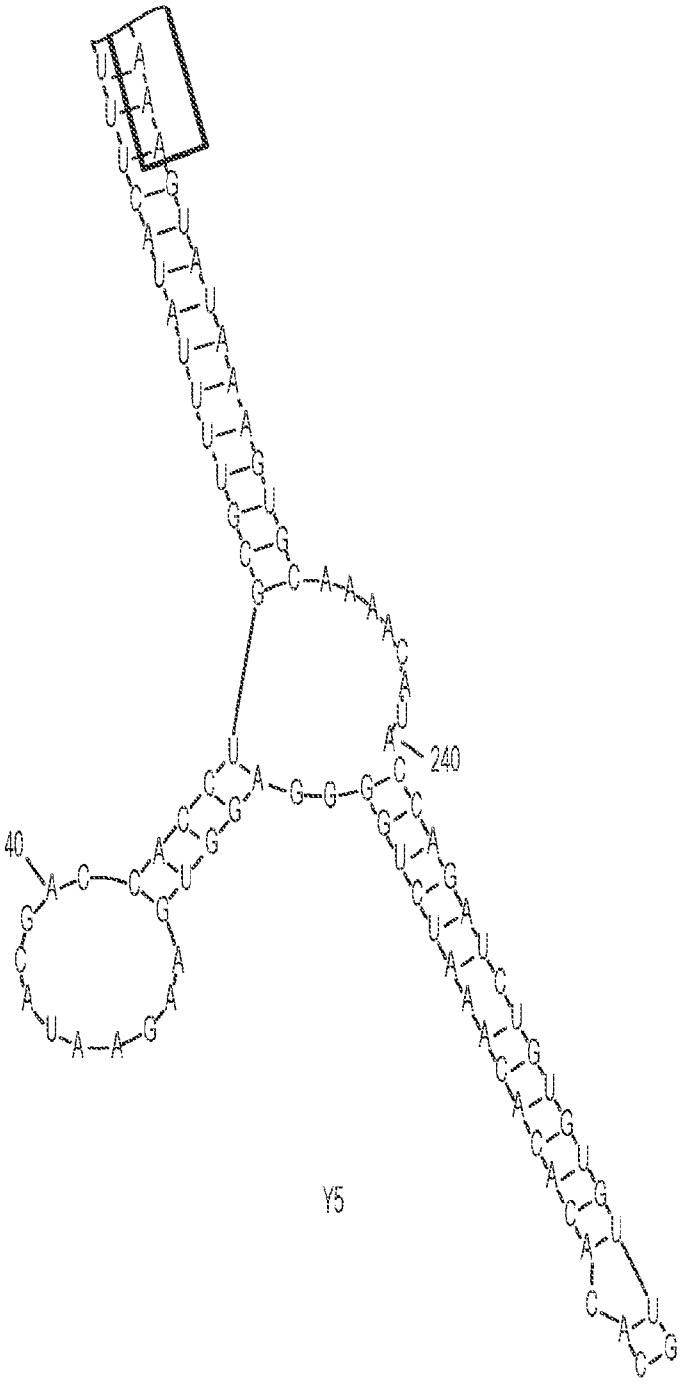


Figure 3B
CONTINUED

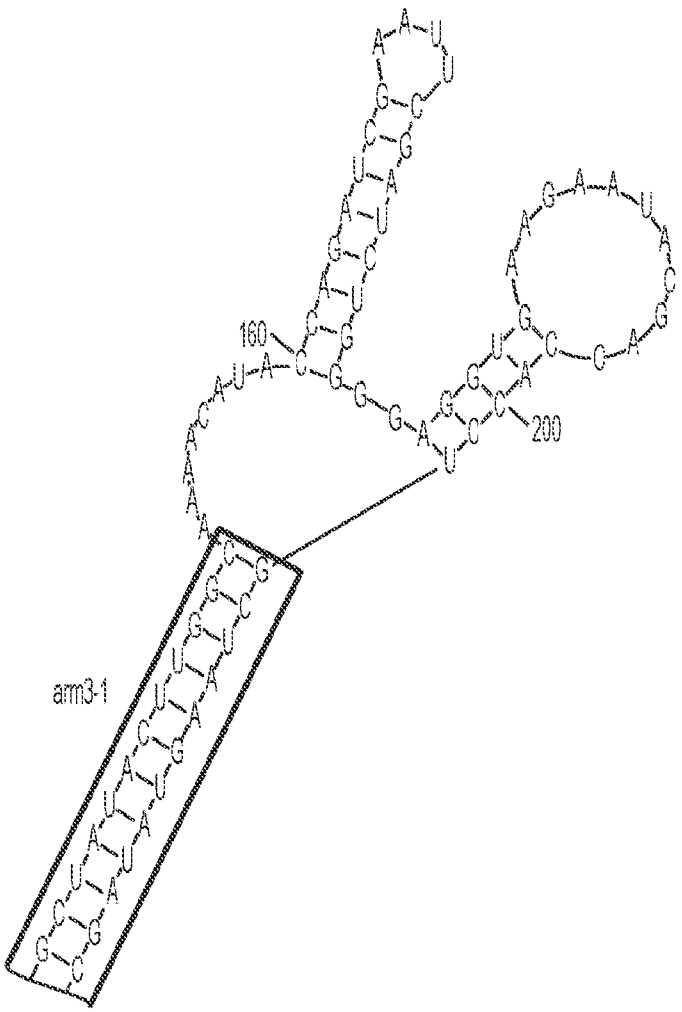


Figure 3B
CONTINUED

	Y6	Y6mut
Induction (5 ug/mL)	24	15

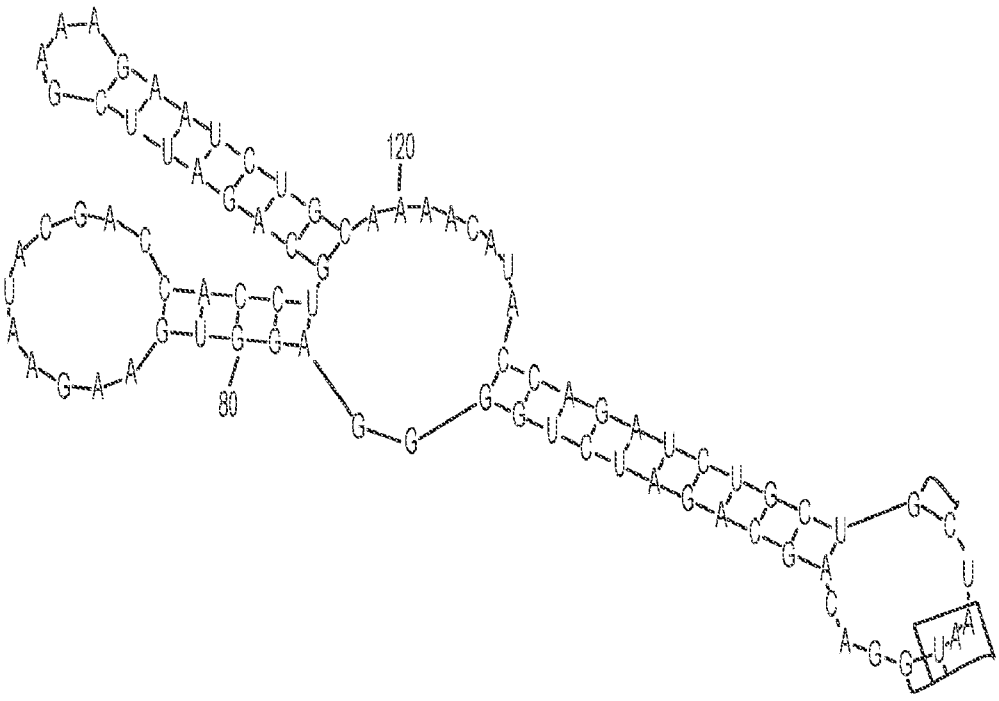


Figure 3C

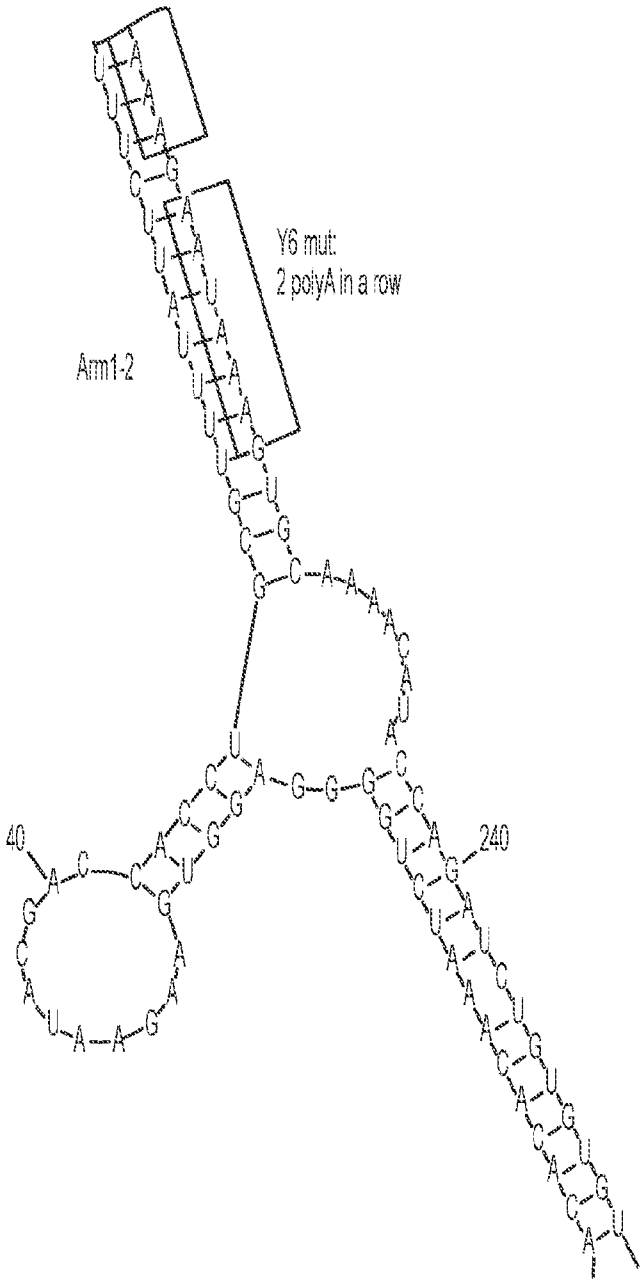


Figure 3C
CONTINUED

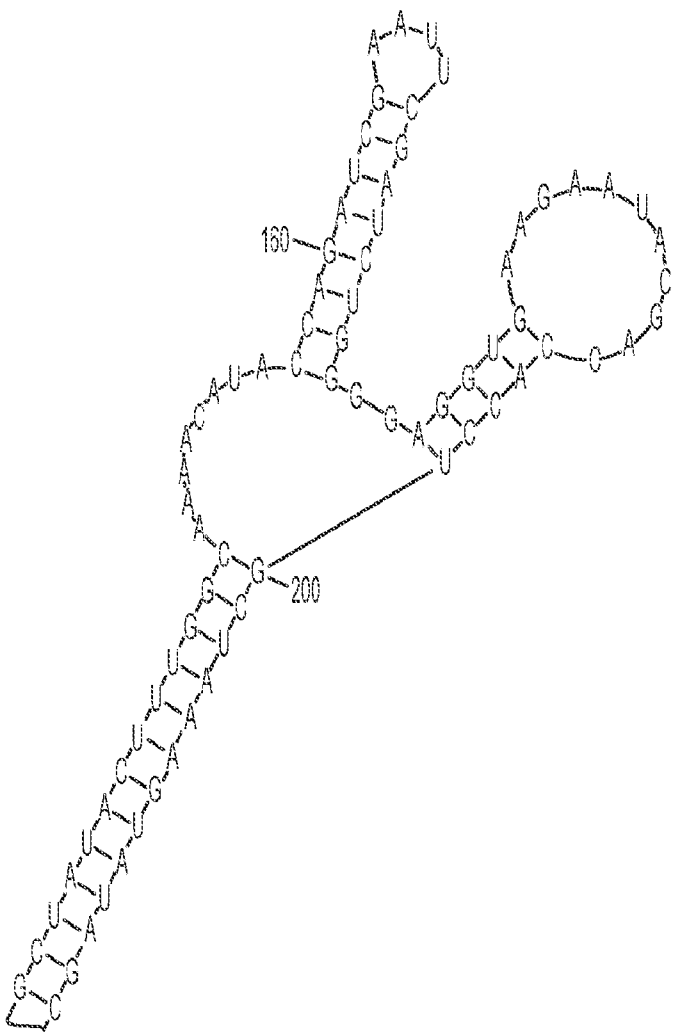
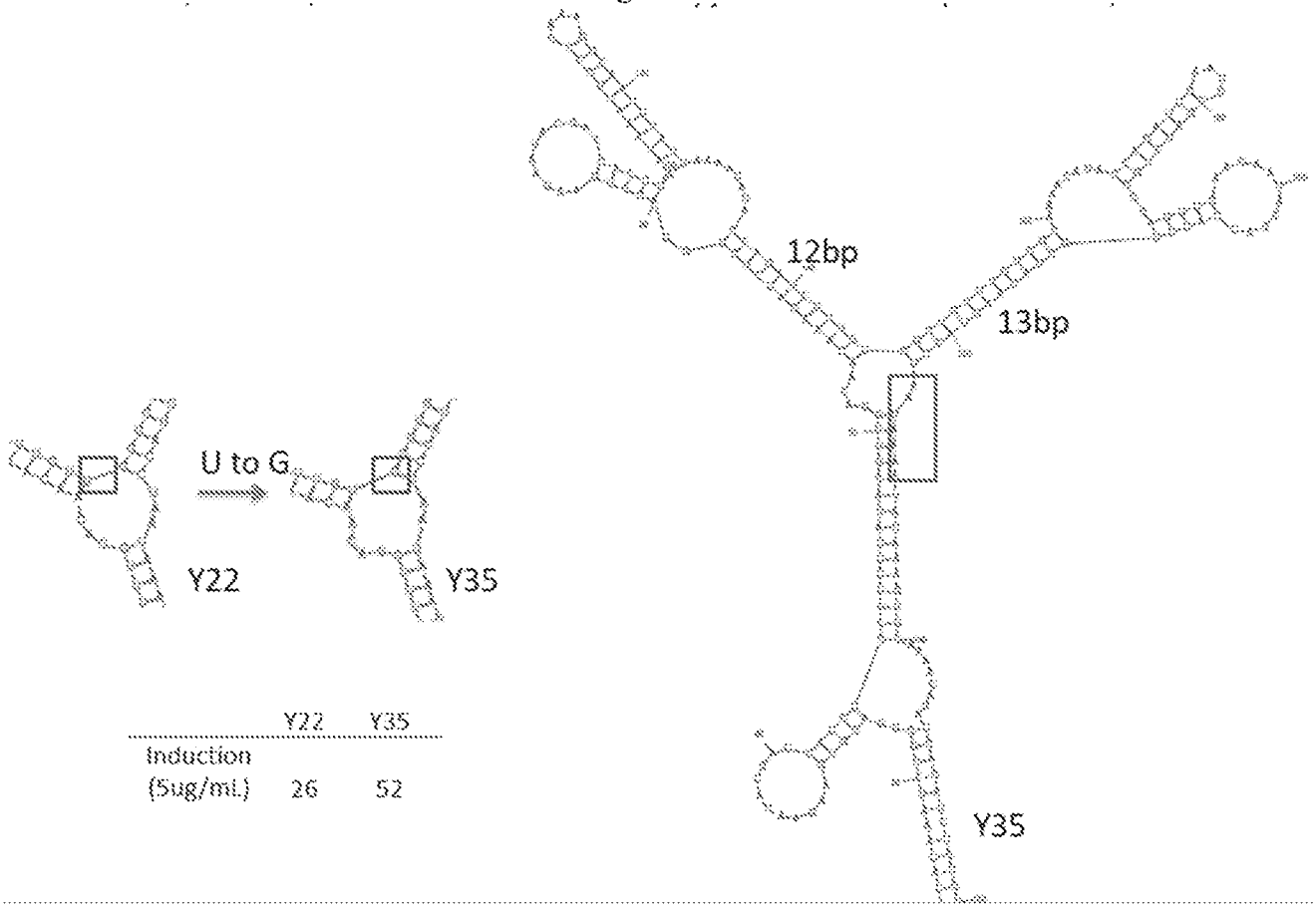


Figure 3C
CONTINUED

Figure 4A



	Y22	Y35
Induction (5ug/mL)	26	52

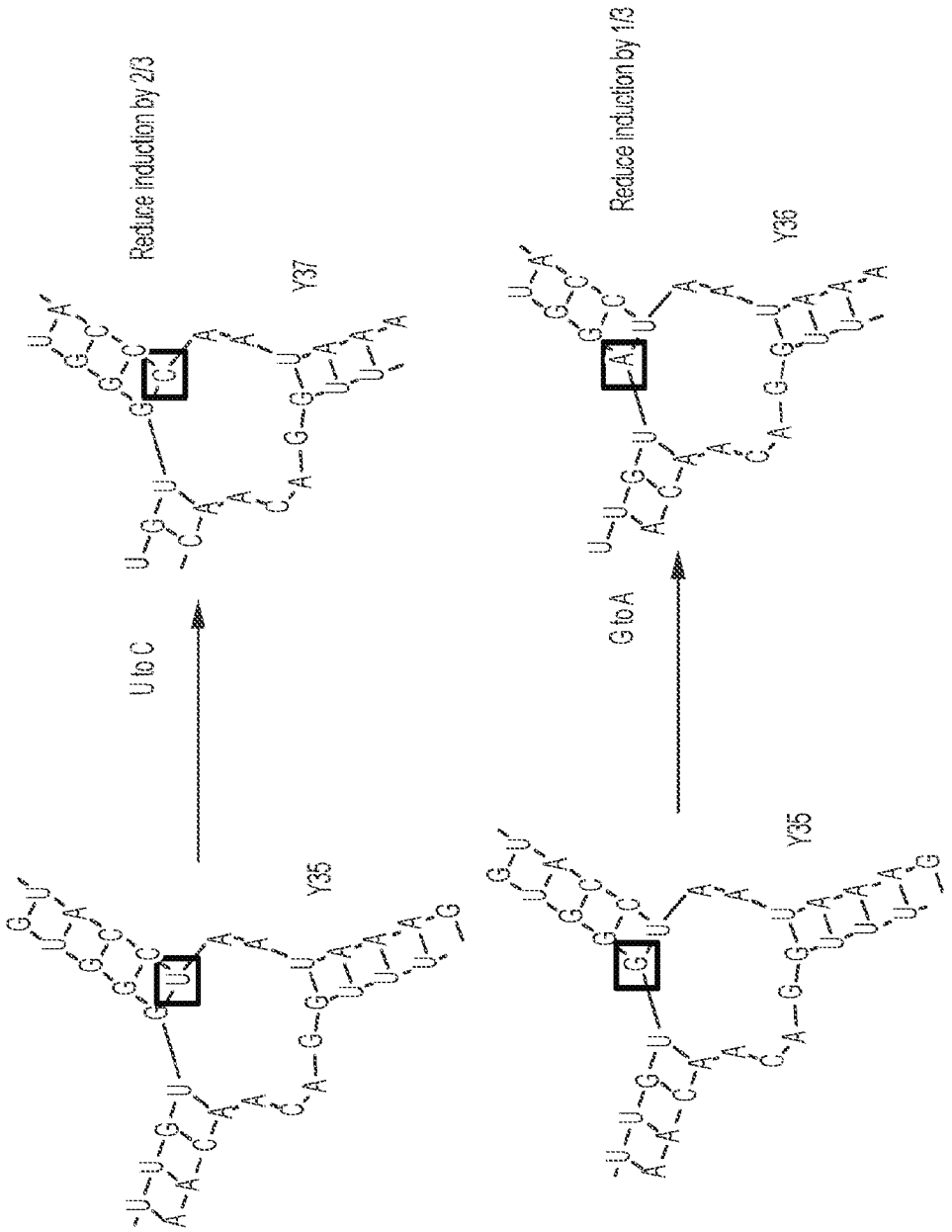
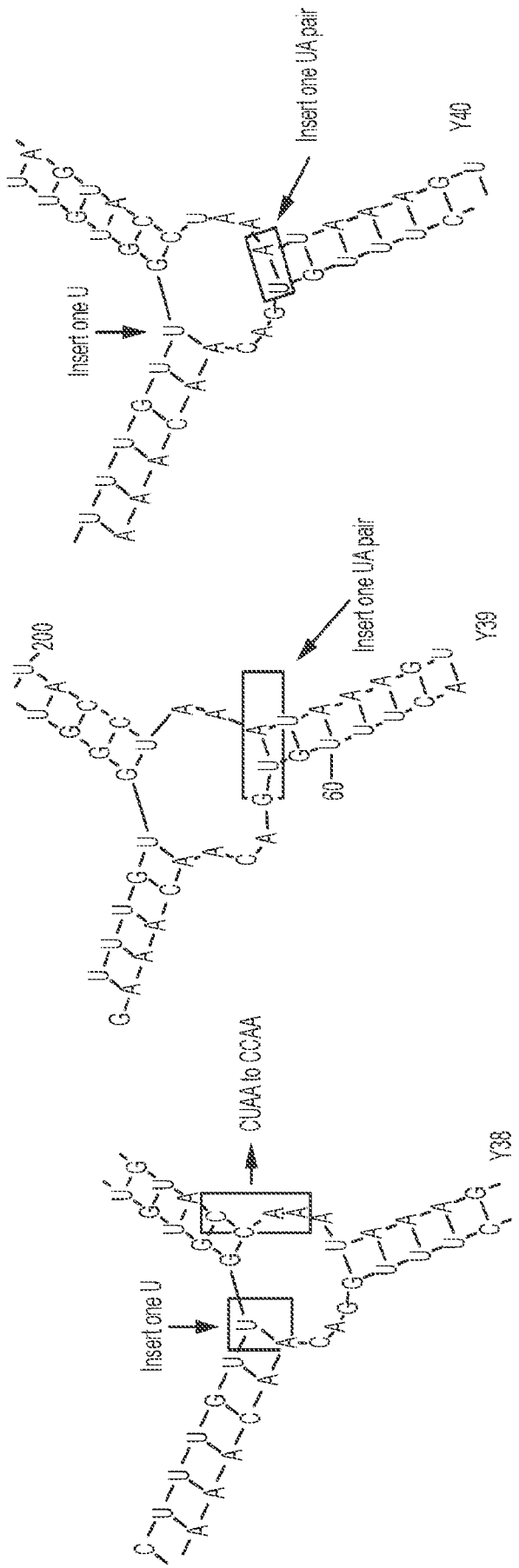
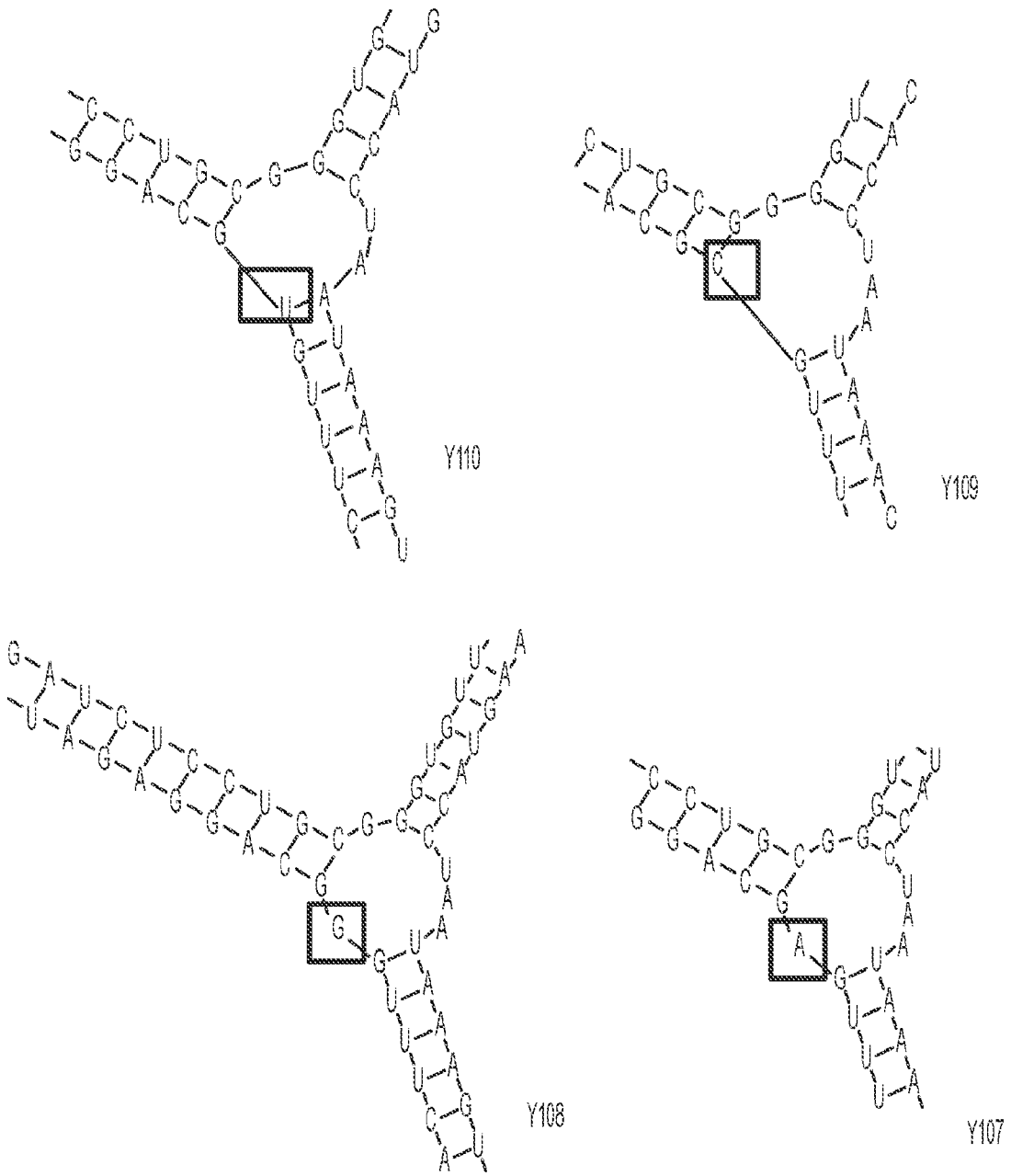

Reduce induction by 2/3
Y37
U to C
Y35
Reduce induction by 1/3
Y36
G to A
Y35

Figure 4B



	Y22	Y35	Y36	Y37	Y38	Y39	Y40
Induction (5 ug/mL)	26	52	35	17	14	36	18

Figure 4B
CONTINUED



	Y79	Y107	Y108	Y109	Y110
Induction (5ug/mL)	56	52	24	36	40

Figure 4C

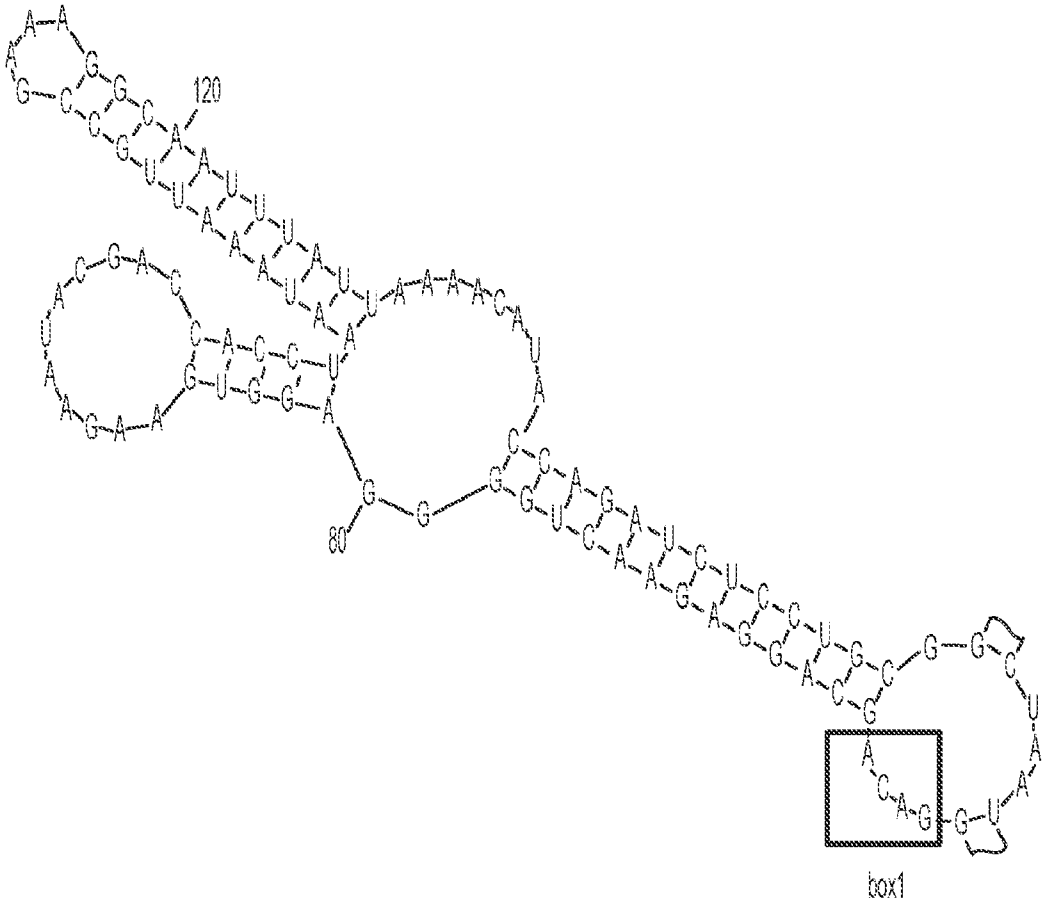


Figure 4C
CONTINUED

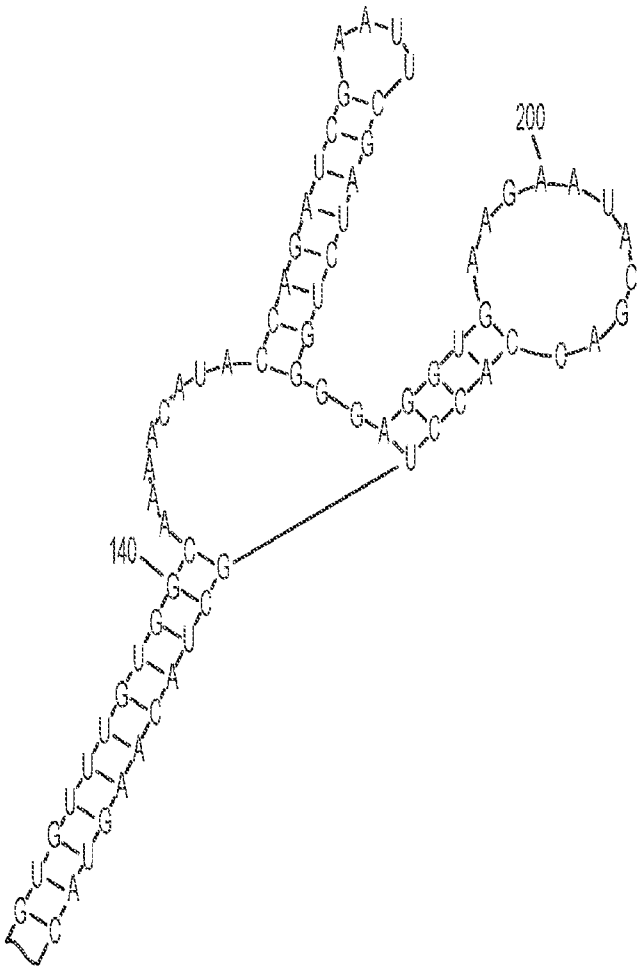


Figure 4C
CONTINUED

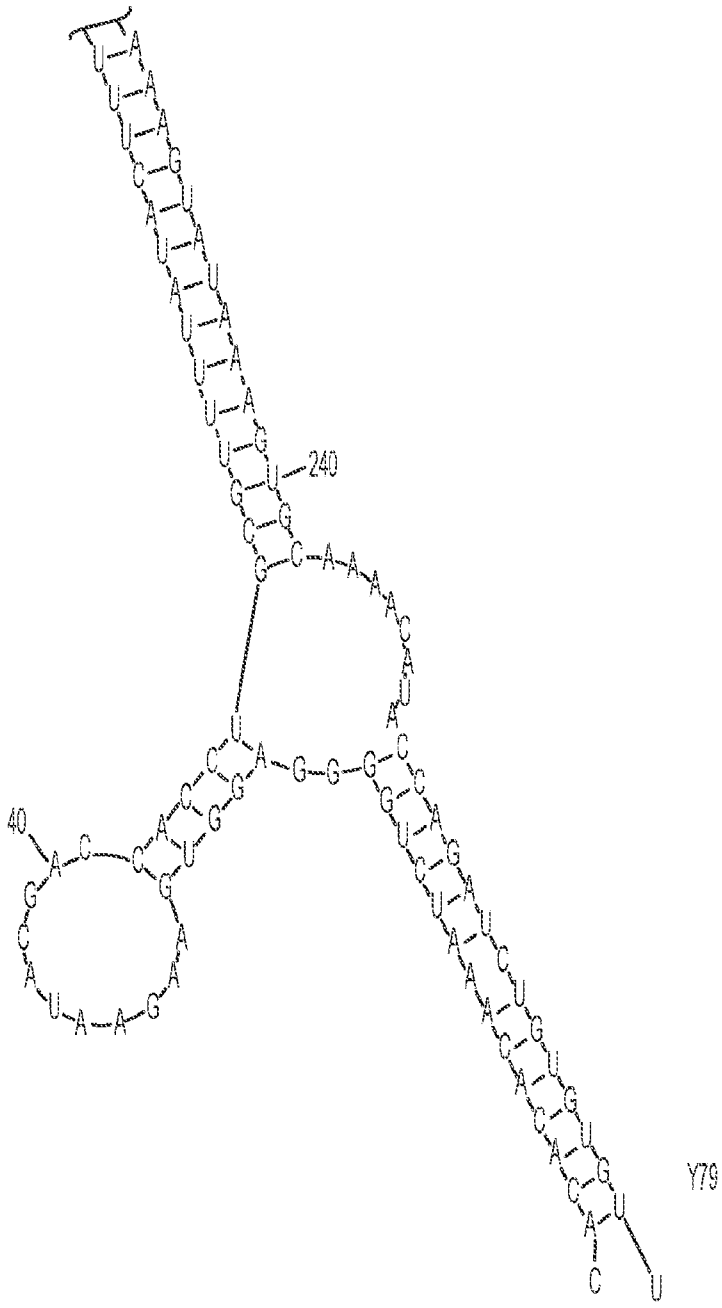
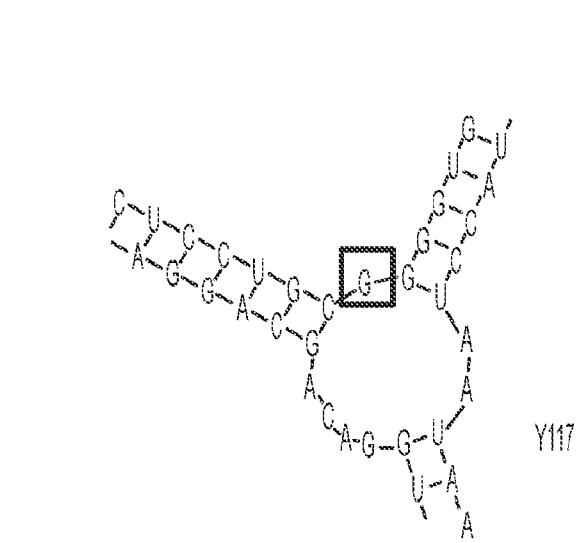
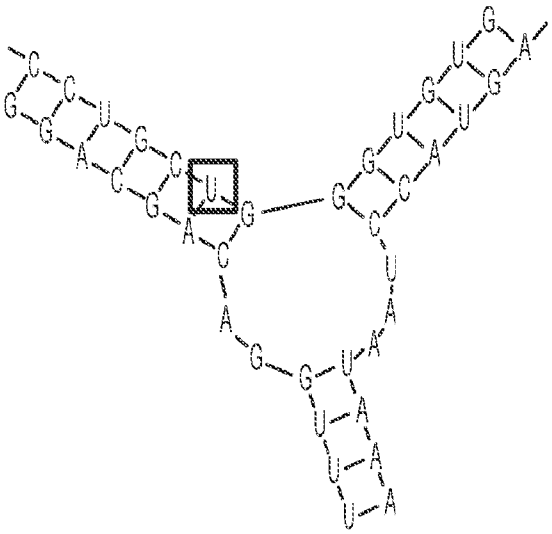


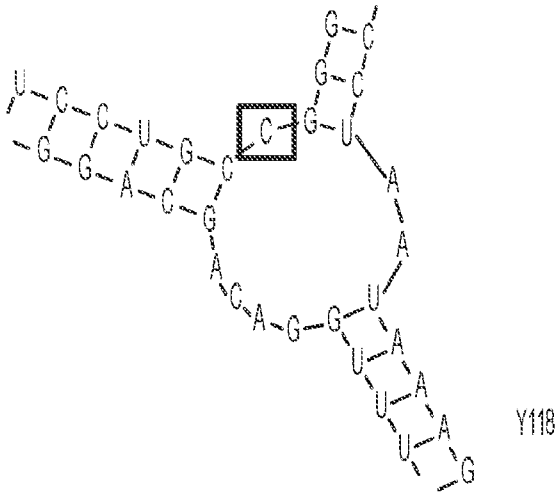
Figure 4C
CONTINUED



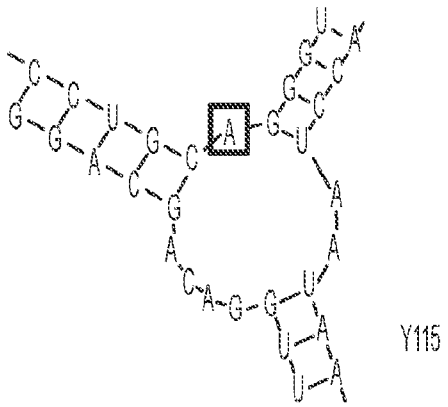
Y117



Y116



Y118



Y115

	Y79	Y115	Y116	Y117	Y118
Induction (5ug/mL)	37	29	25	47	36

Figure 4D

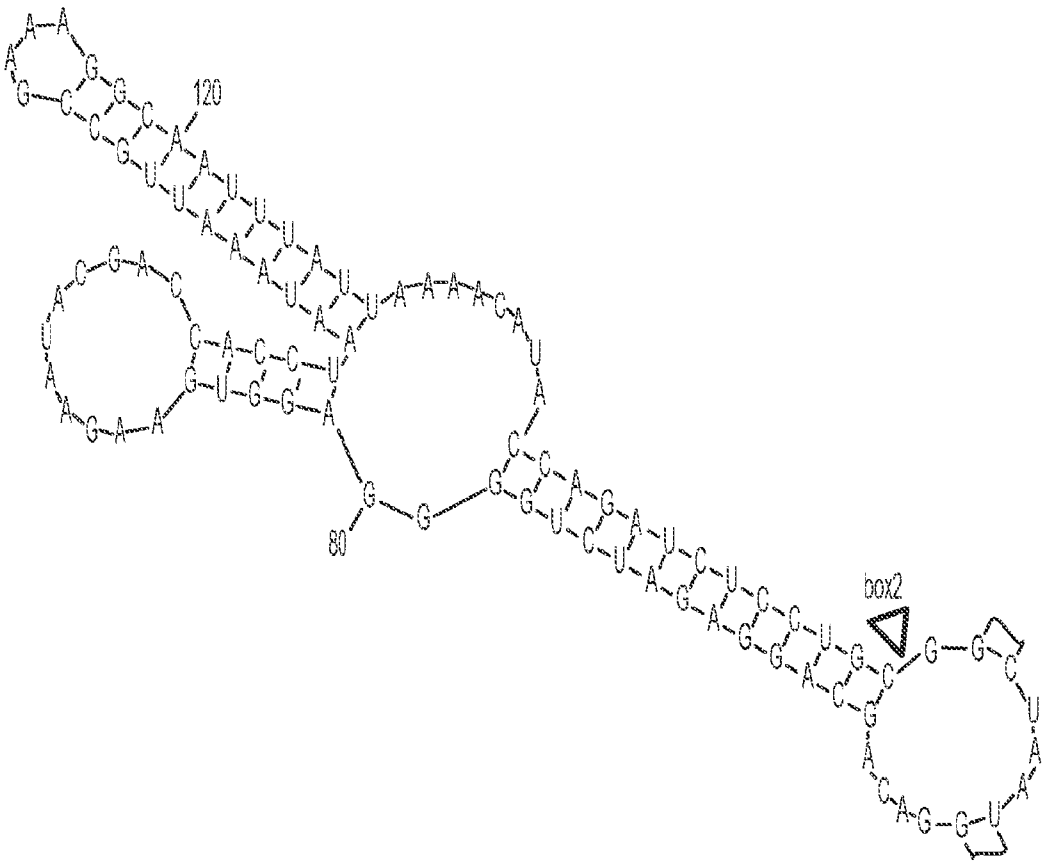


Figure 4D
CONTINUED

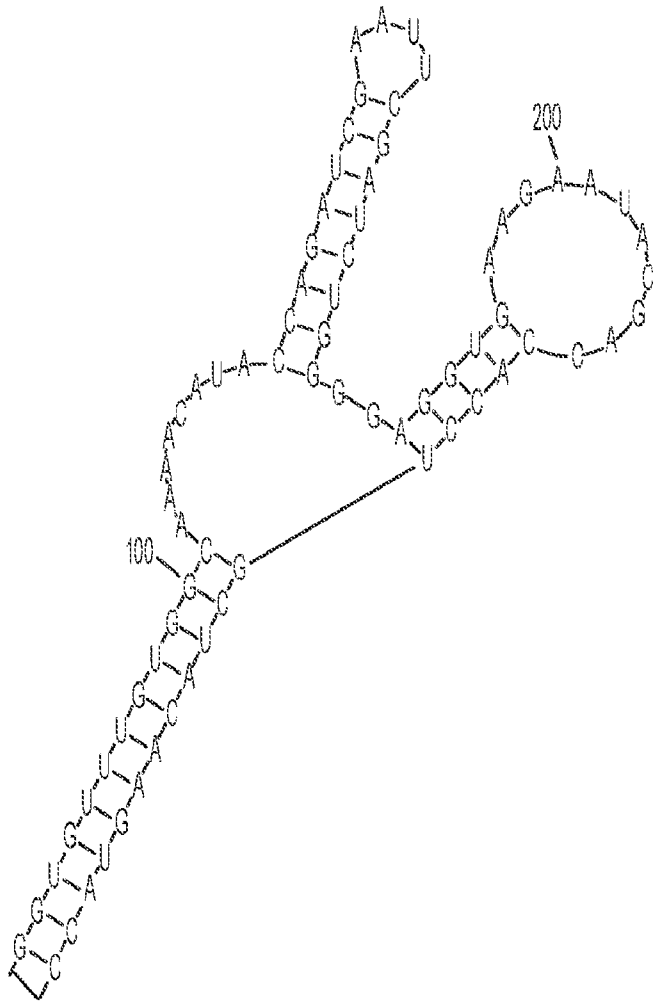


Figure 4D
CONTINUED

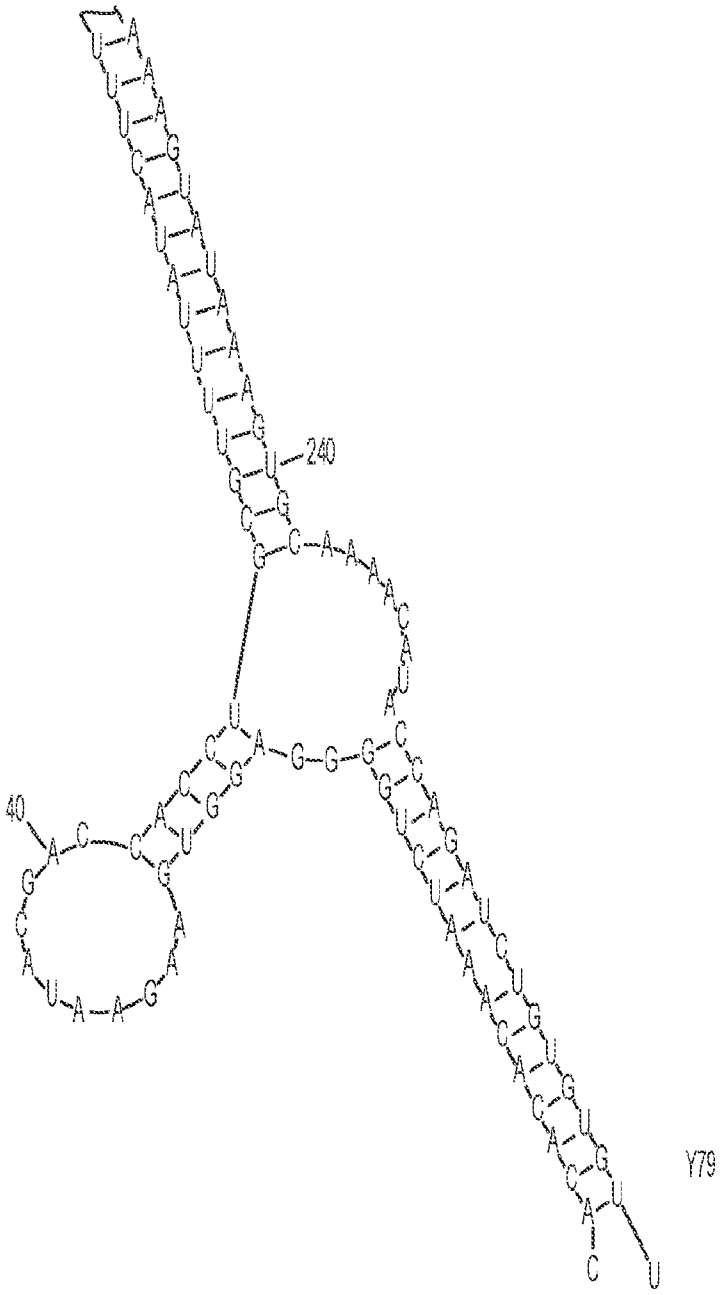
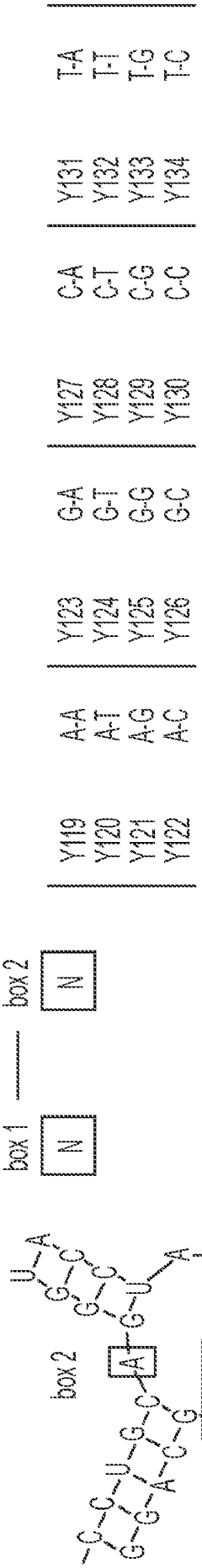
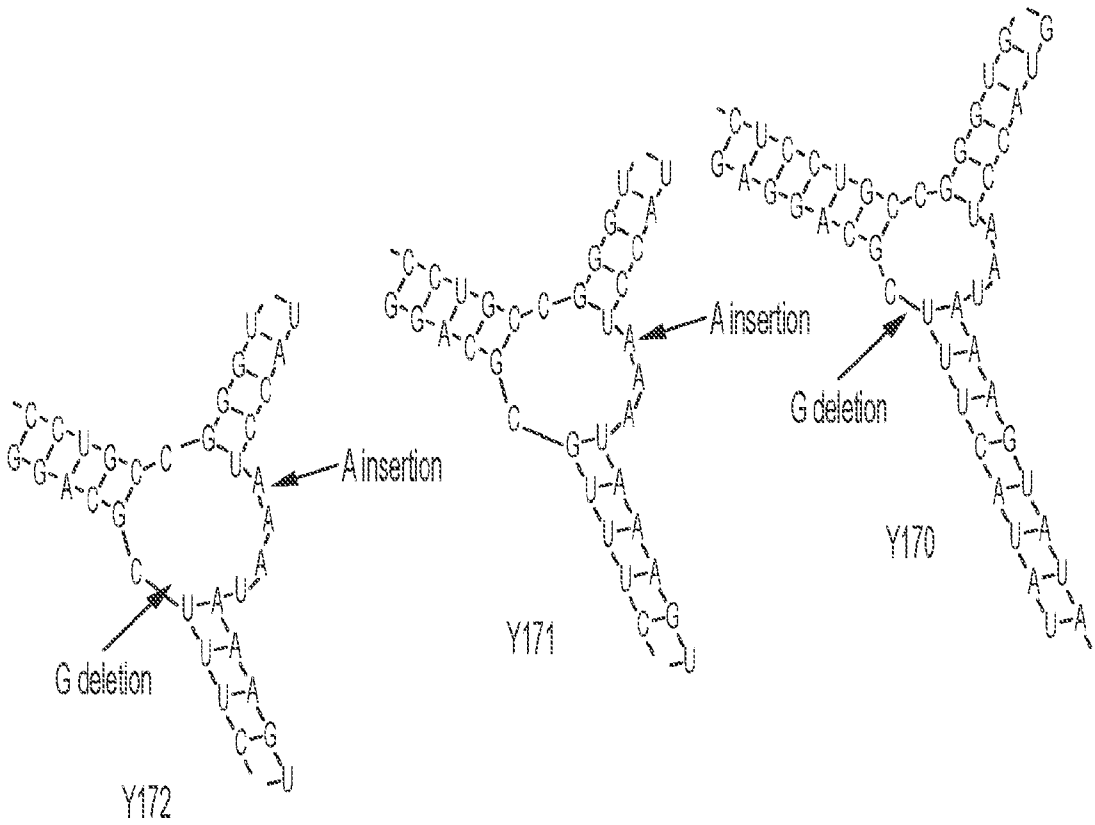


Figure 4D
CONTINUED



	Y119	A-A	Y123	G-A	Y127	C-A	Y131	T-A						
	Y120	A-T	Y124	G-T	Y128	C-T	Y132	T-T						
	Y121	A-G	Y125	G-G	Y129	C-G	Y133	T-G						
	Y122	A-C	Y126	G-C	Y130	C-C	Y134	T-C						
(1ug/mL)	Y79		Y119		Y120		Y121		Y122					
Induction	21		12		9		10		15					
	Y79		Y123		Y124		Y125		Y126					
Induction	16		10		9		5		5					
	Y79		Y127		Y128		Y129		Y130					
Induction	16		15		9		7		18					
	Y79		Y131		Y132		Y133		Y134					
Induction	12		7		10		11		15					

Figure 4E



	Y130	Y170	Y171	Y172
Induction (1ug/mL)	17	17	18	13

Figure 4F

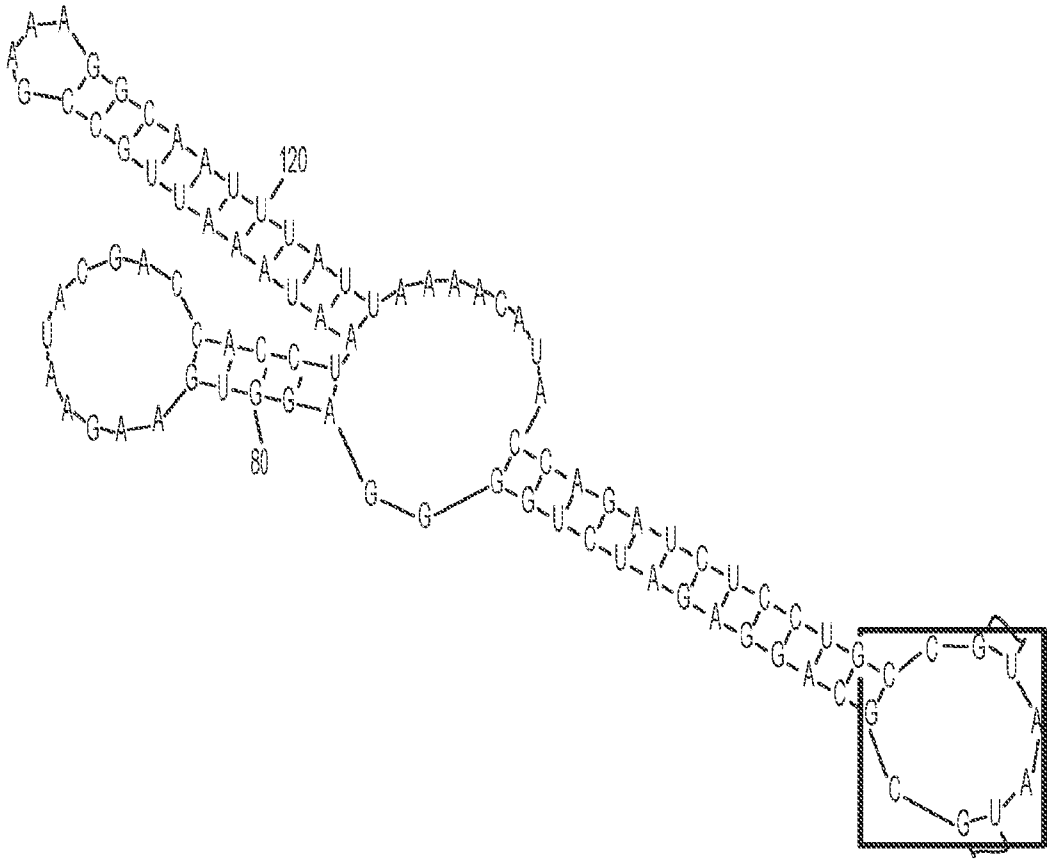


Figure 4F
CONTINUED

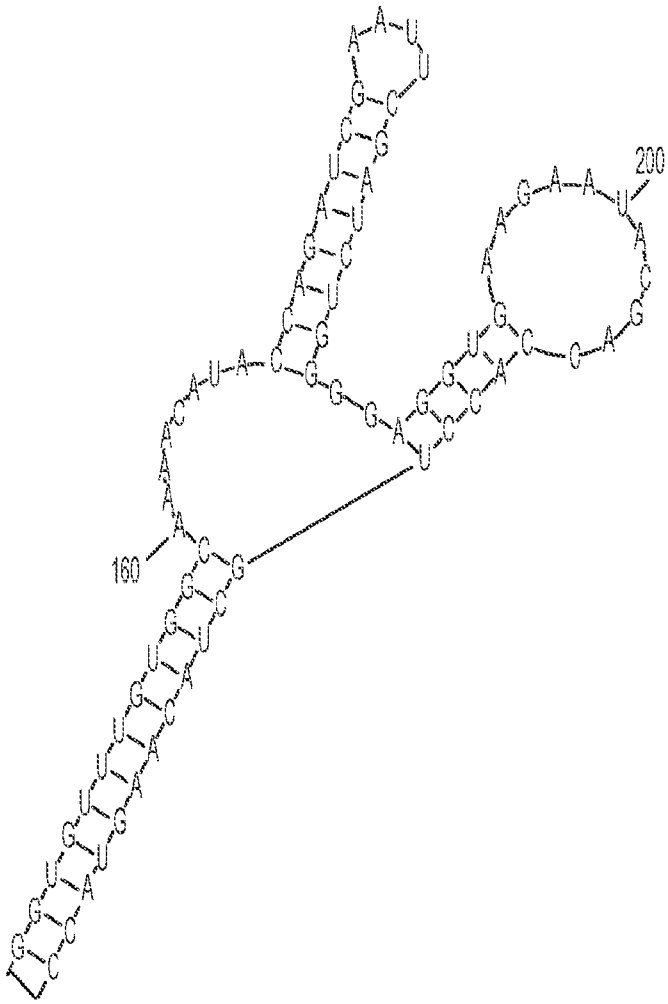


Figure 4F
CONTINUED

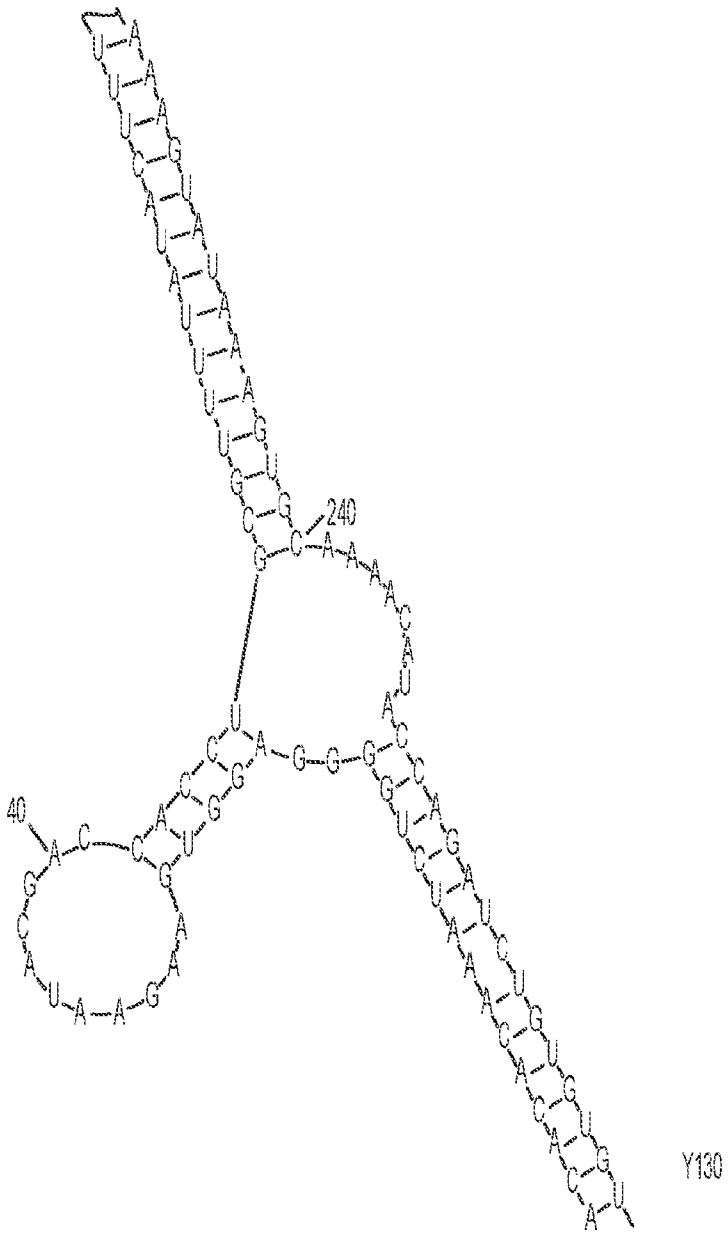


Figure 4F
CONTINUED

	Y143	Y160	Y161
Induction (1 ug/mL)	15	13	16

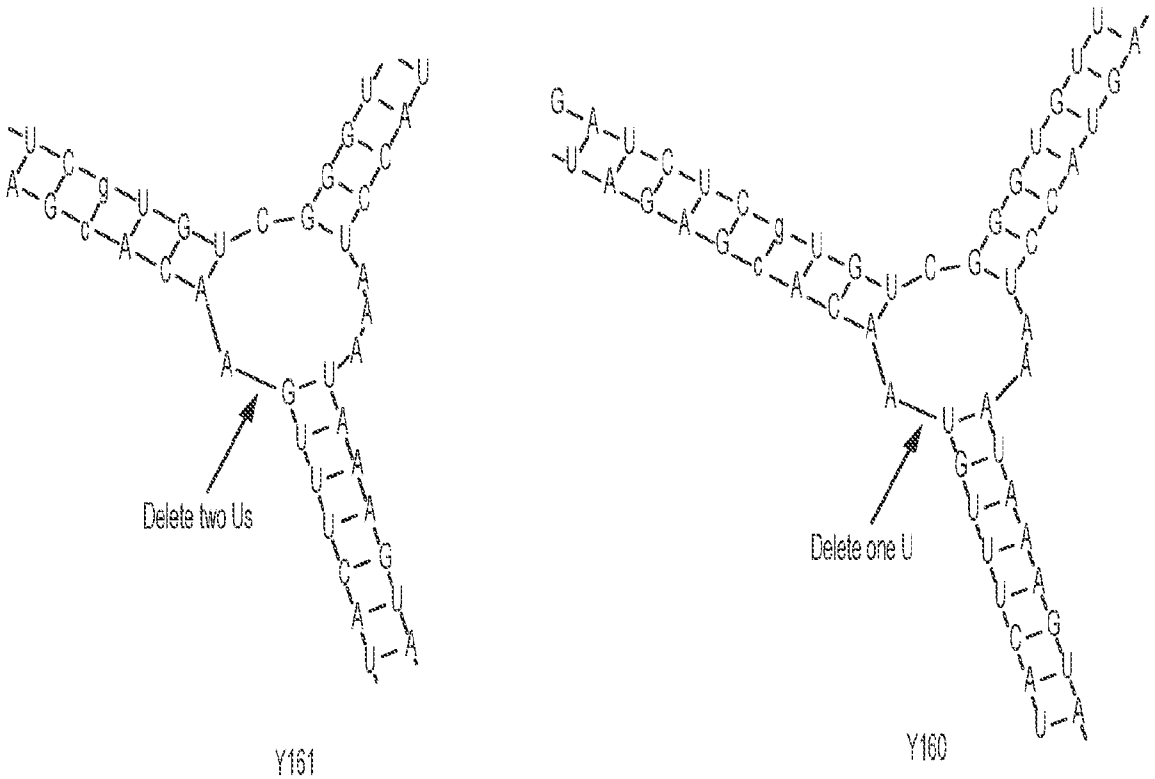


Figure 4G

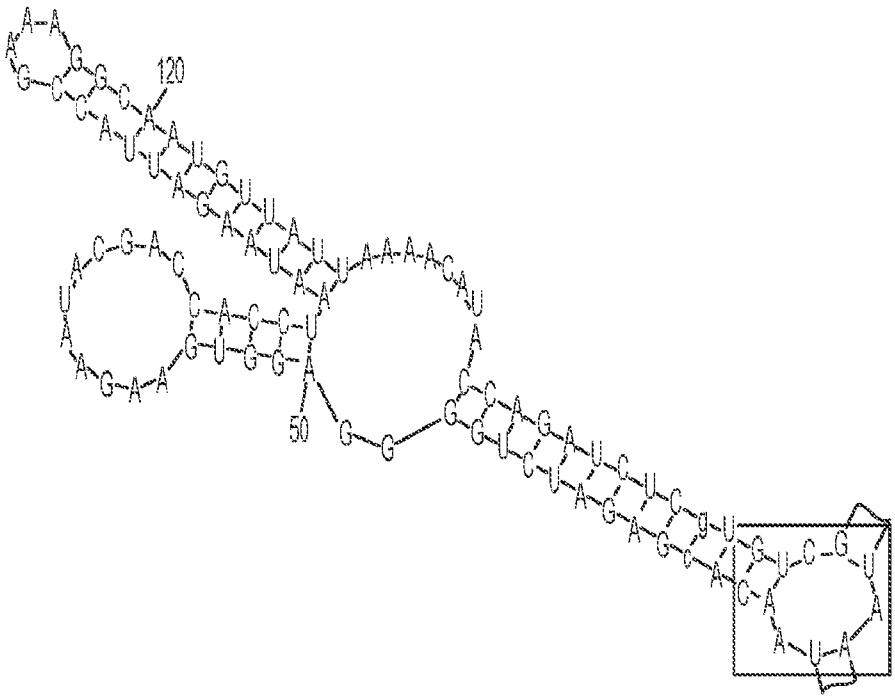


Figure 4G
CONTINUED

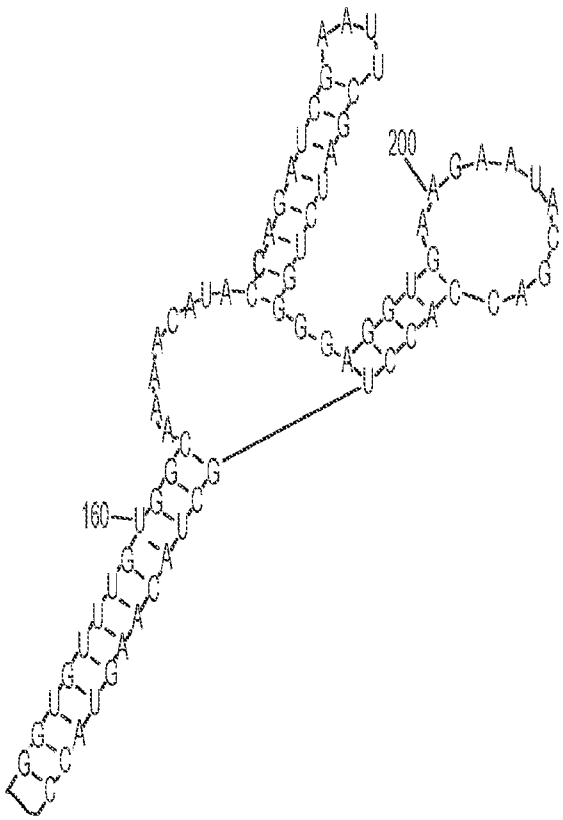


Figure 4G
CONTINUED

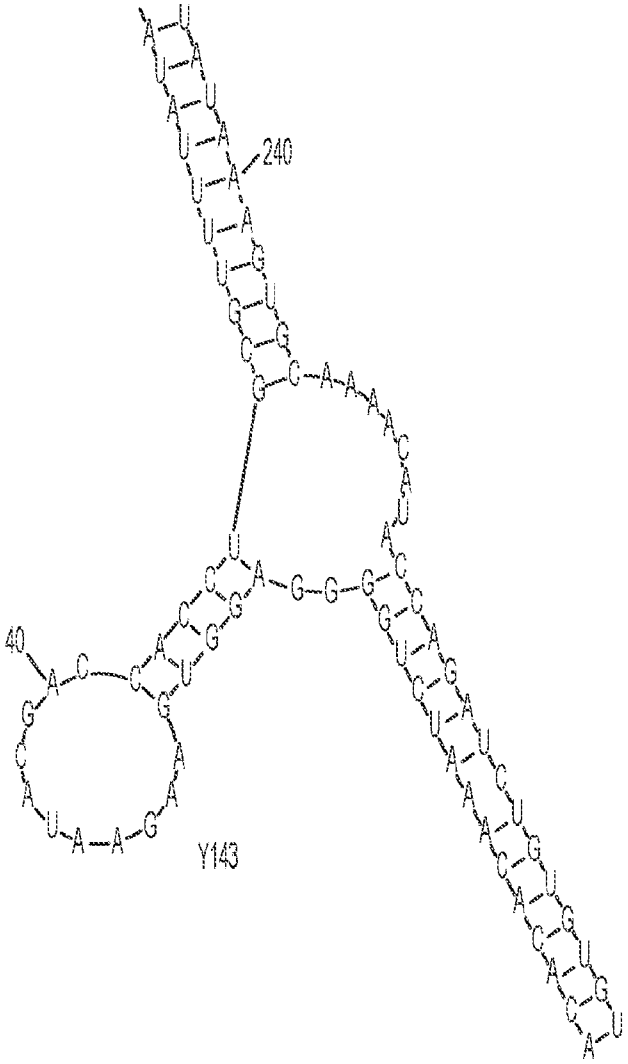
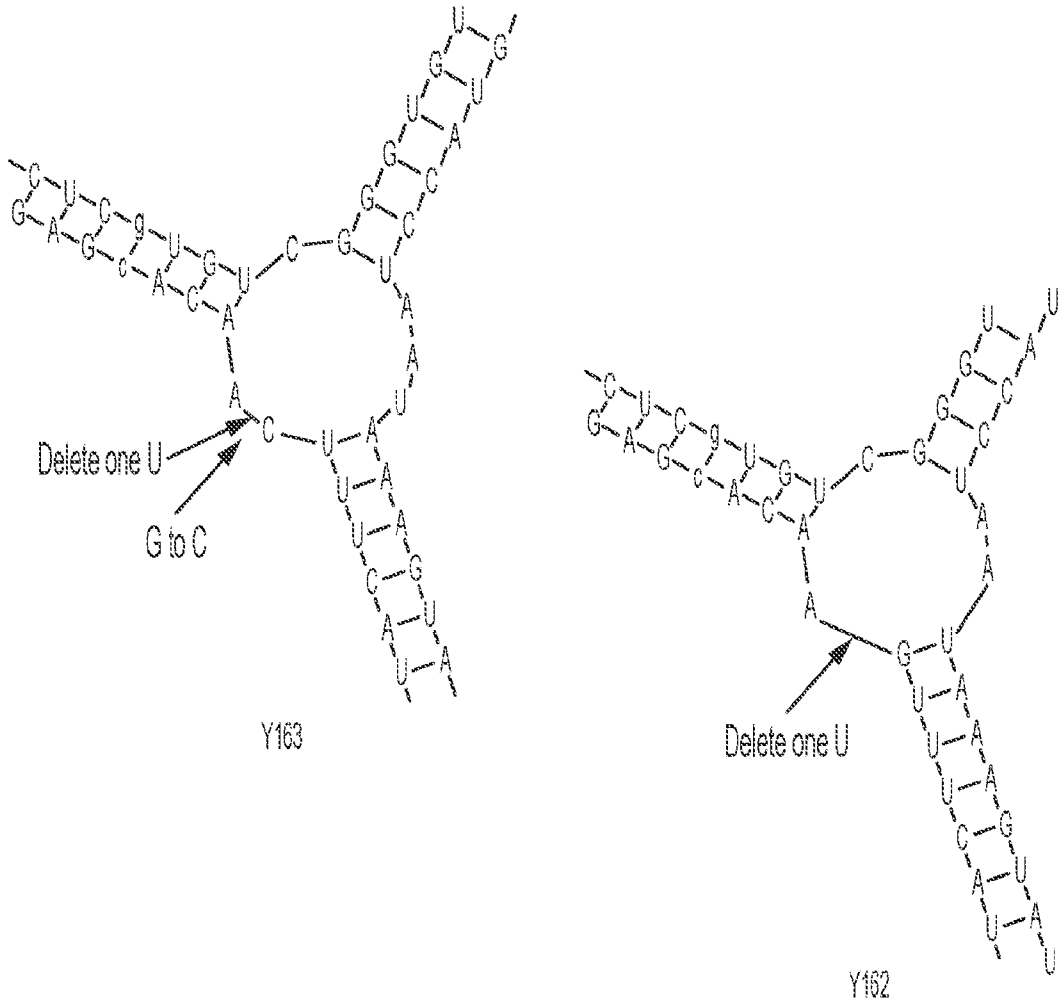


Figure 4G
CONTINUED



	Y147	Y162	Y163
Induction (1 ug/mL)	18	15	20

Figure 4H

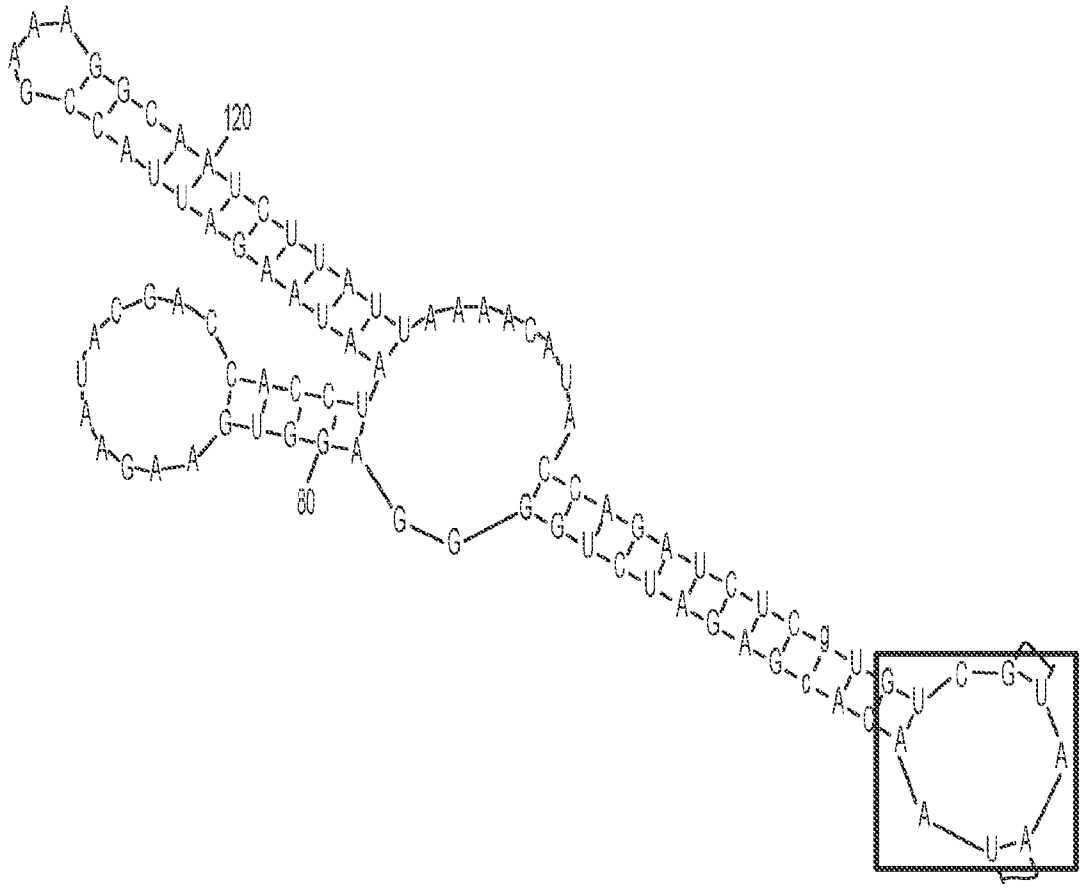


Figure 4H
CONTINUED

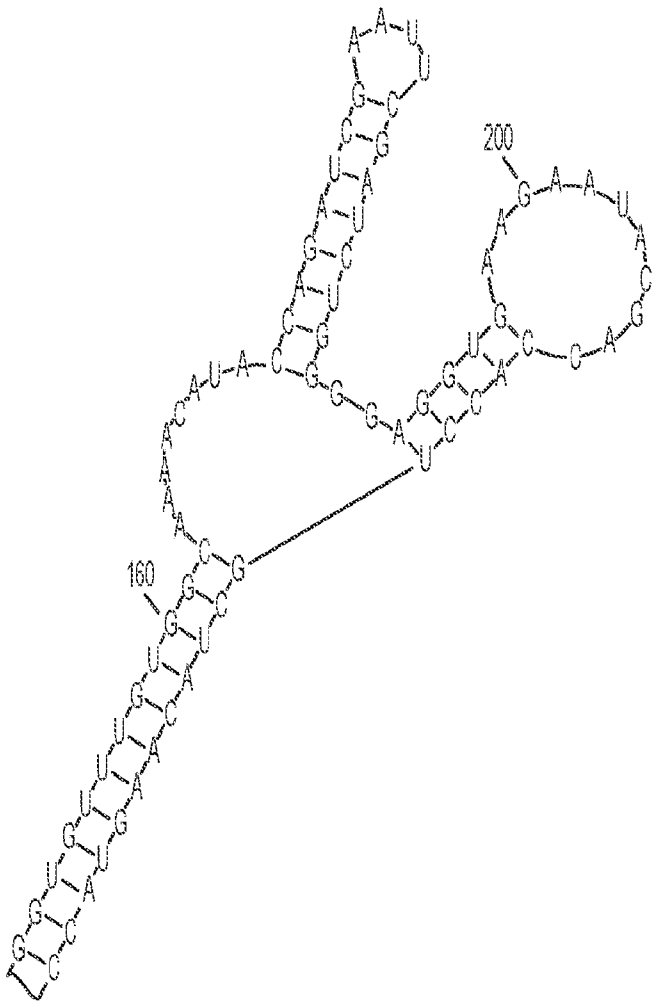


Figure 4H
CONTINUED

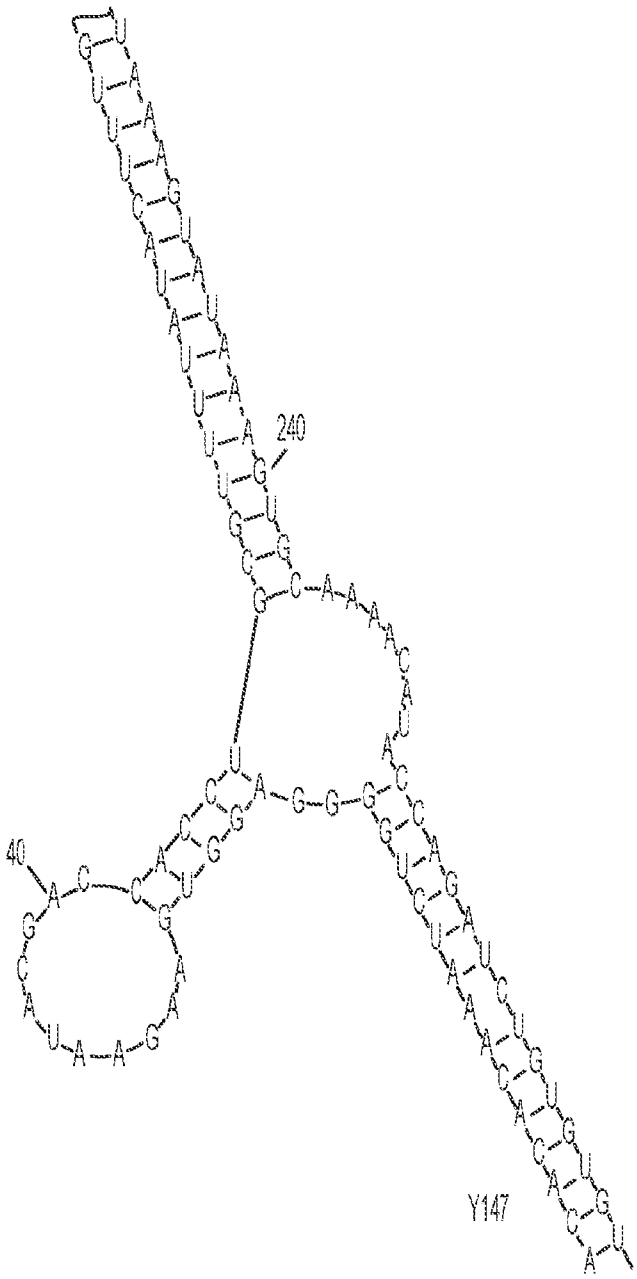
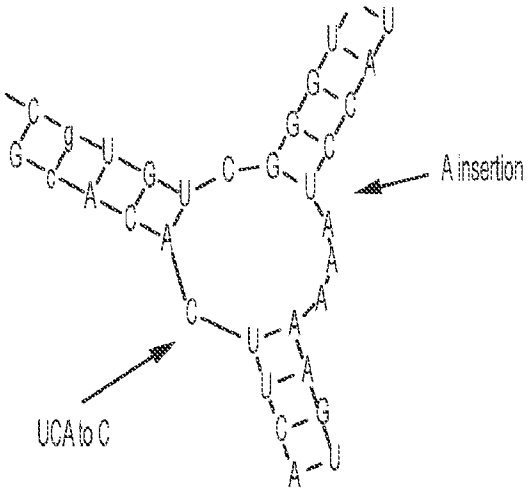
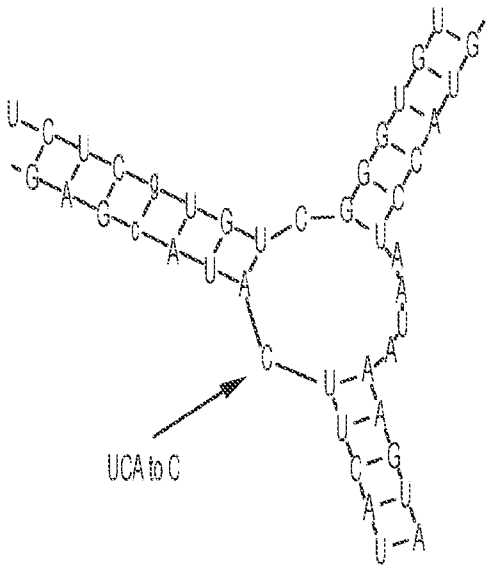


Figure 4H
CONTINUED

	Y163	Y177	Y178
Induction (1ug/mL)	16	23	10



Y178



Y177

Figure 4I

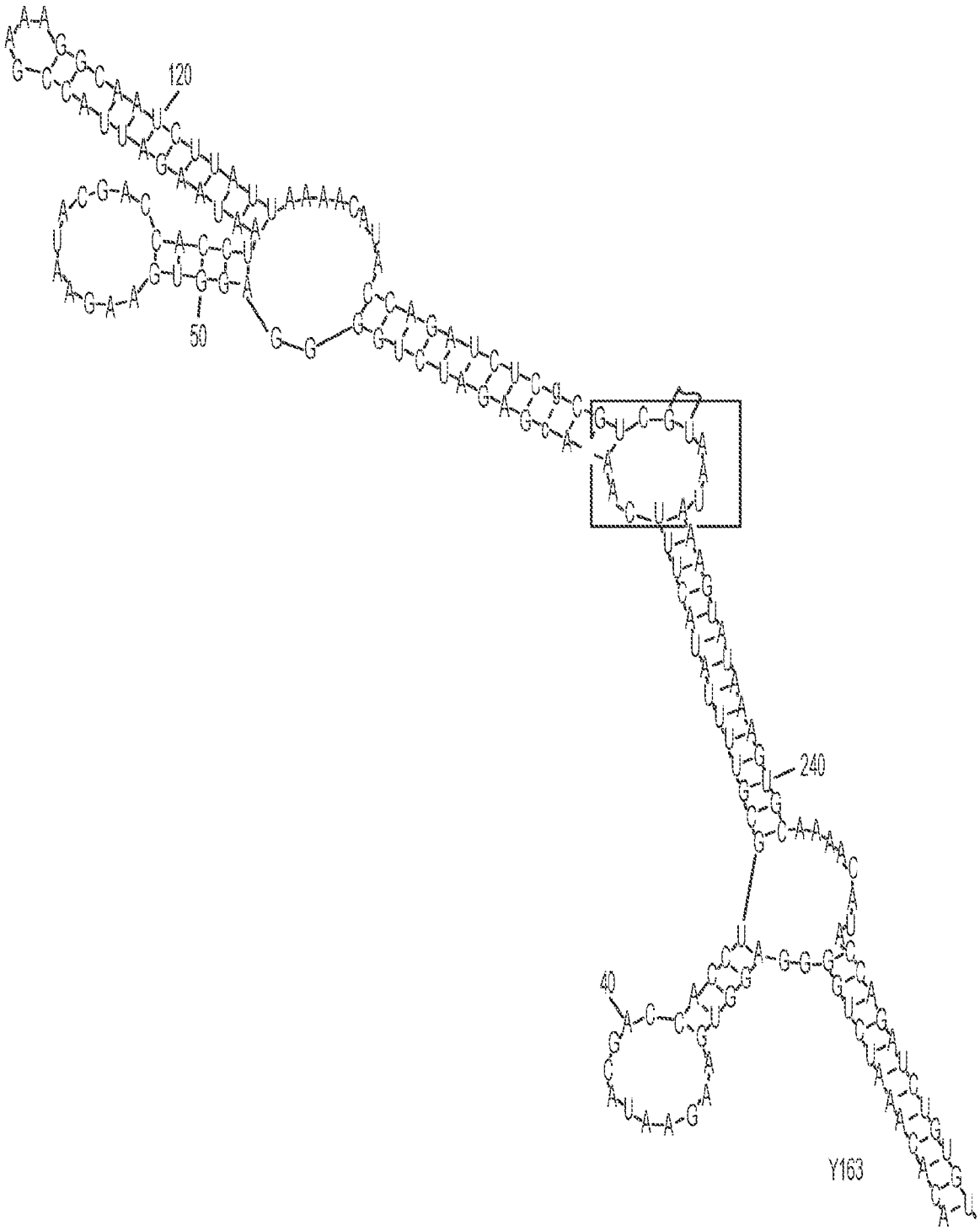


Figure 4I
CONTINUED

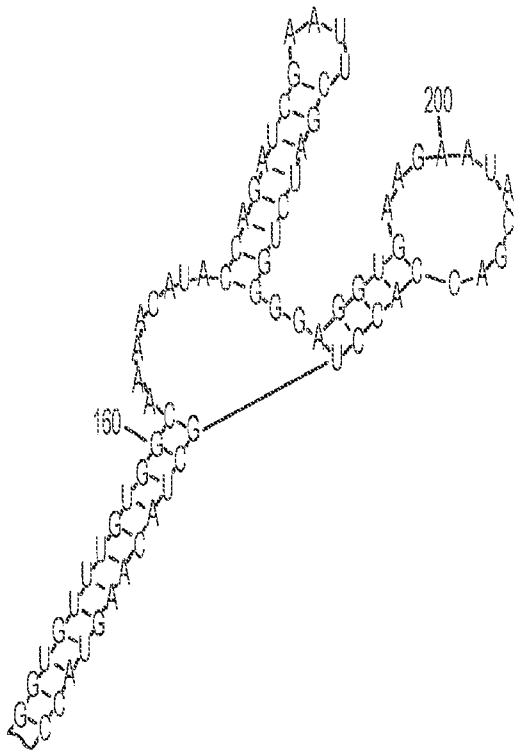


Figure 4I
CONTINUED

	Y152	Y164	Y165	Y166
Induction (1ug/mL	12	16	17	24

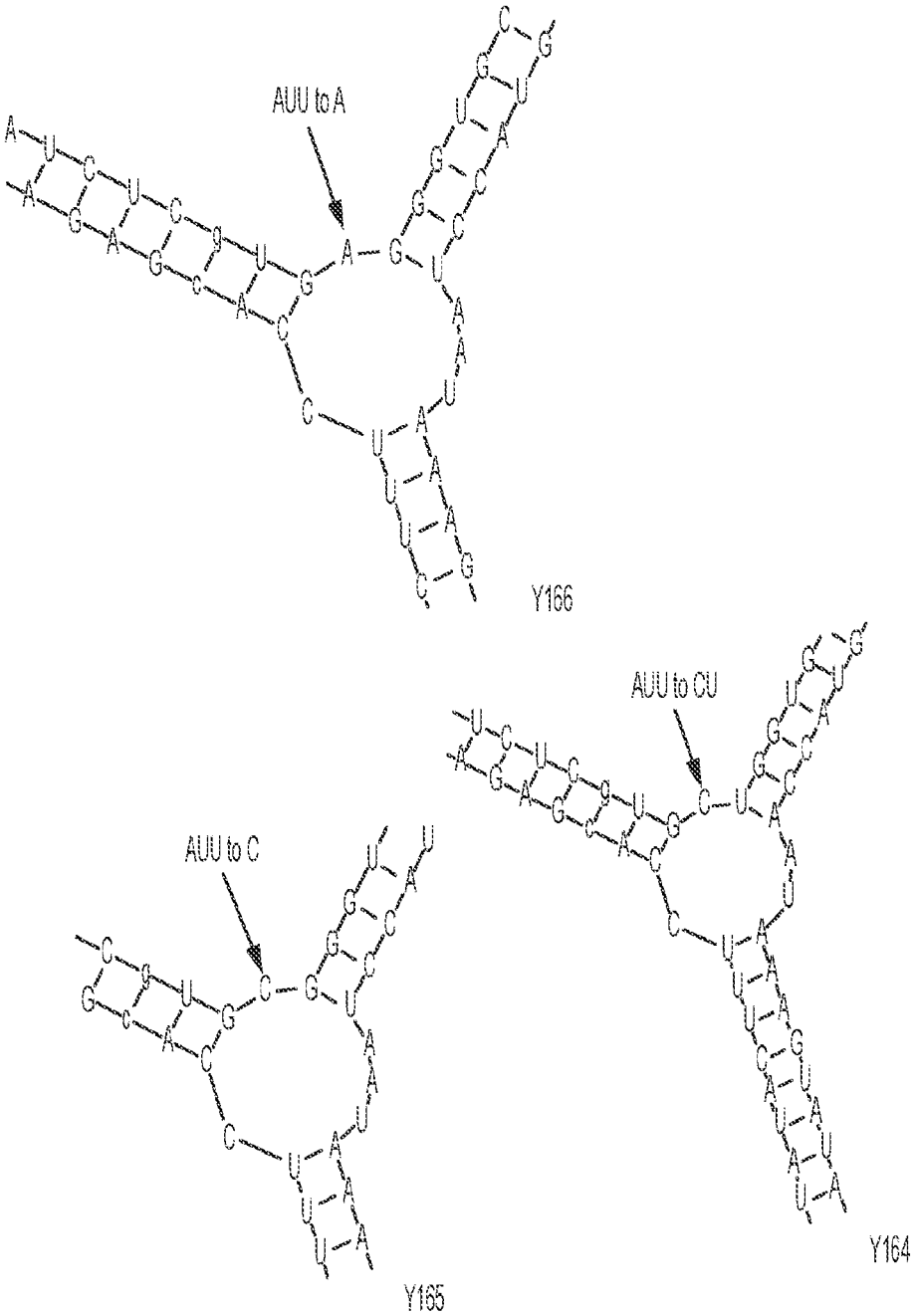


Figure 4J

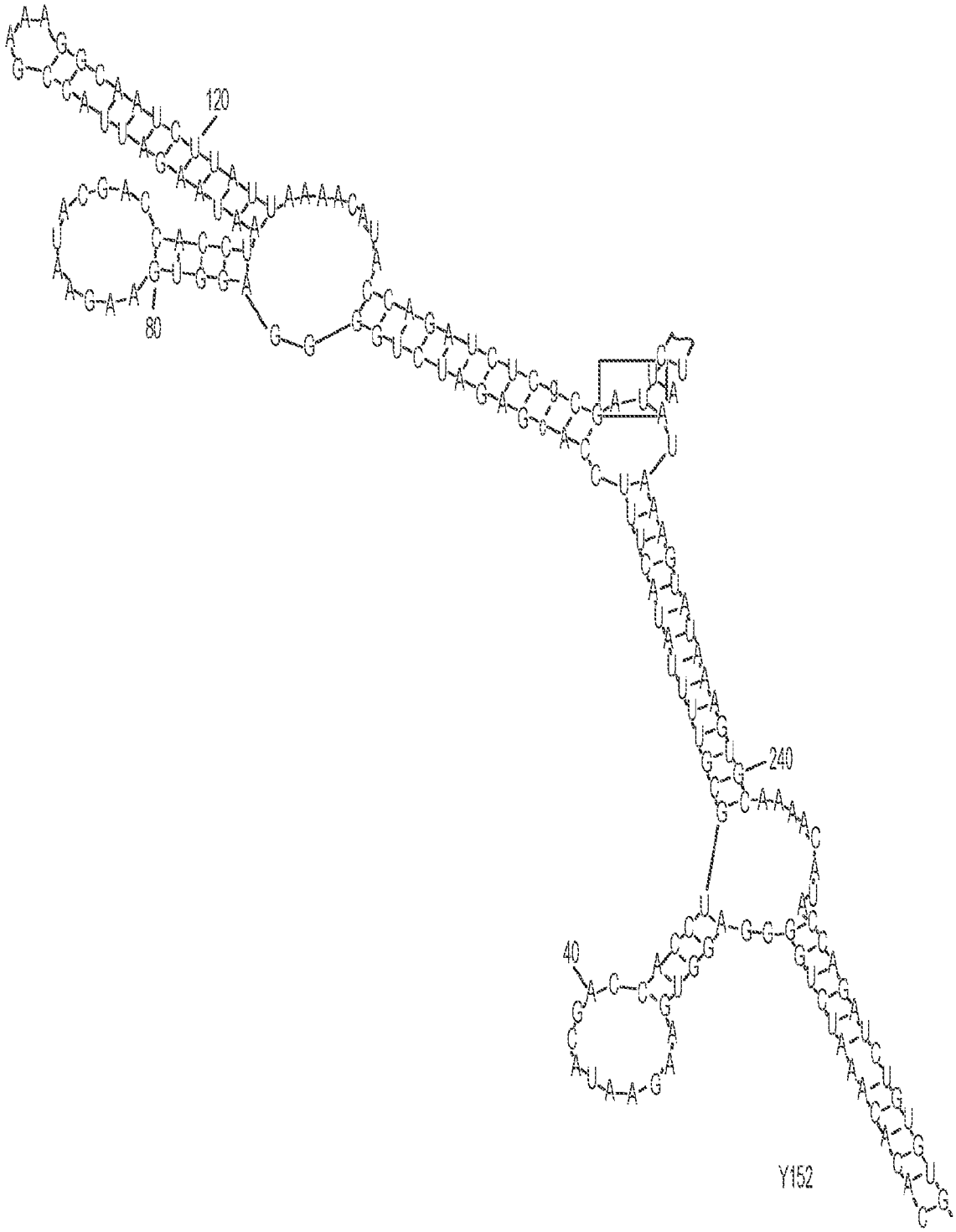


Figure 4J
CONTINUED

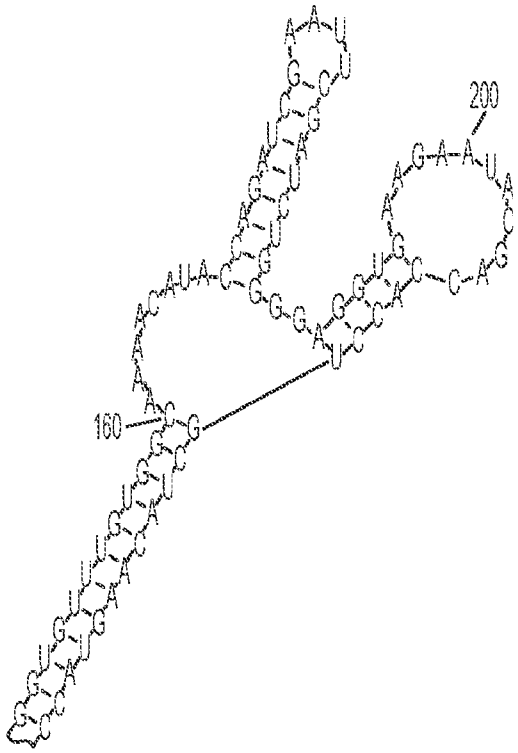


Figure 4J
CONTINUED

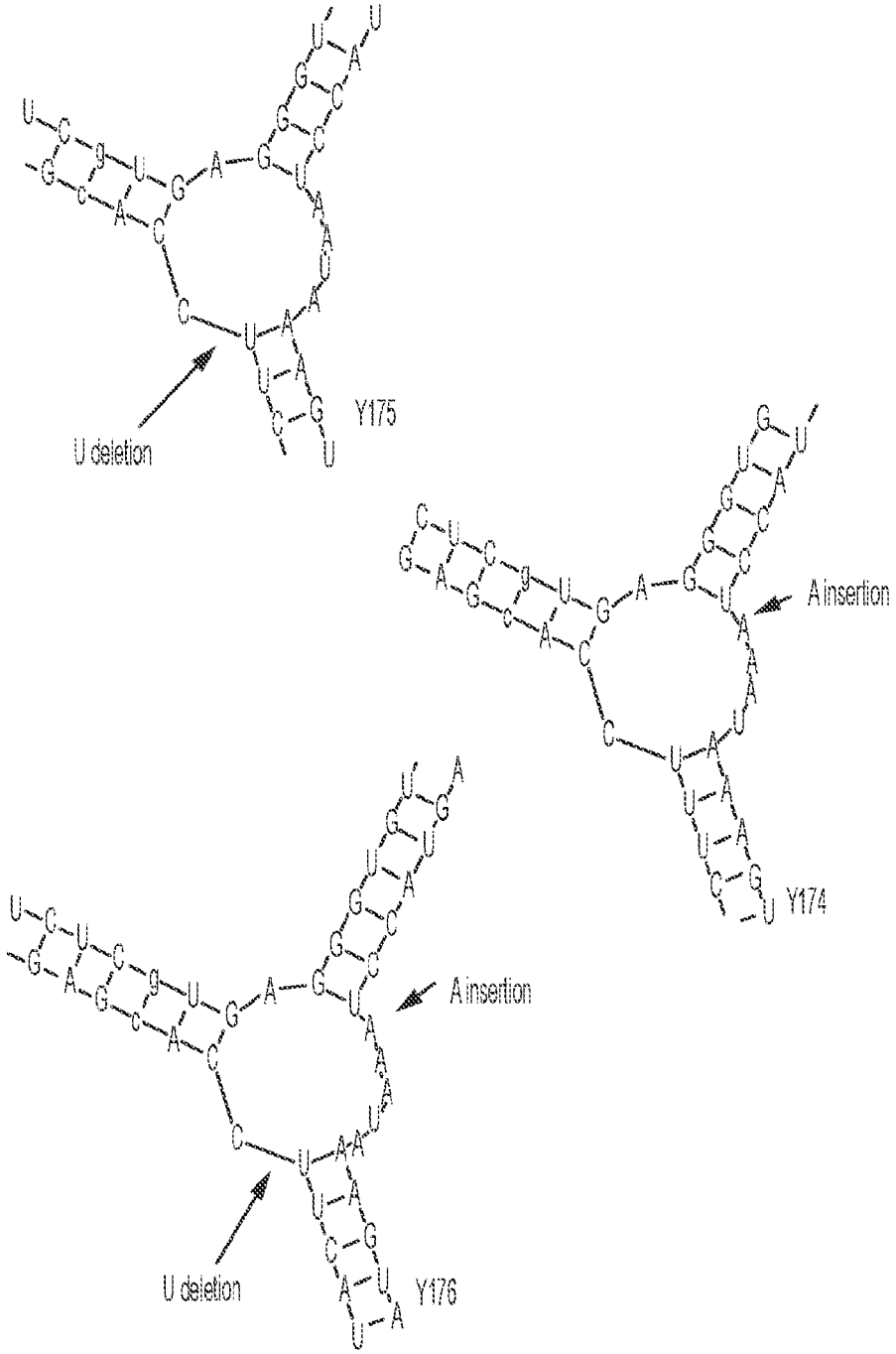
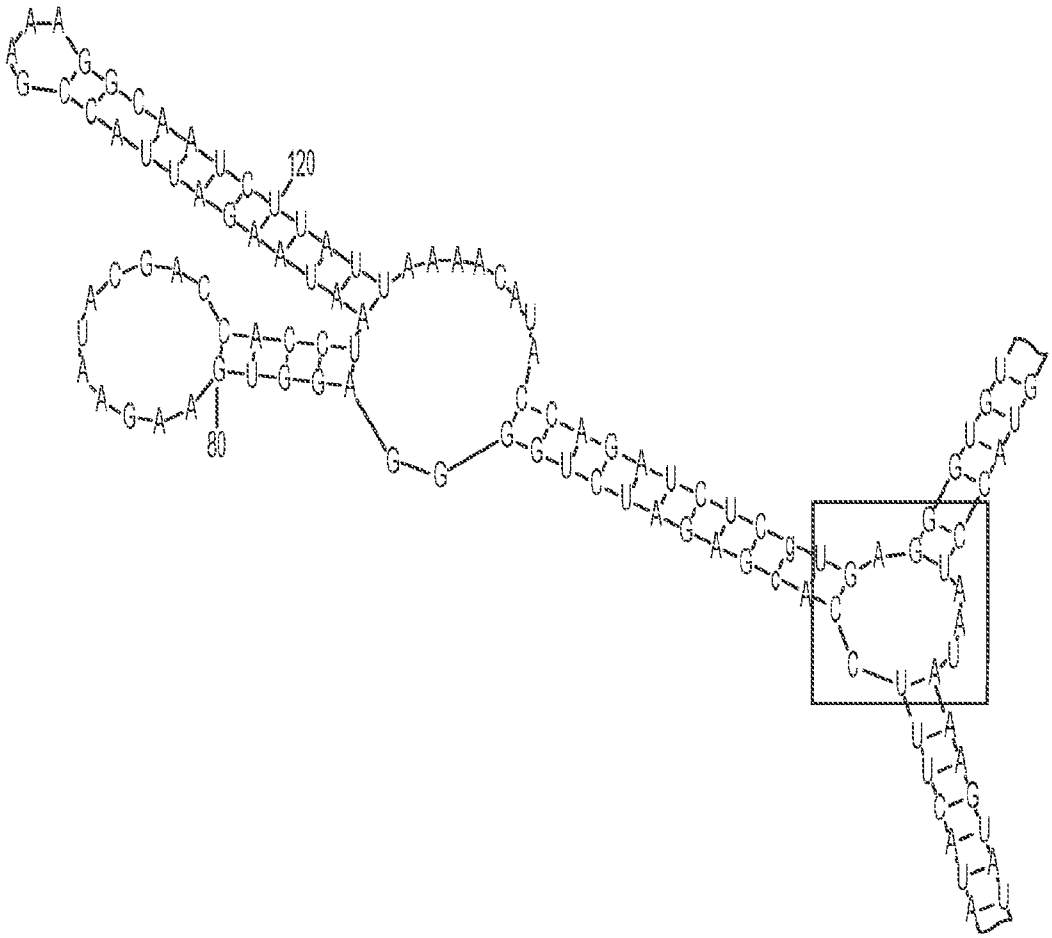

U deletion
Y175
A insertion
Y174
A insertion
U deletion
Y176

Figure 4K



	Y166	Y174	Y175	Y176
Introduction (1ug/mL)	23	35	38	35

Figure 4K
CONTINUED

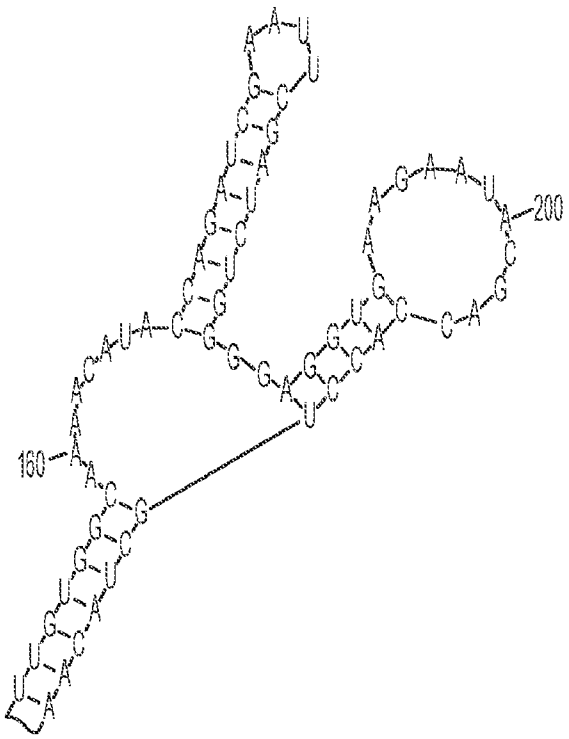


Figure 4K
CONTINUED

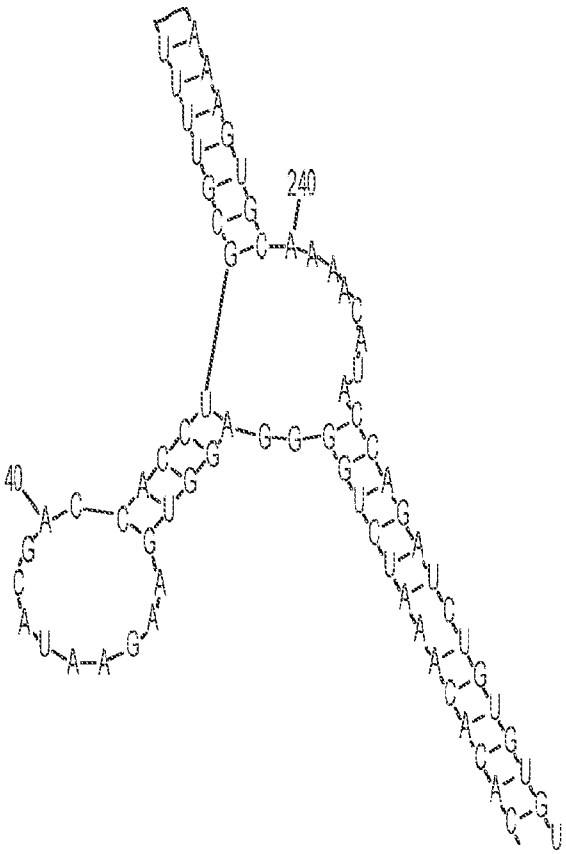
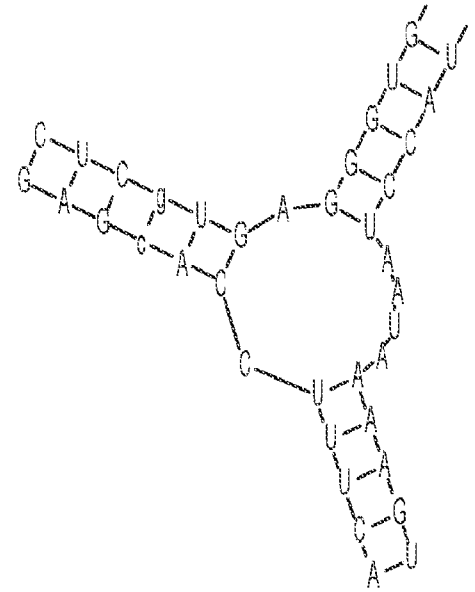
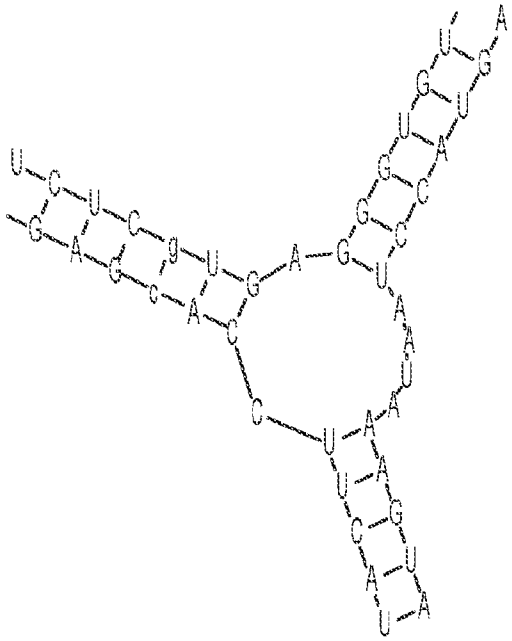


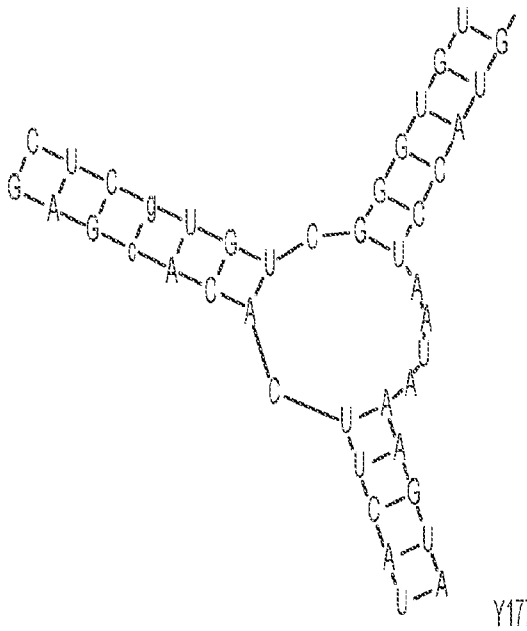
Figure 4K
CONTINUED



Y174



Y176



Y177

Figure 4L

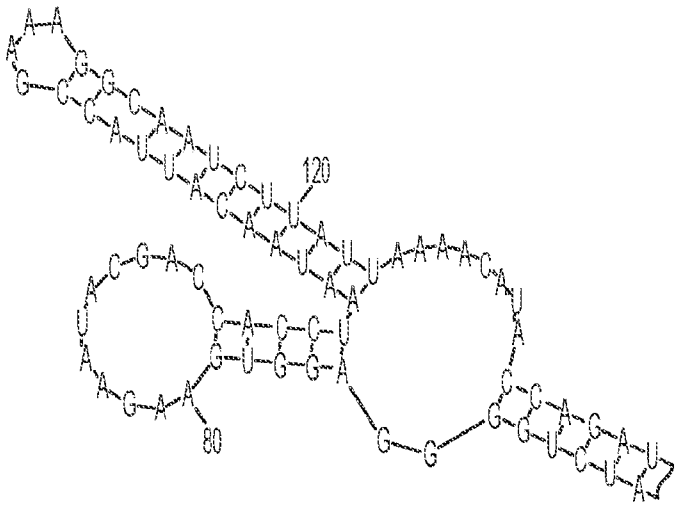


Figure 4L
CONTINUED

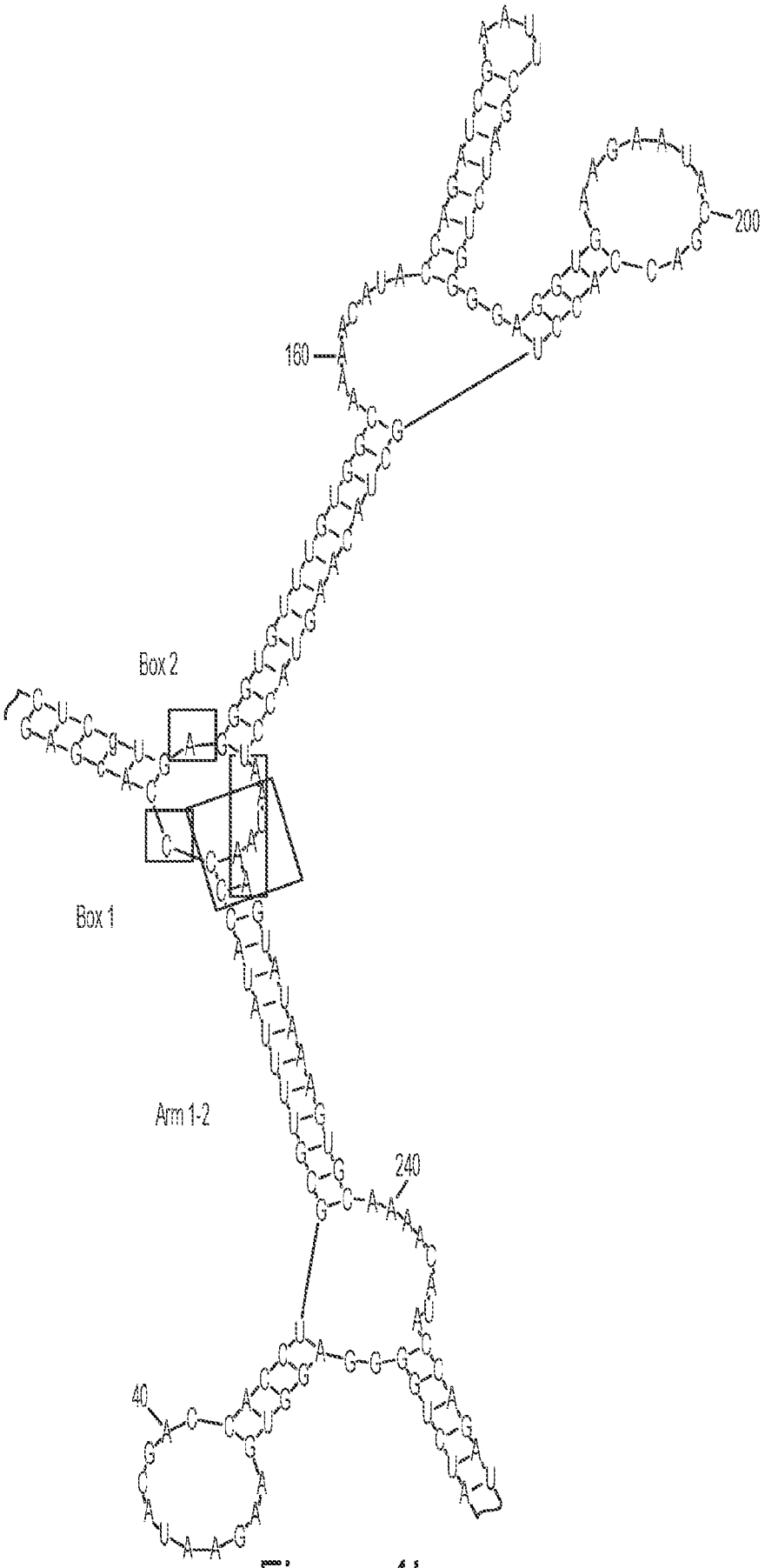


Figure 4L
CONTINUED

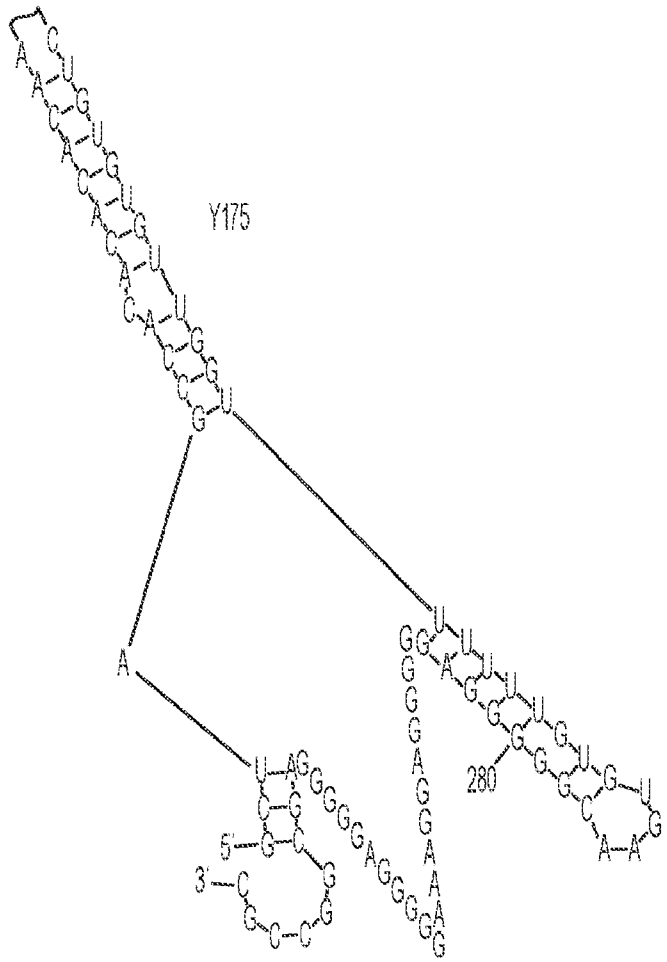
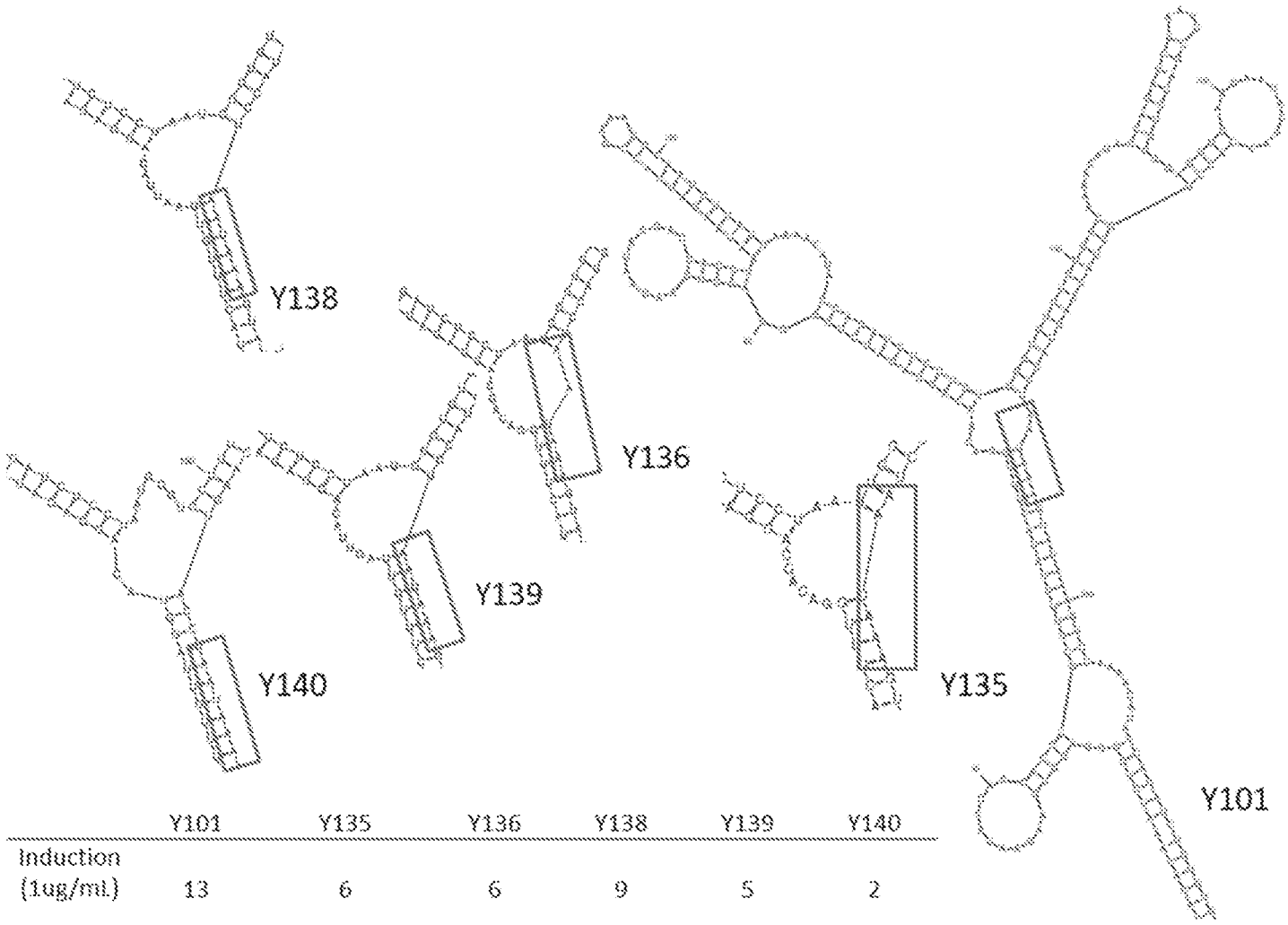


Figure 4L
CONTINUED

Figure 5



	Y101	Y135	Y136	Y138	Y139	Y140
Induction (1ug/mL)	13	6	6	9	5	2

Figure 5 (continued)

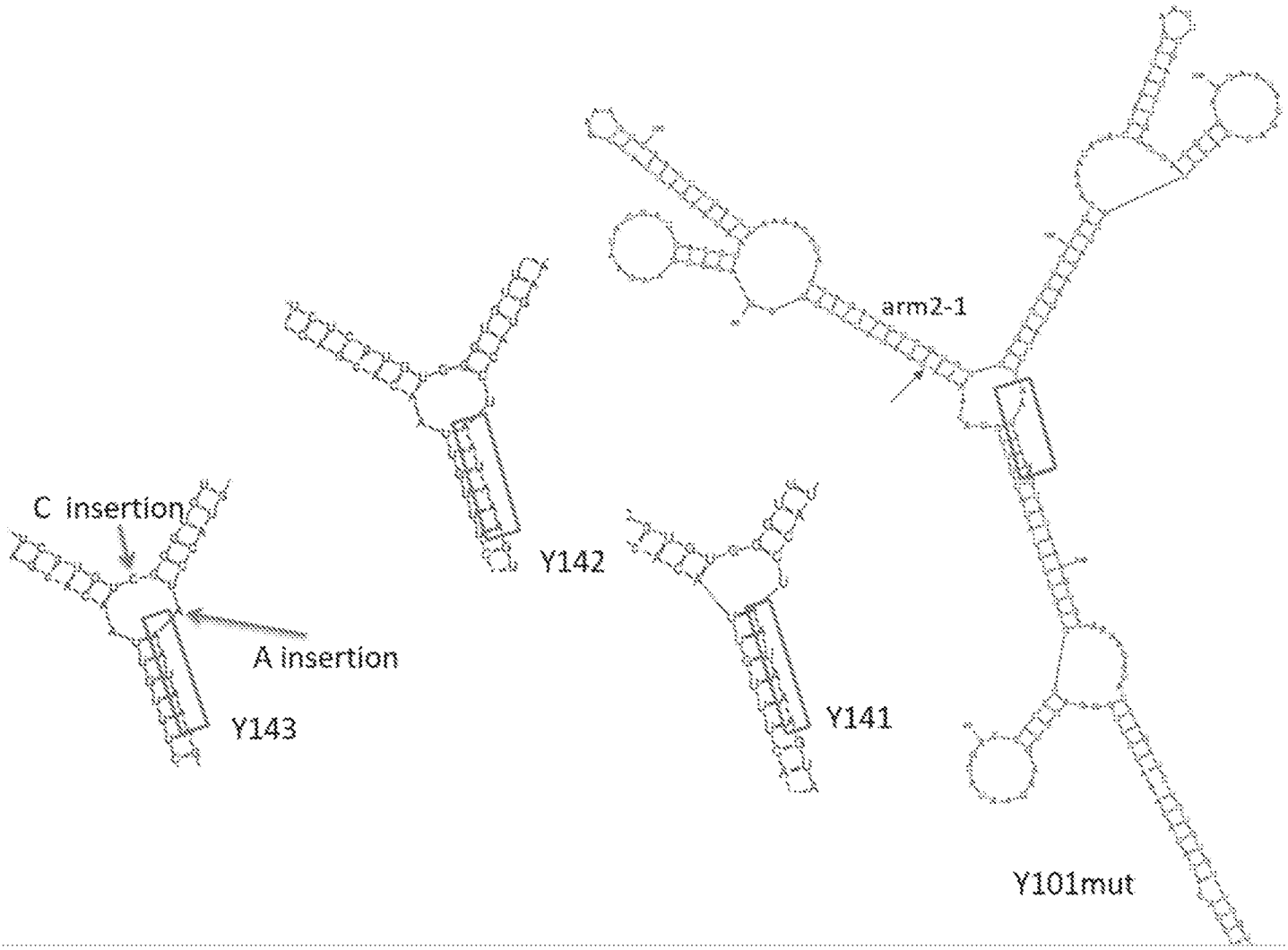
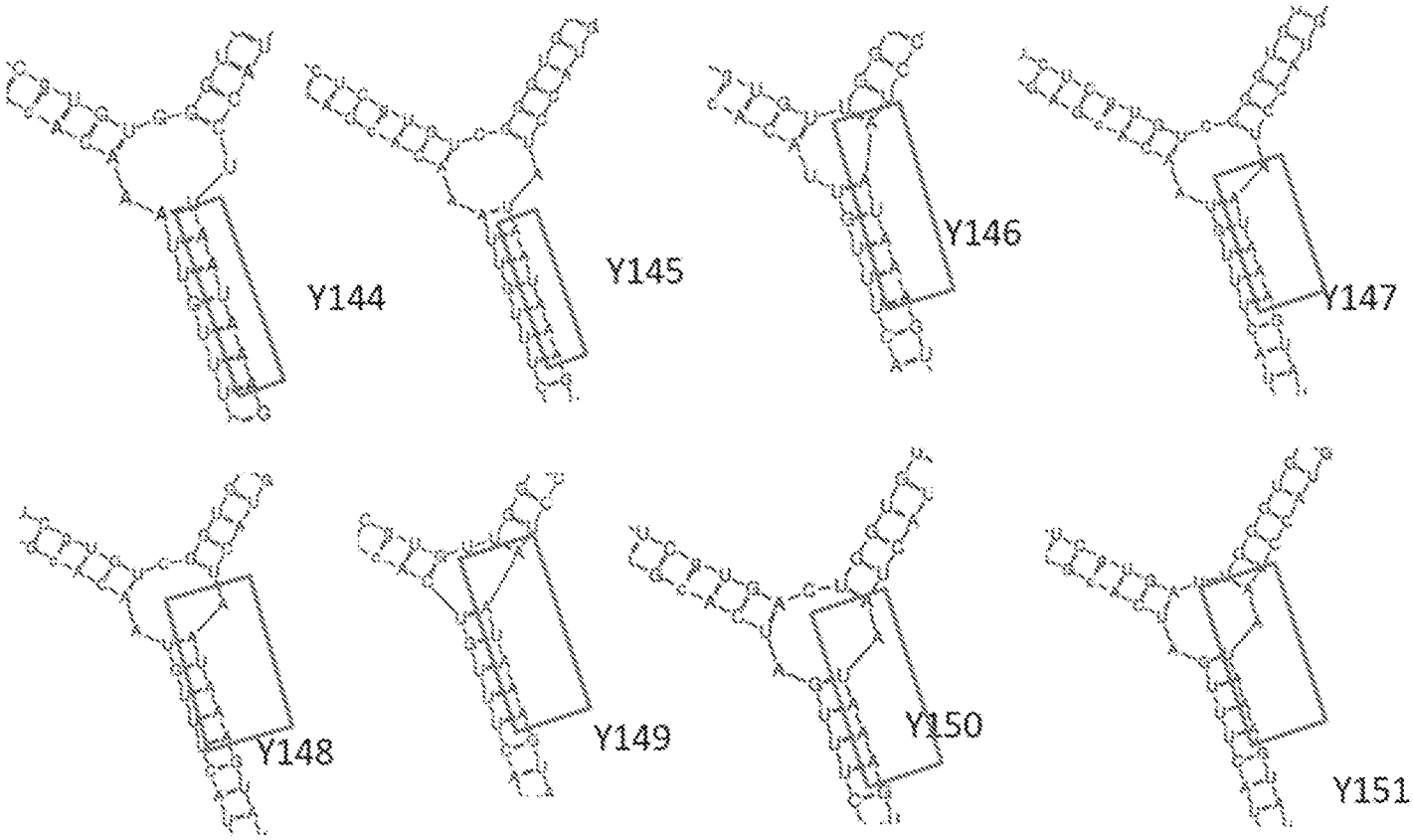


Figure 5 (continued)



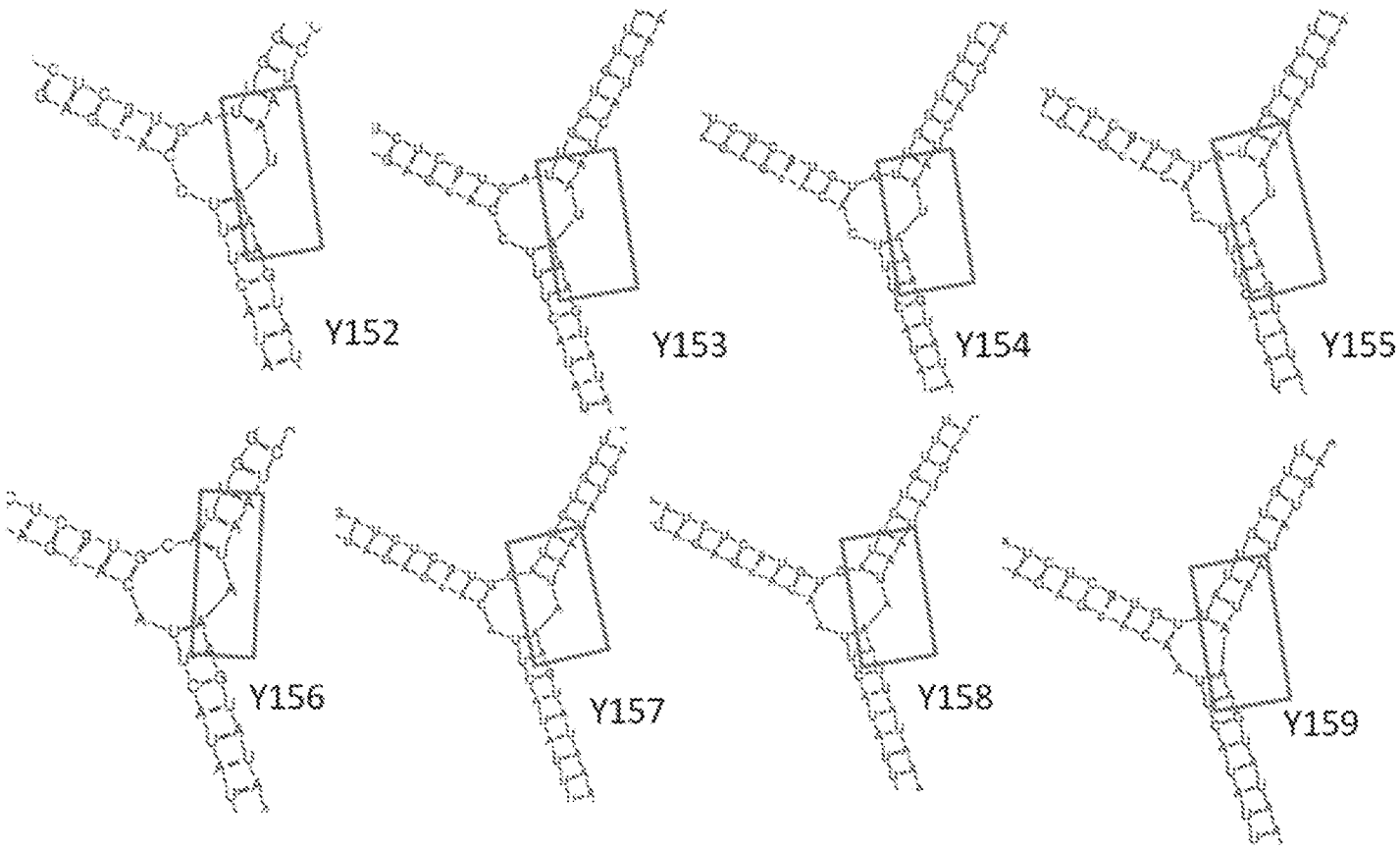


Figure 5 (continued)

	Y101mut	Y141	Y142	Y143	Y144	
Induction (1 ug/mL)	14	10	10	17	6	
	Y101mut	Y145	Y145	Y146	Y147	Y148
Induction (1 ug/mL)	14	10	8	12	19	6
	Y101mut	Y149	Y150	Y151	Y152	
Induction (1 ug/mL)	14	13	7	8	12	
	Y101mut	Y153	Y153	Y154	Y156	
Induction (1 ug/mL)	14	8	10	14	10	
	Y101mut	Y157	Y158	Y159		
Induction (1 ug/mL)	14	2	2	2		

Figure 6A

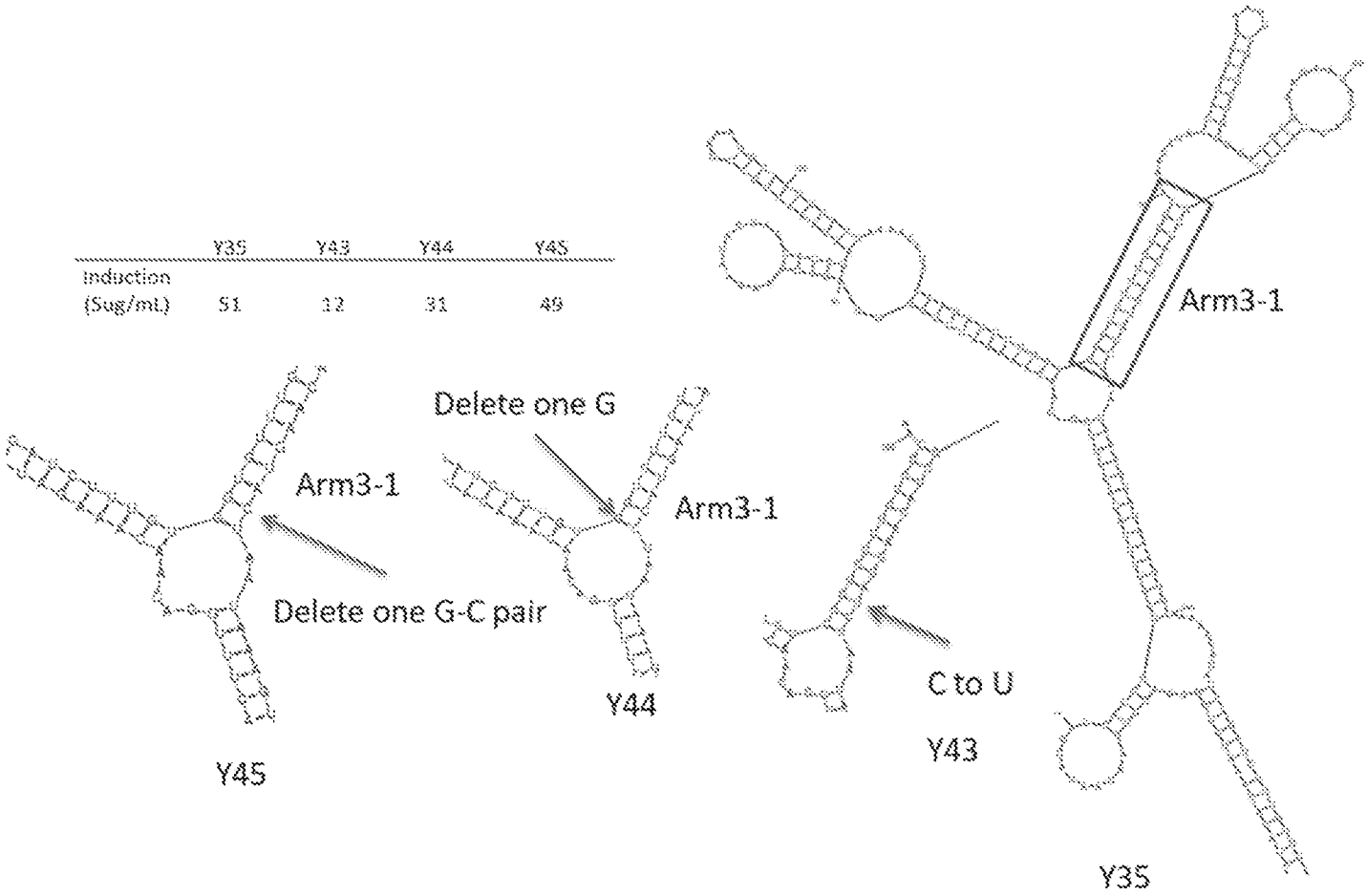
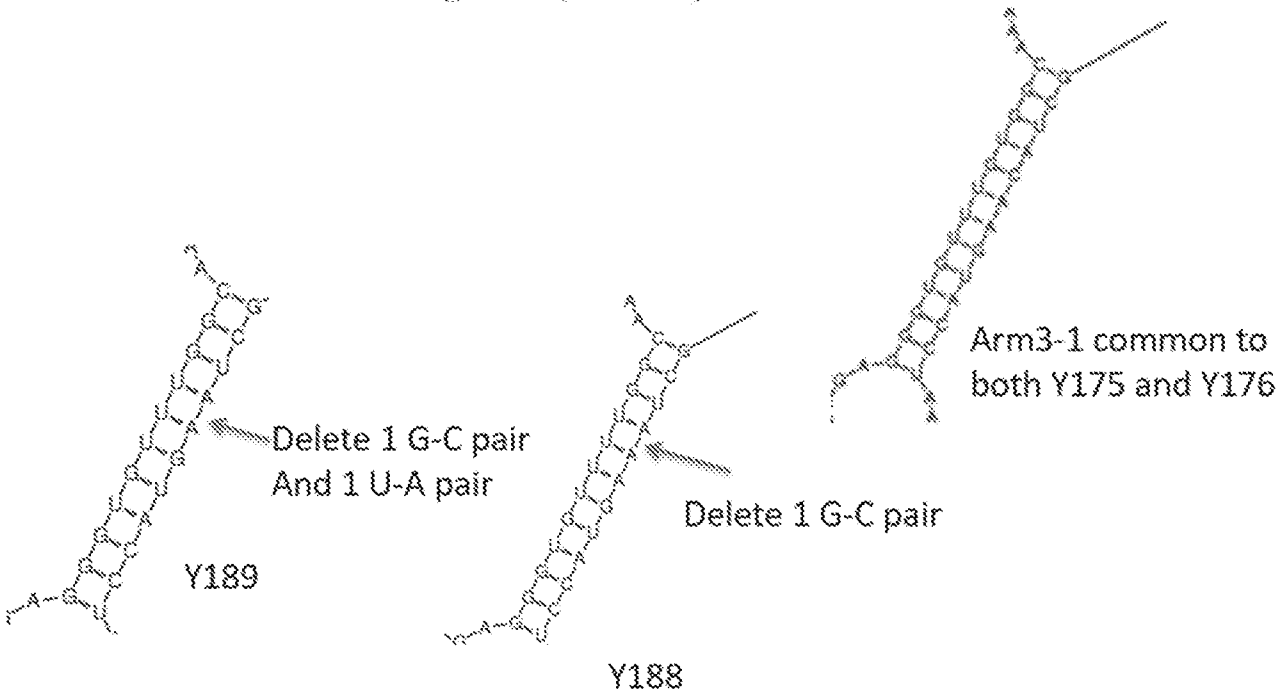
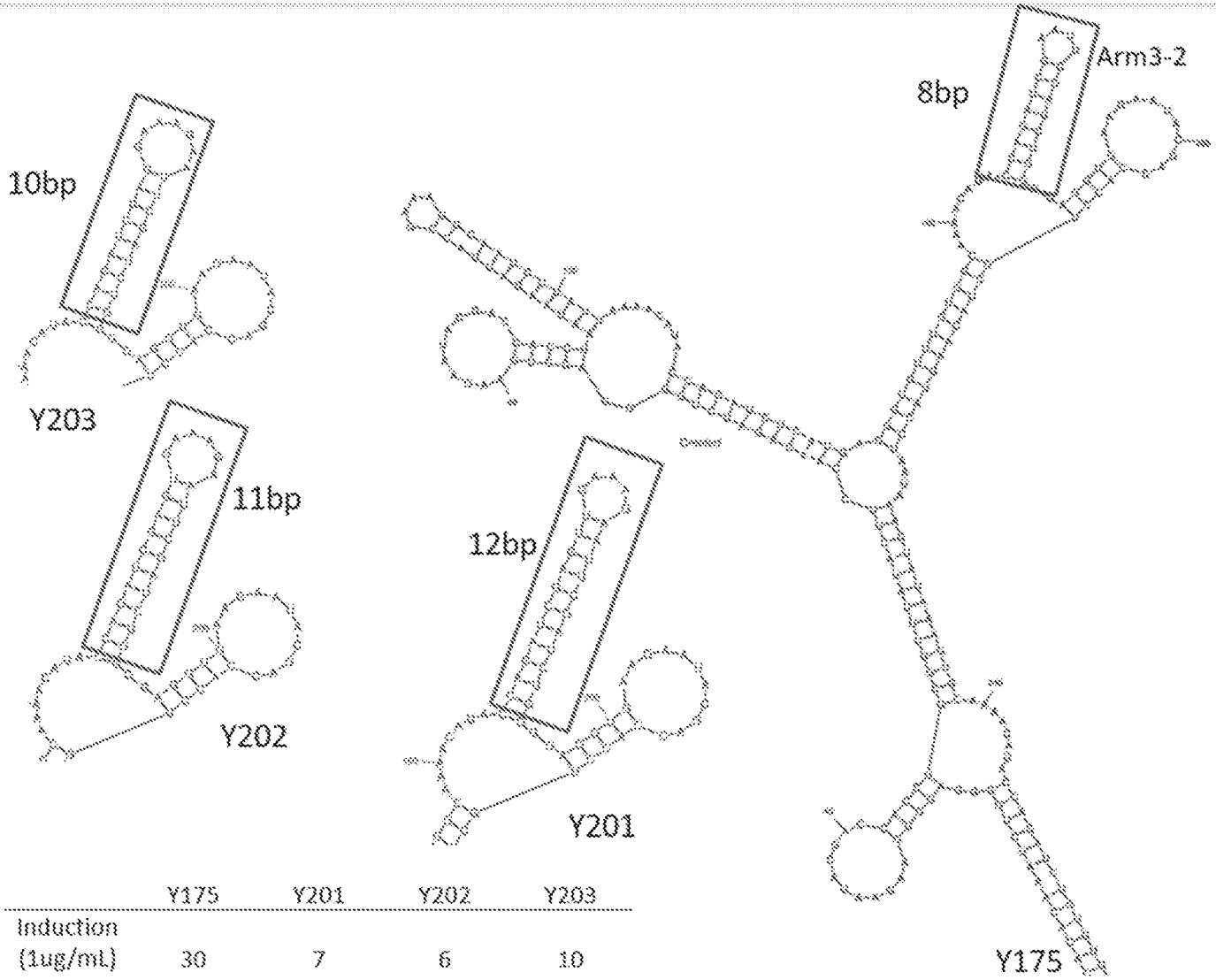

Y35 Y43 Y44 Y45
Induction
(Sug/mL) 51 12 31 49
Arm3-1
Delete one G-C pair
Y45
Delete one G
Arm3-1
Y44
Arm3-1
C to U
Y43
Y35

Figure 6A (continued)



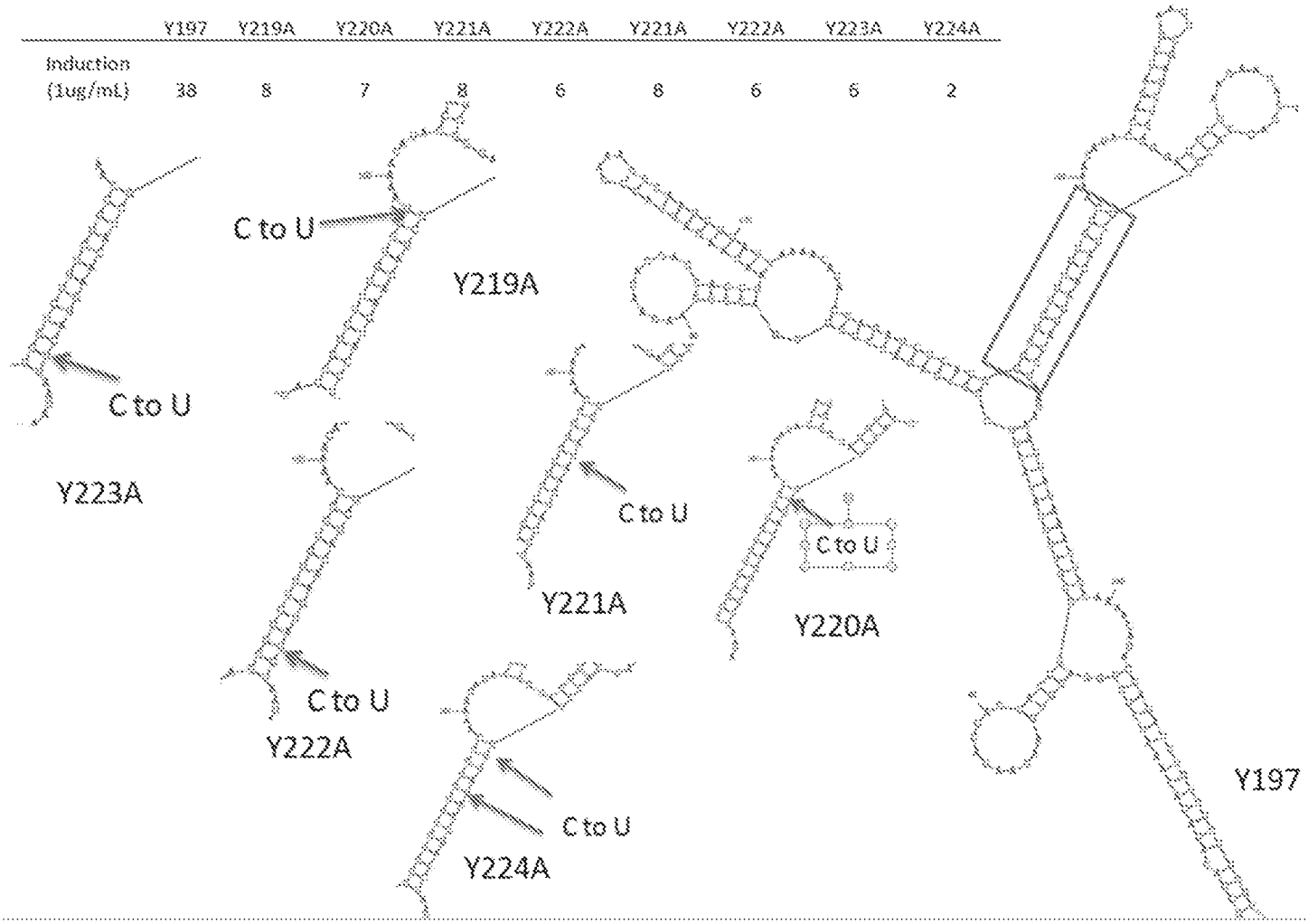
	Y175	Y188C	Y189C	Y176	Y188D	Y189D
Induction (1 ug/mL)	29	5	3	36	9	2

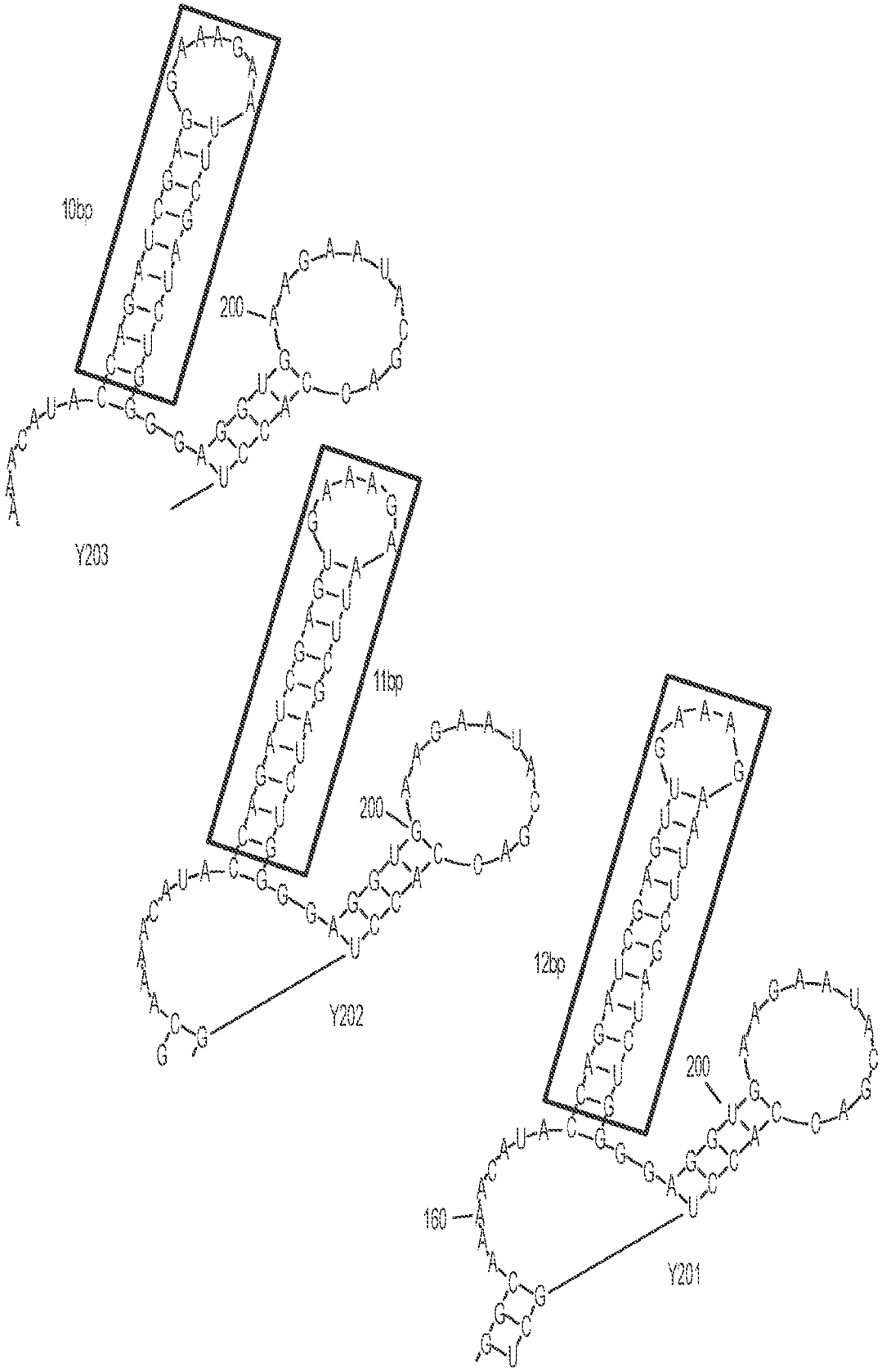
Figure 6A (continued)



	Y175	Y201	Y202	Y203
Induction (1ug/mL)	30	7	6	10

Figure 6A (continued)





	Y175	Y201	Y202	Y203
Induction (1ug/mL)	30	7	6	10

Figure 6B

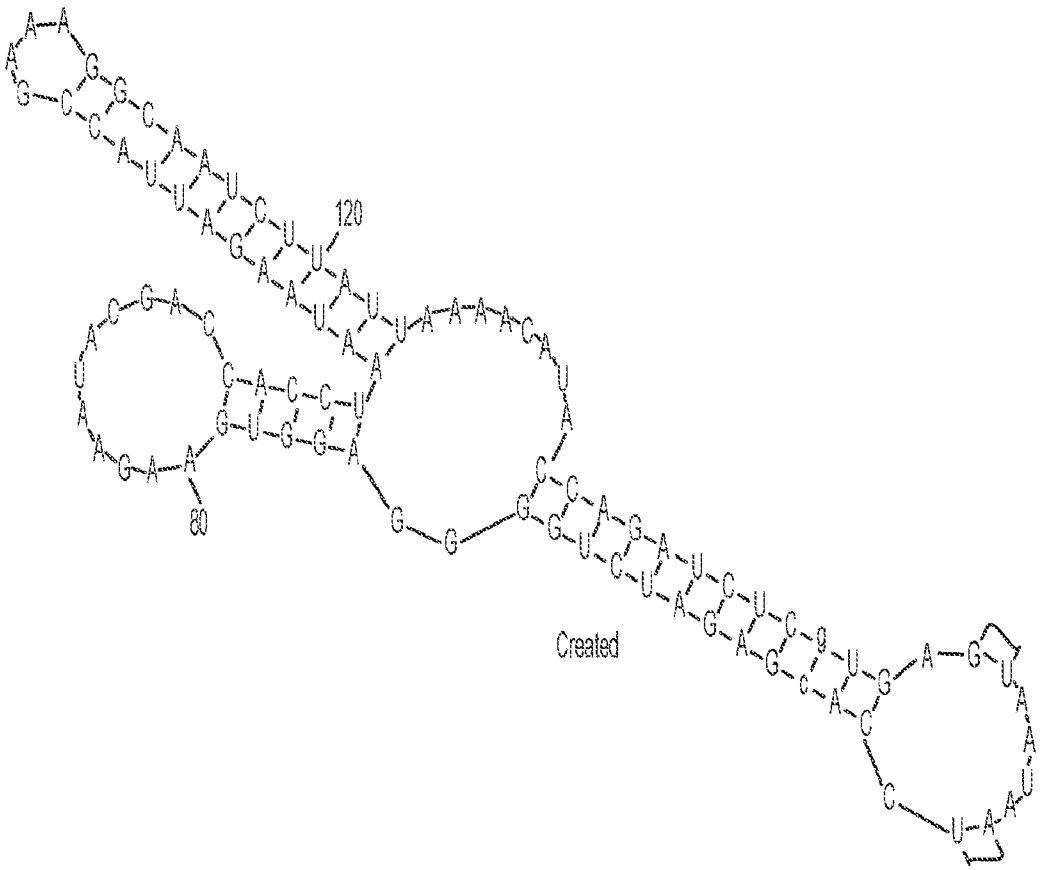


Figure 6B
CONTINUED

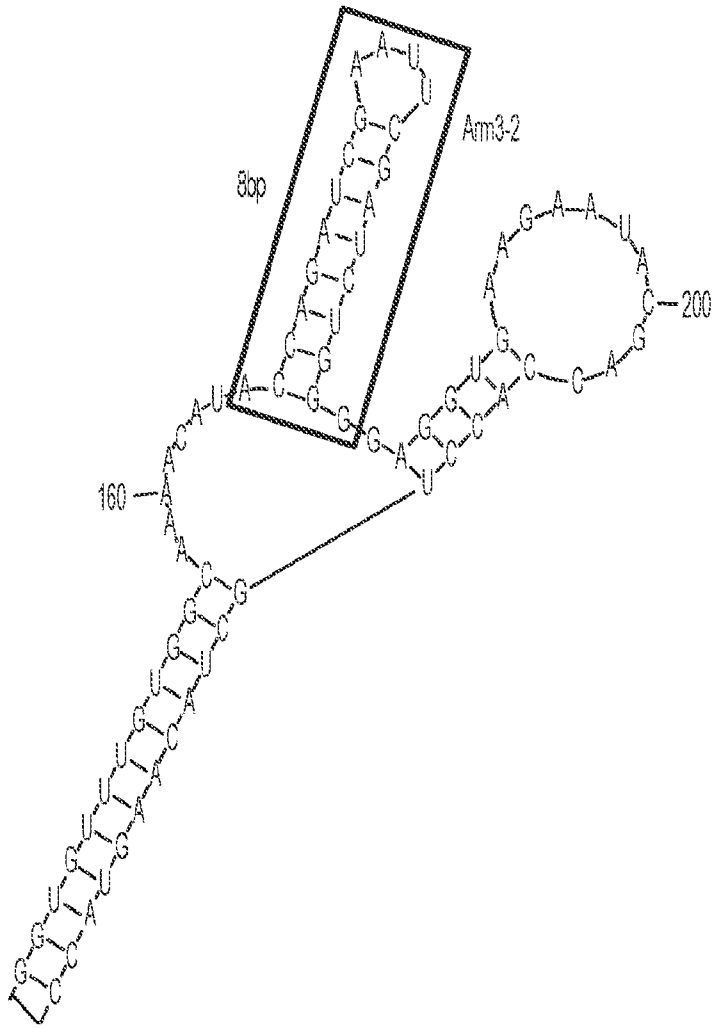


Figure 6B
CONTINUED

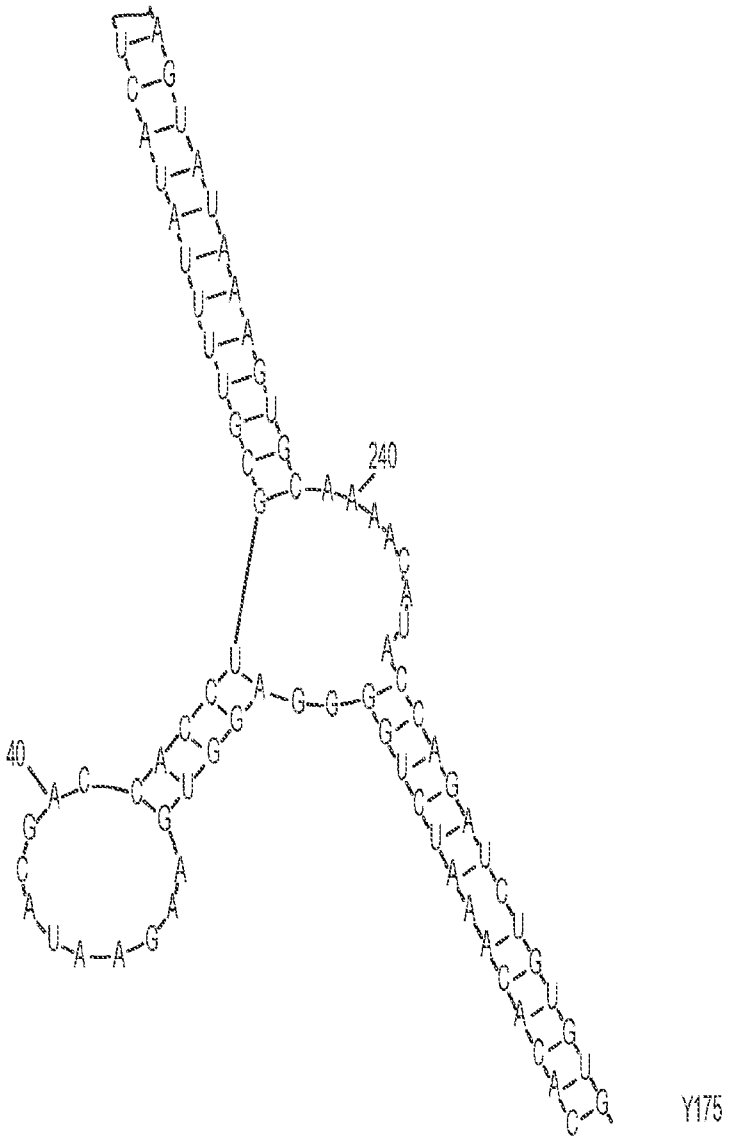
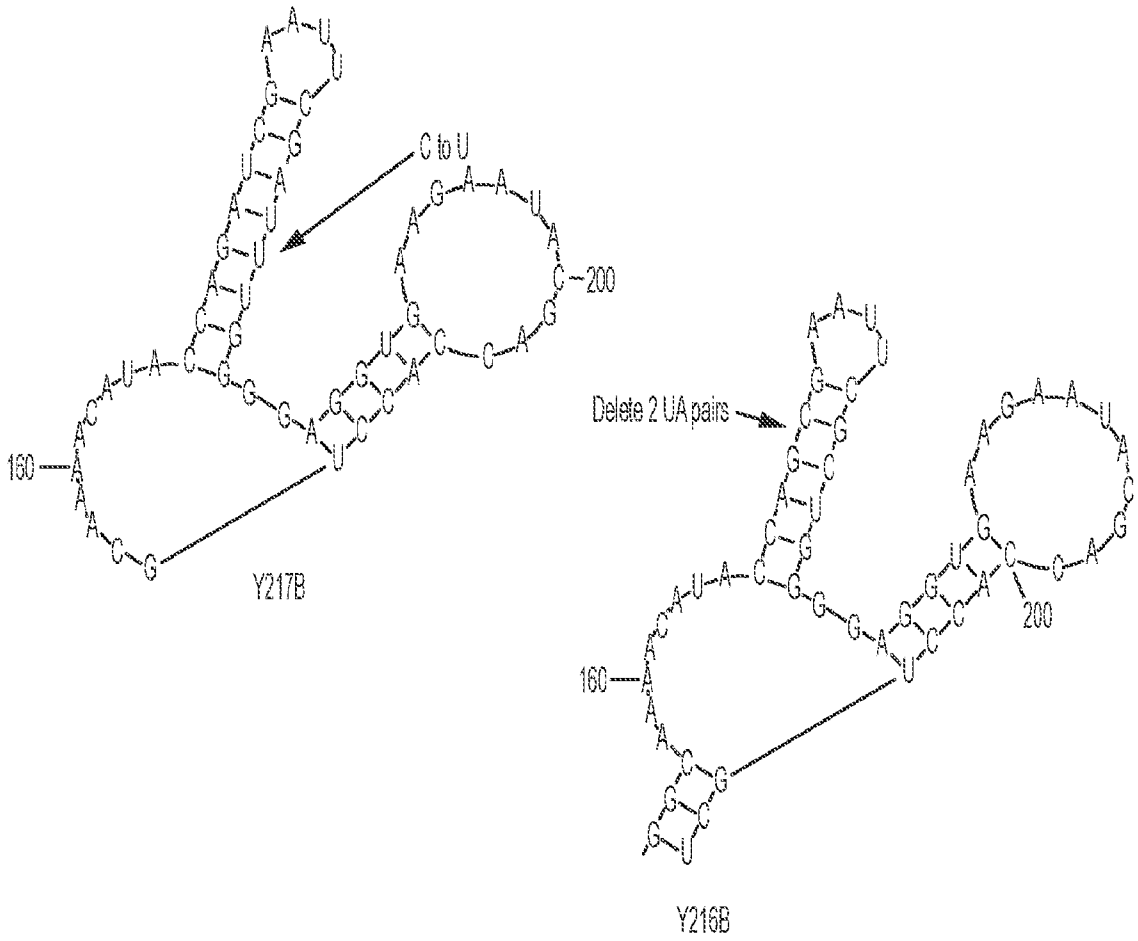


Figure 6B
CONTINUED



	Y208	Y216B	Y217B
Introduction (1ug/mL)	51	13	18

Figure 6B
CONTINUED

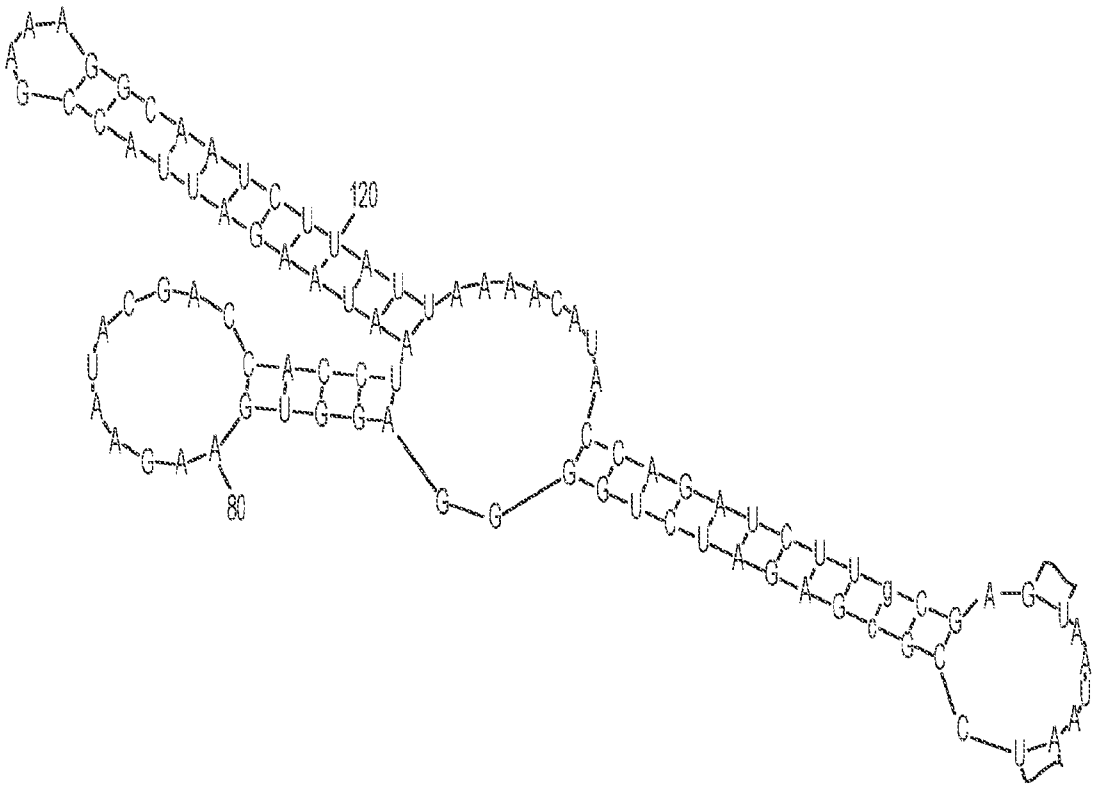


Figure 6B
CONTINUED

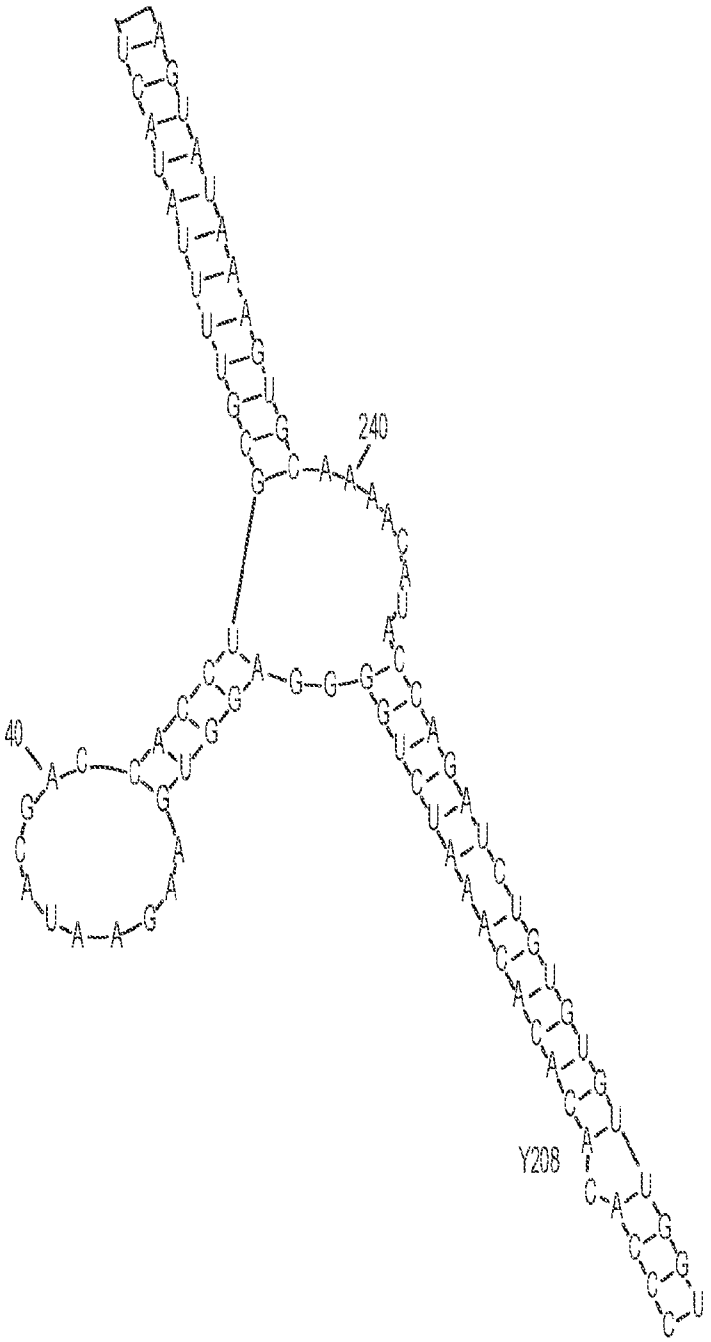


Figure 6B
CONTINUED

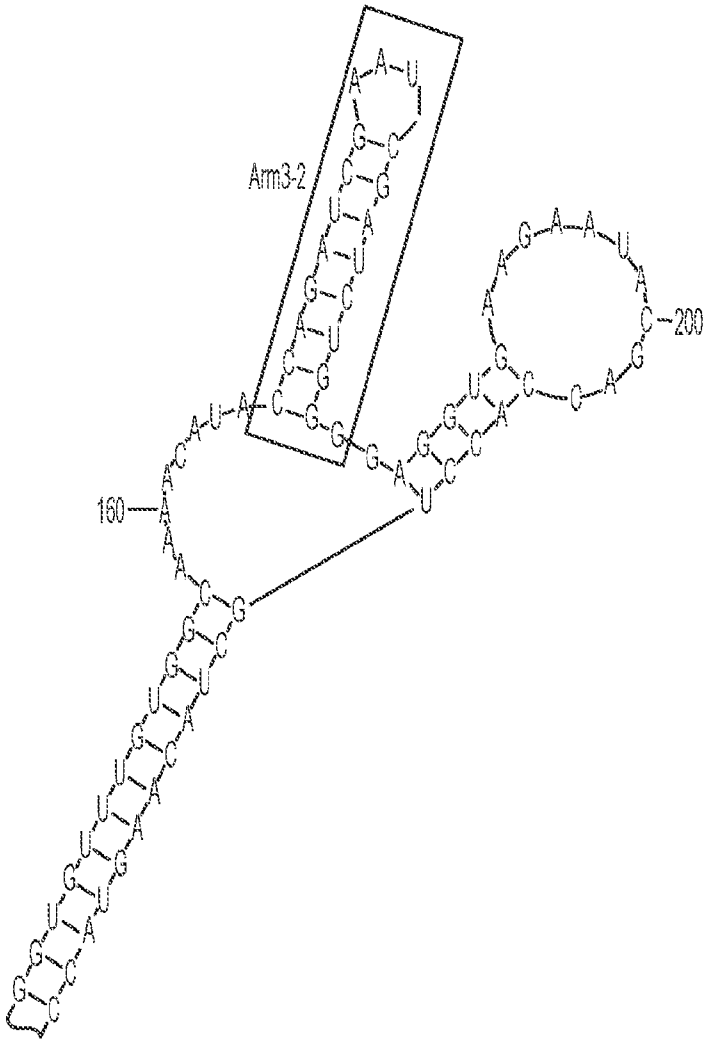
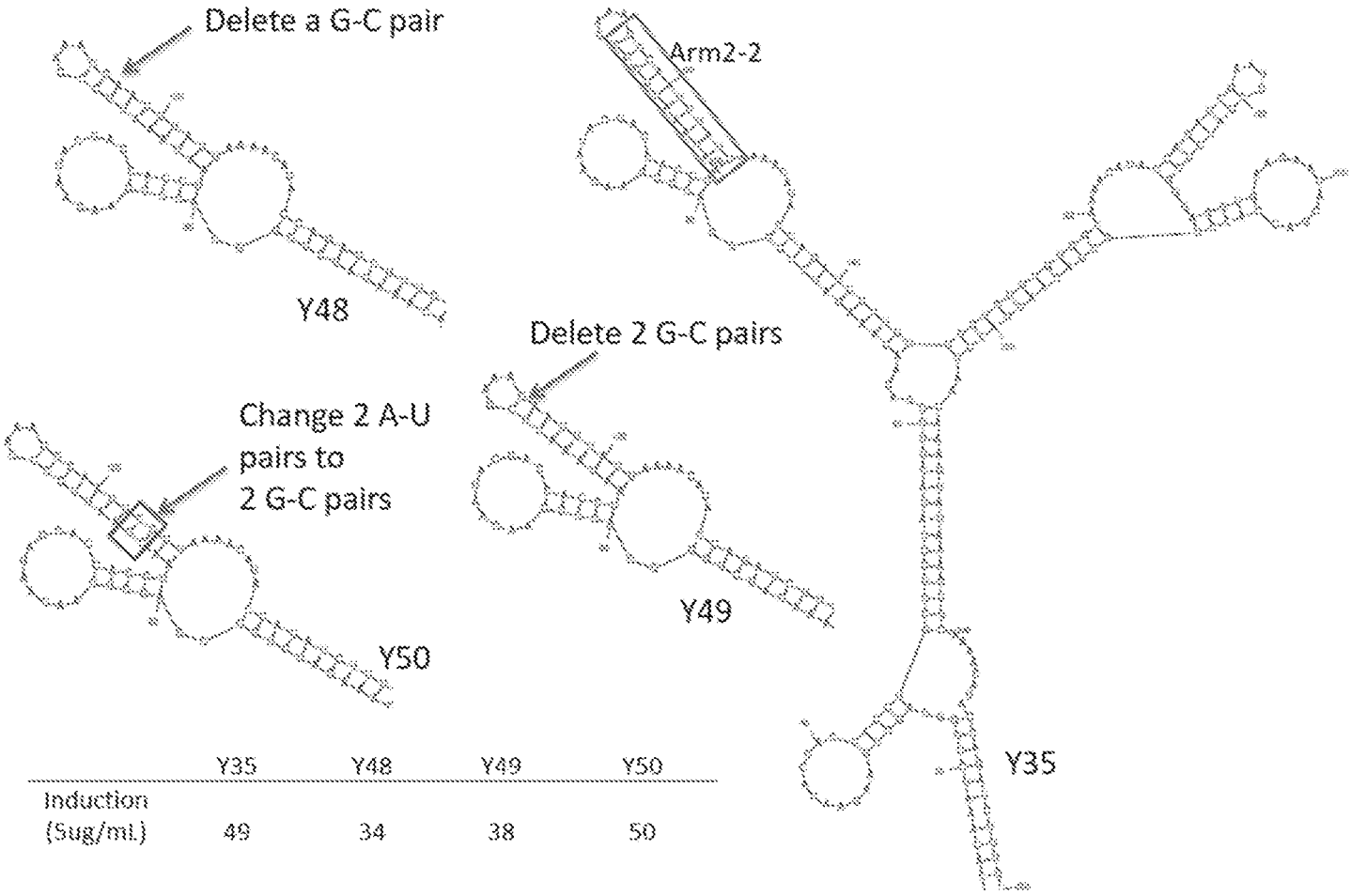


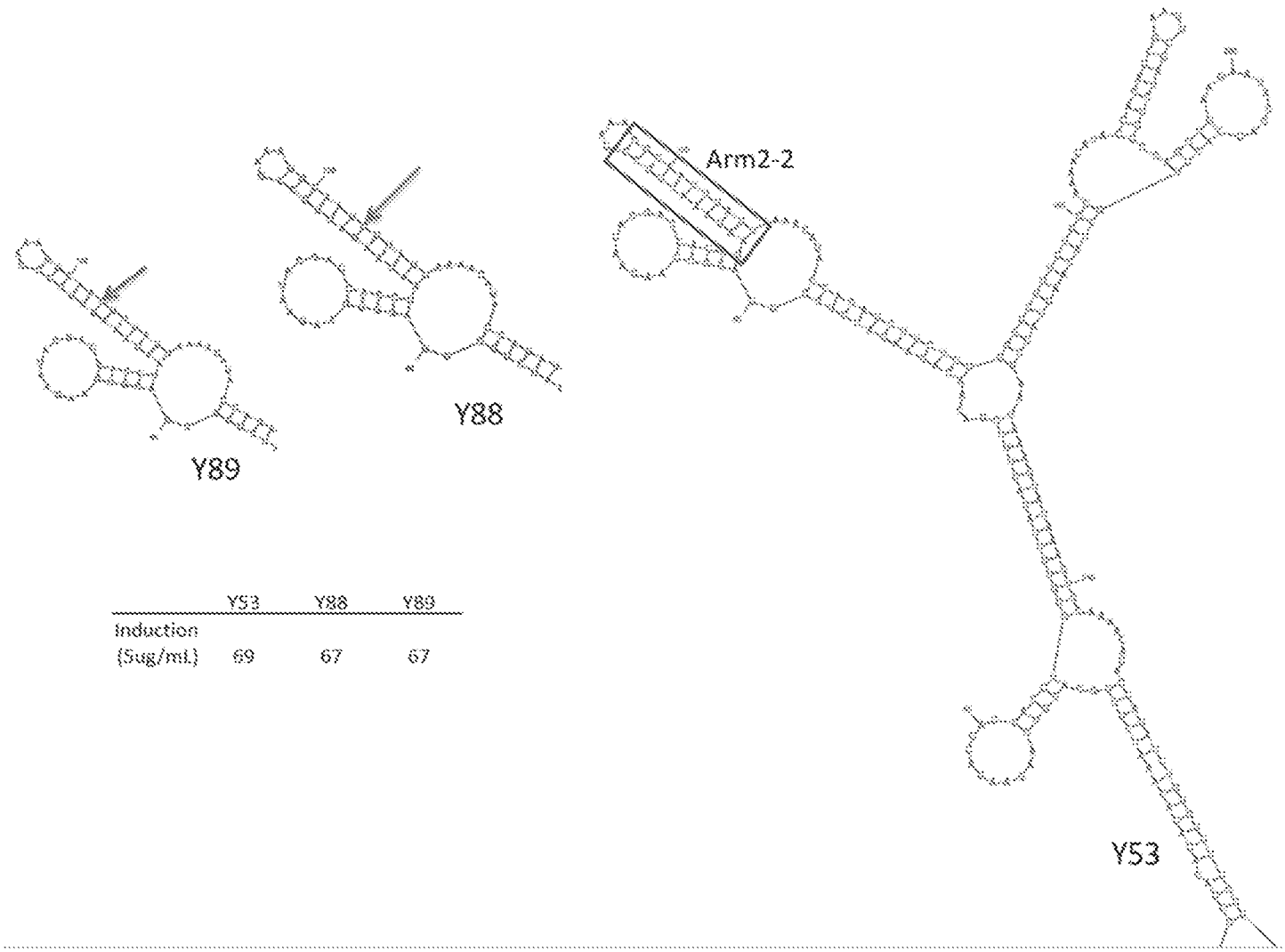
Figure 6B
CONTINUED

Figure 7A



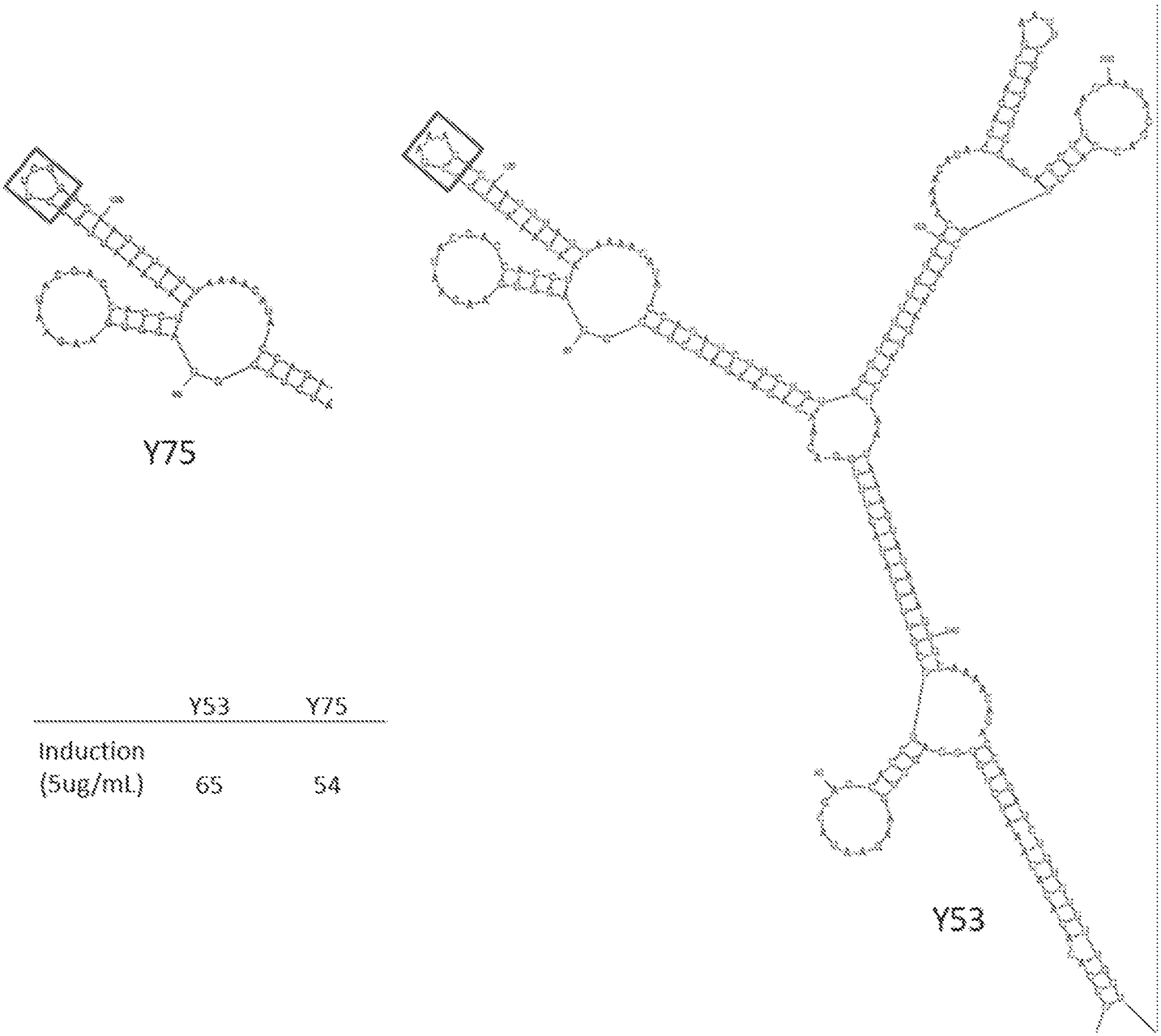
	Y35	Y48	Y49	Y50
Induction (Sug/mL)	49	34	38	50

Figure 7A (continued)



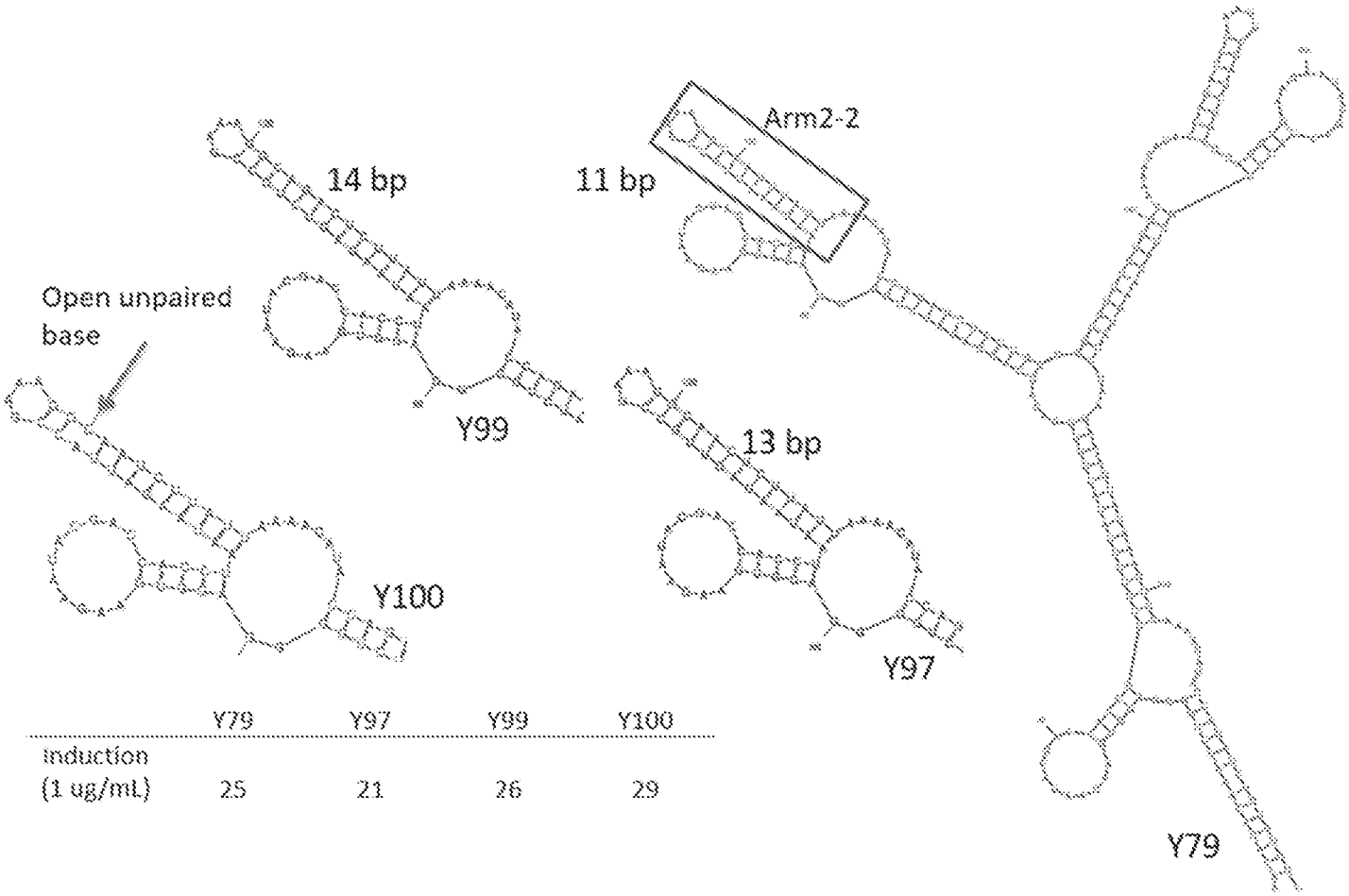
	Y53	Y88	Y89
Induction (5ug/mL)	69	67	67

Figure 7A (continued)



	Y53	Y75
Induction (5ug/mL)	65	54

Figure 7A (continued)



	Y79	Y97	Y99	Y100
induction (1 ug/mL)	25	21	26	29

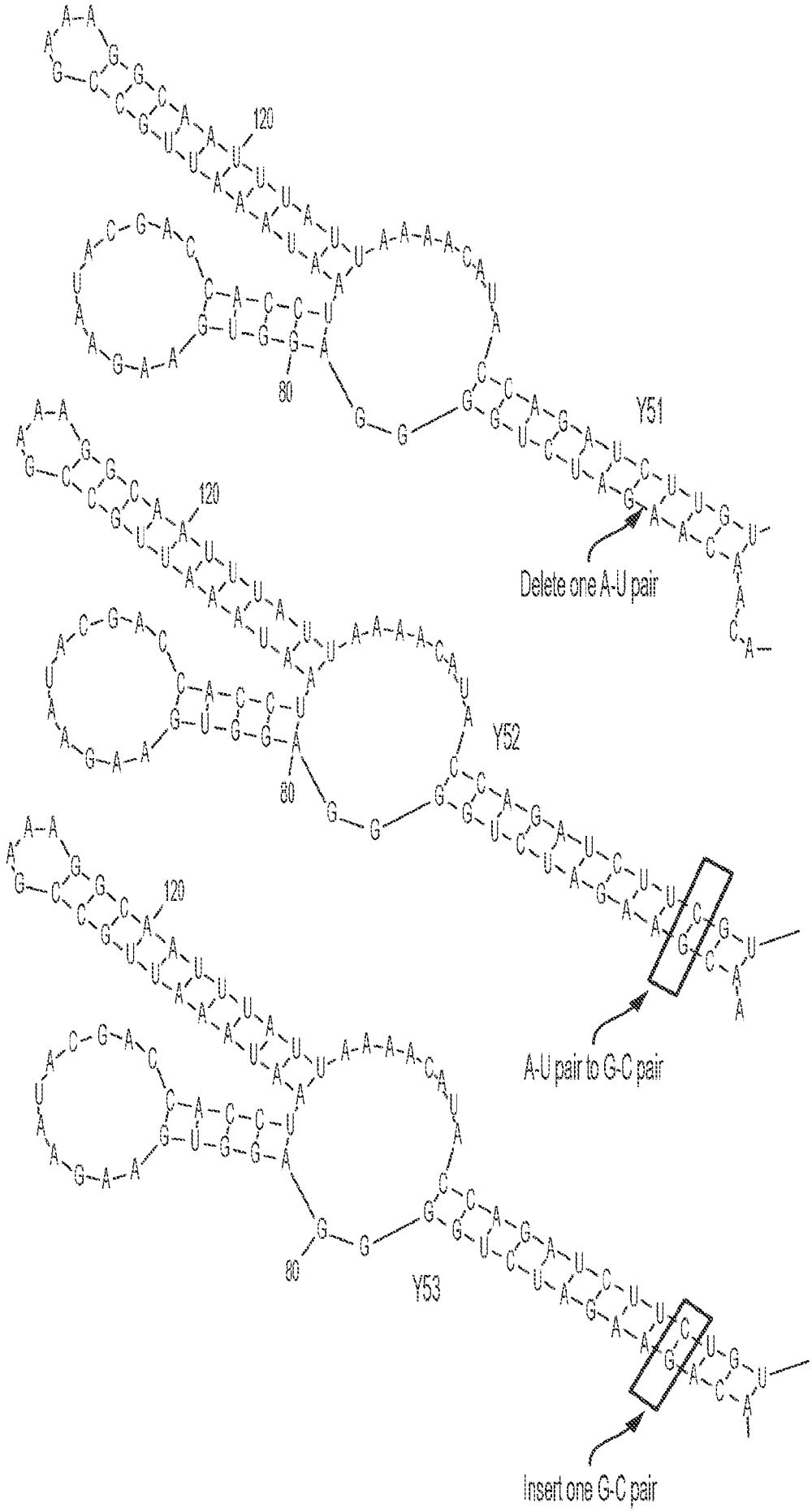

120
80
Y51
Delete one A-U pair
Y52
A-U pair to G-C pair
Y53
Insert one G-C pair

Figure 7B

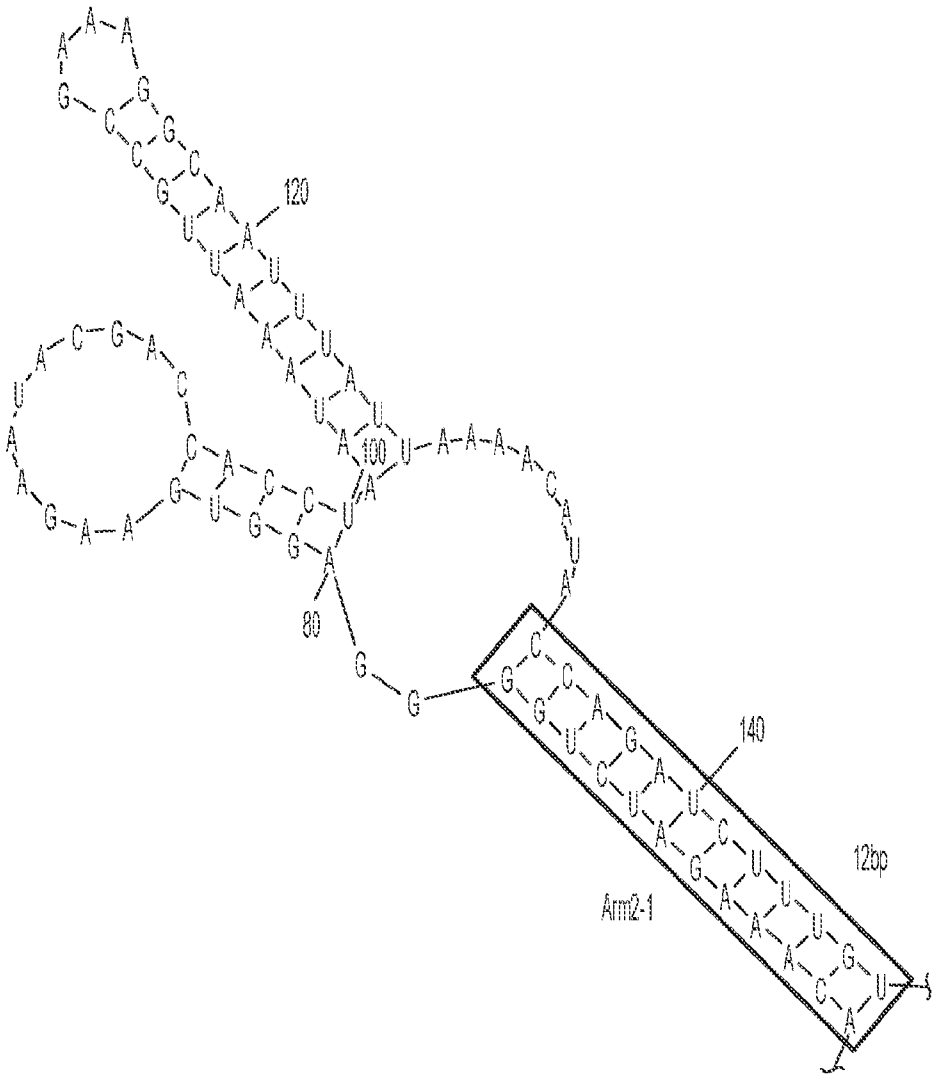


Figure 7B
CONTINUED

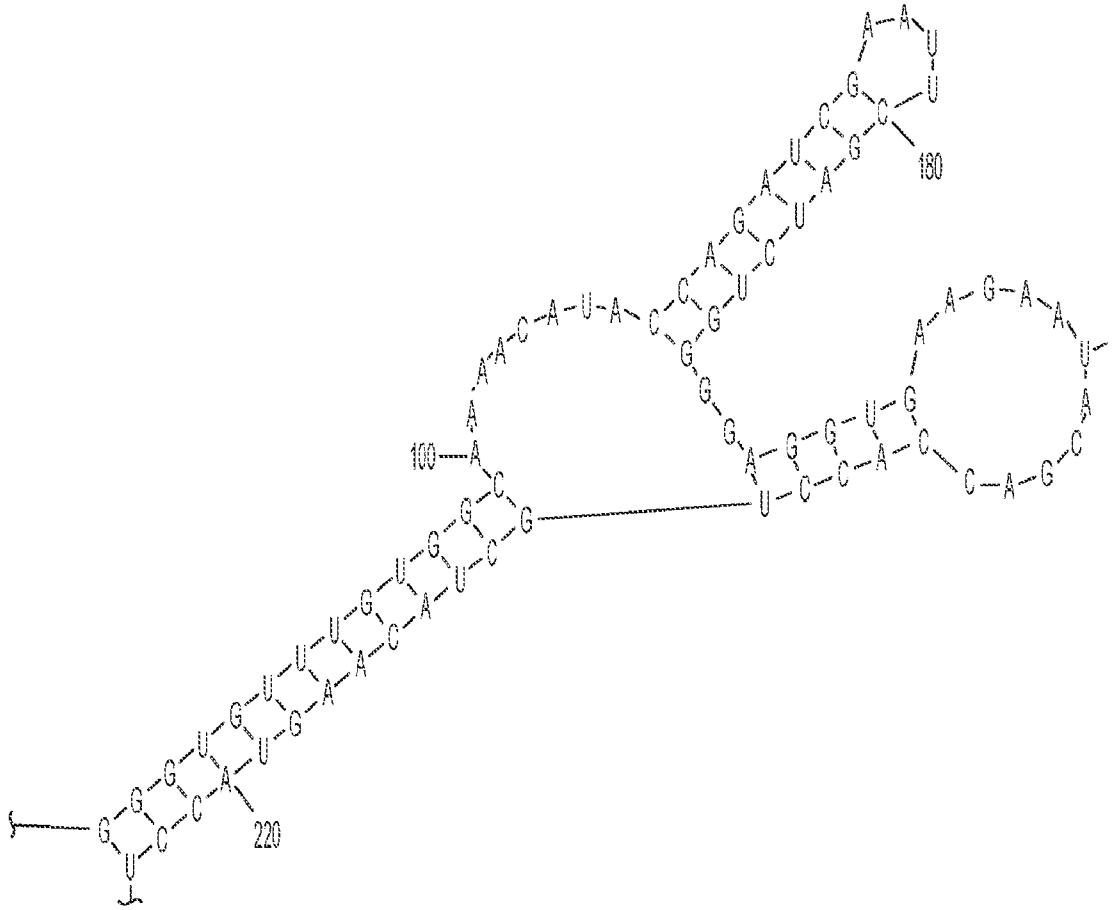
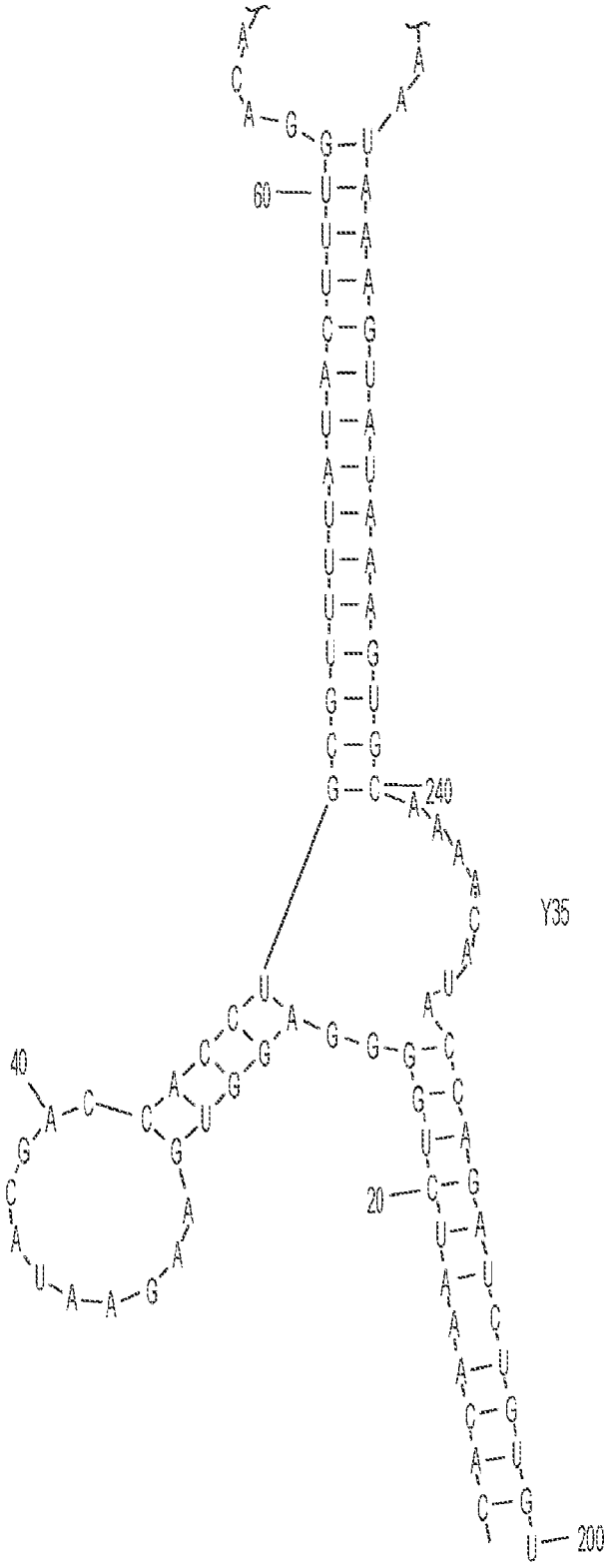


Figure 7B
CONTINUED



	Y35	Y51	Y52	Y53
Induction (5ug/mL)	49	45	64	51

Figure 7B
CONTINUED

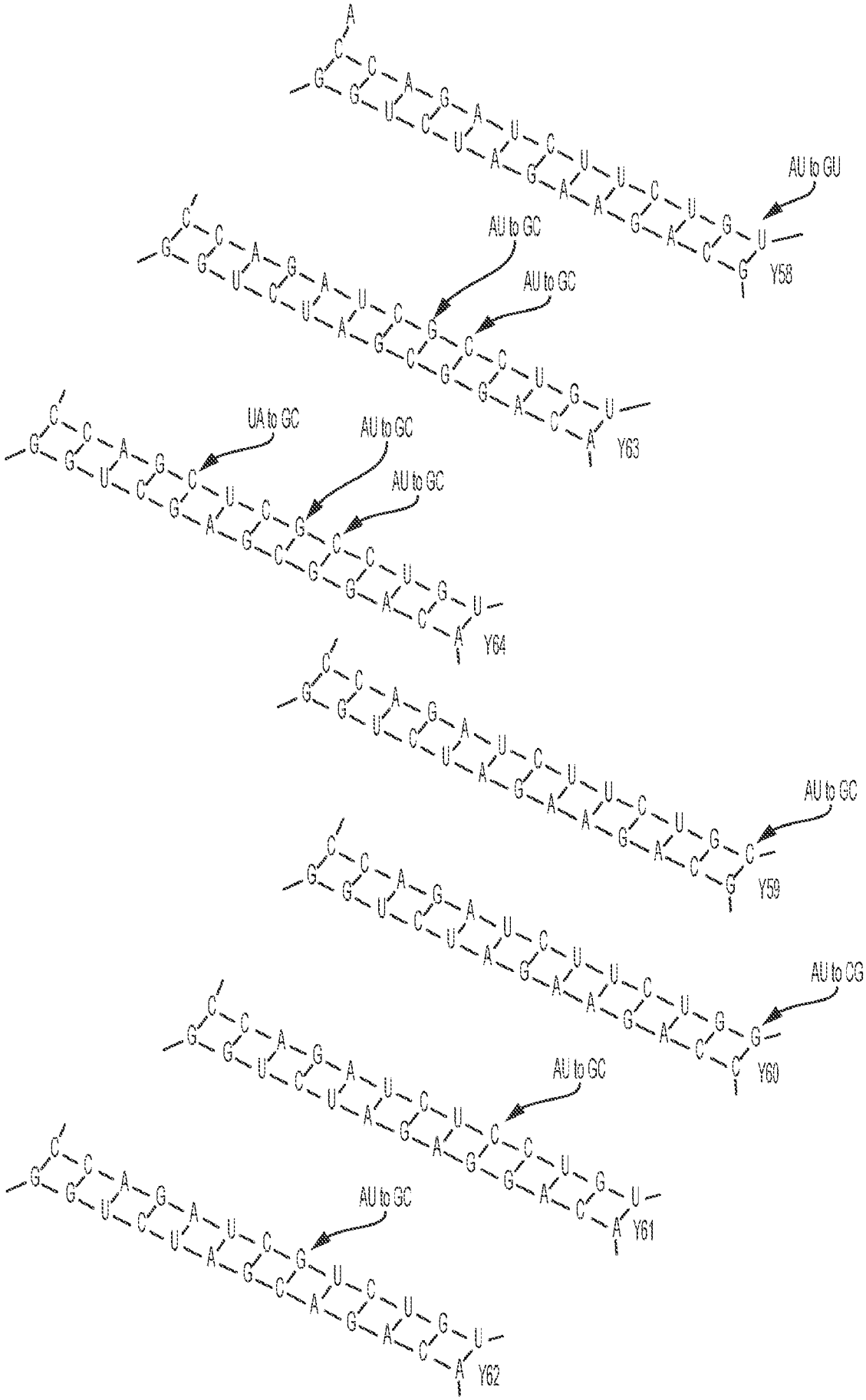


Figure 7B
CONTINUED

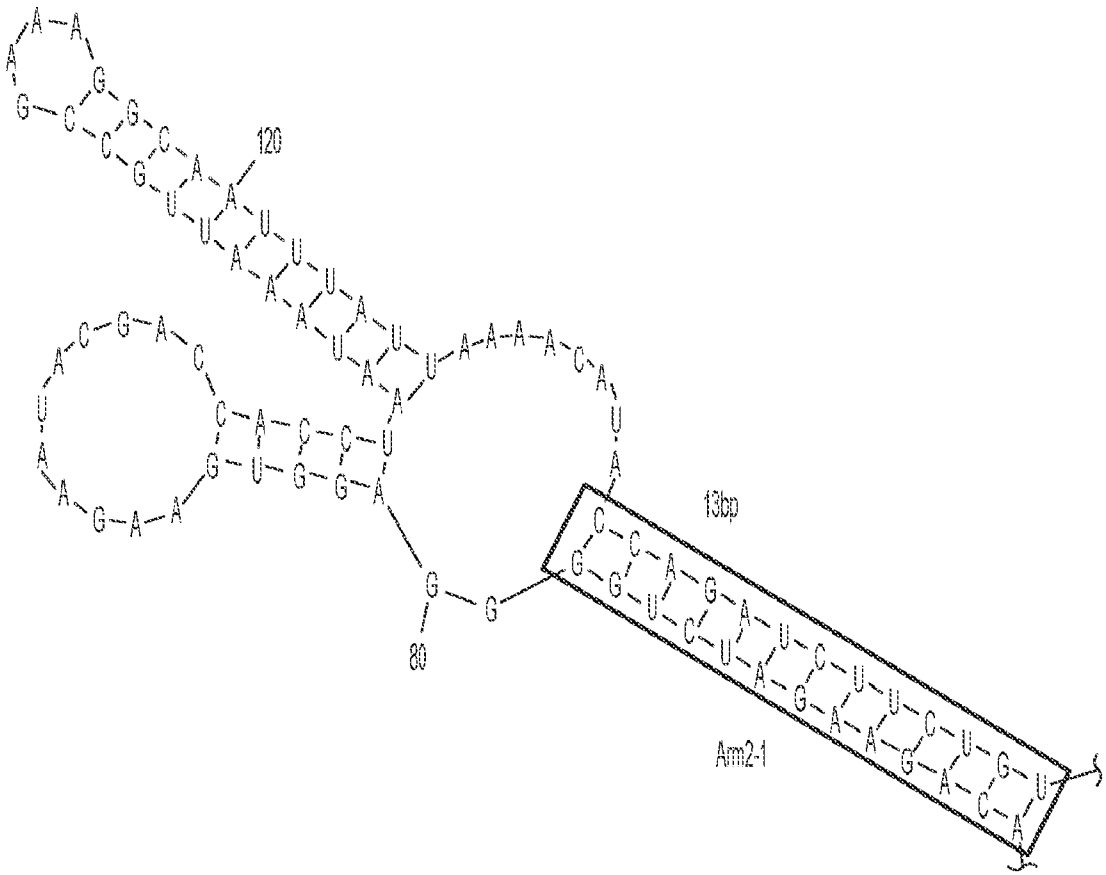


Figure 7B
CONTINUED

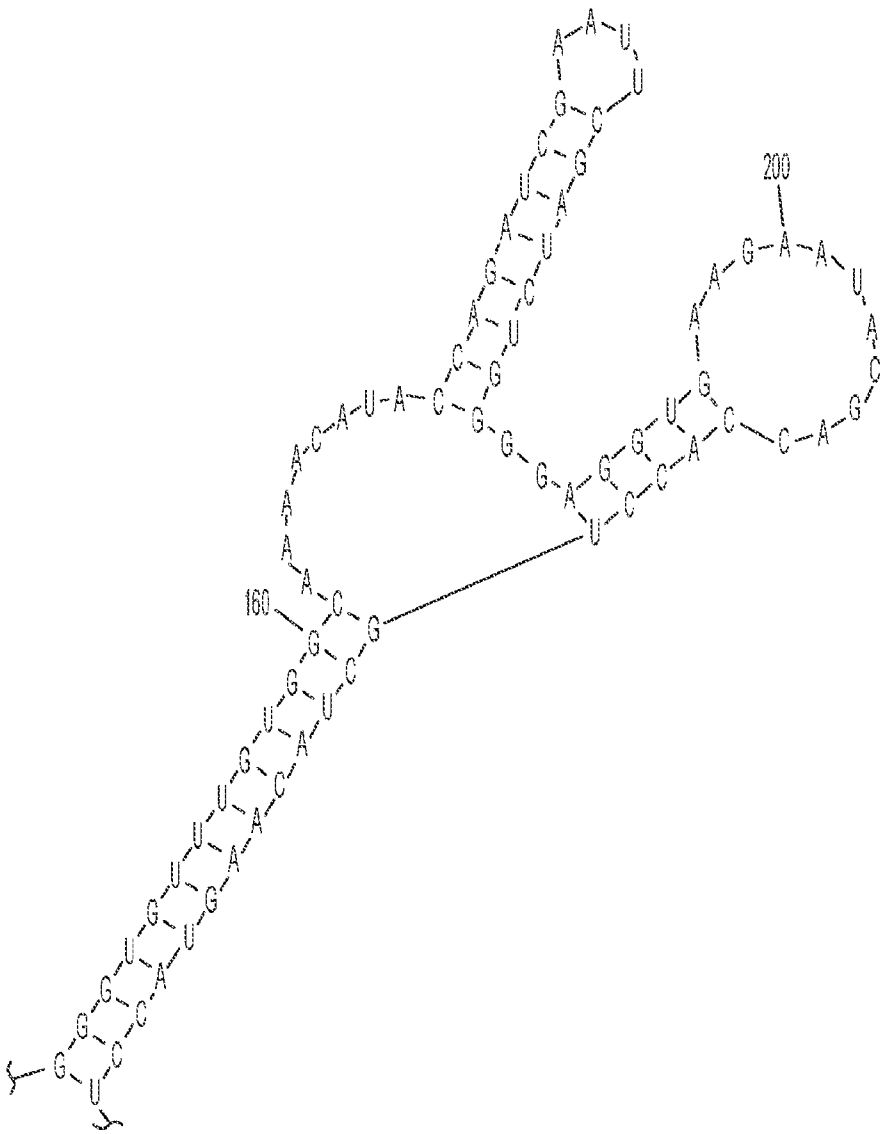
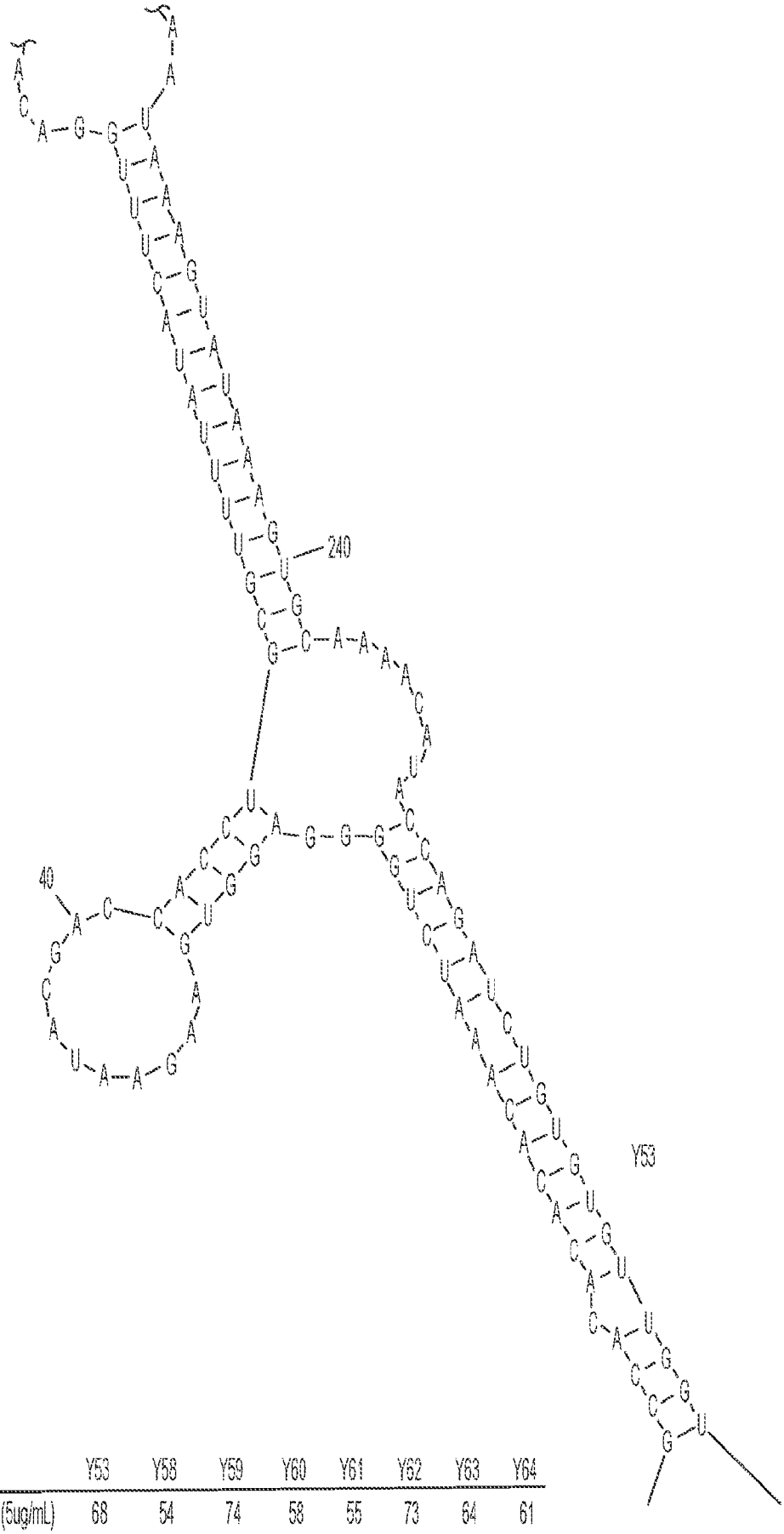
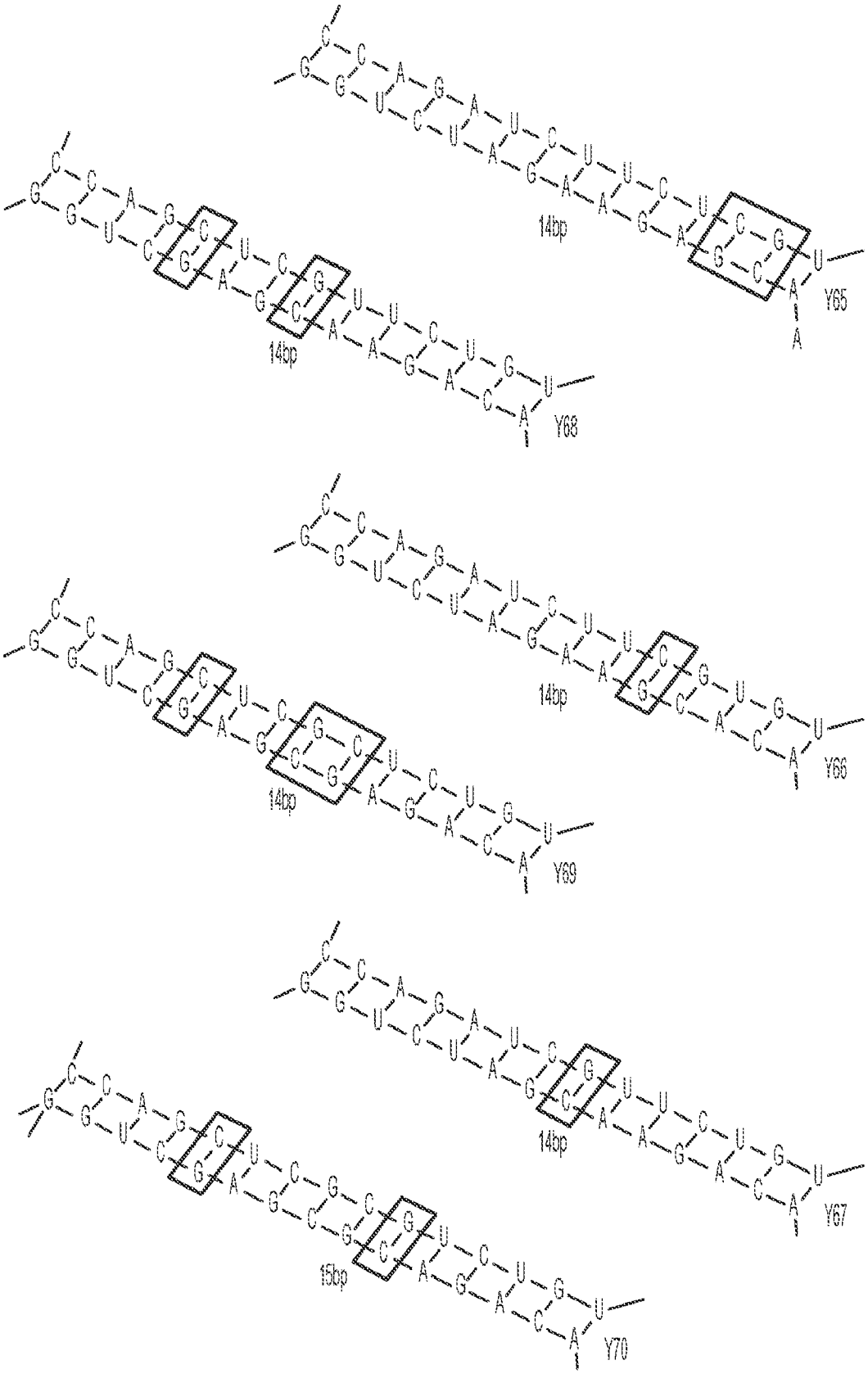


Figure 7B
CONTINUED



	Y53	Y58	Y59	Y60	Y61	Y62	Y63	Y64
Induction (5ug/mL)	68	54	74	58	55	73	64	61

Figure 7B
CONTINUED



	Y53	Y65	Y66	Y67	Y68	Y69	Y70
Induction (5ug/mL)	66	76	75	24	43	54	32

Figure 7B
CONTINUED

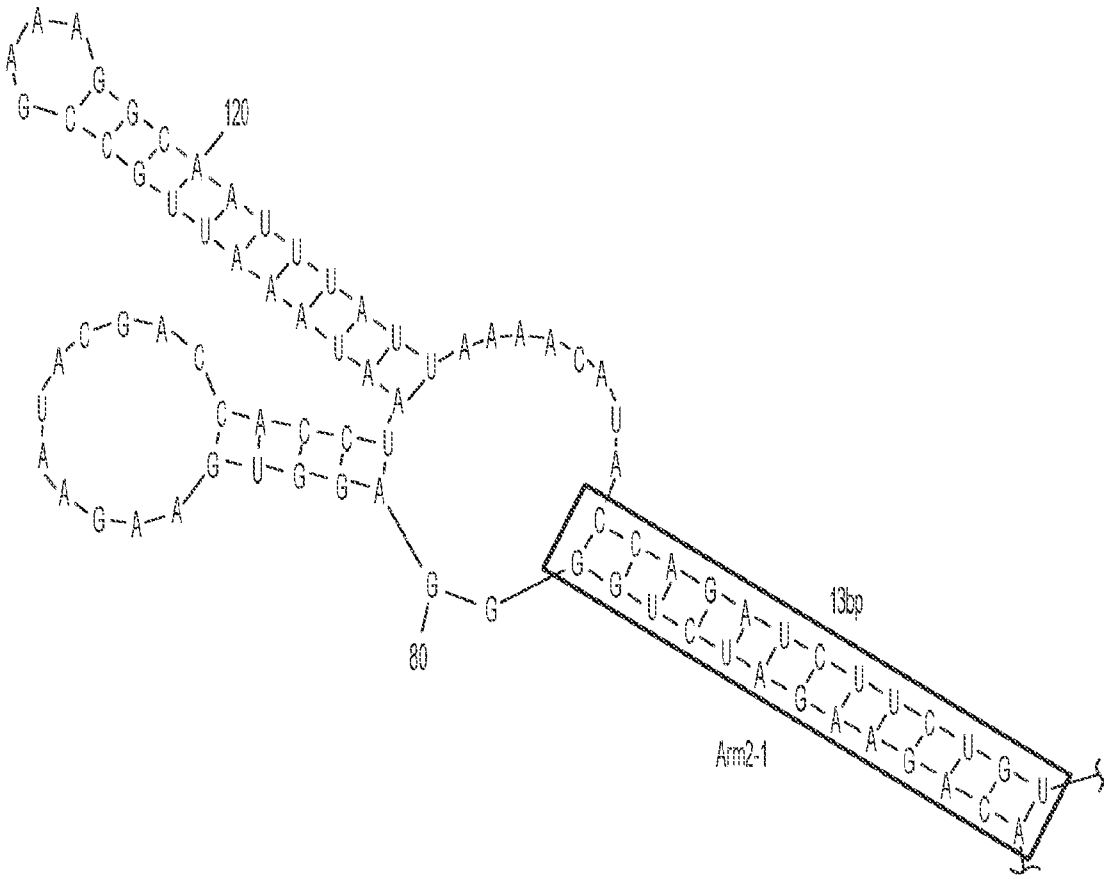


Figure 7B
CONTINUED

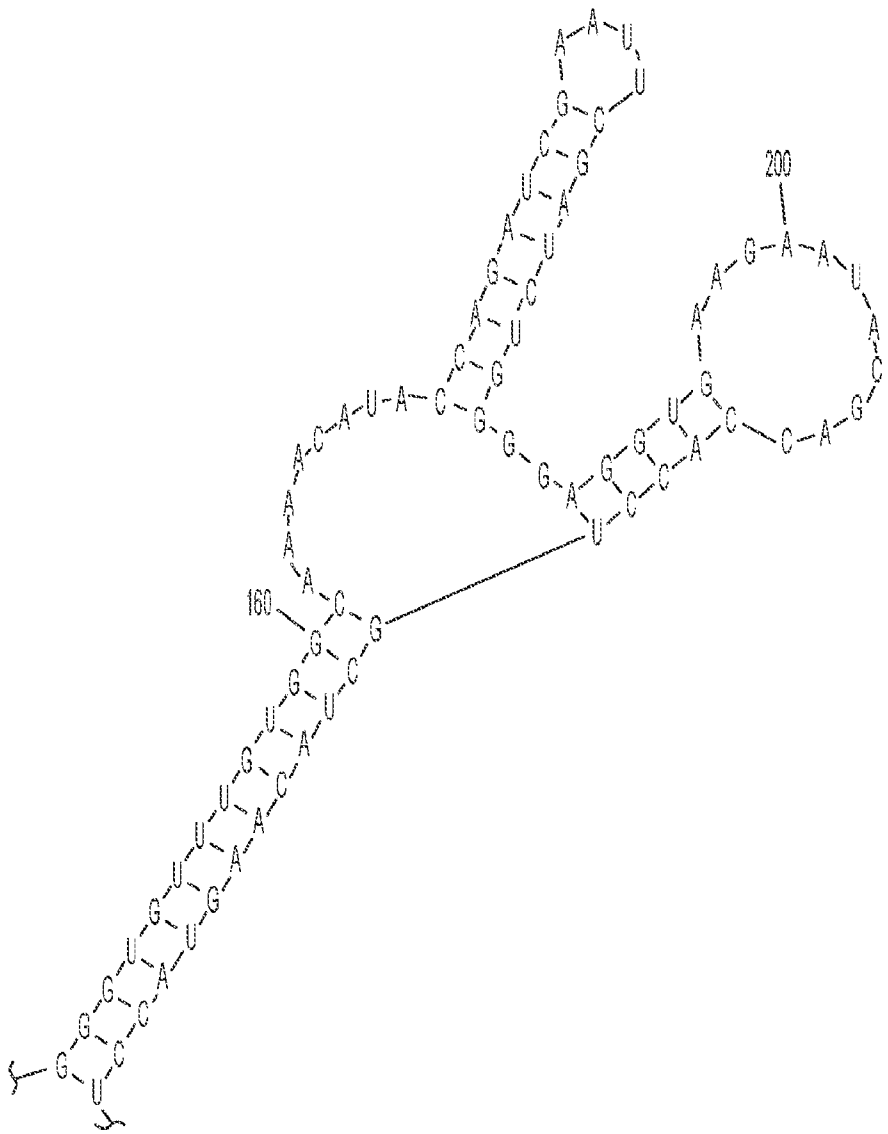


Figure 7B
CONTINUED

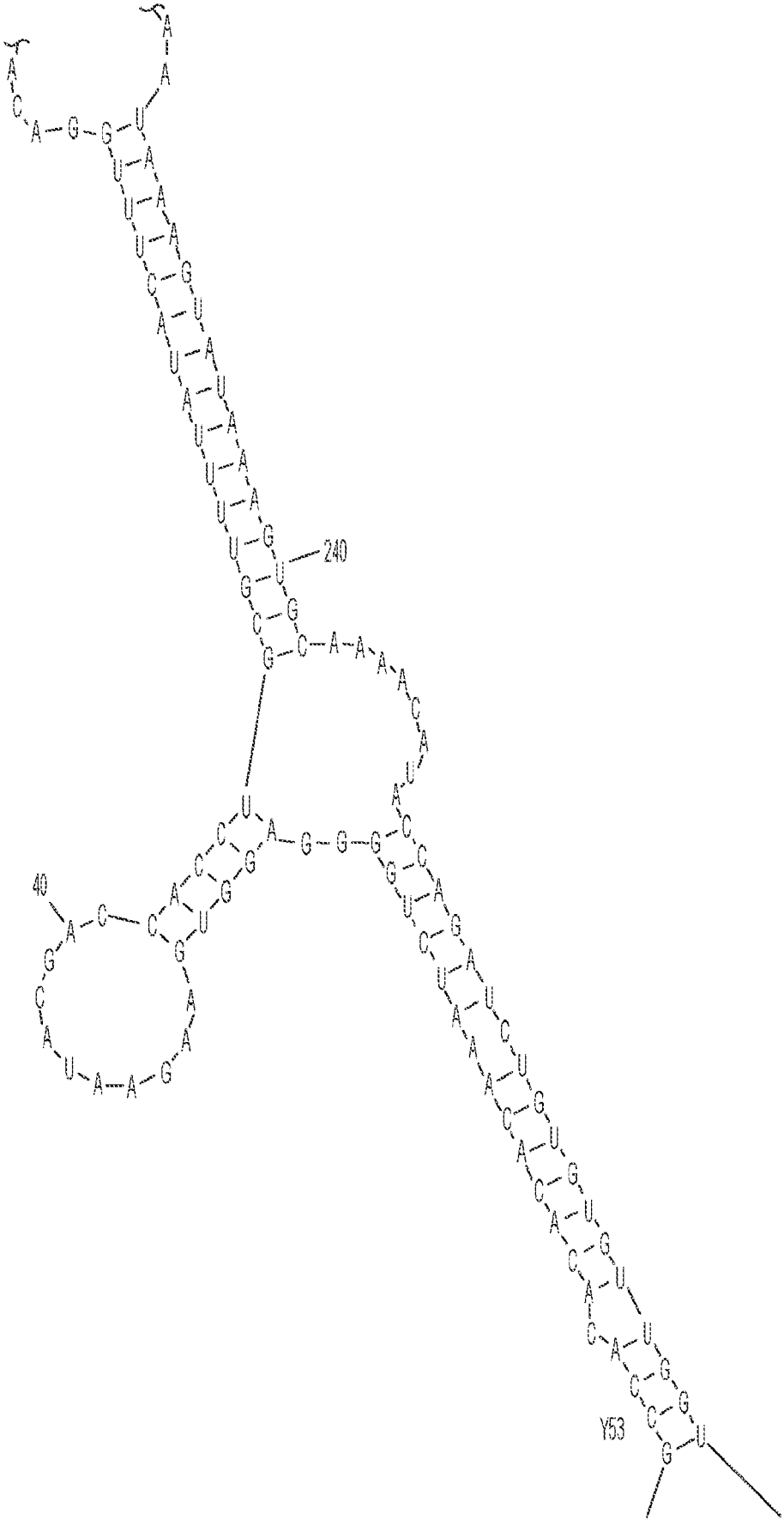


Figure 7B
CONTINUED

	Y175	Y190	Y191
Induction (1ug/mL)	32	14	25

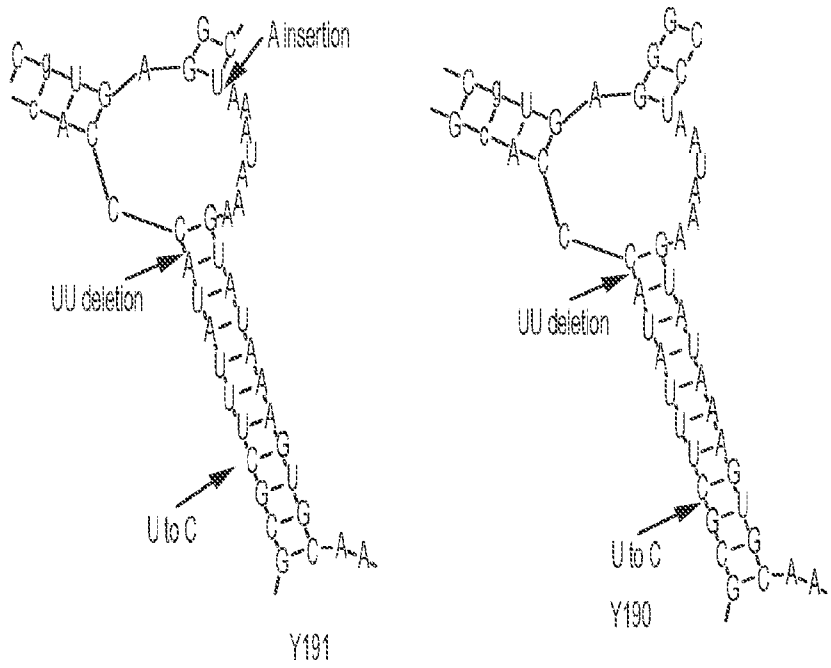


Figure 8

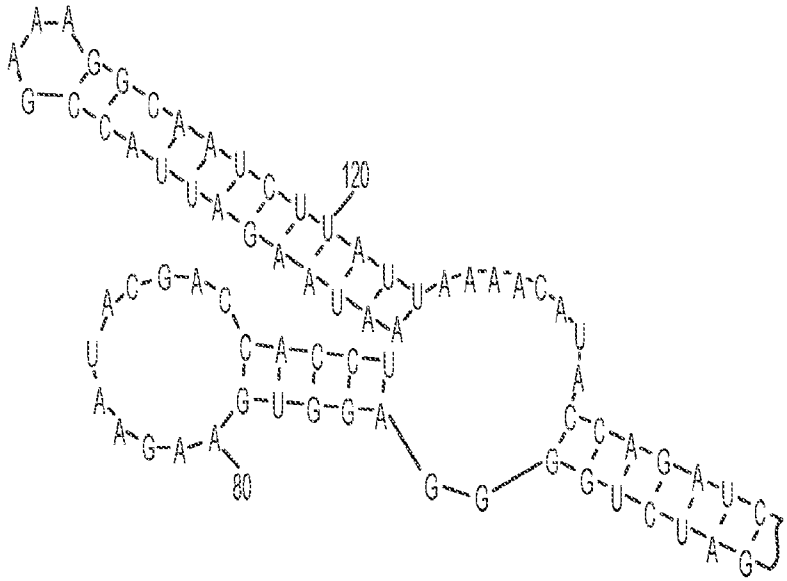


Figure 8
CONTINUED

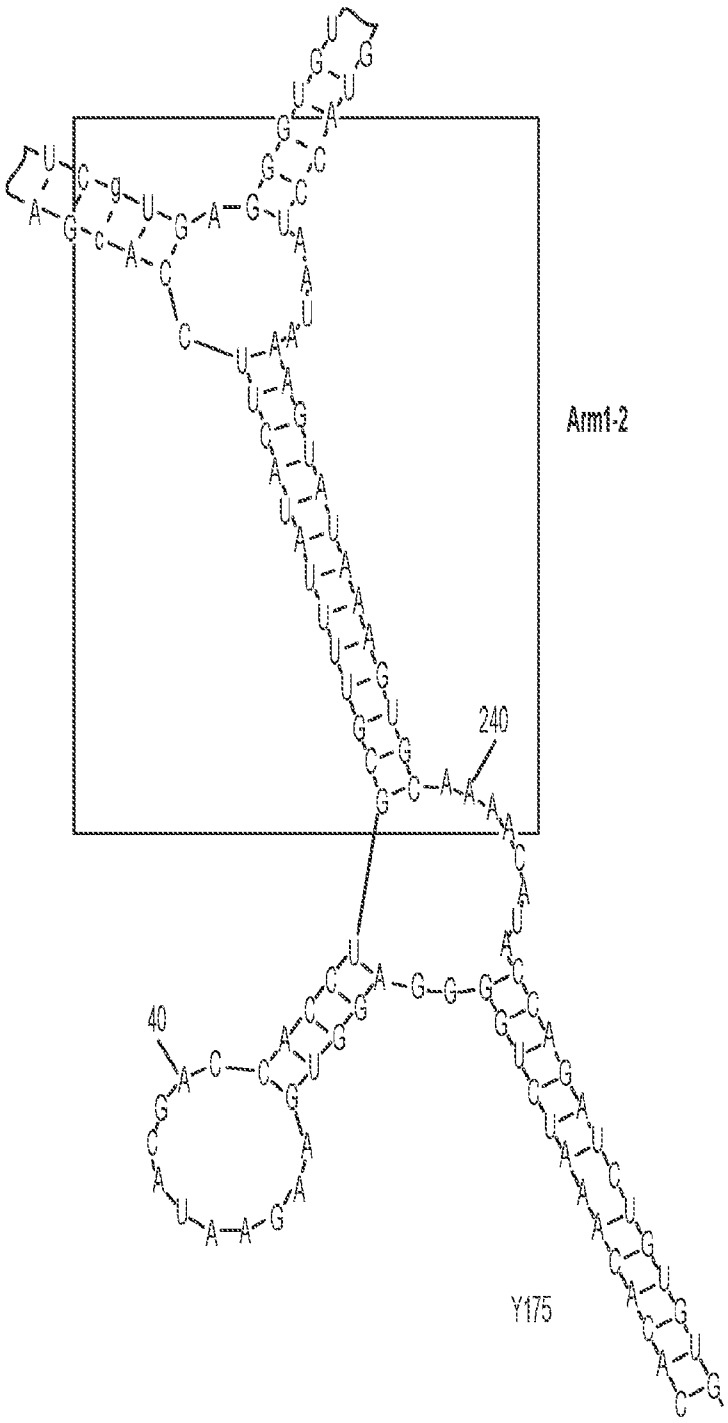


Figure 8
CONTINUED

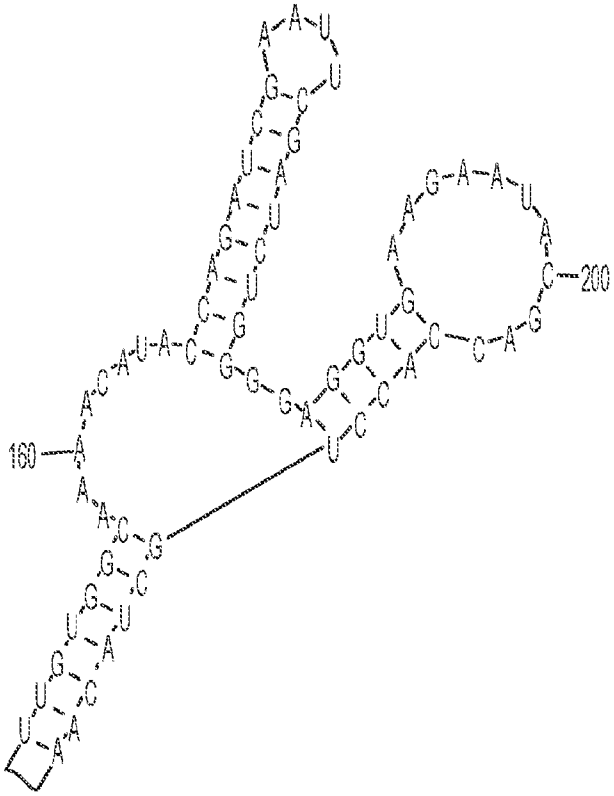


Figure 8
CONTINUED

	Y196CAA	Y234	Y235	Y236
1ug/mL	55	36	15	16

	Y196CAA	Y234	Y235	Y236
5ug/mL	204	206	92	109

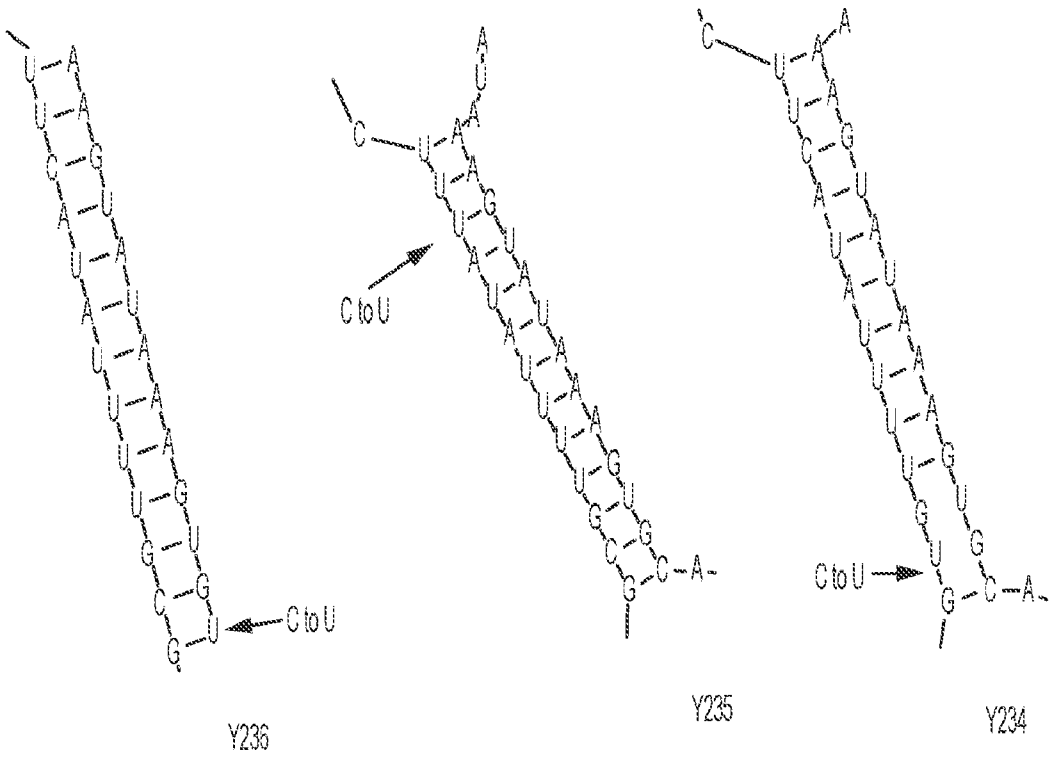


Figure 8
CONTINUED

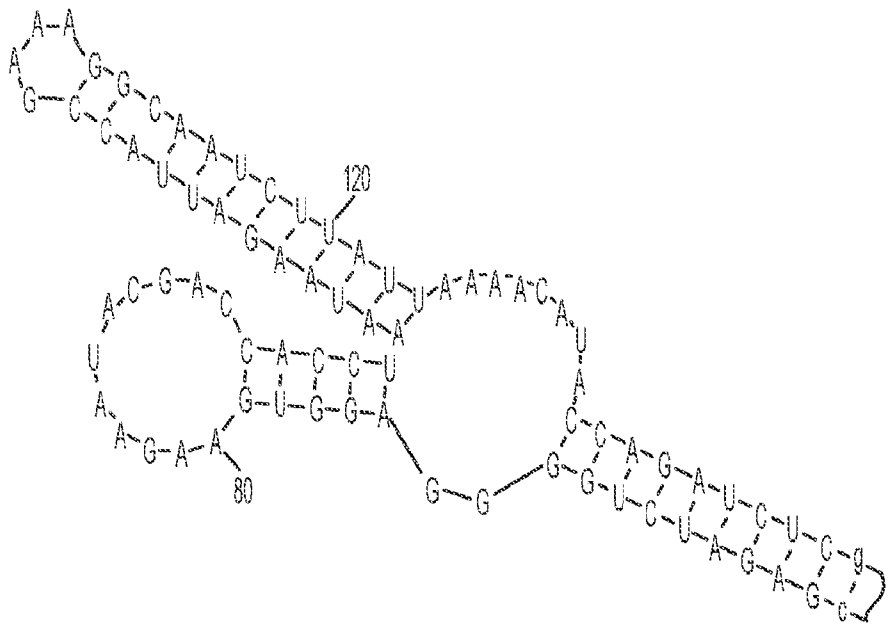


Figure 8
CONTINUED

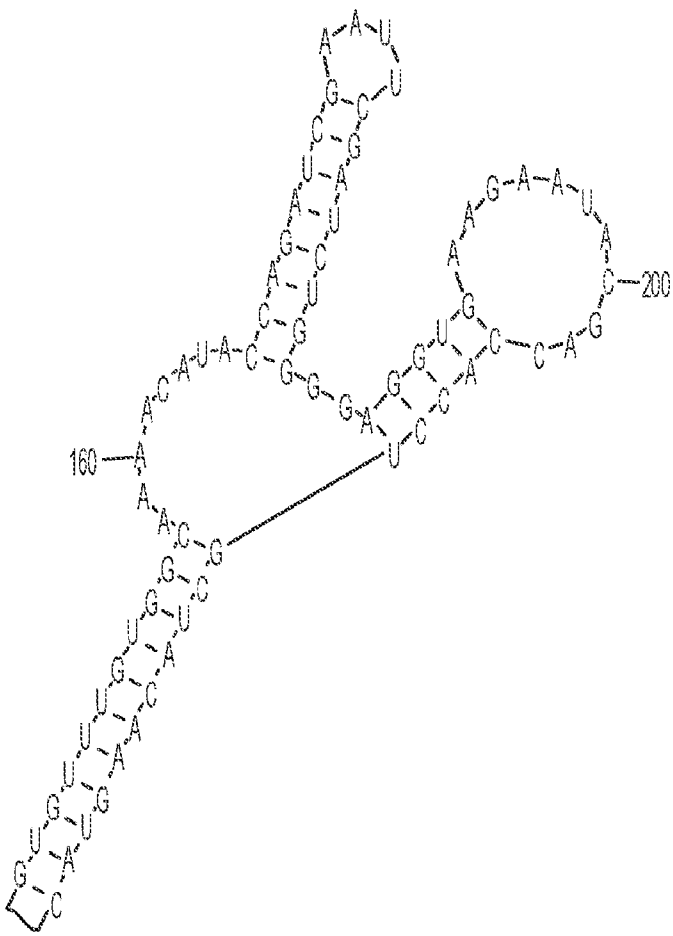


Figure 8
CONTINUED

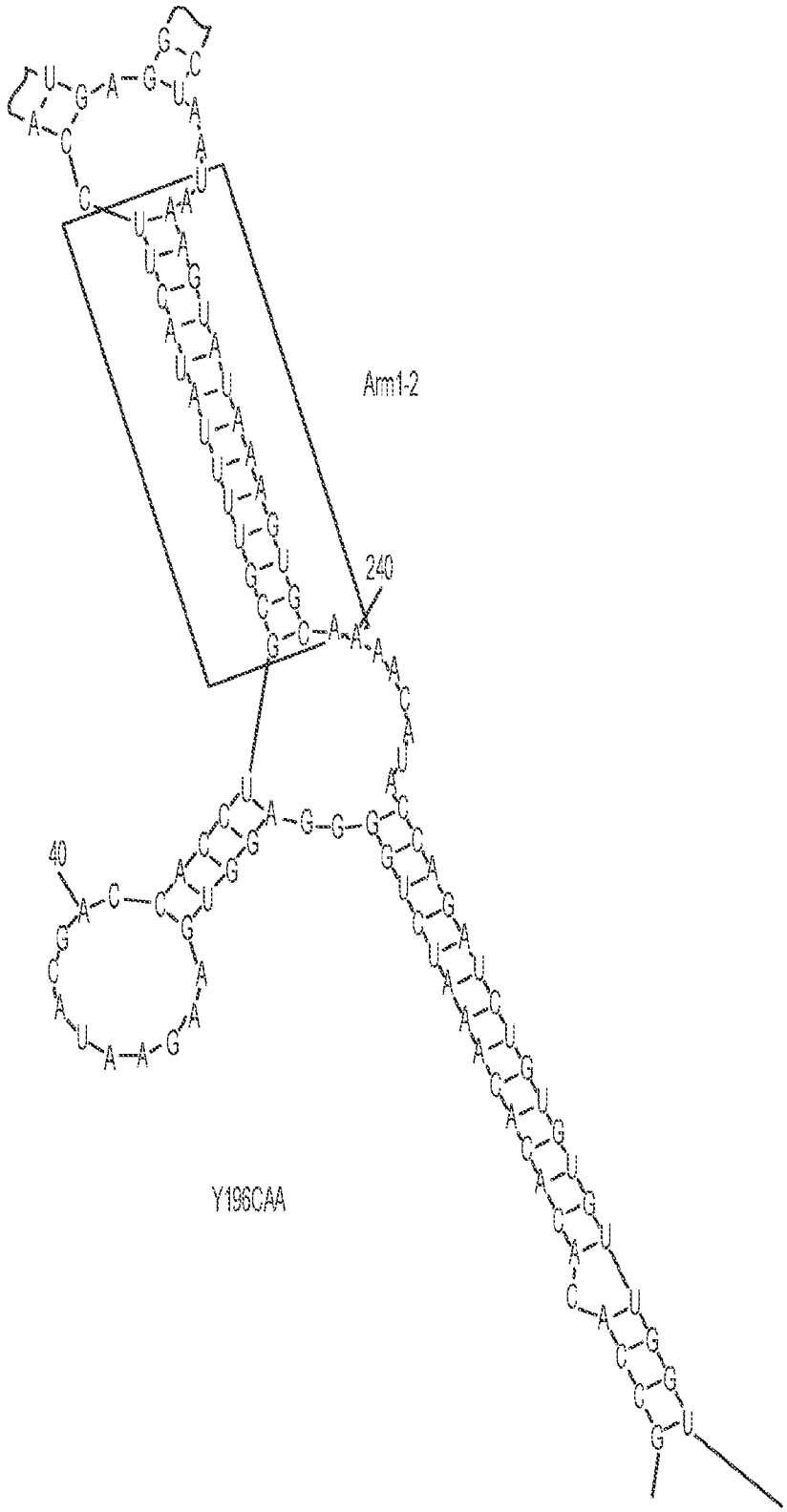


Figure 8
CONTINUED

	Y210	Y212C	Y213C	Y214C
Introduction (1ug/mL)	32	11	22	6

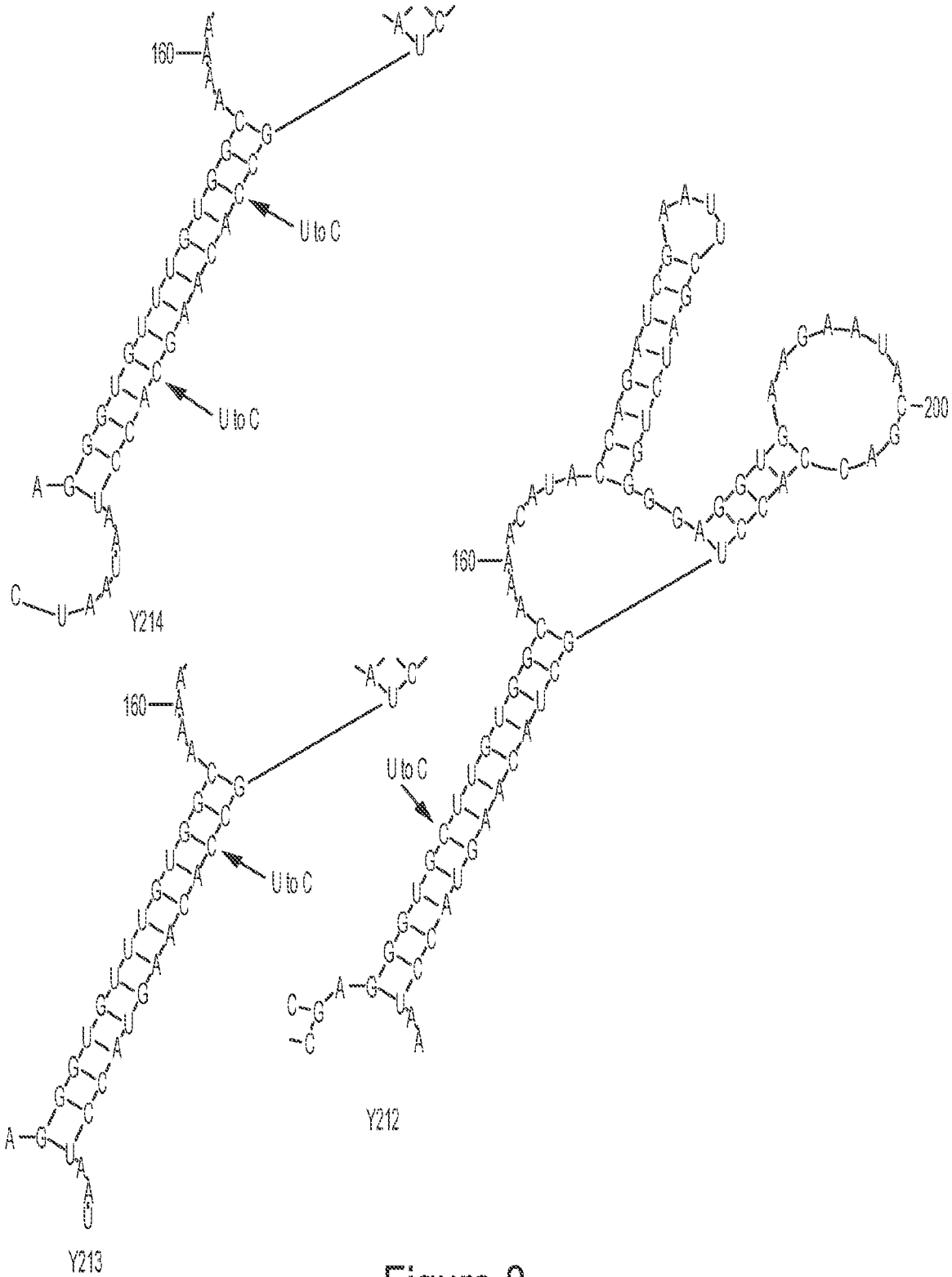


Figure 8
CONTINUED

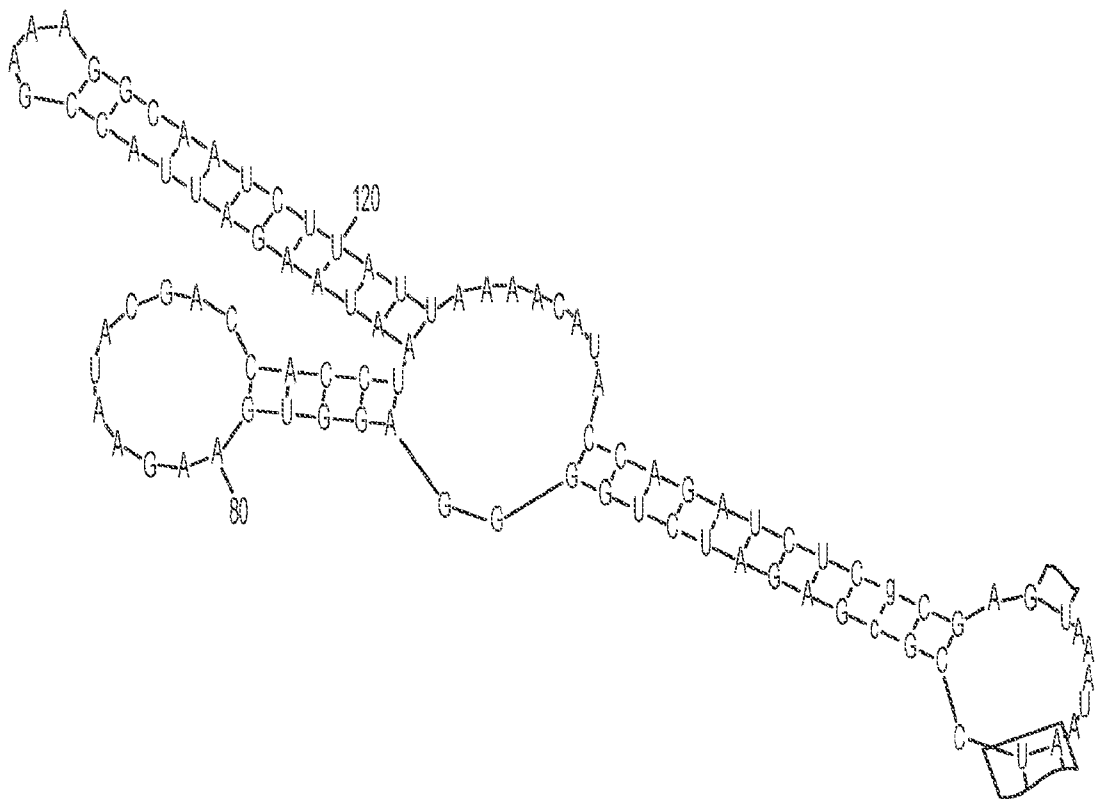


Figure 8
CONTINUED

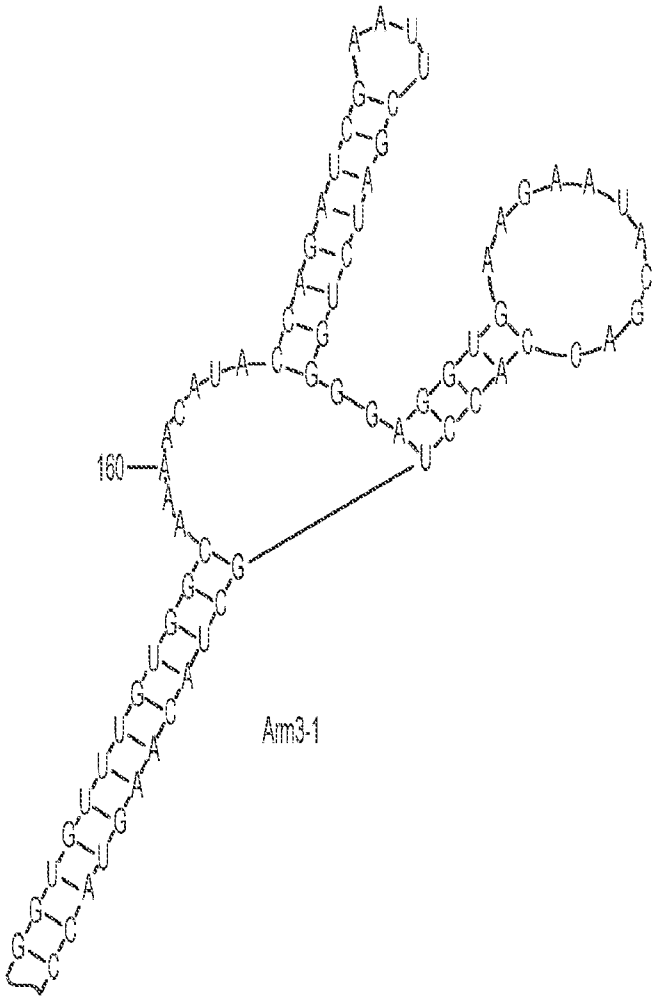


Figure 8
CONTINUED

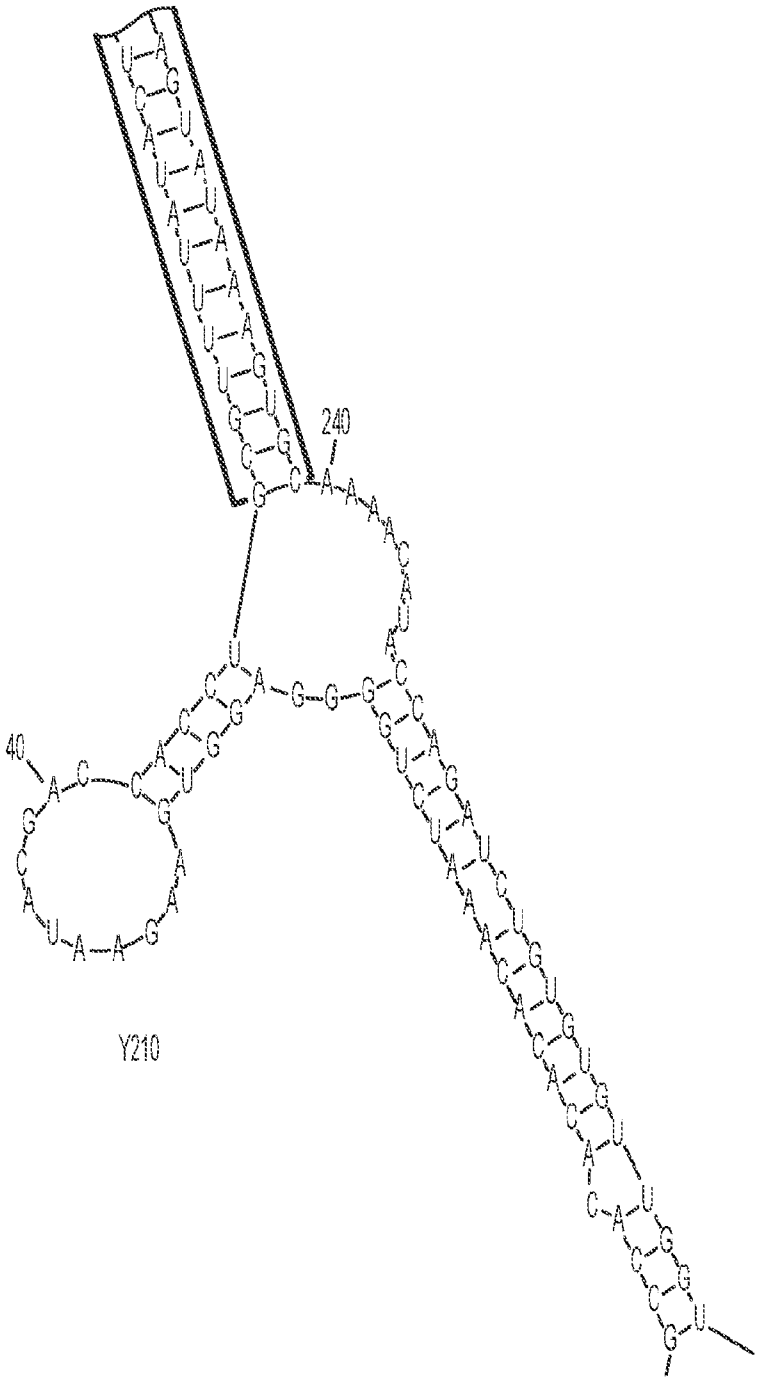
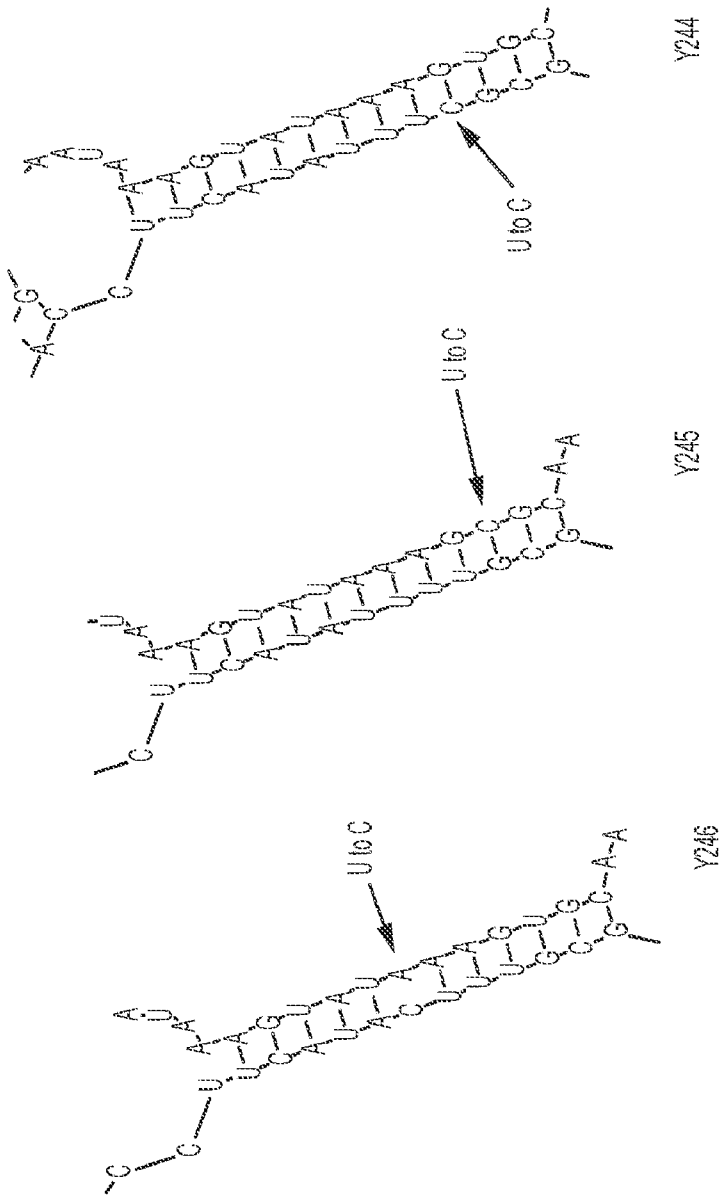


Figure 8
CONTINUED



	Y198CAA	Y244	Y245	Y246
Induction (5ug/mL)	250	172	95	127

Figure 8
CONTINUED

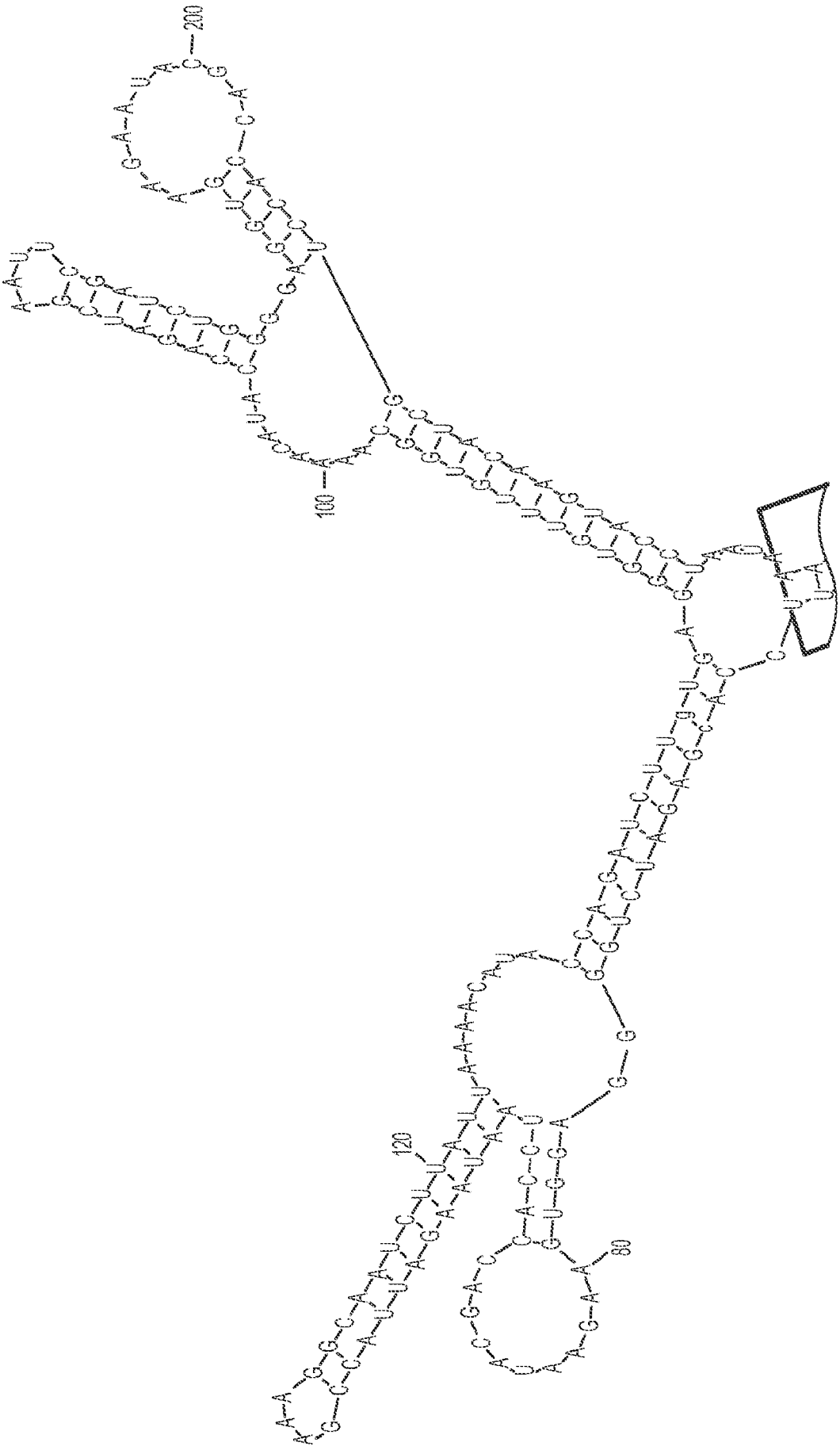


Figure 8
CONTINUED

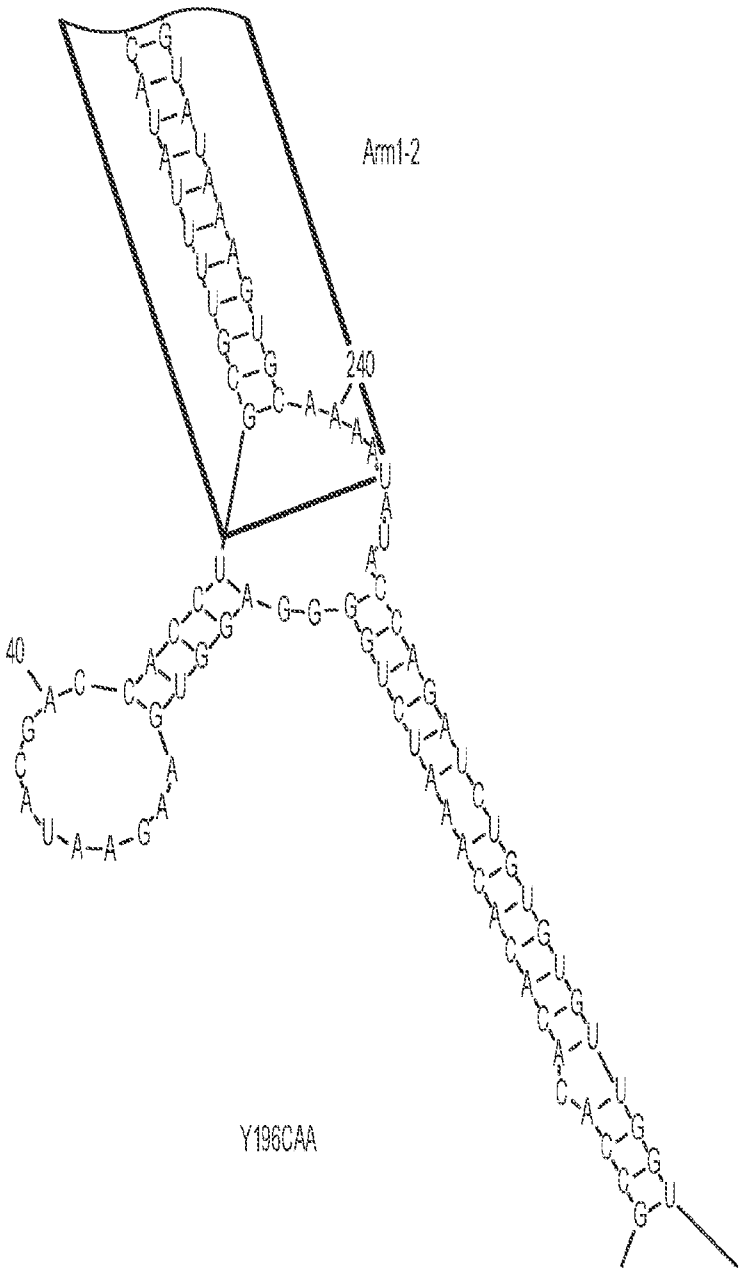
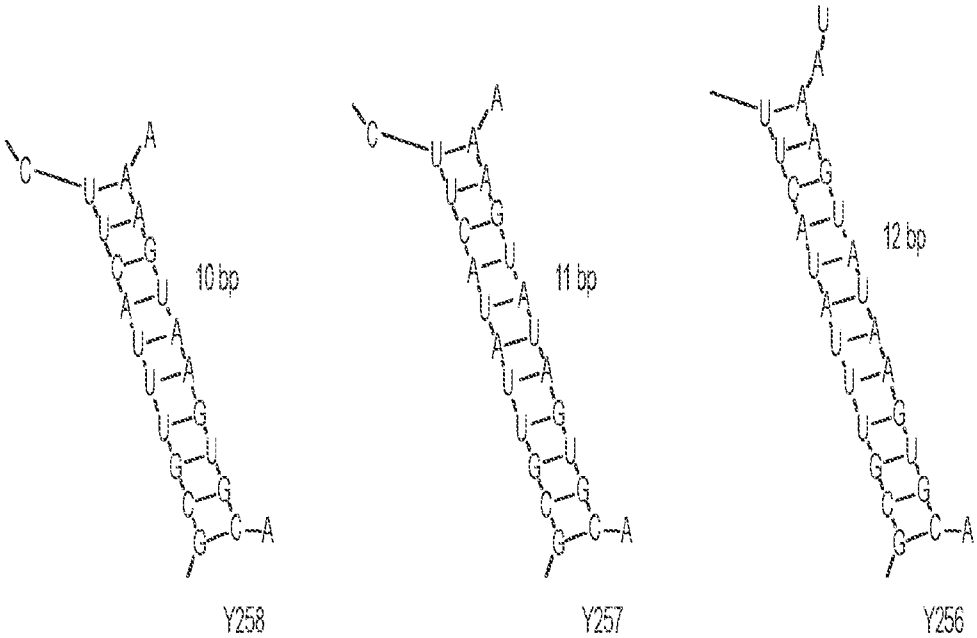


Figure 8
CONTINUED



	Y196CAA	Y256	Y257	Y258
Introduction (5ug/mL)	250	175	216	28

Figure 8
CONTINUED

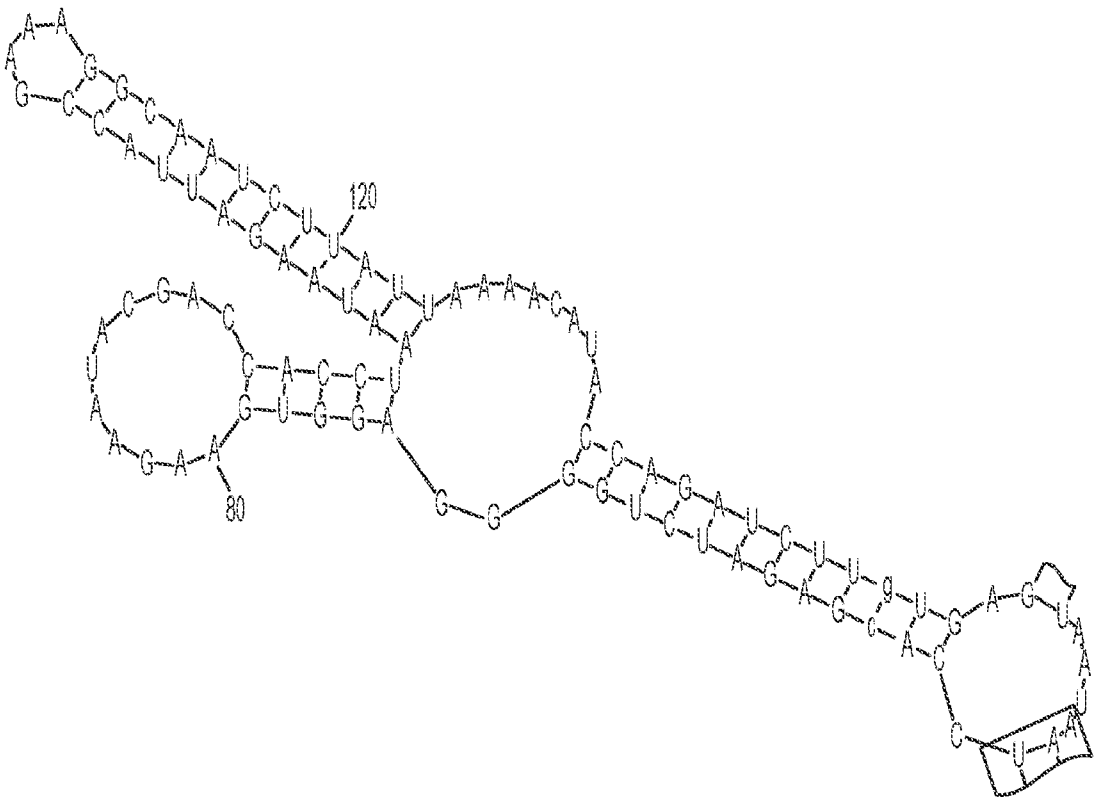


Figure 8
CONTINUED

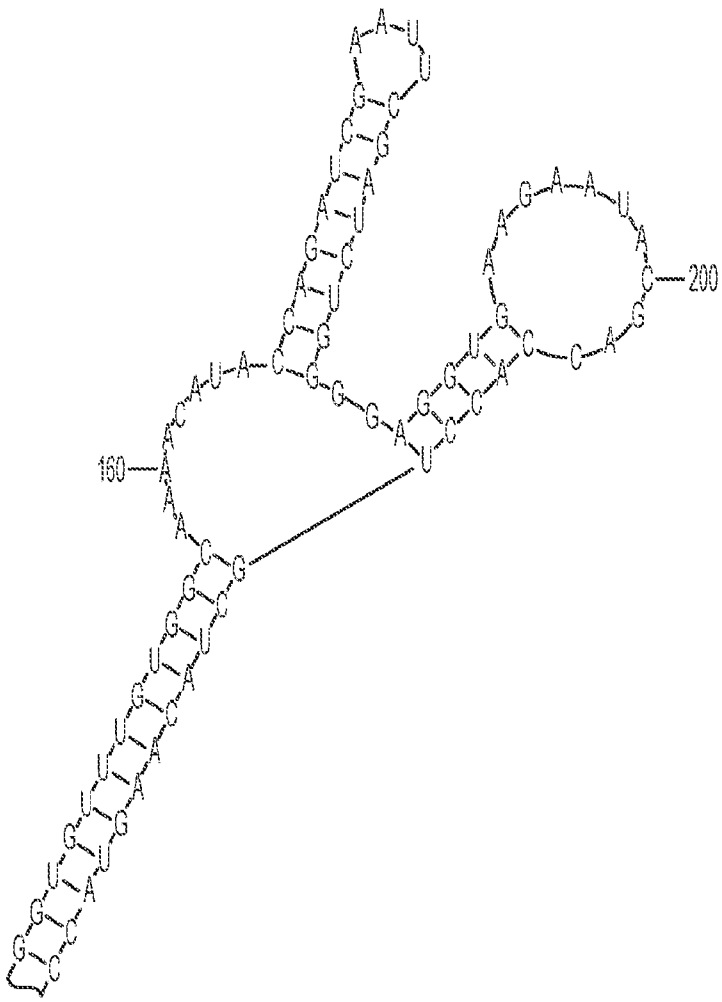


Figure 8
CONTINUED

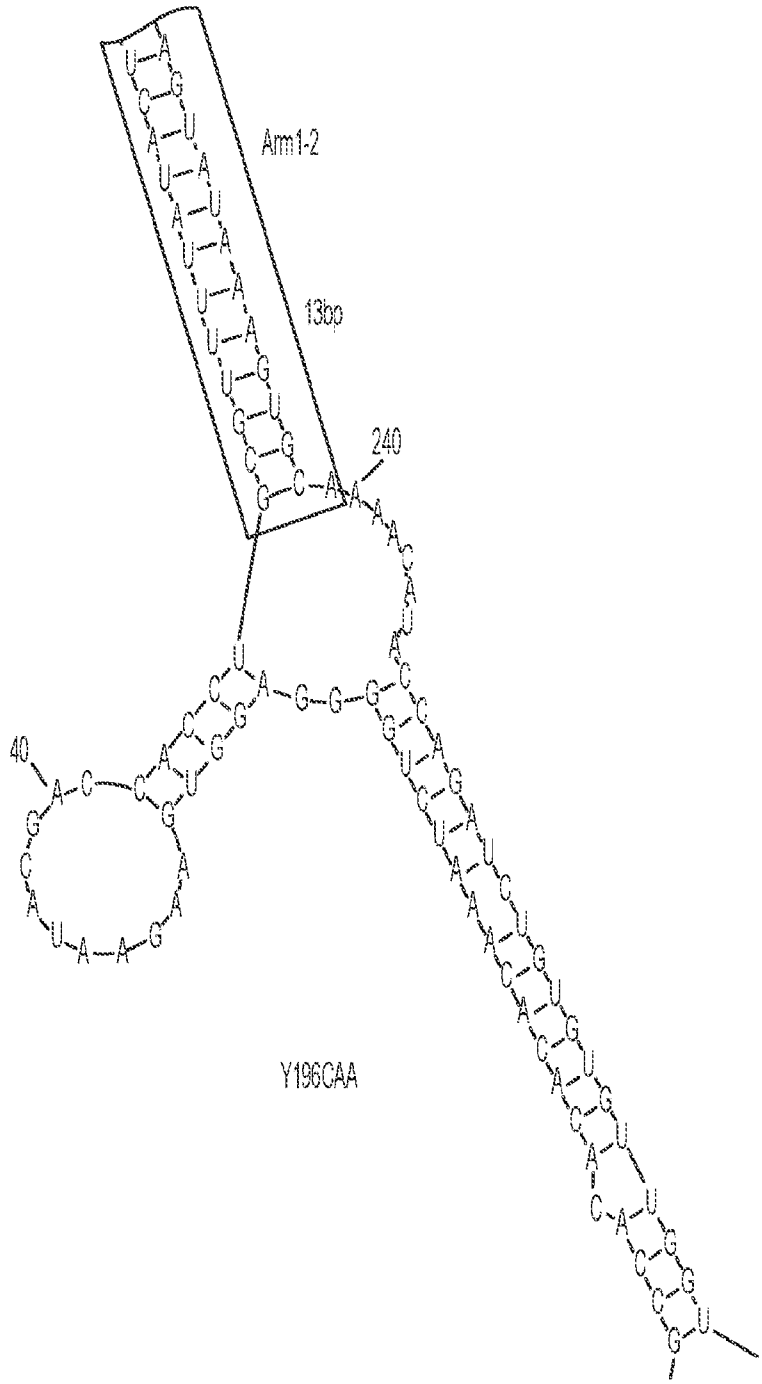
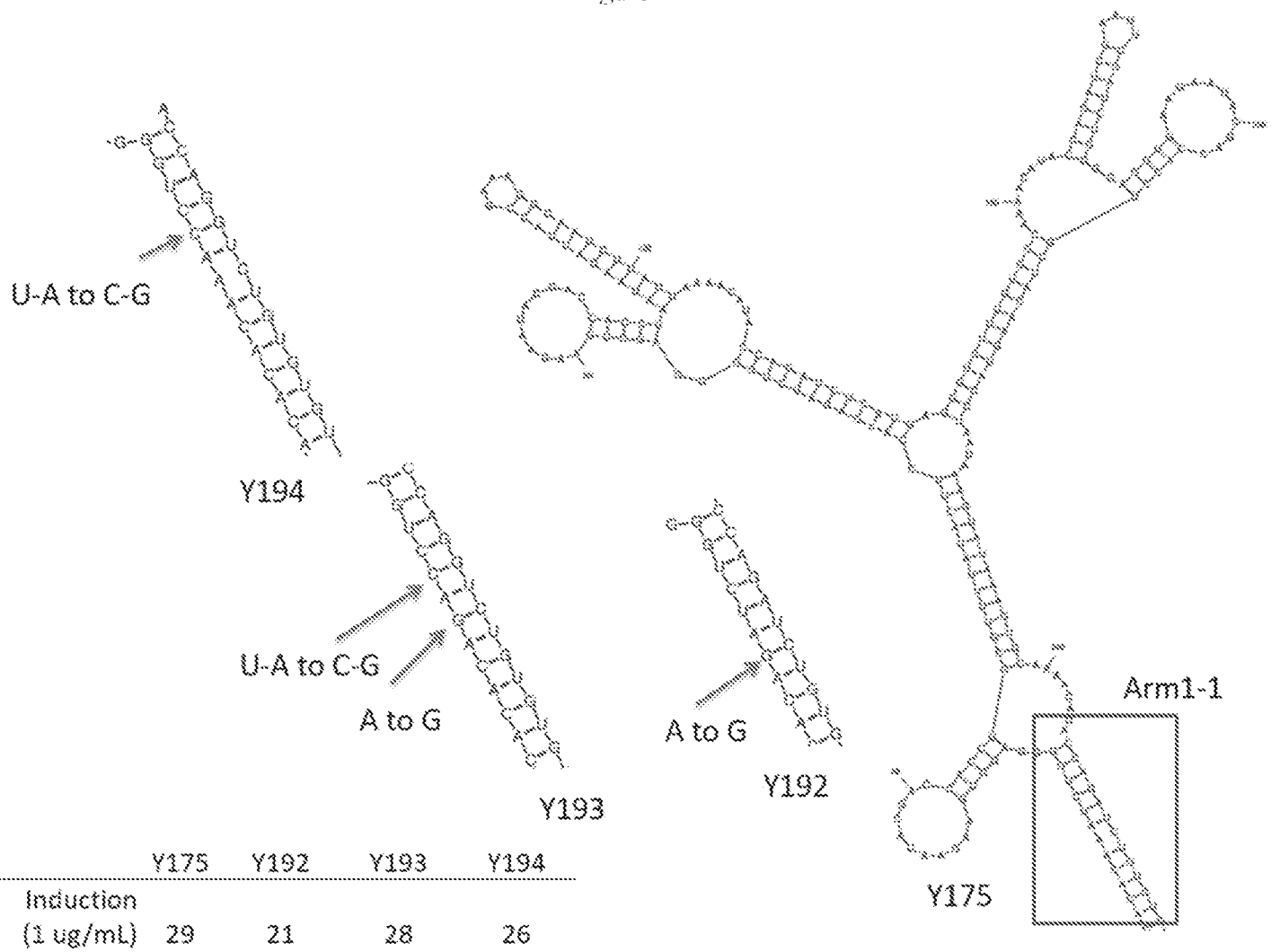


Figure 8
CONTINUED

Figure 9



	Y175	Y192	Y193	Y194
Induction (1 ug/mL)	29	21	28	26

Figure 9 (continued)

	Y196	Y226	Y211	Y227
Induction (Sug/mL)	162	160	180	171

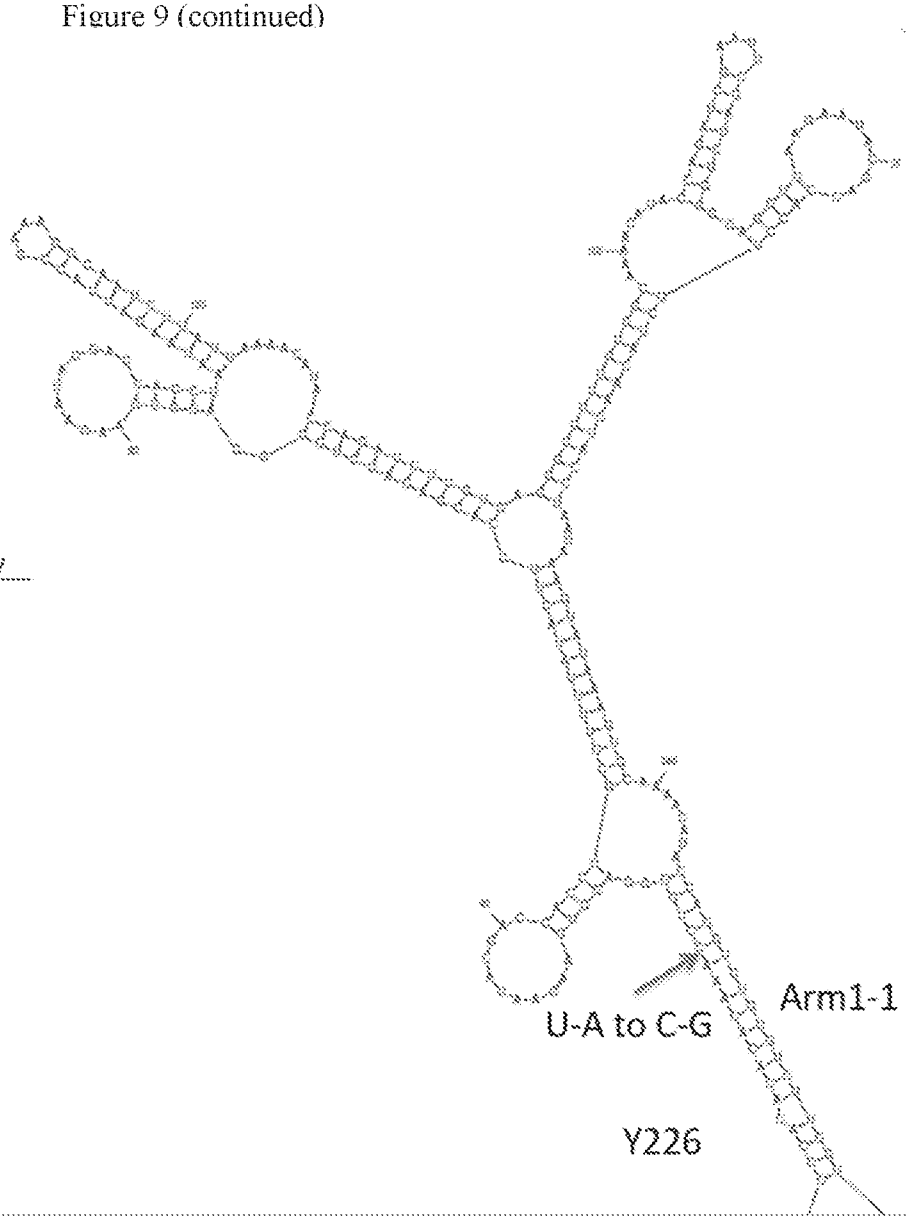
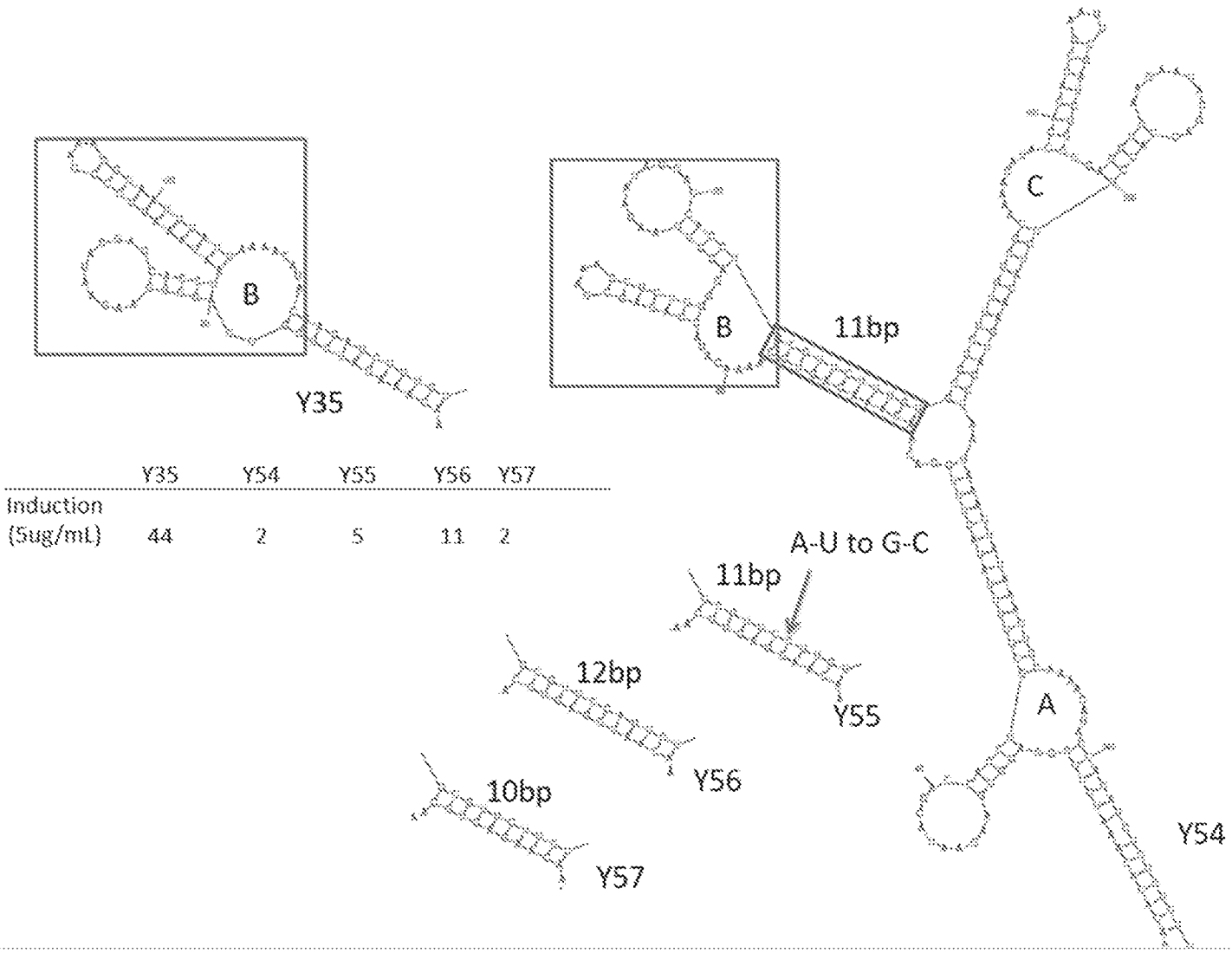


Figure 10A



	Y35	Y54	Y55	Y56	Y57
Induction (5ug/mL)	44	2	5	11	2

	Y196CAA	Y240	Y241	Y242	Y243	Y251
Induction (1 ug/mL)	36	1	1	1	1	1

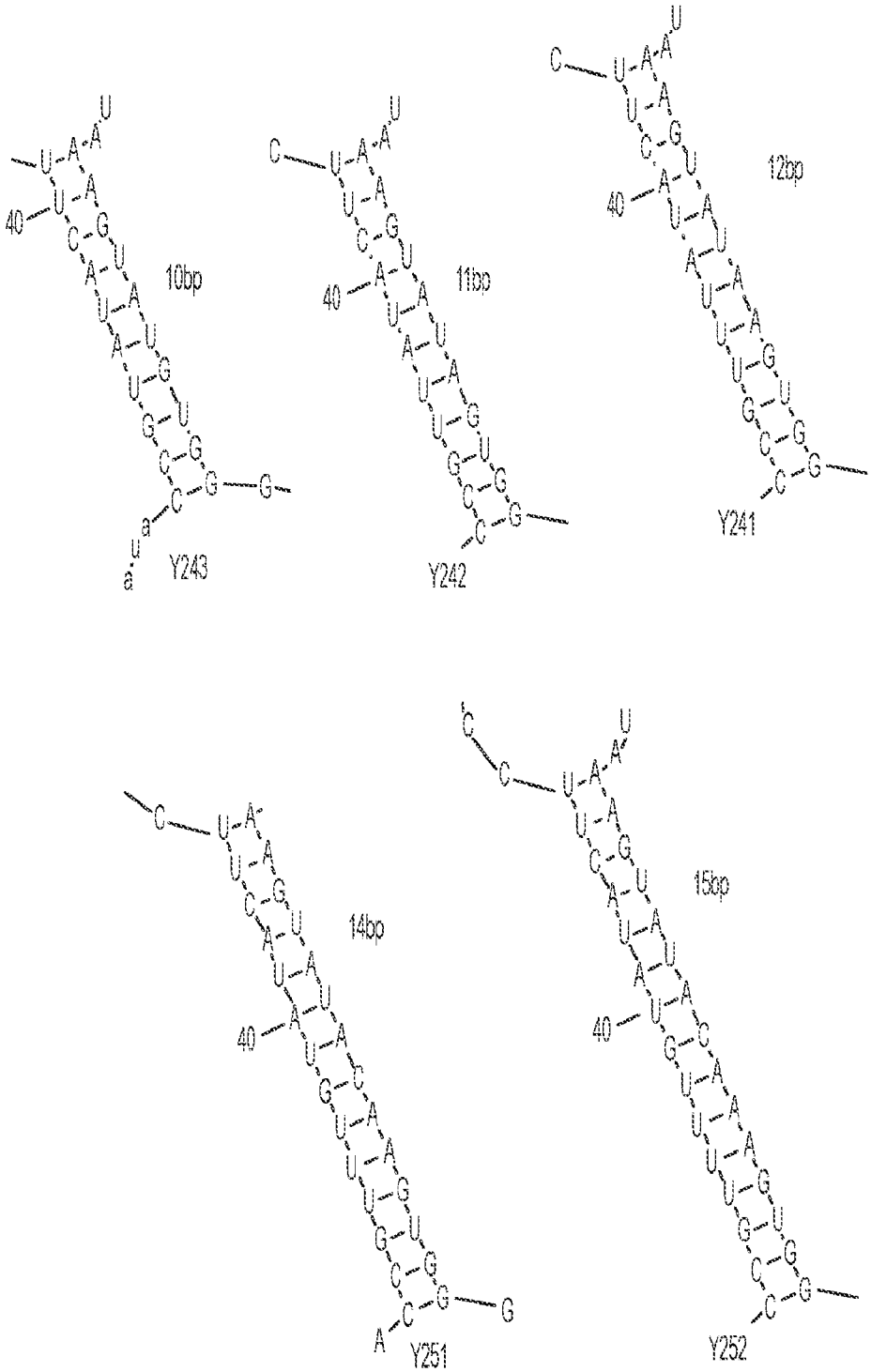


Figure 10B

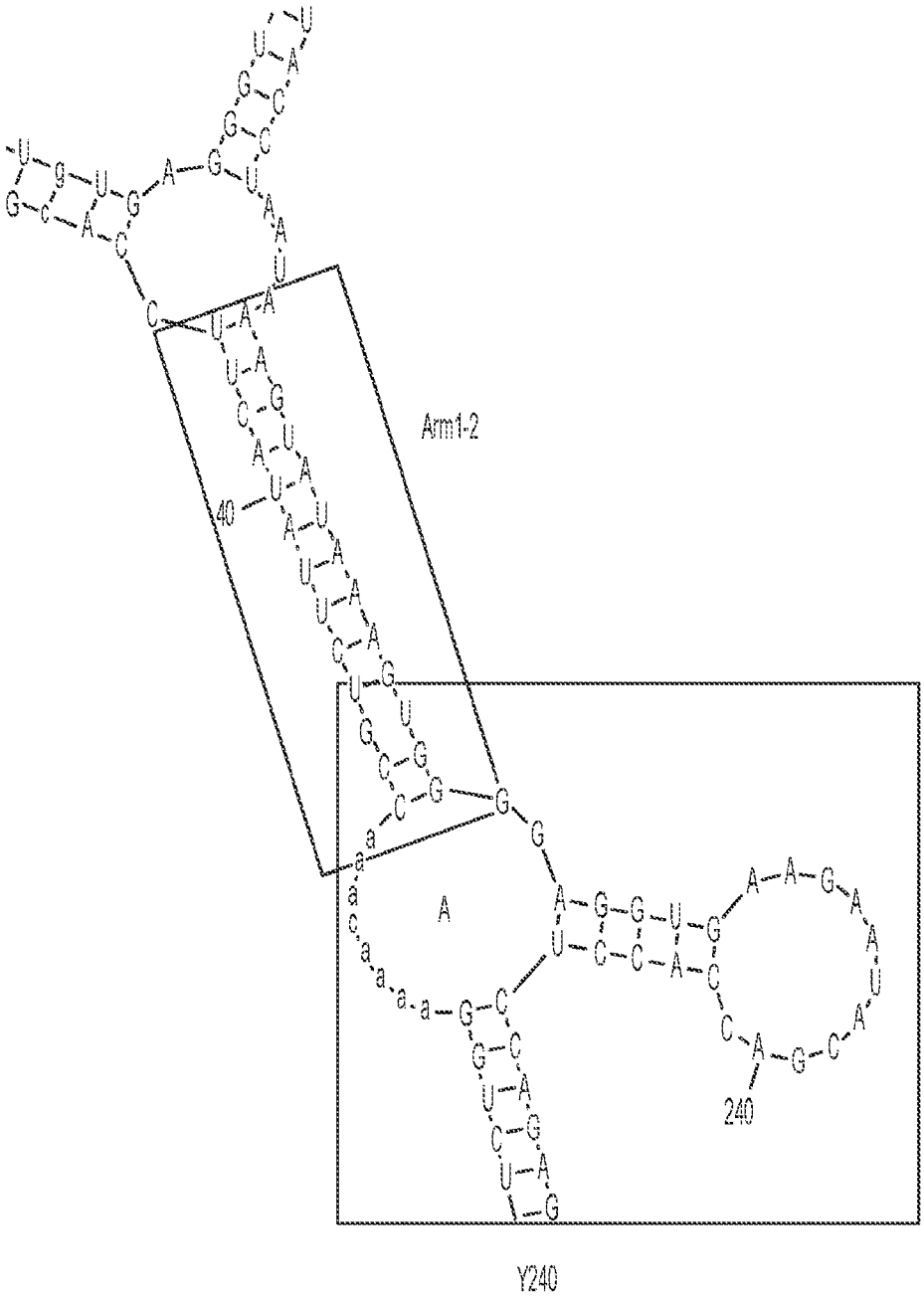


Figure 10B
CONTINUED

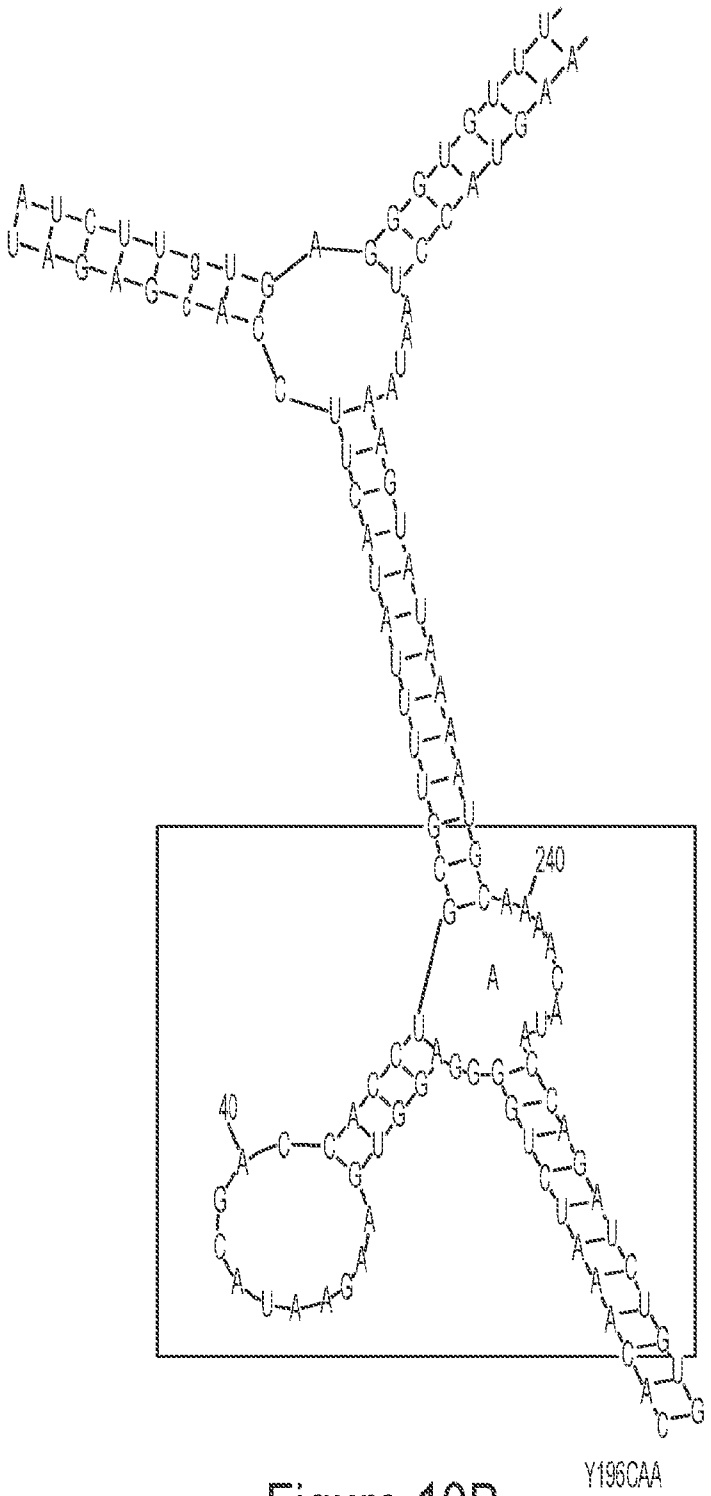


Figure 10B
CONTINUED

	Y196CAA	NA	NB	NC	NAB	NBC	NAC
Induction (1 ug/mL)	35	37	9	8	8	4	5
	Y196CAA	NA	NB	NC	NAB	NBC	NAC
Induction (5ug/mL)	273	220	80	60	38	10	32

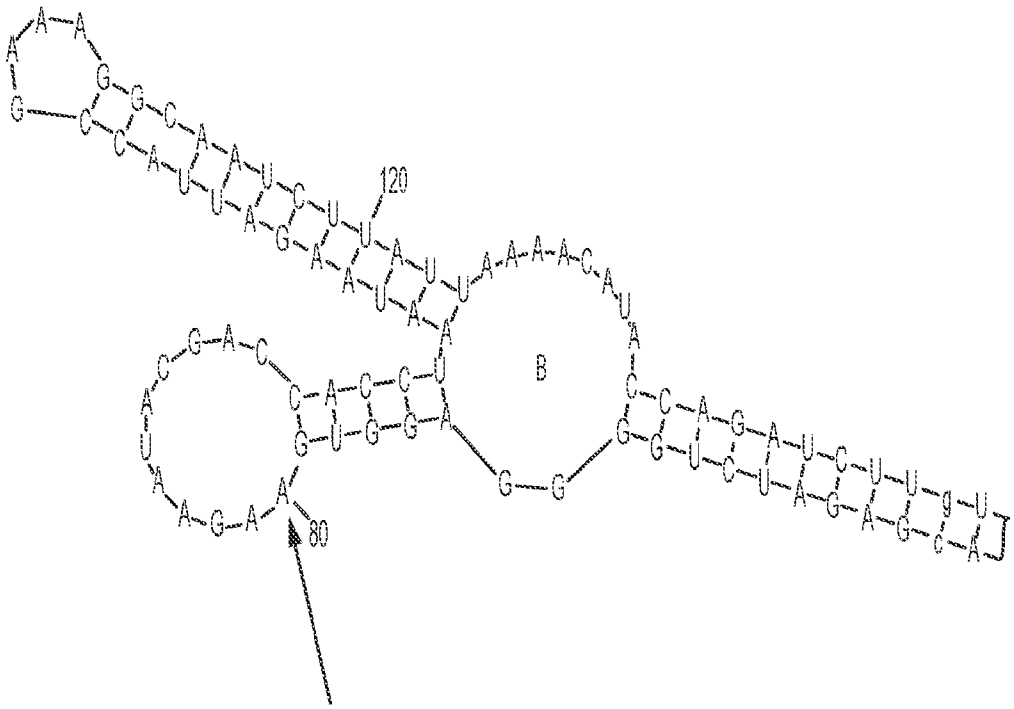


Figure 11A

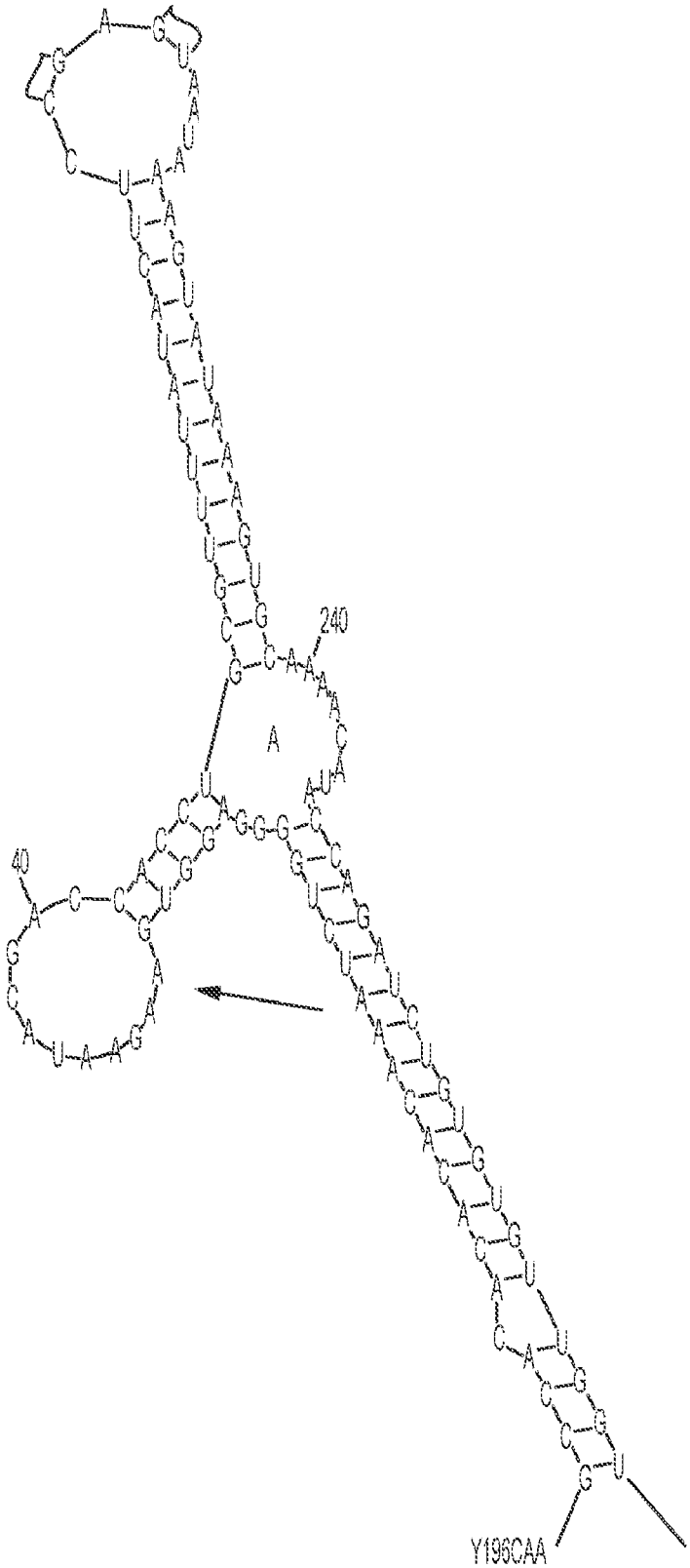


Figure 11A
CONTINUED

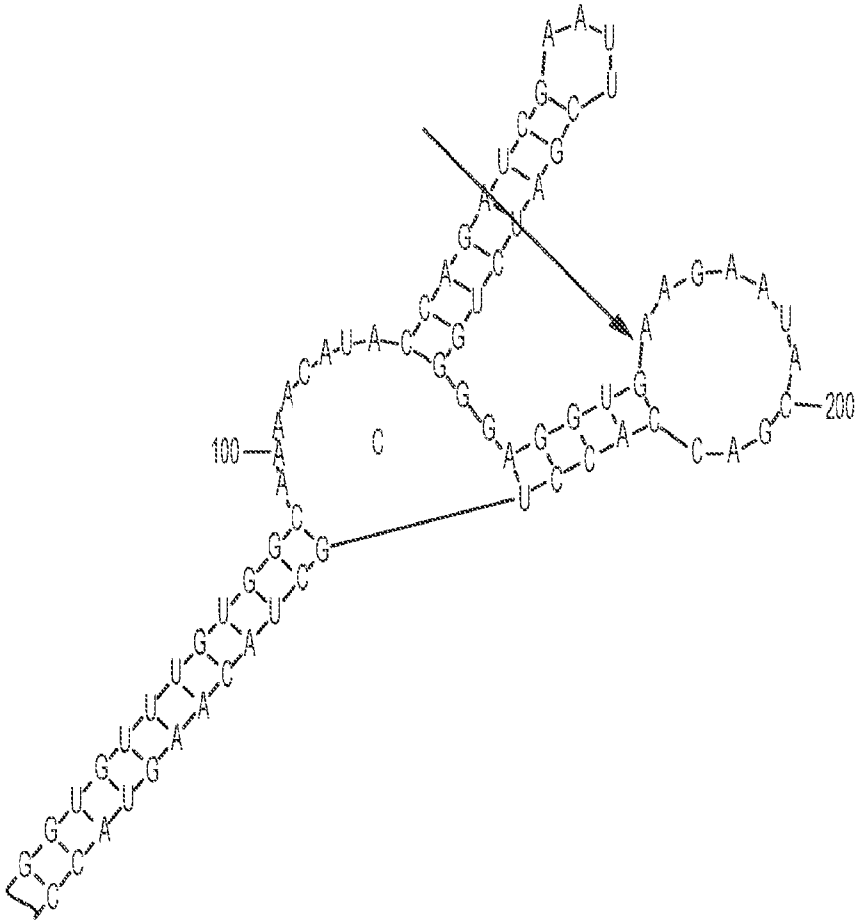


Figure 11A
CONTINUED

	Y196CAA	Y230	Y231
Induction (1 ug/mL)	39	25	23
	Y196CAA	Y230	Y231
Induction (5ug/mL)	279	100	151

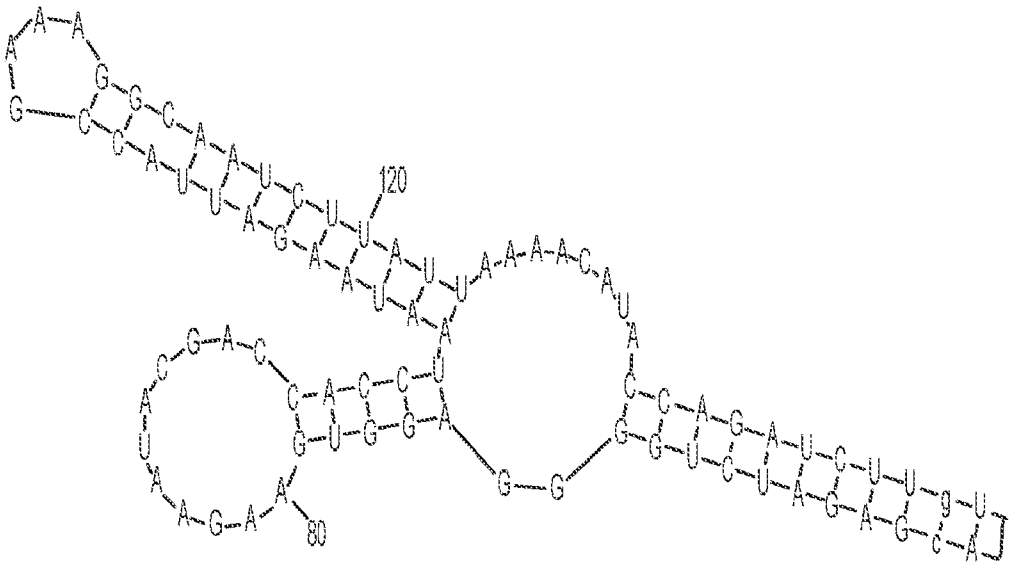


Figure 11B

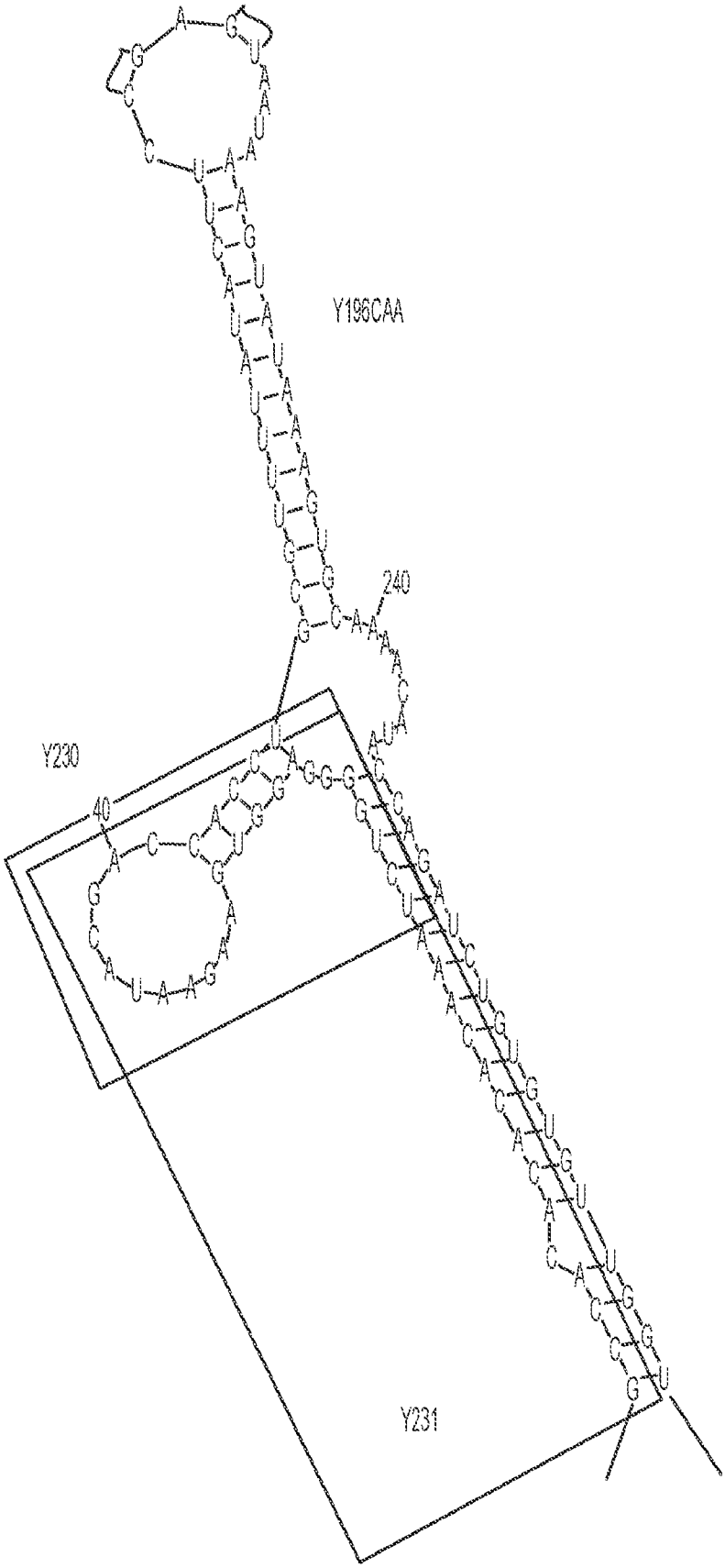


Figure 11B
CONTINUED

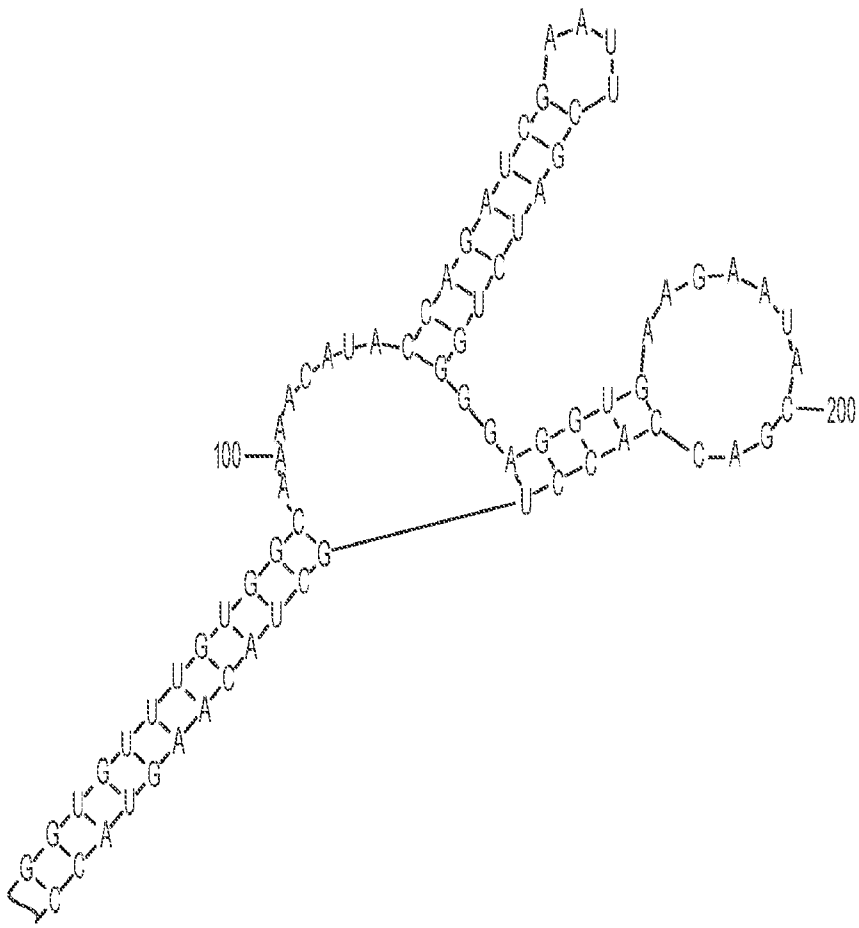


Figure 11B
CONTINUED

Insert CAA repeats into the 5'UTR of Y196, Y208, Y209 and Y211.

Seq: gcggccgccaacaacaac

Testing new 5'UTR sequence with strong 3'ss. These are based on S56. A few examples are shown:

S70:
GCGGCCGCCAACAACAACAATTCCTGCTCCTCTTCTGCCCAGGAACACGCTTGCCTTCCCCAAGGCTTCCAGAAGCTCTGAGGCAGGAGGCACCAAGTTCTACCTC
ACGTTTGGAGGATCTTGCTAGCT**ATG**
S71:
GCGGCCGCCCAGCAGATCCAGTGCTTCCTGCTCCTCTTCTGCCCAGGAACACGCTTGCCTTCCCCAAGGCTTCCAGAAGCTCTGAGGCAGGAGGCACCAAGTTCTACCT
CACGTTTTGGAGGATCTTGCTAGCT**ATG**
S74:
GCGGCCGCCCAACAACAACAATTCCTGCTCCTCTTCTGCCCTGGAACACGCTTGCCTTCCCCAAGGCTTCCAGAAGCTCTGAGGCAGGAGGCACCAAGTTCTACCTC
ACGTTTTGGAGGATCTTGCTAGCT**ATG**
S75:
GCGGCCGCCCAGCAGATCCAGTGCTTCCTGCTCCTCTTCTGCCCTGGAACACGCTTGCCTTCCCCAAGGCTTCCAGAAGCTCTGAGGCAGGAGGCACCAAGTTCTACCT
CACGTTTTGGAGGATCTTGCTAGCT**ATG**
S78:
GCGGCCGCCAACAACAACAACAACAACAACACGCTTGCCTTCCCCAAGCTTCCACAAGCAACCTCAAACAGACACC**ATG**
S79:
GCGGCCGCTACTAACAACACGCTTTGCCTTCCCCAAGCTTCCACAAGCAACCTCAAACAGACACC**ATG**
S80:
GCGGCCGCGCTTCCTCTTCCTGCCCAGGAACACGCTTTGCCTTCCCCAAGCTTCCACAAGCAACCTCAAACAGACACCATG
S81:
GCGGCCGCGCTTCCTGCTTCTTCTGCCCTGGAACACGCTTGCCTTCCCCAAGCTTCCACAAGCAACCTCAAACAGACACC**ATG**
S82:
GCGGCCGCTACTAACGCTCCTCTTCTGCCCTGGAACACGCTTGCCTTCCCCAAGCTTCCACAAGCAACCTCAAACAGACACC**ATG**

	S56-e	S70-e	S71-e	S74-e	S75-e
Induction (5ug/mL)	188	3	76	95	55
	S56-e	S78-e	S80-e	S81-e	S82-e
Induction (5ug/mL)	188	25	183	121	21

Figure 12B

New unstructured 5'UTR modified from Y305

Y324: ttaattaacagtgttcactagcatcTAACCACTTACATACCATCTACCACCGACACC**ATG**
Y325: ttaattaacagtgttcactagcatcTAACCACTTACATACCATCTACCACCGCCACC**ATG**
Y326: ttaattaacagtgttcactagcatcTAACCACTTACATAttctgccgcctacc**ATG**

New unstructured 5'UTR modified from Y300

Y328: gcggccgcCataacagtgttcactagcaTccCcaacagACCATCTACCACCGACACC**ATG**
Y329: gcggccgcCataacagtgttcactagcaTccCccagACCATCTACCACCGACACC**ATG**
Y330: gcggccgcCataacagtgttcactagcaTccCcTACCATCTACCACCGACACC**ATG**
Y331: gcggccgcCataacagtgttcactagcaTccCcaacagACTTACATACCATTGACACC**ATG**
Y332: gcggccgcCataacagtgttcactagcaTccCcACTTACATACCATTGACACCATG

	Y305	Y324	Y325	Y326	Y328	Y329	Y330	Y331	Y332
Induction (1ug/mL)	599	6	4	3	73	354	33	346	46

Figure 12C

Y322:
GCGGCCGCCTTAATTAA**CAACAACAACAACAACAACAACAACAACAA**CCAGTGTTCACTAGCATCGACAC
CATGATCCCCGGGG

Y323:
GCGGCCGCC**CAACAACAACAACAACAACAACAACAACAA**CTTAATTAACAGTGTTCACTAGCATCGACAC
CATGATCCCCGGGG

	Y305	Y322	Y323
Induction (1ug/mL)	599	67	12

Figure 12D

Parental: Y196CAA (which has 2MAZ (in red))
GTGTGTTGGTTTTTTGTGTGAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGCGGCCGCC

Y232: identical to Y196CAA, with modified CD44 G-quad (in green) placed before 2MAZ
GTGTGTTGGTTTTTTGTGTG*GTGGTGGATTGGT*TAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGCGGCCGC

Y233: identical to Y196CAA, with modified CD44 G-quad (in green) placed after 2MAZ
GTGTGTTGGTTTTTTGTGTGAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGA*GTGGTGGTGGATTGGT*GCGGCCGC

Y196CAA-3MAZ: identical to Y196CAA, with one additional MAZ
GTGTGTTGGTTTTTTGTGTGAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGCGGCCGCC

Y196CAA-4MAZ: identical to Y196CAA, with two additional MAZ
GTGTGTTGGTTTTTTGTGTGAACGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGAGGAAAGGGGGAGGGGGGAGCGGCCGCC

	Y196CAA	Y232	Y233	Y196CAA3MAZ	Y196CAA4MAZ
Induction (5ug/mL)	258	261	238	275	569

Figure 13A

S177: based on S56, replacing 4MAZ with one CD44 G-quad: TGGTGGTGGAATGGT
S178: based on S56, replacing 4MAZ with two CD44 G-quad:
TGGTGGTGGAATGGTAAATGGTGGTGGAATGGT
S179: based S56, replacing 4MAZ with four CD44 G-quad:
TGGTGGTGGAATGGTAAATGGTGGTGGAATGGTA
AATGGTGGTGGAATGGTAAATGGTGGTGGAATGGT

	S56-e	S177-e	S178-e	S179-E
Induction (1ug/mL)	39	2	2	10
	S56-e	S177-e	S178-e	S179-E
Induction (5ug/mL)	139	2	2	46

Figure 13B

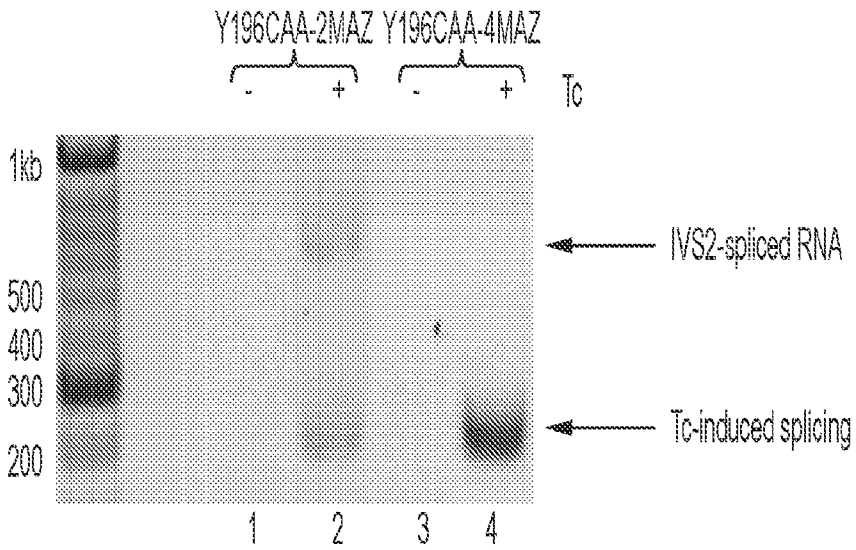
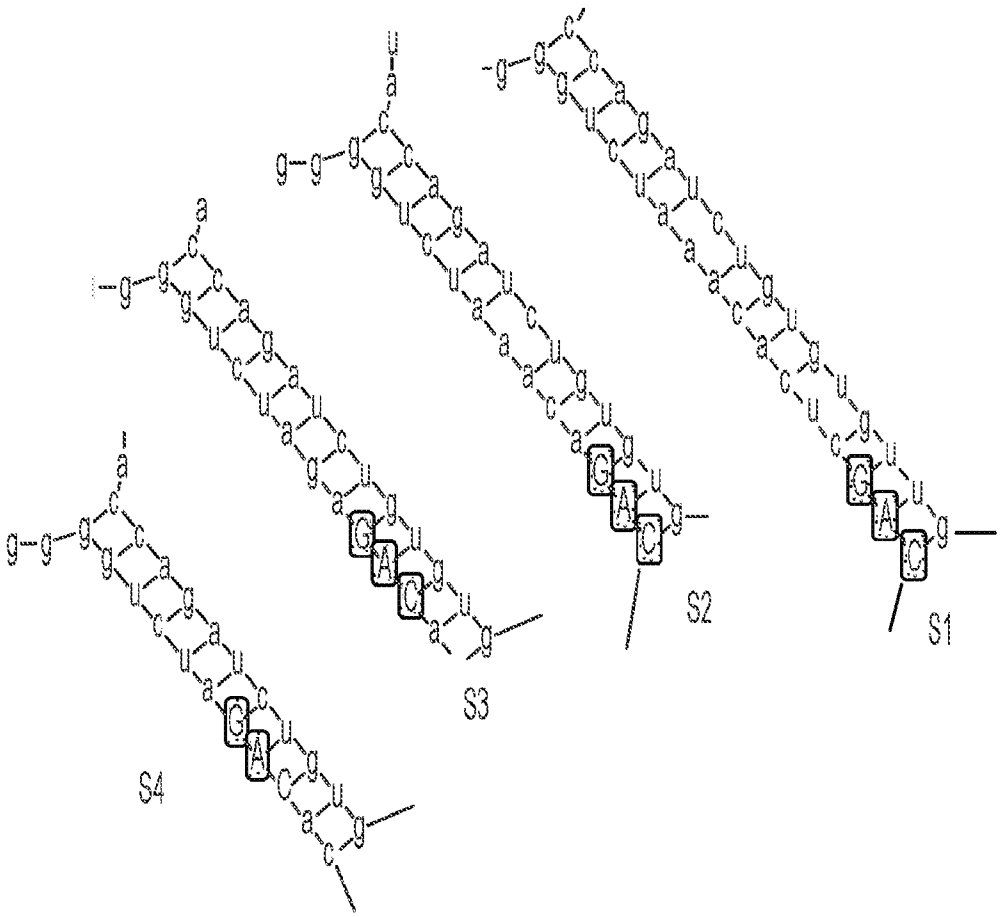

Y196CAA-2MAZ
Y196CAA-4MAZ
-
+
-
+
Tc
1kb
500
400
300
200
IVS2-spliced RNA
Tc-induced splicing
1
2
3
4

Figure 14



S1e-S5e are based on a Y196CAA-4MAZ construct that uses eGFP-luciferase fusion protein as the reporter (indicated by "e") which has shorter half life therefore lower induction levels as compared to luciferase reporter

	luciferase	With 4MAZ and eGFP/luc					
	Y196CAA-4MAZ	Y196CAA-4MAZ-e	S1e	S2e	S3e	S4e	S5e
Induction (5ug/mL)	698	196	219	44	177	235	2

Figure 15A

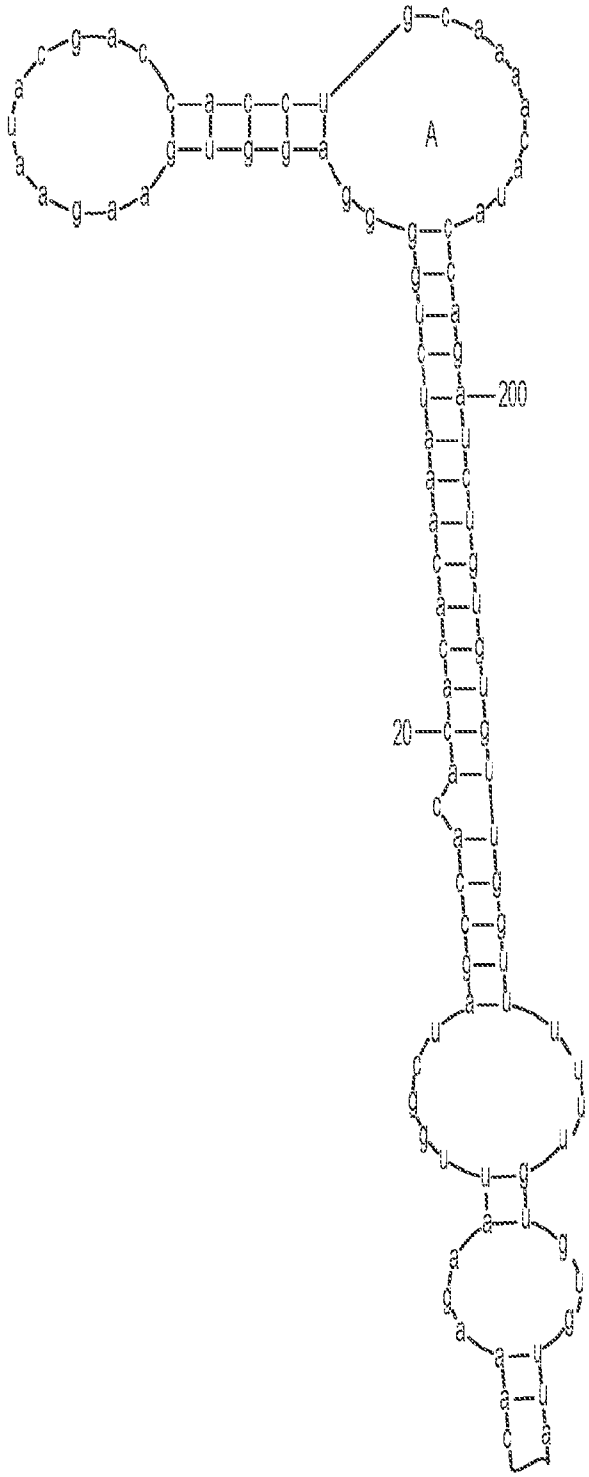


Figure 15A
CONTINUED

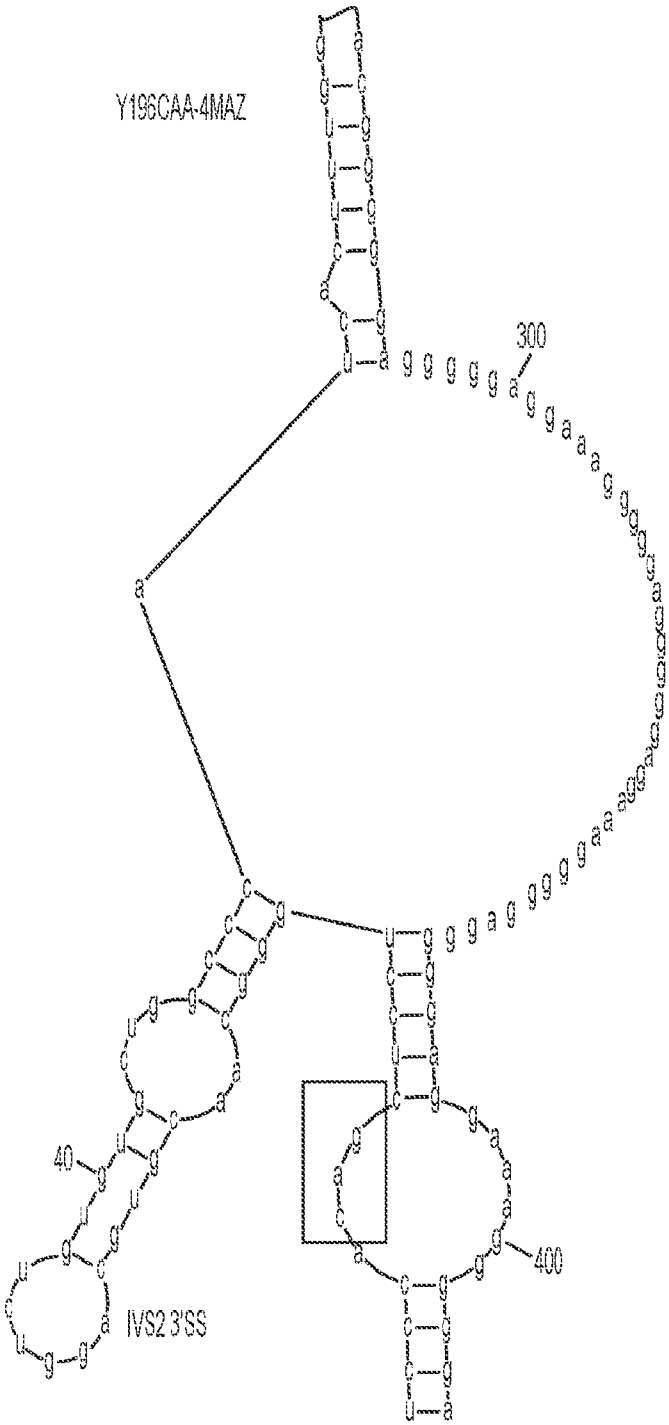
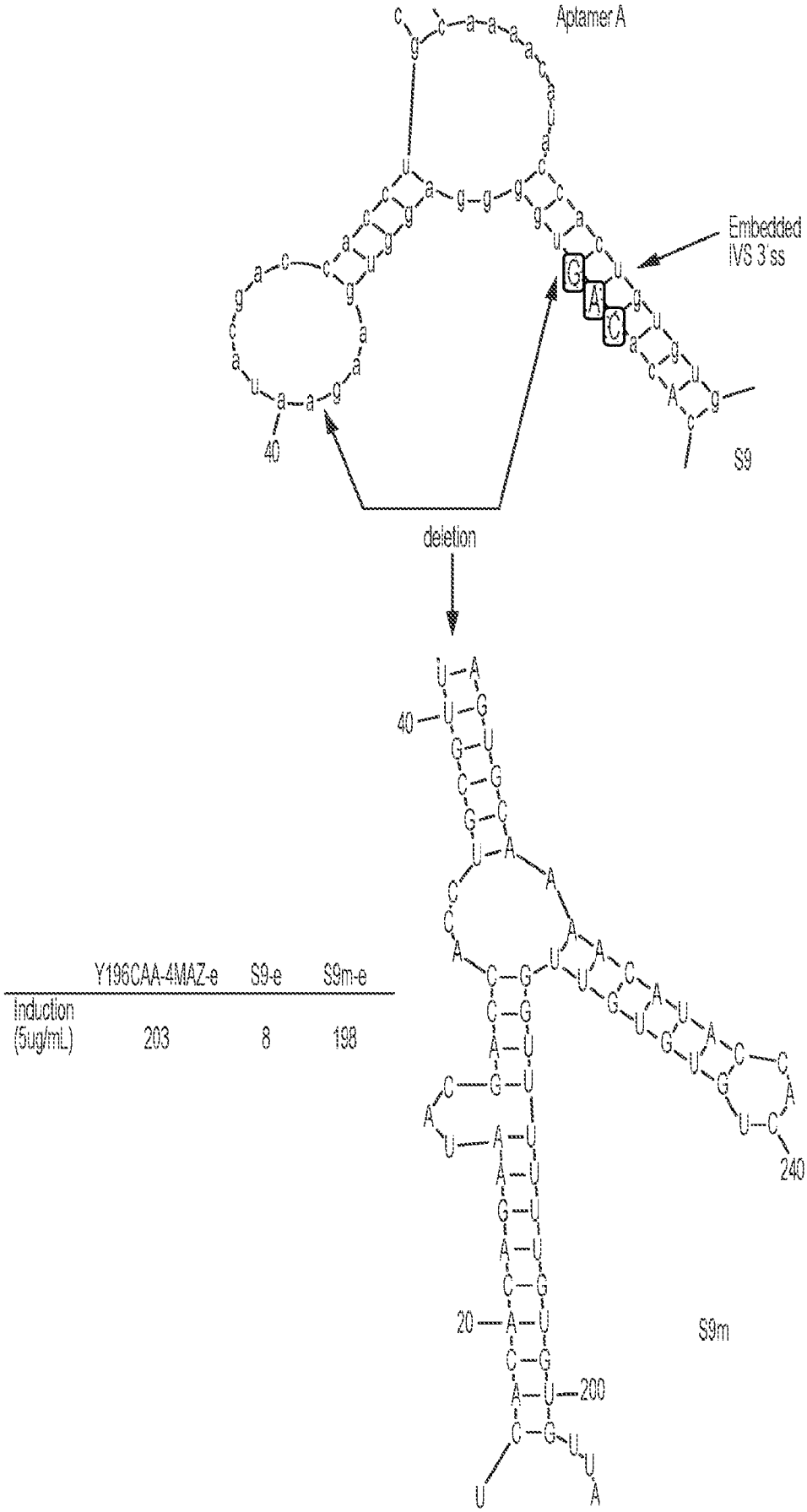
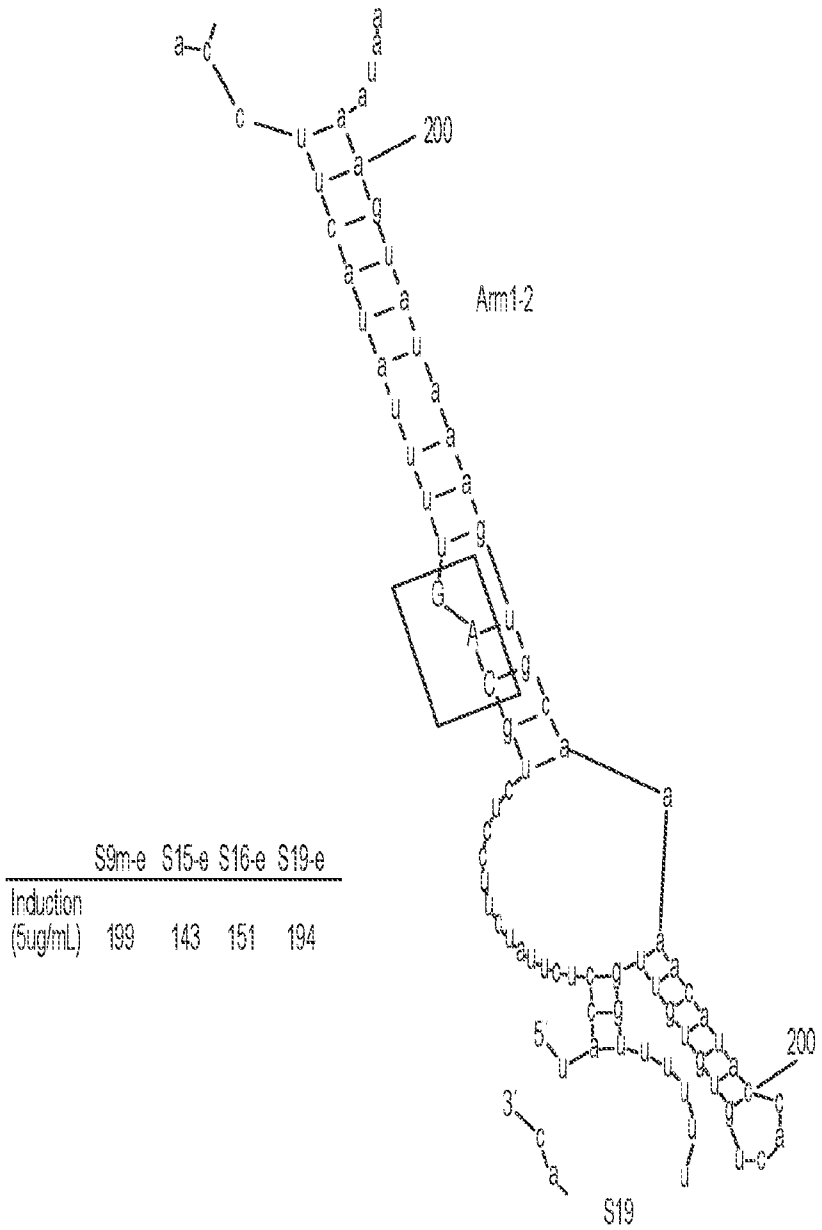


Figure 15A
CONTINUED



	Y196CAA-4MAZ-e	S9-e	S9m-e
Induction (5ug/mL)	203	8	198

Figure 15B



	S9m-e	S15-e	S16-e	S19-e
Induction (5ug/mL)	199	143	151	194

Figure 15C

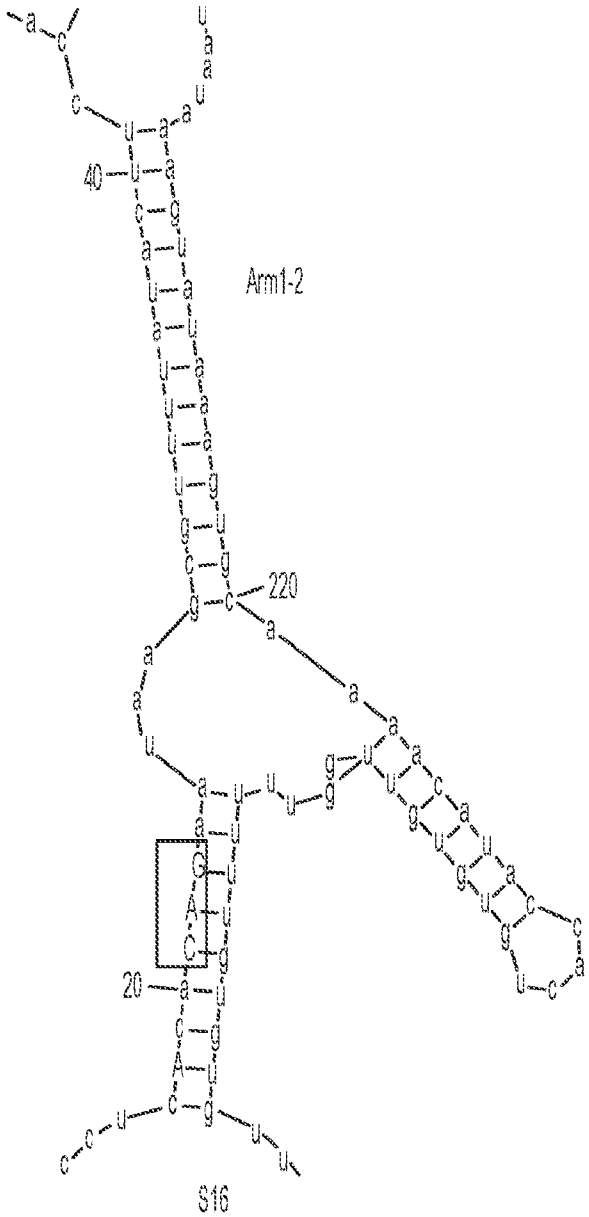


Figure 15C
CONTINUED

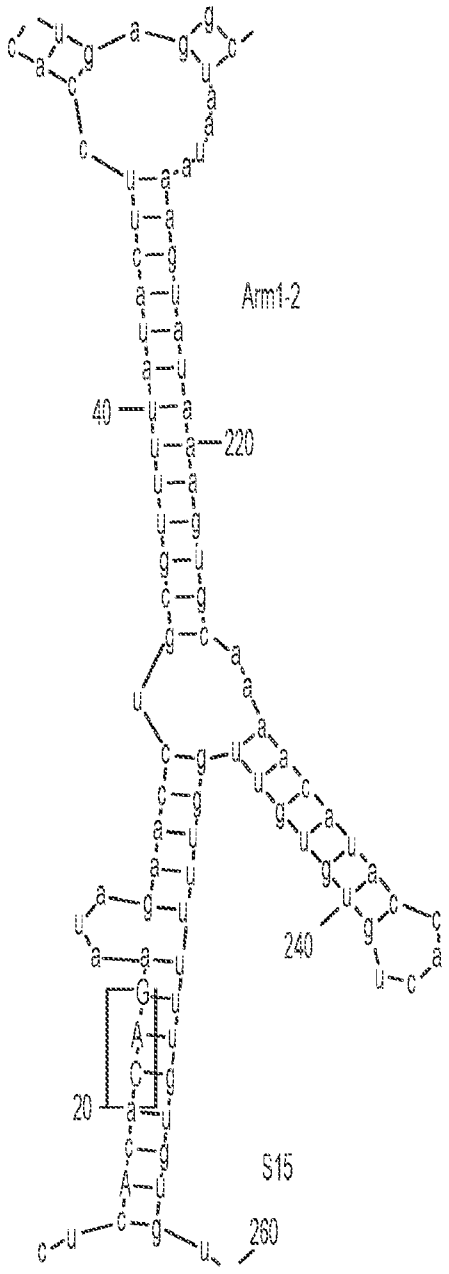


Figure 15C
CONTINUED

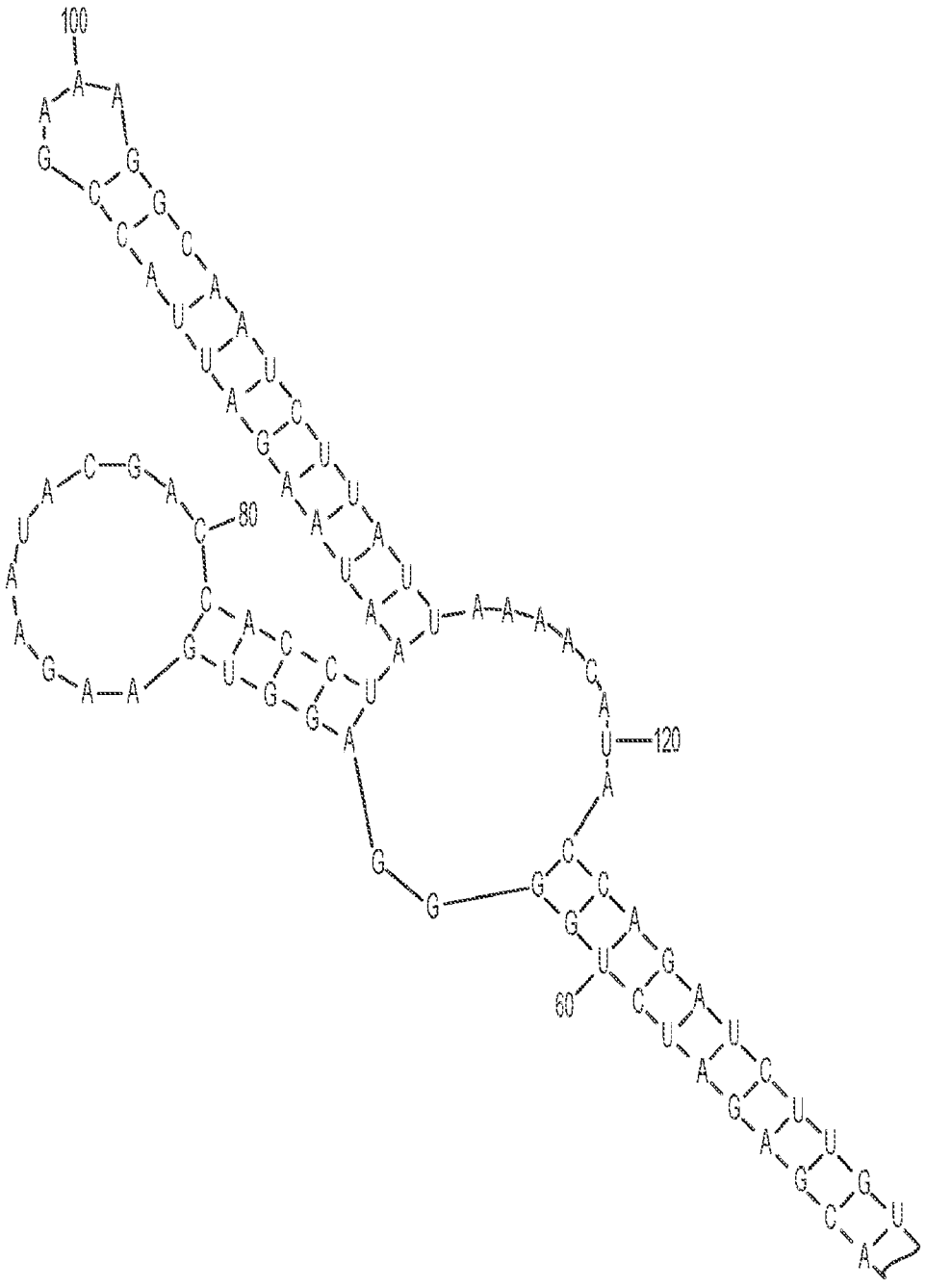


Figure 15C
CONTINUED

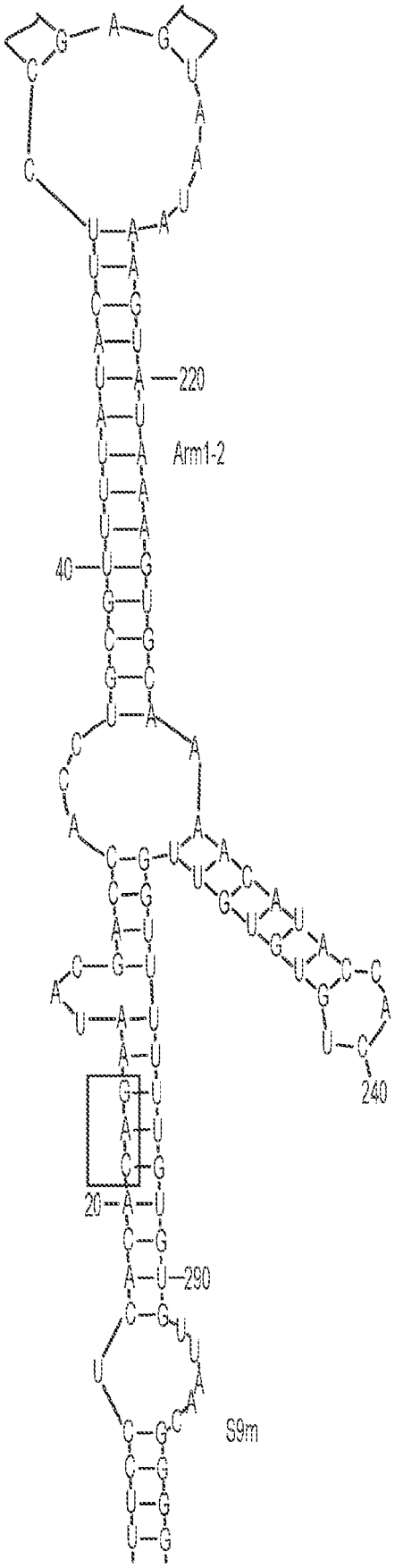


Figure 15C
CONTINUED

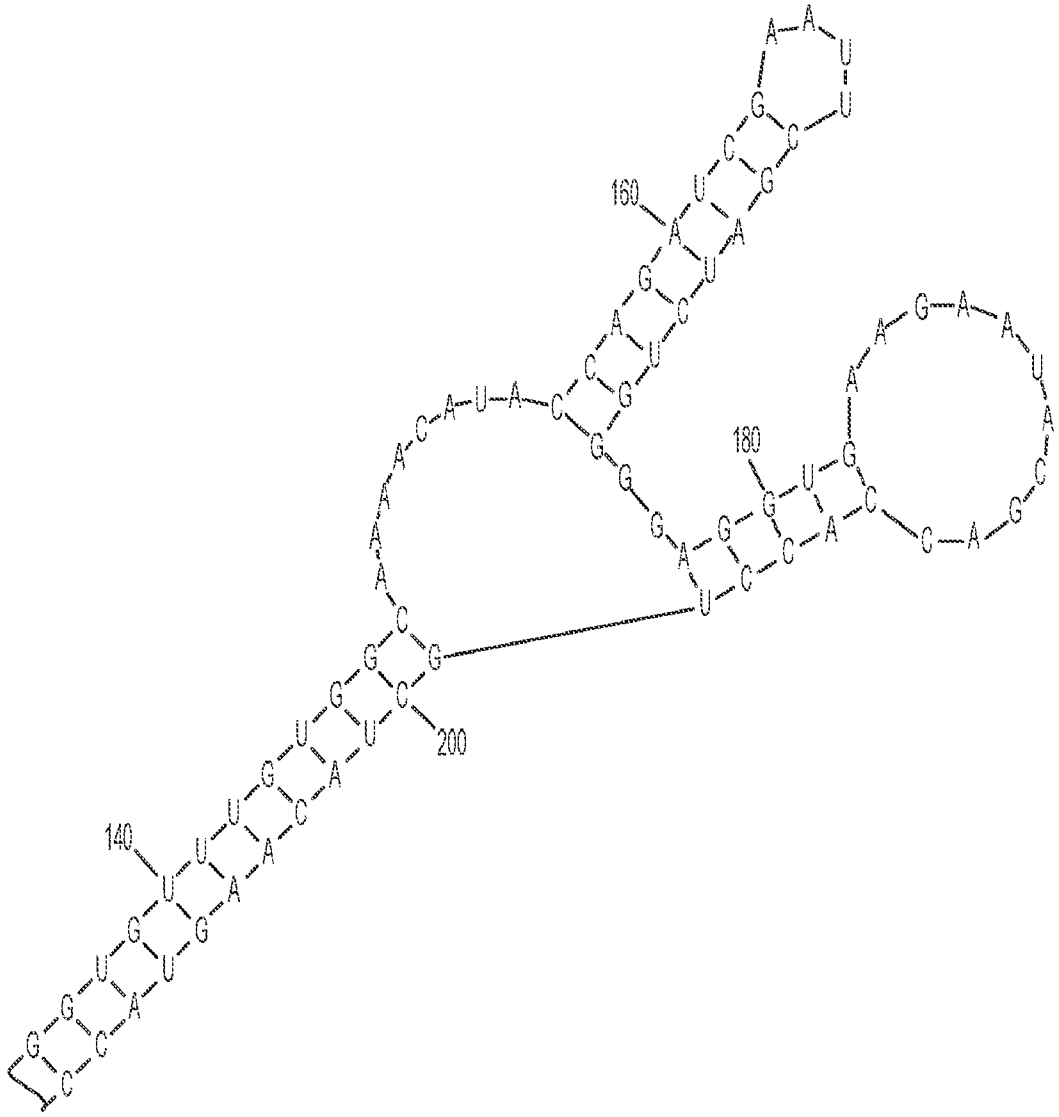


Figure 15C
CONTINUED

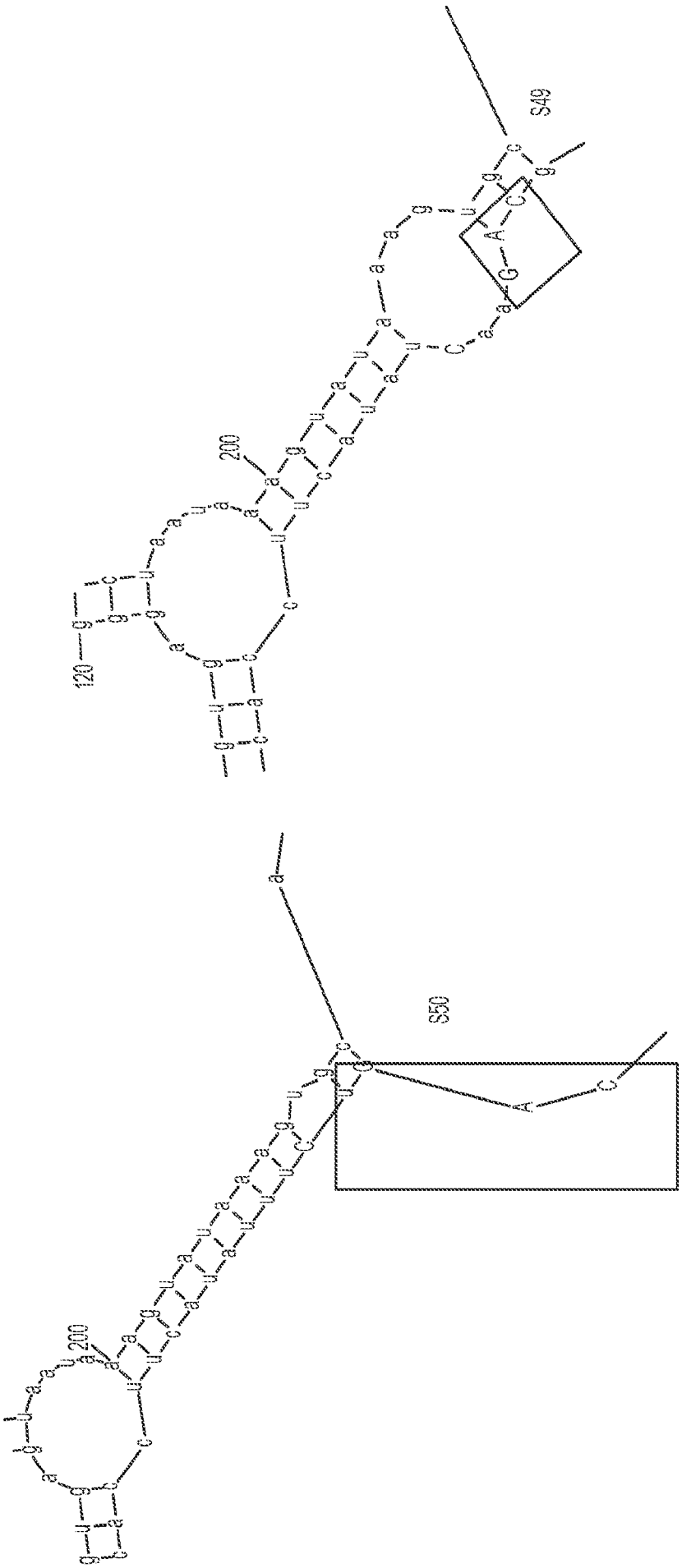


Figure 15D

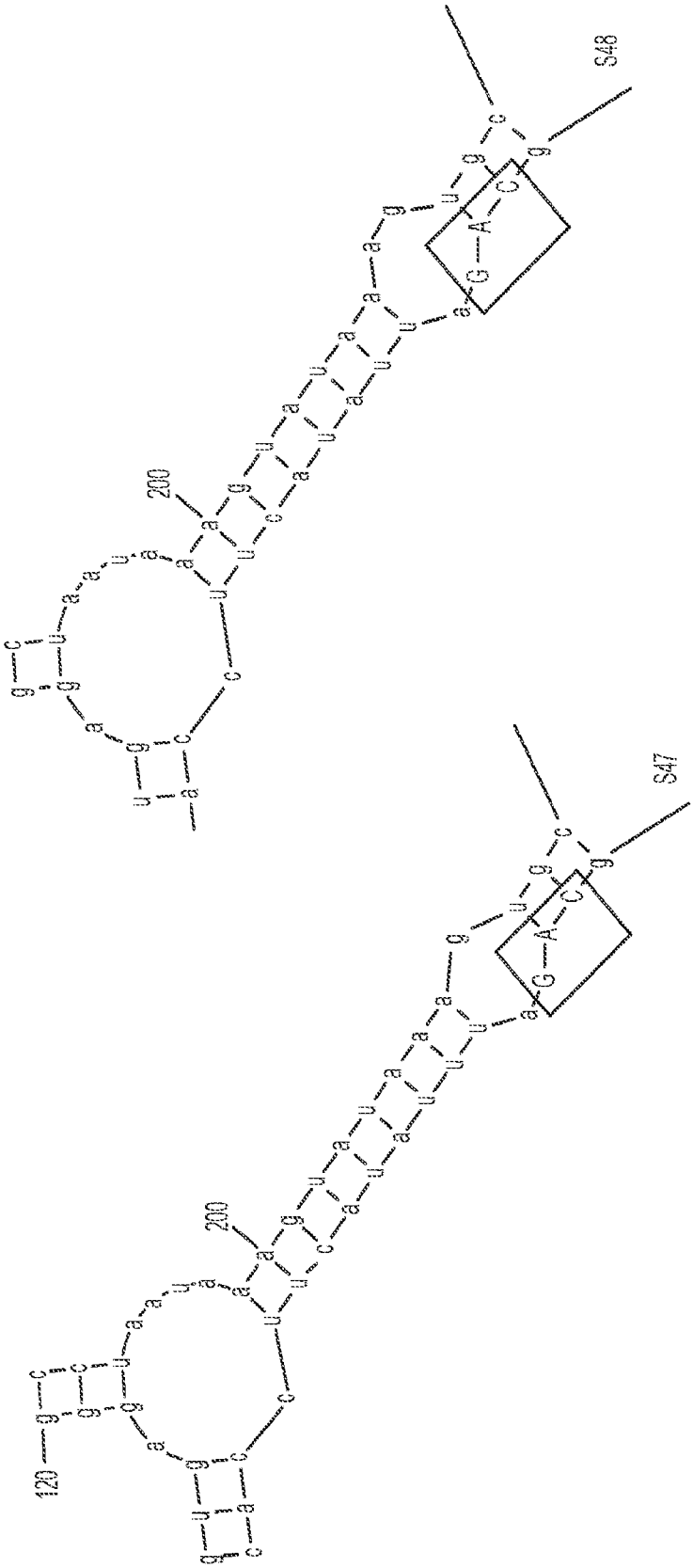


Figure 15D
CONTINUED

	S19	S47	S48	S49	S50
Induction (1 ug/mL)	65	15	29	22	36
	S19	S47	S48	S49	S50
Induction (5ug/mL)	112	172	164	134	41

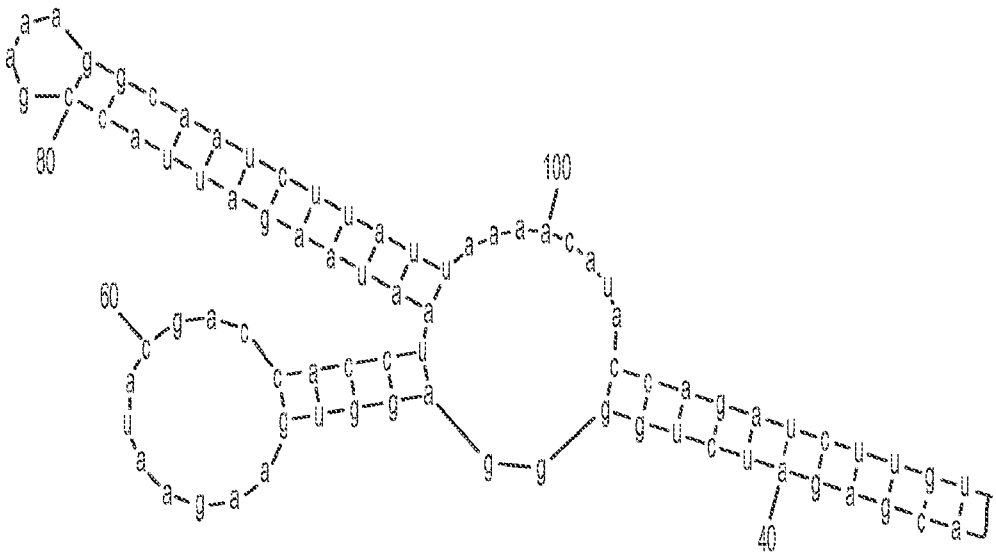


Figure 15D
CONTINUED

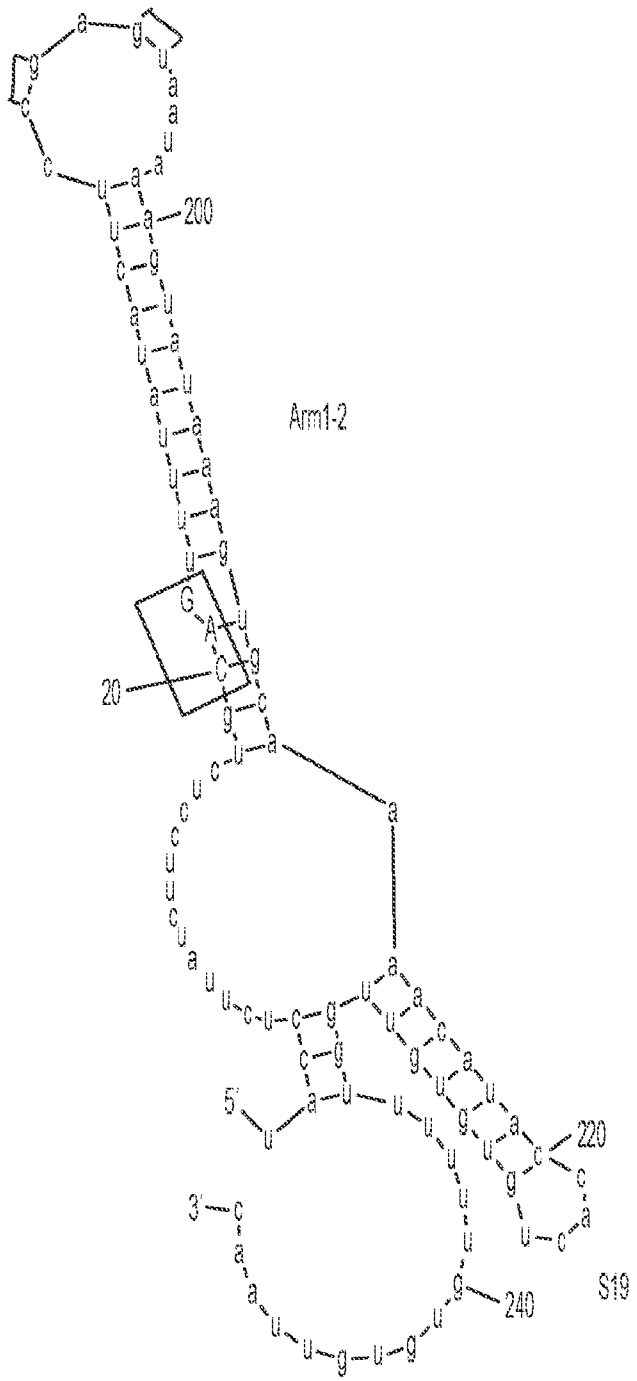


Figure 15D
CONTINUED

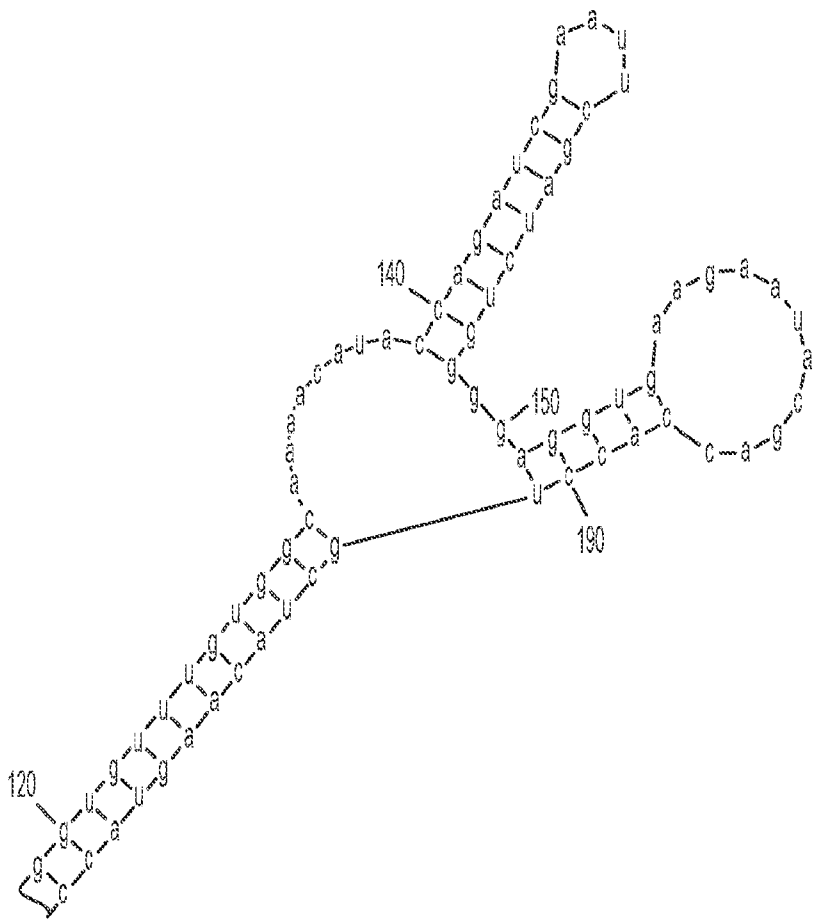


Figure 15D
CONTINUED

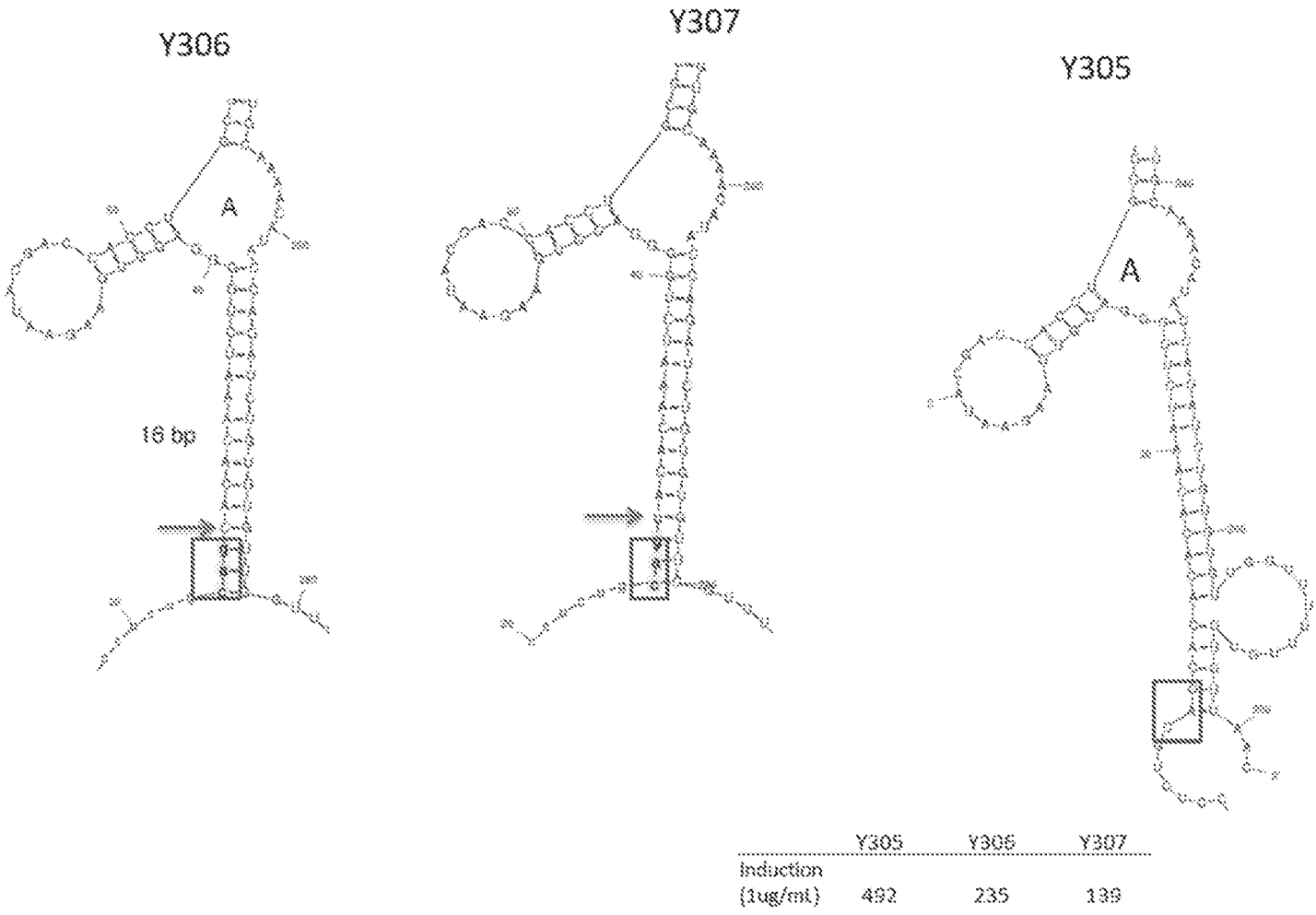
S20 to S22 are based on S9m. The predicted splicing strength is indicated by the numbers in blue:
S9m: CACACAGAATACG: 0.76
S20: CACACAGGATACG: 0.88
S21: CACACAGCATACG: 0.46
S22: CACACAGATTACG: 0.92
S23 to S24 are based on Y196CAA-4MAZ, the predicted splicing strength indicated by the numbers in blue:
Y196: ttatctttcctCCCACAGCTCCTGGGCAACgtgctgtctgtgctggcccatcactttggcaaagaattgggctagc 0.99
S23: ttatctttcctCACACAGAATACGACCgtctgtgctggcccatcactttggcaaagaattggctagc: 0.89
S24: ttatctttcctCACACAGCATACGACCgtctgtgctggcccatcactttggcaaagaattggctagc: 0.74

Y196CAA-4MAZ-e	S9m-e	S20-e	S21-e	S22-e	S23-e	S24-e
Induction (5ug/mL)	182	197	178	202	171	191	195

Figure 15E

Figure 15F

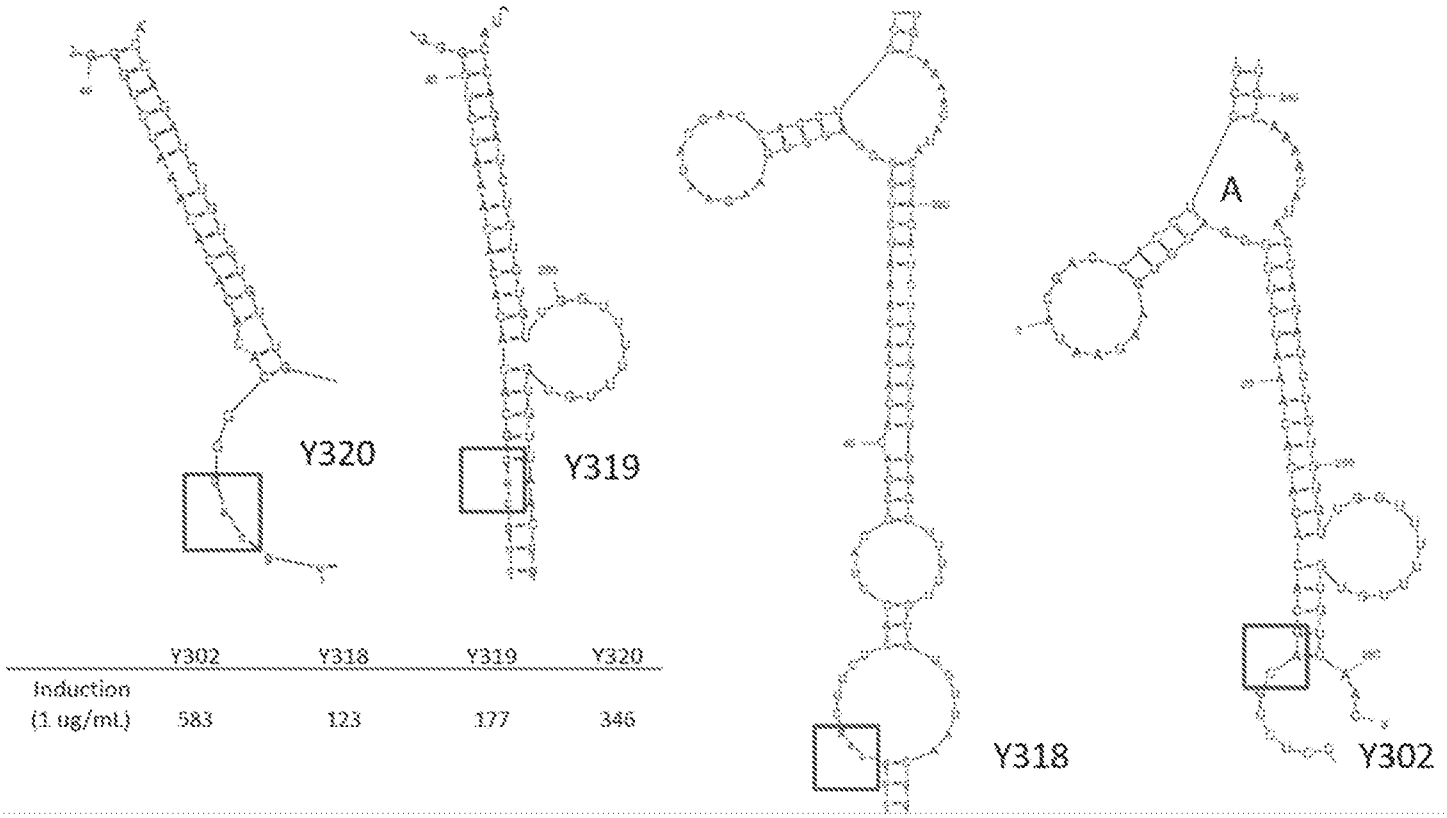
Y306: ctg**cag**CA**CA**CAAATCTGGGGA
Y307: ctg**cag**TA**CA**CAAATCTGGGGA



	Y305	Y306	Y307
Induction (1ug/ml)	492	235	139

Figure 15F (continued)

Y318: ctg**cag**CTCCTCACCTACGC**CACACACAC**AAATCTGGGGA
Y319: ctg**cagGCACACACAC**AAATCTGGGGA
Y320: ctg**cagGGCACACACAC**AAATCTGGGGA



	Y302	Y318	Y319	Y320
Induction (1 ug/mL)	583	123	177	346

Min-IVS2 Intron 120 bases:
GTgagtcctatgccagctaccattcctgcttttatttatggttgggataaggctggattattctgagtccaagctaggcccttttgctaaatcatCttcatacctctta
tcttcctctg**CAGNNN**aca

Best NNN sequences identified by testing:
Y344-based: Y359 (CAT), Y360 (TTT), Y361 (TGA), Y362 (TCT)
Y358-based: Y363 (CAT), Y366 (TAC), Y367 (TTT)

	Y344-based					Y358-based			
	Y344	Y359	Y360	Y361	Y362	Y358	Y363	Y366	Y367
Induction (0.5ug/mL)	200	171	186	156	204	144	172	233	140
	Y344	Y359	Y360	Y361	Y362	Y358	Y363	Y366	Y367
Induction (1ug/mL)	397	445	350	385	688	503	315	574	605

Figure 15G

S10: Insert a branch point: caacaac**TACTAAC**ataacagtgttcactagcaacctcaaacagacaccATG

S11: Insert a PPT: caacaac**TCCTGTGCTT**ataacagtgttcactagcaacctcaaacagacaccATG

S12: Mutate CAG to CCC: caacaacataa**CCC**tgttcactagcaacctcaaacagacaccATG

S13: mutate CAG to AAG: caacaacataa**AAG**tgttcactagcaacctcaaacagacaccATG

	Y196CAA	S10	S11	S12	S13
Induction (5ug/mL)	620	277	73	157	161

Figure 16A

5'UTR sequence: gcggccgccttaattaacagtgttcactagNNNcaacaacaacaacaacaacaacaacaacgacaccATG

Best NNN sequences identified by testing:

R1-4	TAG	R10-3	TAG
R2-4	ACA	R10-4	ACC
R4-1	AAC	R11-4	AGC
R5-2	TGC	R11-5	TGA
R5-5	TTG	R12-5	TGC
R6-1	ACC	R13-3	TCC
R6-2	CCC	R14-2	AGC
R6-3	AGC	R14-4	TGC
R7-3	AGG	R15-4	CCT
		R15-5	TCT

Induction in fold, 1ug/mL

Y305	R1-4	R2-4	R4-1	R5-2	R5-5	R6-1	R6-2
293	311	158	432	242	230	187	329
Y305	R6-3	R7-3	R10-4	R10-5	R11-5	R12-1	R12-5
216	224	195	248	185	191	162	223
Y305	R14-2	R14-4	R15-5				
250	215	173	239				

Y305R6-3 renamed as Y358

Example of NNN sequences that lead to much reduced induction:

		GGC	GTG	GTC
GGC, GTG, GTC	Y305	NR7-1	NR8-1	NR8-2
Induction (1ug/mL)	433	49	58	64

Figure 16B

inserting the best tagNNN in 5'UTR tested in Y305 into Y329

Parental sequence:
Y329 5'UTR: GcggccgcCataacagtgttcactagCATccCccagACCATCTACCACCGACACC**ATG**
Best tagNNN tested in Y305 backbone: TAG, AAC, ACC, CCC, AGC, CCT, TCT

Y341 (identical to Y329 but with NNN= TAG)
gcggccgcCataacagtgttcactagTAGccCccagACCATCTACCACCGACACC**ATG**

Y342 (identical to Y329 but with NNN= AAC)
gcggccgcCataacagtgttcactagAACccCccagACCATCTACCACCGACACC**ATG**

Y343 (identical to Y329 but with NNN=ACC)
gcggccgcCataacagtgttcactagACCccCccagACCATCTACCACCGACACC**ATG**

Y344 (identical to Y329 but with NNN=CCC)
gcggccgcCataacagtgttcactagCCCccCccagACCATCTACCACCGACACC**ATG**

Y345 (identical to Y329 but with NNN= AGC)
gcggccgcCataacagtgttcactagAGCccCccagACCATCTACCACCGACACC**ATG**

Y346 (identical to Y329 but with NNN=CCT)
gcggccgcCataacagtgttcactagCCTccCccagACCATCTACCACCGACACC**ATG**

Y347 (identical to Y329 but with NNN= TCT)
gcggccgcCataacagtgttcactagTCTccCccagACCATCTACCACCGACACC**ATG**

	Y305	Y341	Y342	Y343	Y344	Y345	Y346	Y347
Induction (1 ug/mL)	276	332	406	481	525	246	231	74

Figure 16C

S159 to S163: Shorten the IVS2 intron length of S56. S56 contains the original full-length IVS2 intron of 476 bases.

S159: Intron Size 120 bases

GTgagtctatgccagctaccattctgcttttatttttatggttgggataaggctggattattctgagtccaagctaggccctttgctaaatcatCttcatacctcttatcttcctc tgCAG

S160: Intron Size 140 bases

GTgagtctatgttgctaatagcagctacaatccagctaccattctgcttttatttttatggttgggataaggctggattattctgagtccaagctaggccctttgctaaatc atCttcatacctcttatcttcctctgCAG

S161: Intron Size 160 bases

GTgagtctatgctgatgtaagaggtttcatattgctaatagcagctacaaatccagctaccattctgcttttatttttatggttgggataaggctggattattctgagtccaa gctaggccctttgctaatcatCttcatacctcttatcttcctctgCAG

S162: Intron Size 180 bases

GTgagtctatgtttctgcatataaattgtaactgatgtaagaggtttcatattgctaatagcagctaccaatccagctaccattctgcttttatttttatggttgggataaggct ggattattctgagtccaagctaggccctttgctaatcatCttcatacctcttatcttcctctgCAG

S163: Intron Size 200 bases

GTgagtctatgaatatttctgcatataaatatttctgcatataaattgtaactgatgtaagaggtttcatattgctaatagcagctacaatccagctaccattctgcttttatt ttatggttgggataaggctggattattctgagtccaagctaggccctttgctaatcatCttcatacctcttatcttcctctgCAG

S164 to S168: similar to S159-S163 but have a branch point TACTAAC inserted at the same location before IVS2 3'ss. The intron sequence of S164 is shown as an example:

GTgagtctatgccagctaccattctgcttttatttttatggttgggataaggctggattattctgagtccaagctaggccctttgctaatcatCttcaTACTAACct cttatcttcctctgCAG

S169 to S173: similar to S159-S163 but have a branch point TACTAAC and one more 3'ss CAG inserted at the same location before IVS2 3'ss. The intron sequence of S169 is shown as an example:

GTgagtctatgccagctaccattctgcttttatttttatggttgggataaggctggattattctgagtccaagcTACTAACttttcctgtgcttcttcagacctcttatC ttcctctgCAG

Figure 17A

	S56-	S159	S160	S161	S162	S163
Induction (5ug/mL)	315	181	150	122	114	119

	S56	S164	S166	S168	S169	S170	S171	S172	S173
Induction (5ug/mL)	267	69	45	25	26	12	14	10	11

Figure 17A
CONTINUED

Previously we have showed that CAA repeats may alter the splicing strength of the 3'ss in 5'UTR. Here the CAA repeats (in red) are to be removed from S159, S164, and S169:

Gcggccgc[**caacaacaacaacaacaacaacaacaacaacaacaa**]cataa**cag**tgttcactagcaacctcaaacagacaccATG

	S56	S192	S193	S194
Induction (5ug/mL)	116	180	58	20

	S56	S192	S193	S194
Induction (5ug/mL)	298	279	86	30

Figure 17B

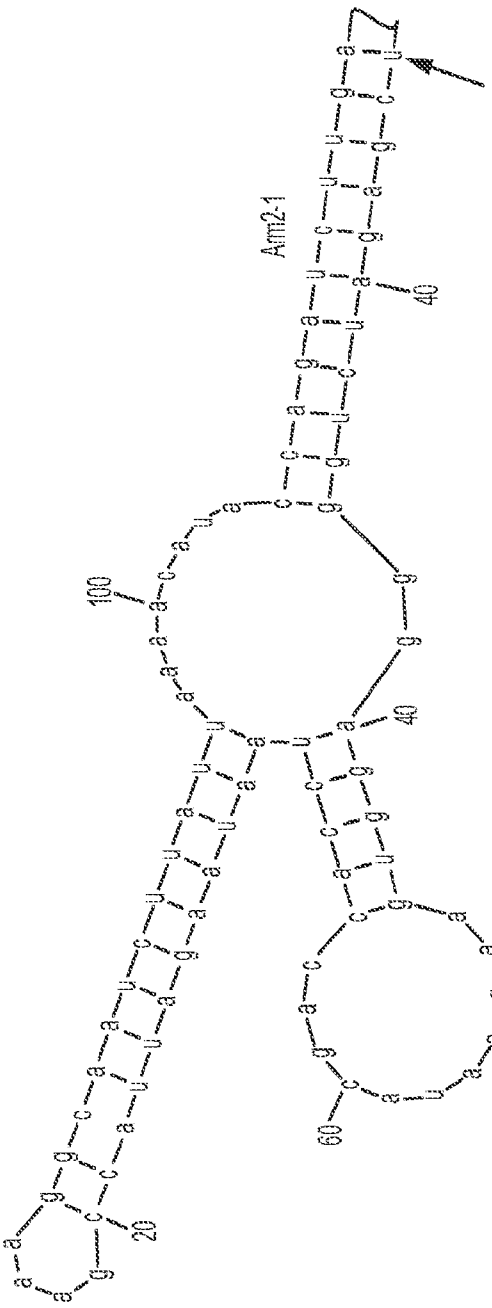
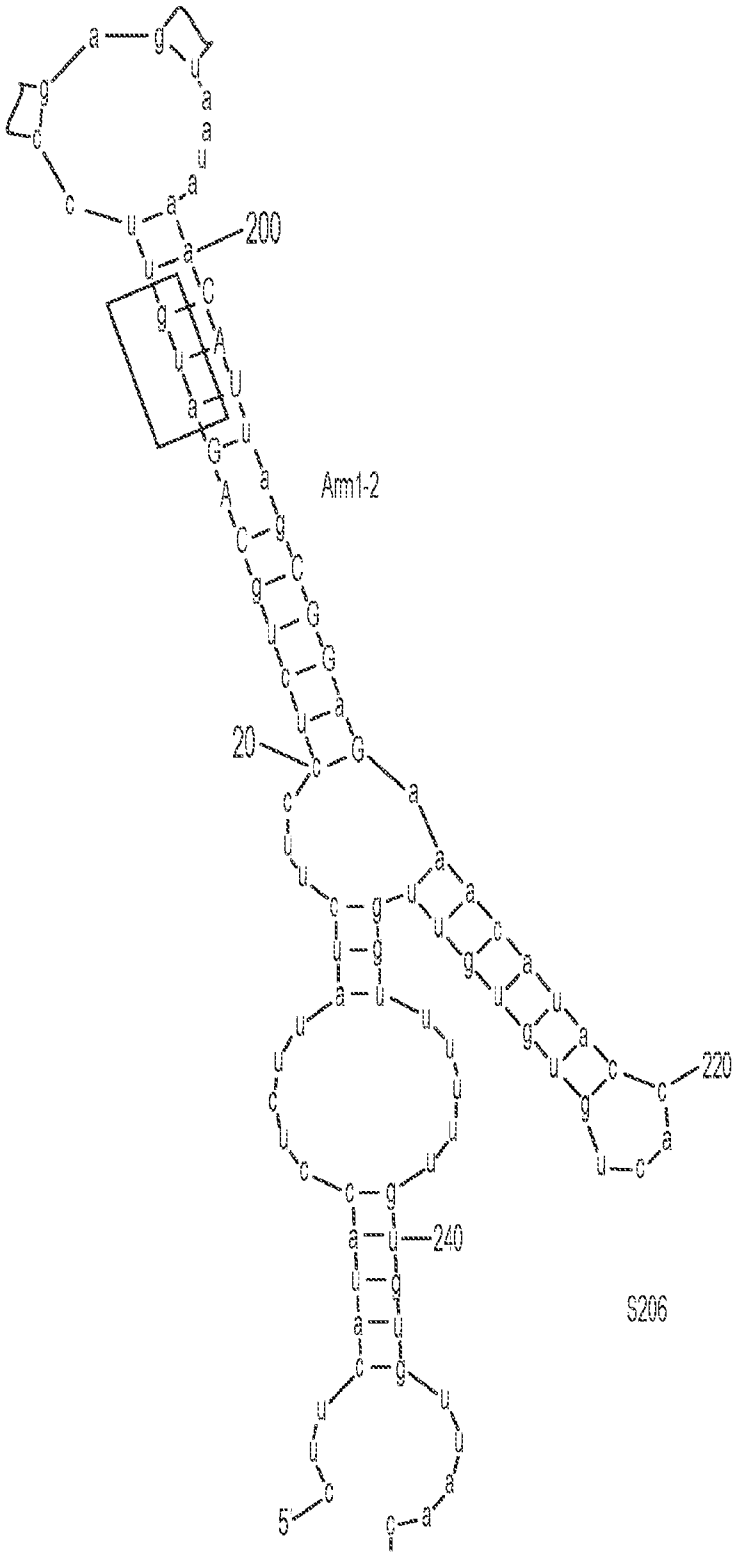


Figure 18A



	S56	S192	S206
Induction (5ug/mL)	211	319	779

Figure 18A
CONTINUED

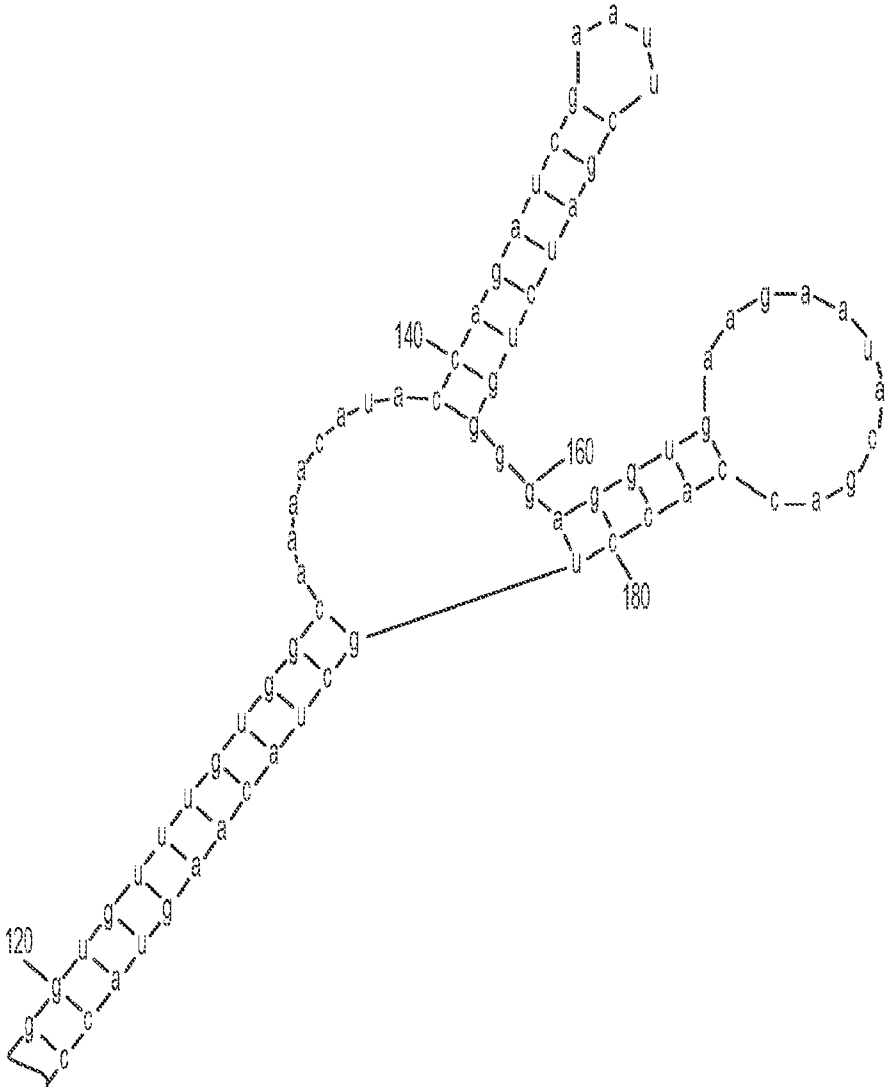


Figure 18A
CONTINUED

S206: gcggccgcc**ata**acagtgttcac**tag**caaccTcaaa**cag**acacc**ATG**

S209: gcggcc**gcCACCATG**ataacagtgttcac**tag**caacctaaa**cag**acacc**ATG**

S210: gcggccgccataacagtgttcac**tag**caaccCcaaa**cag**acacc**ATG**

S211: gcggcc**gcCACC**ataacagtgttcac**tag**caaccCcaaa**cag**acacc**ATG**

S212: gaggcc**gcCACCATG**ataacagtgttcac**tag**caaccCcaaa**cag**acacc**ATG**

S213: gcggcc**gcCACCACG**ataacagtgttcac**tag**caaccCcaaa**cag**acacc**ATG**

S214: gcggcc**gcCACC**ataacagtgttcac**tag**caTccCcaaa**cag**acacc**ATG**

S215: gcggcc**gcCACC**ataacagtgttcac**Cag**caTccCcaaa**cag**acacc**ATG**

S222: gcggcc**gcc**ataacagtgttcac**tag**caTccCcaaa**cag**acacc**ATG**

S223: gcggcc**gcc**ataacagtgttcac**Cag**caTccCcaaa**cag**acacc**ATG**

	S192	S209	S210	S211	S212	S213	S214	S215	S222	S223
Induction (1ug/mL)	231	144	290	304	352	279	313	190	402	390
	S192	S209	S210	S211	S212	S213	S214	S215	S222	S223
Induction (5ug/mL)	330	545	730	624	554	644	696	488	538	759

Figure 18B

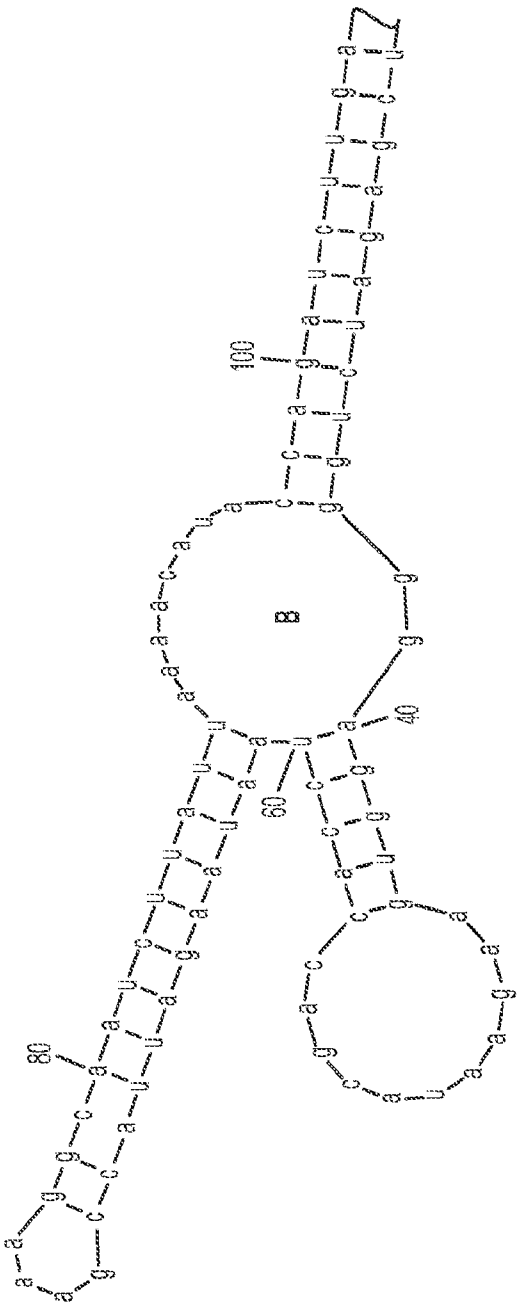


Figure 18C

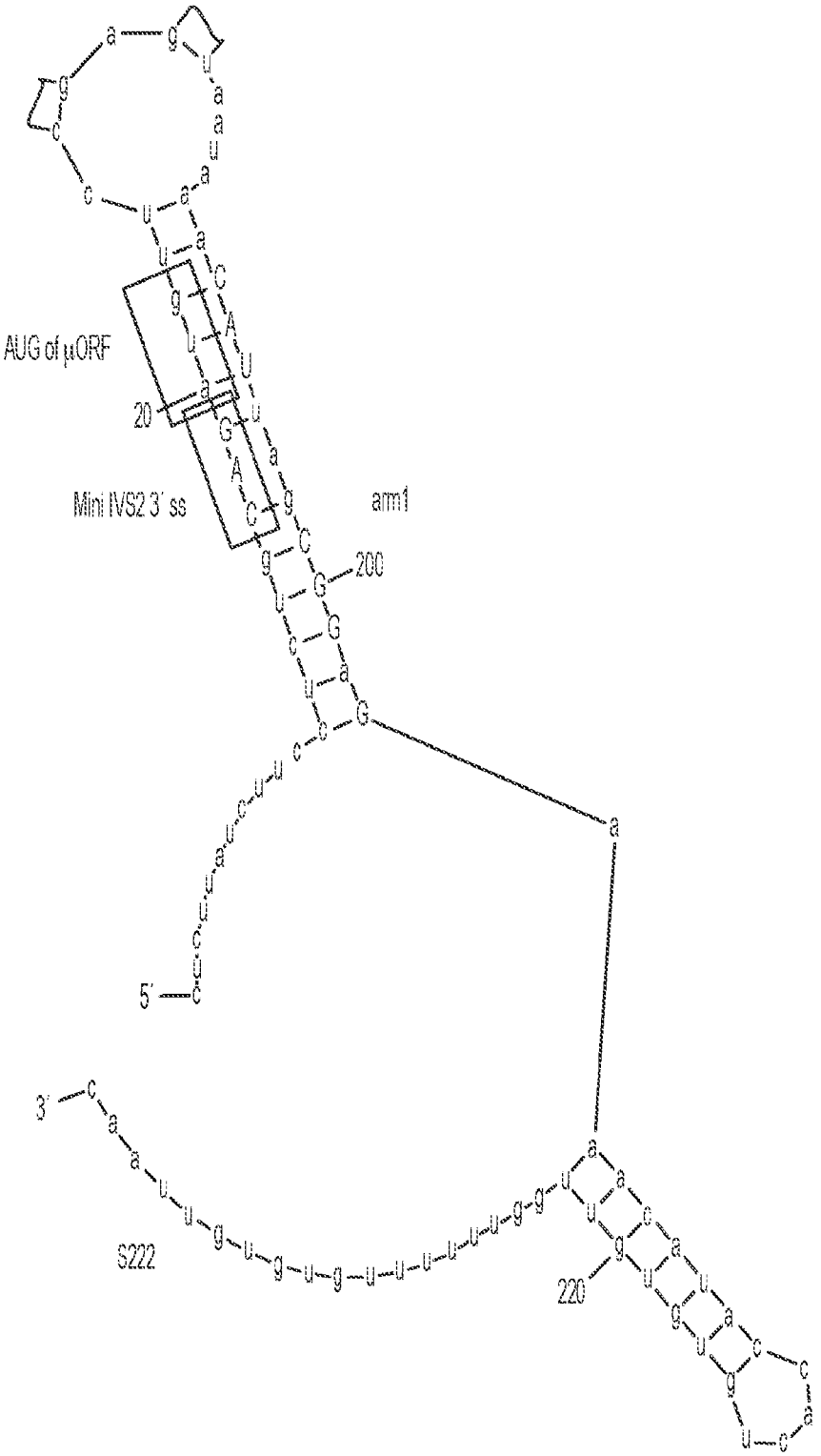


Figure 18C
CONTINUED

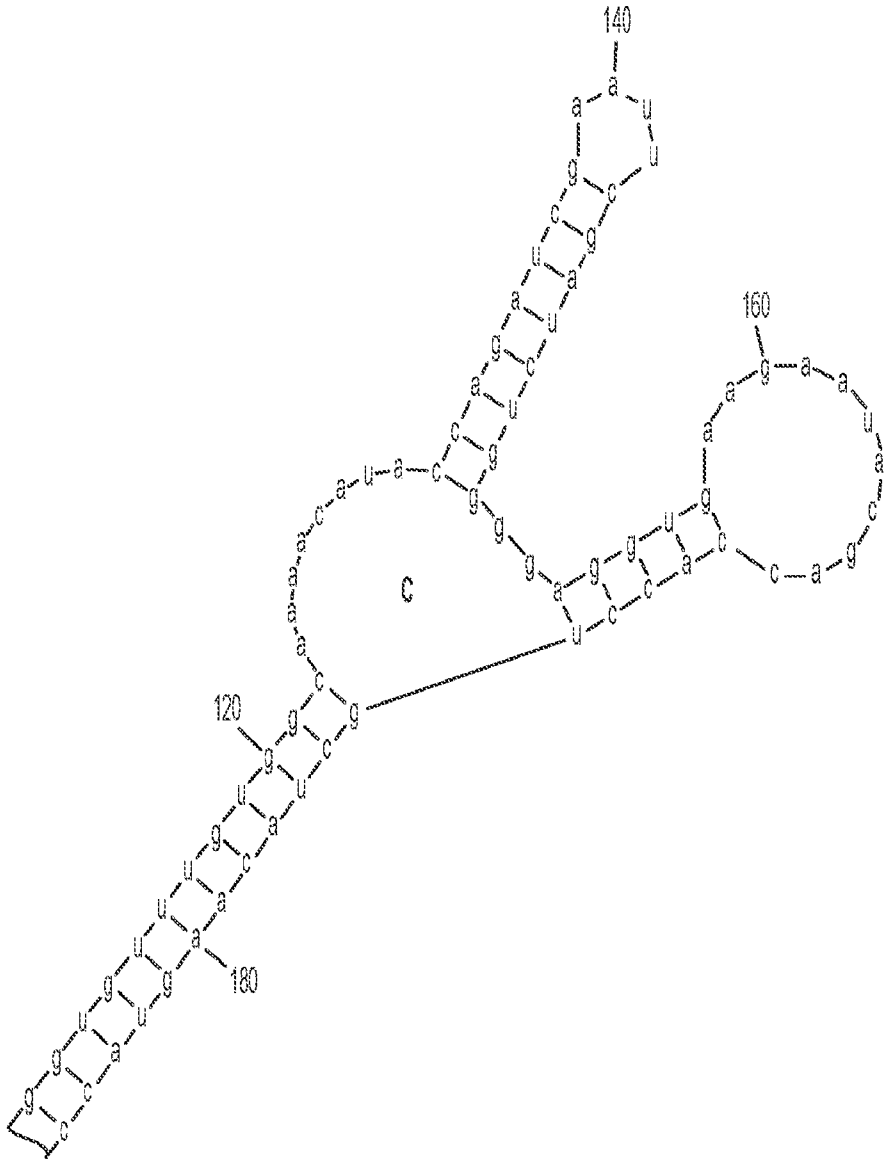


Figure 18C
CONTINUED

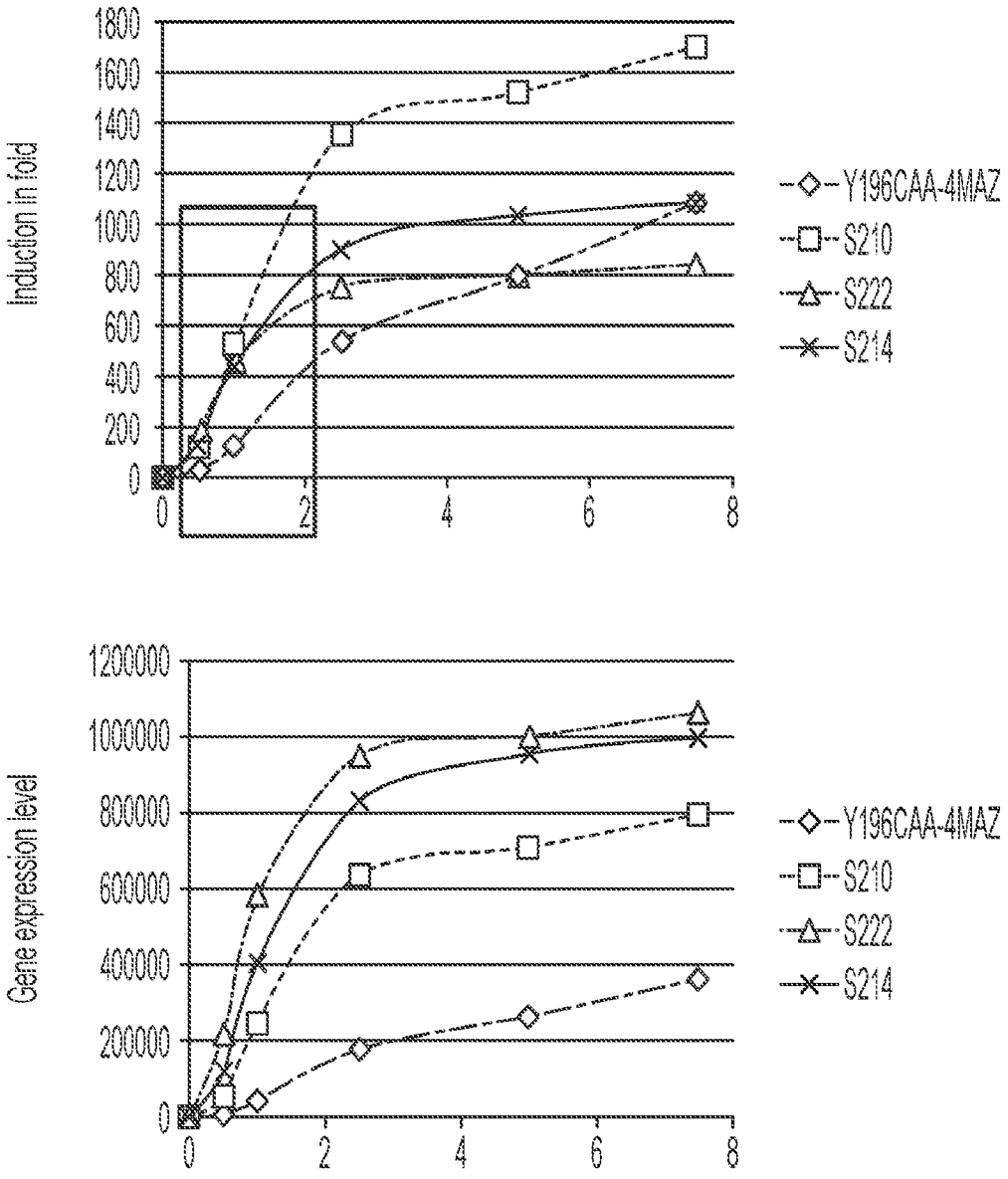

Induction in fold
1800
1600
1400
1200
1000
800
600
400
200
0
0
2
4
6
8
Y196CAA-4MAZ
S210
S222
S214
Gene expression level
1200000
1000000
800000
600000
400000
200000
0
0
2
4
6
8
Y196CAA-4MAZ
S210
S222
S214

Figure 19A

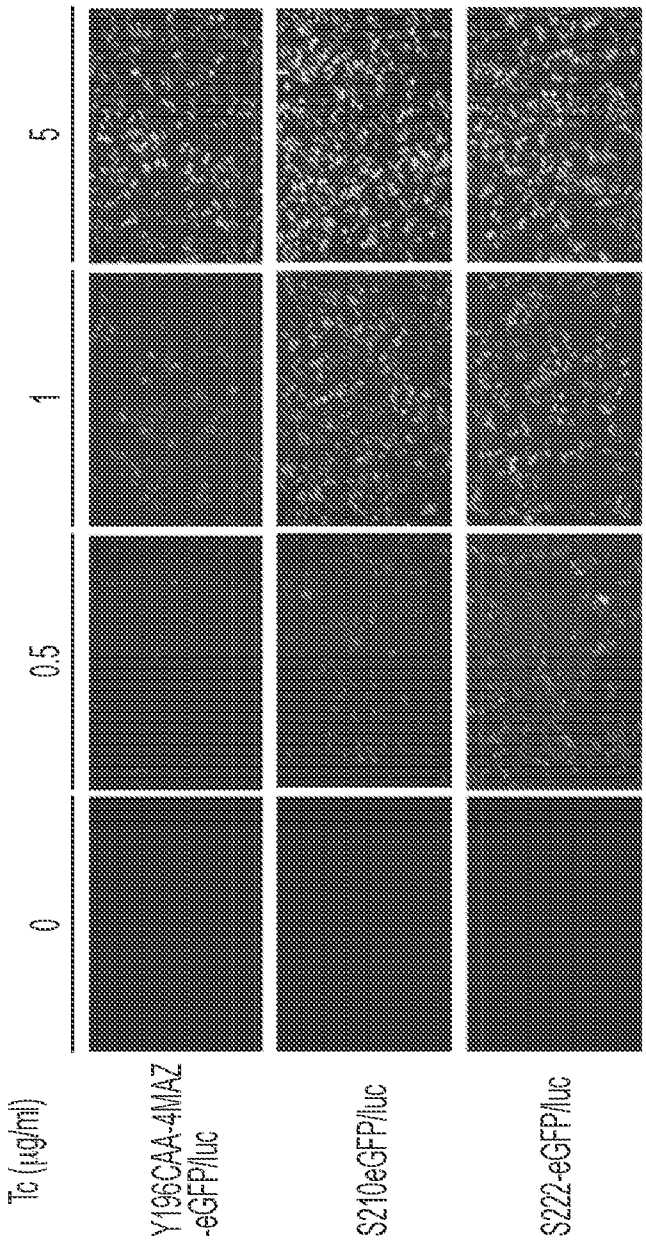
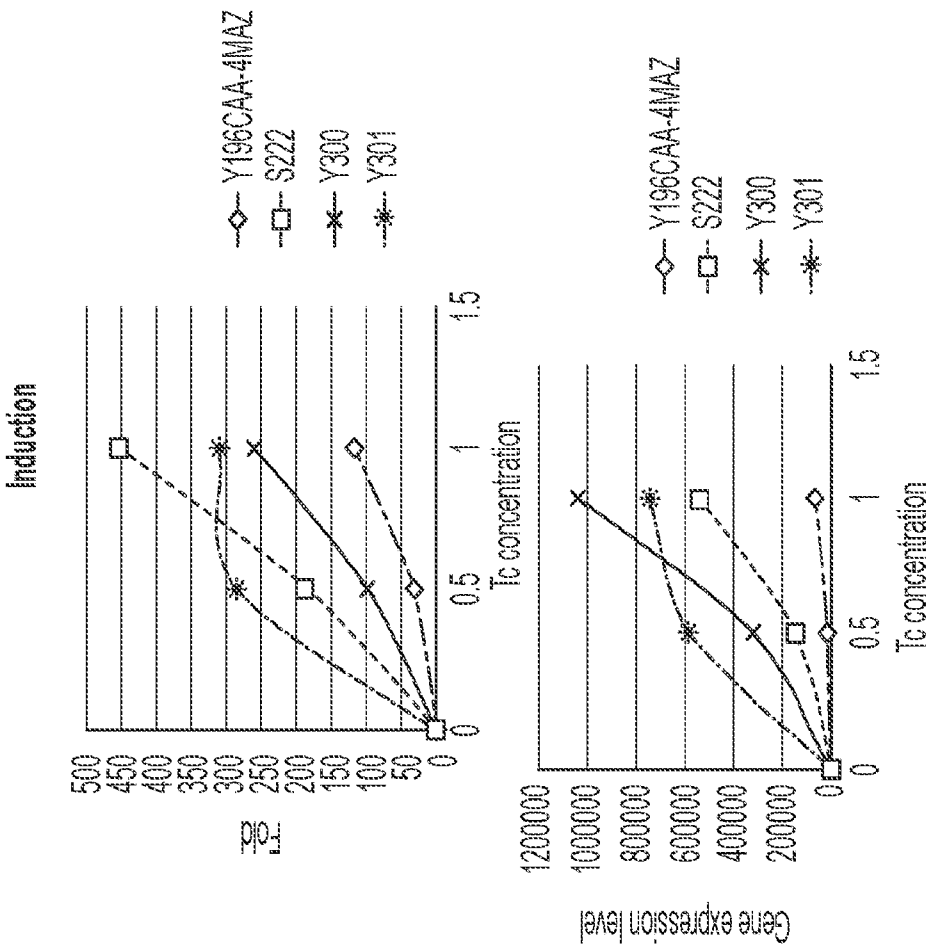

Tc (ug/ml)
0
0.5
1
5
Y196CAA-4MAZ -eGFP/luc
S210eGFP/luc
S222-eGFP/luc
0.5 ug DNA, 18 h after transfection

Figure 19B



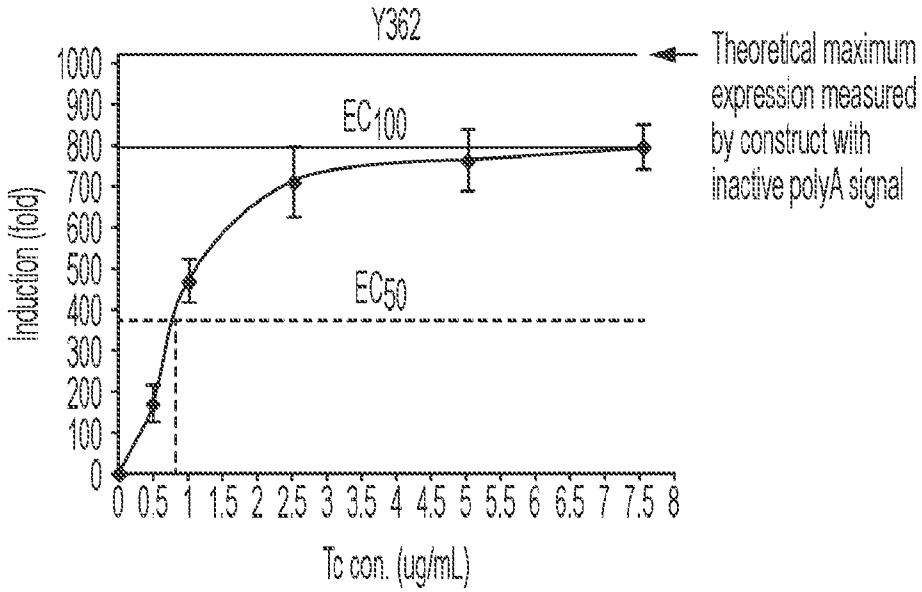
	Y196CAA-4MAZ	S222	Y300	Y301
Induction (0.5ug/mL)	33	187	100	285

	Y196CAA-4MAZ	S222	Y300	Y301
Induction (1ug/mL)	117	454	259	311

Figure 19C

Y362

Tc con.	0	0.5	1	2.5	5	7.5
Mean induction	1	172	471	713	766	797
std	0	44	54	86	77	54



Y367

Tc con.	0	0.5	1	2.5	5	7.5
Mean induction	1	138	496	917	1071	1147
std	0	36	96	161	153	147

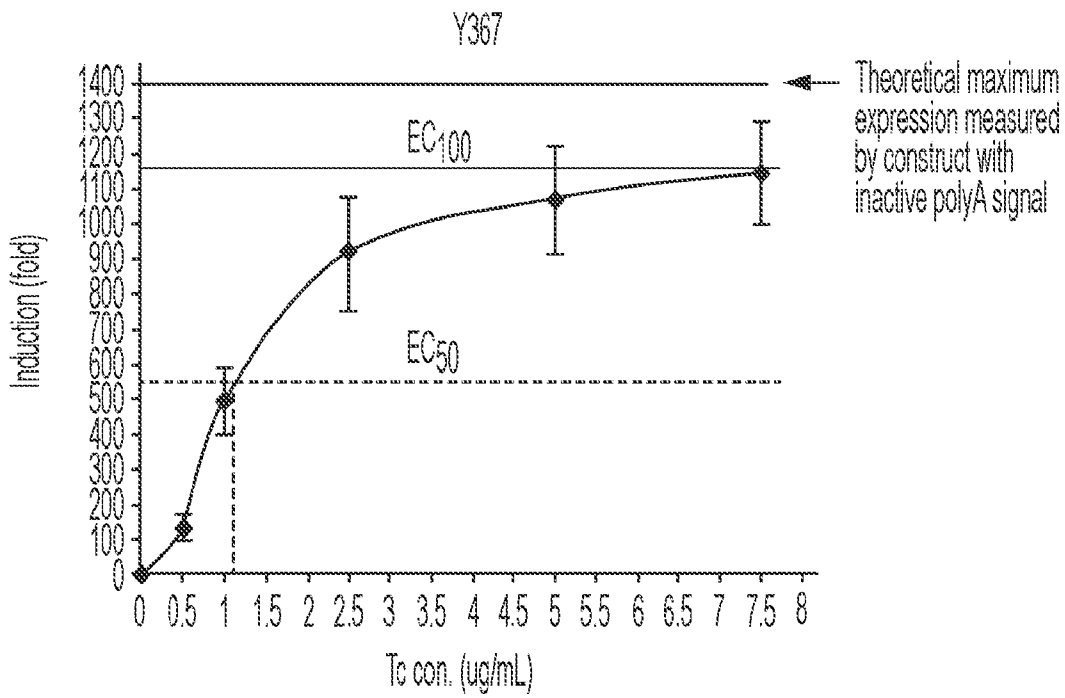


Figure 19D

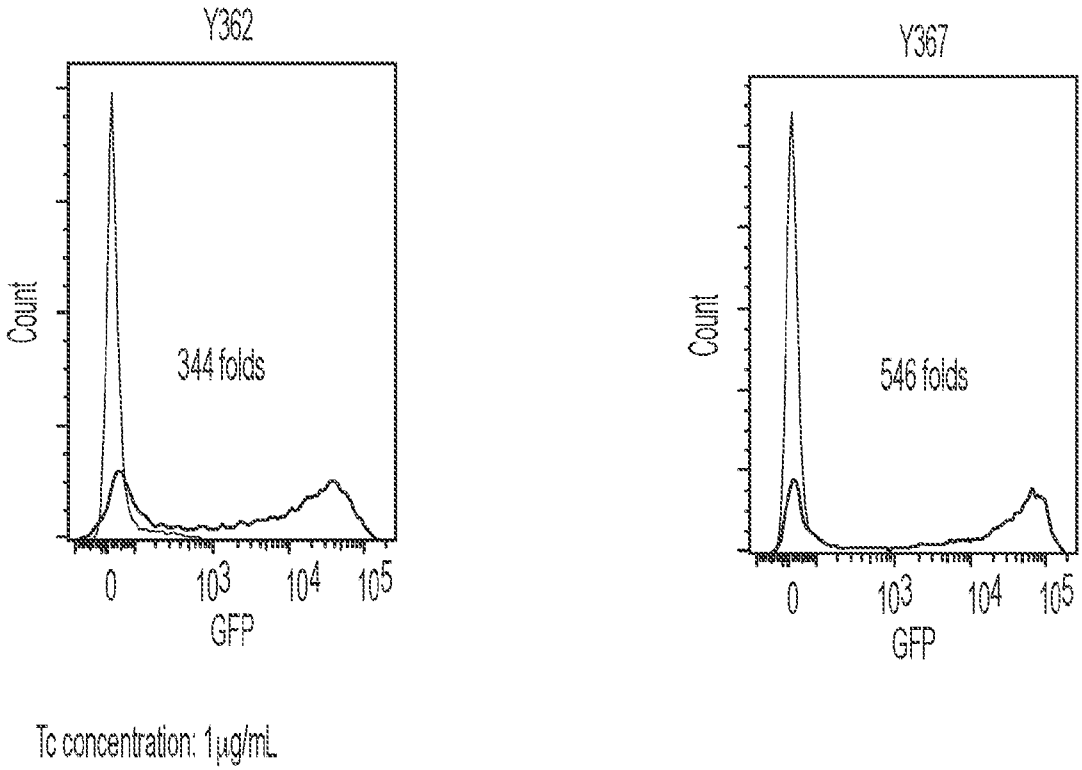

Y362
Count
344 folds
0
10^3
10^4
10^5
GFP
Y367
Count
546 folds
0
10^3
10^4
10^5
GFP
Tc concentration: 1µg/mL

Figure 19E

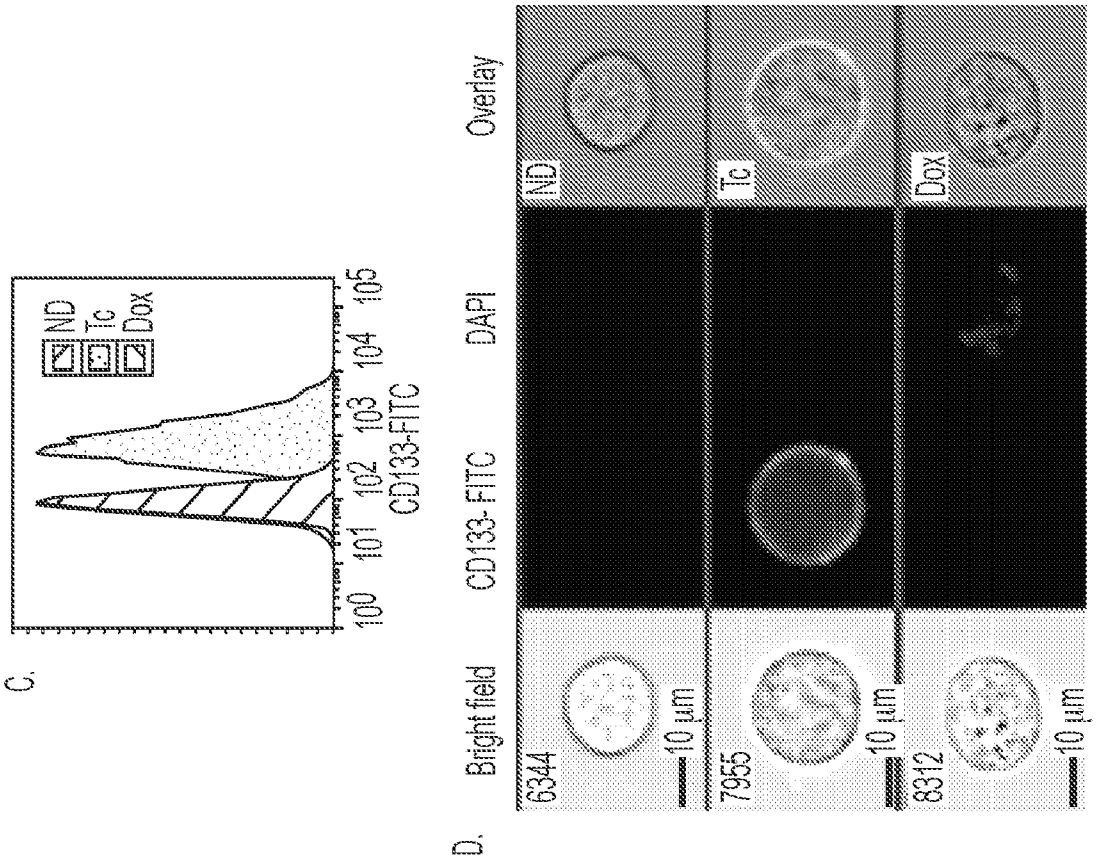

C.
ND
Tc
Dox
100 101 102 103 104 105
CD133-FITC
D.
Bright field
CD133- FITC
DAPI
Overlay
ND
Tc
Dox
6344
7955
8312
10 µm
A.
CD133 gene in human genome
5' arm
CMV
IVS2
CY196
g1
g4
g3
Intron 1
3' arm
CD133 repair matrix
B.
Parental cell
Riboswitch insertion
Engineered cell
+ Tc
Conditional gene expression from the endogenous CD133 gene

Figure 20

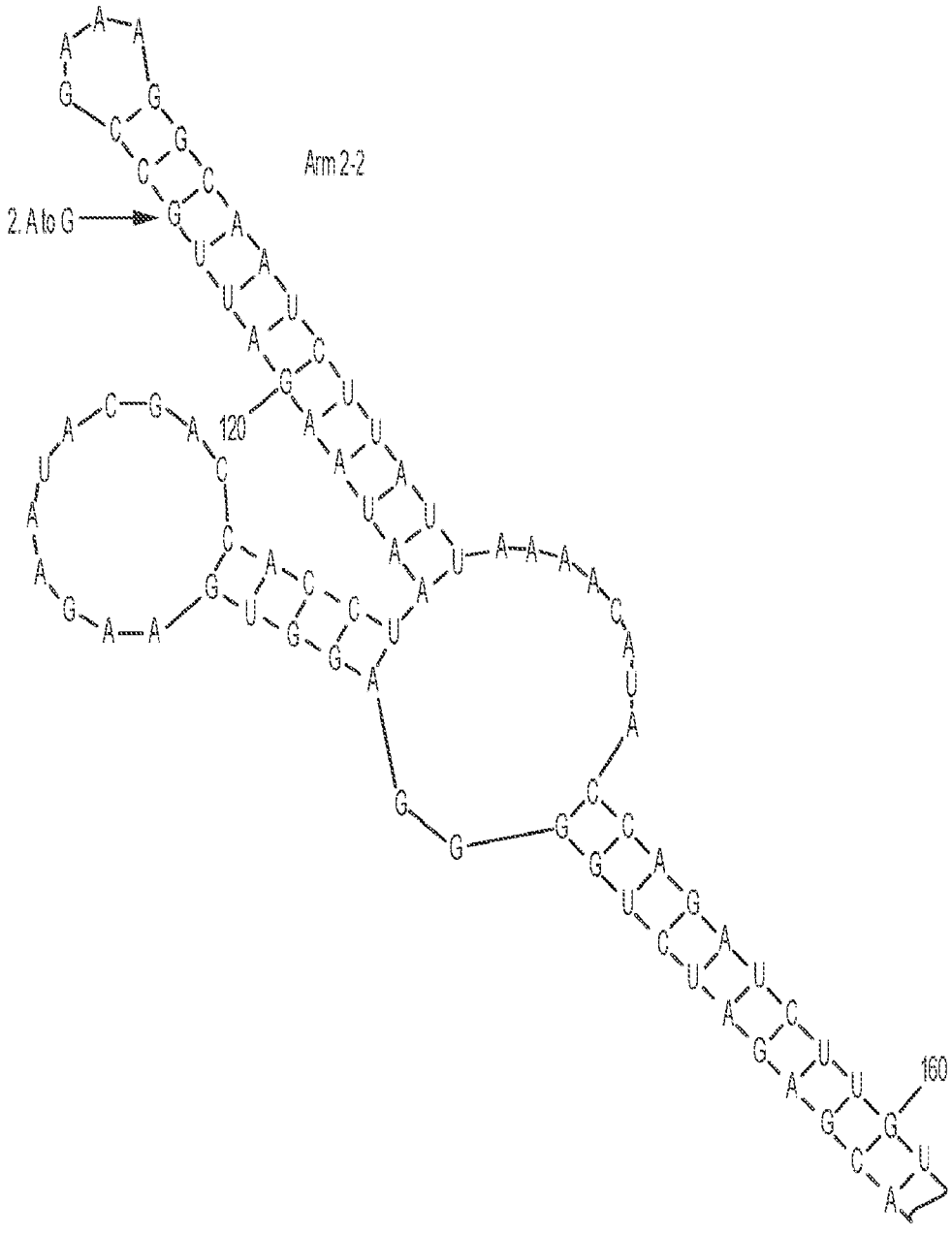


Figure 21

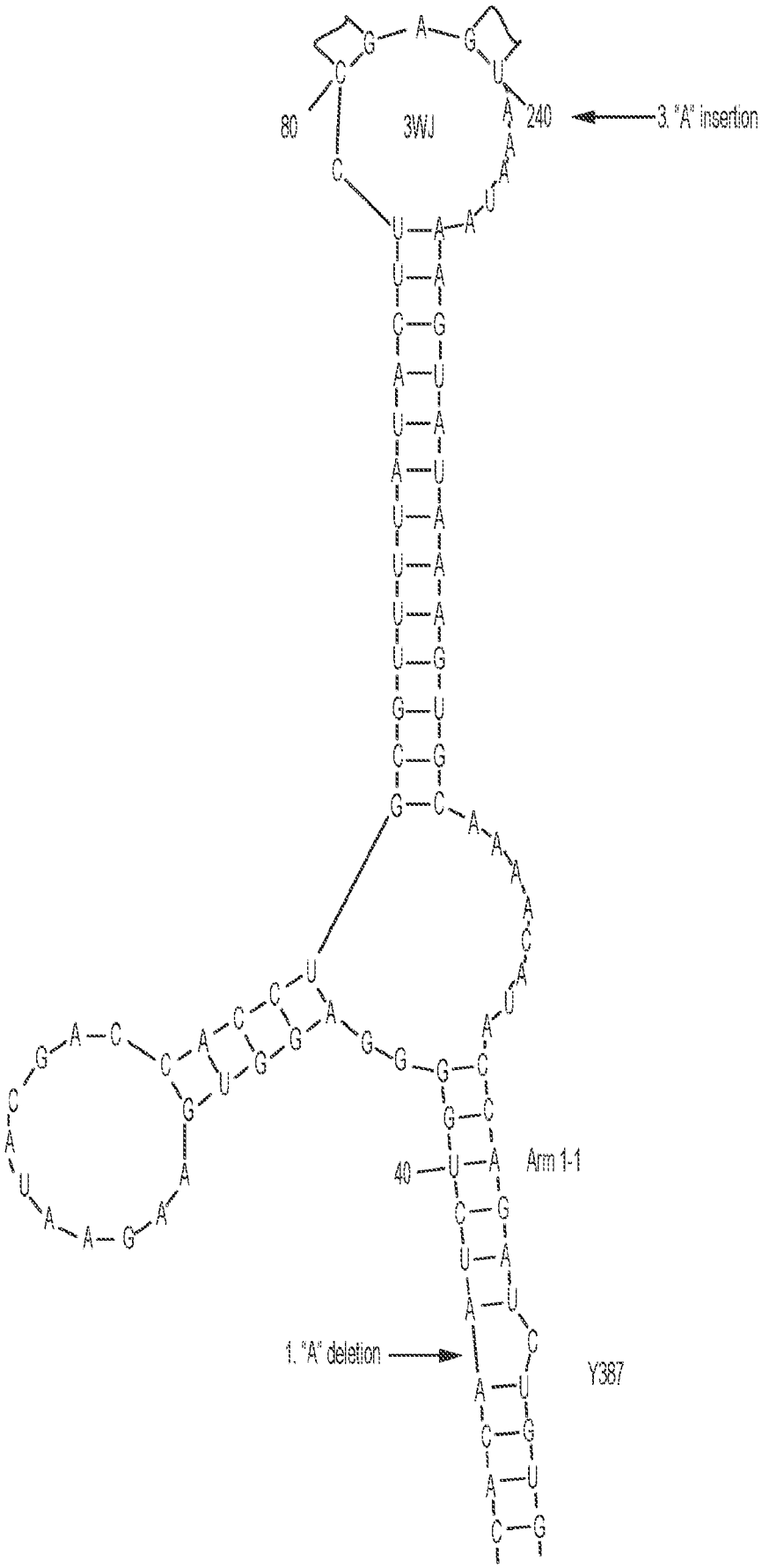


FIG. 21
CONTINUED

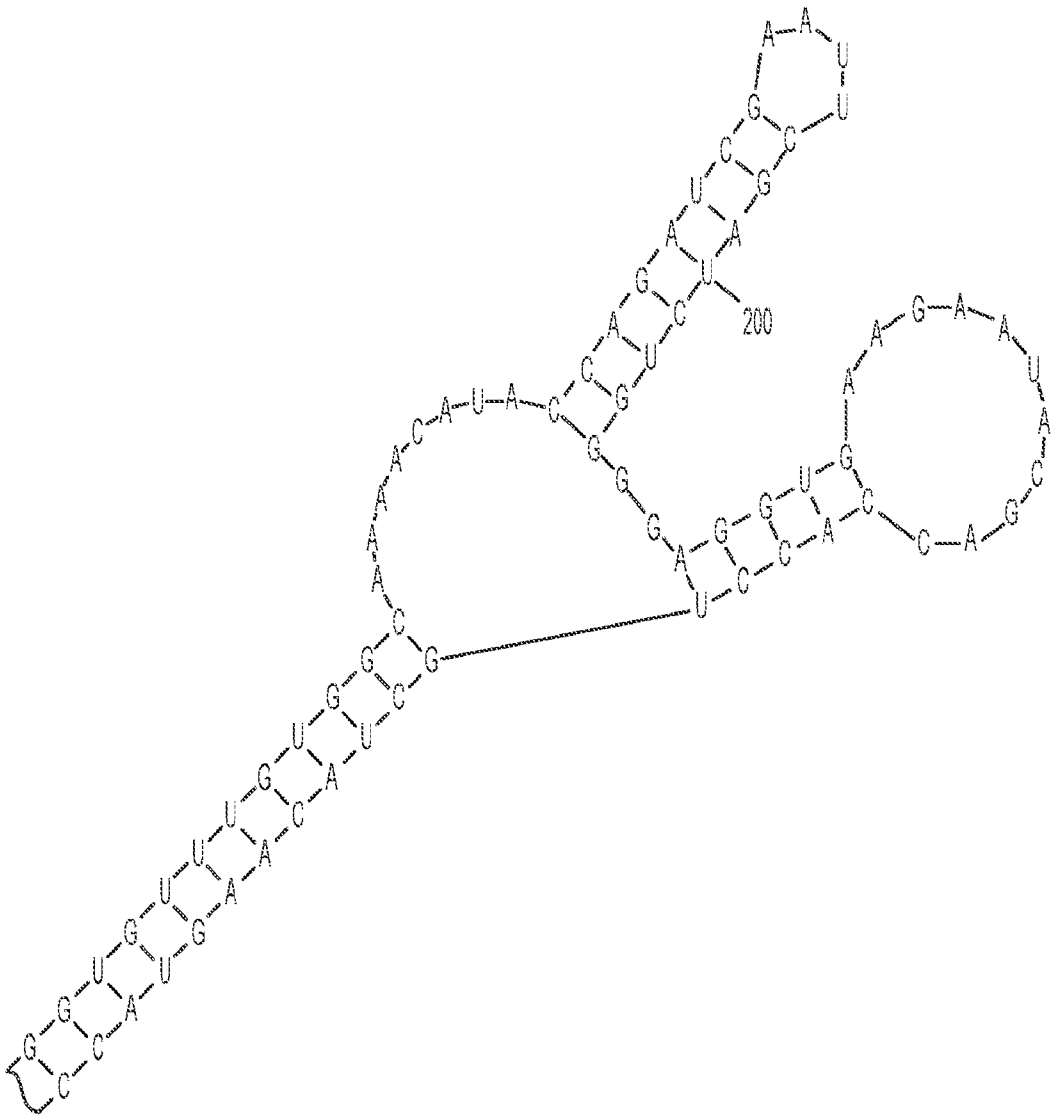


Figure 21
CONTINUED

	Y359	Y392	Y395	Y360	Y393	Y396	Y361	Y394	Y397	Y362C	Y362	Y387
Mean Induction	472	631	925	407	555	775	424	483	922	260	627	735
std	55	64	88	45	45	48	63	53	57	20	19	84

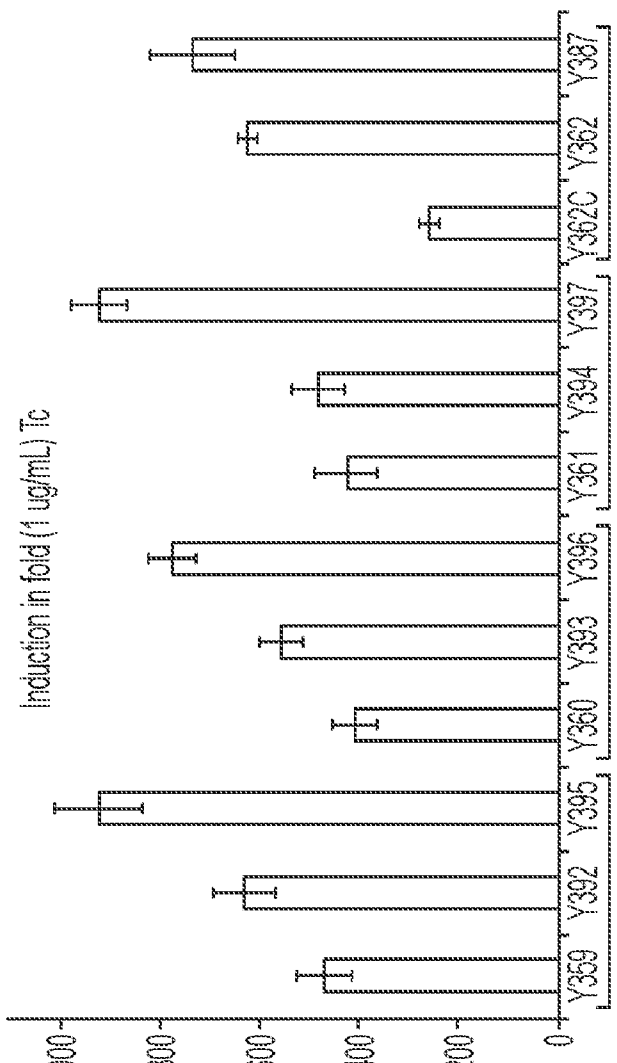


Figure 22

Tc [ug/ml]	0.0	0.5	1.0	2.5	5.0	7.5
Y362	1	332	654	918	896	891
Y387	1	464	710	854	835	798
std-Y362	0	45	73	139	173	158
std-Y387	0	27	38	74	108	90

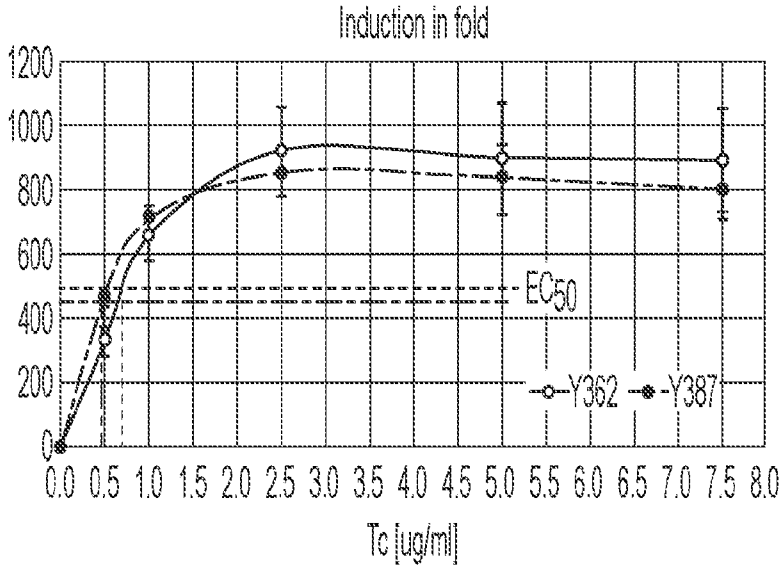


Figure 23A

Tc [ug/ml]	0.0	0.5	1.0	2.5	5.0	7.5
Y362	0.074%	23.3%	46.2%	65.1%	63.8%	65.5%
Y387	0.104%	47.5%	71.7%	87.0%	85.6%	83.3%
std-Y362	0.013%	3.8%	5.2%	6.1%	8.8%	8.9%
std-Y387	0.014%	7.5%	8.8%	10.8%	9.4%	8.5%

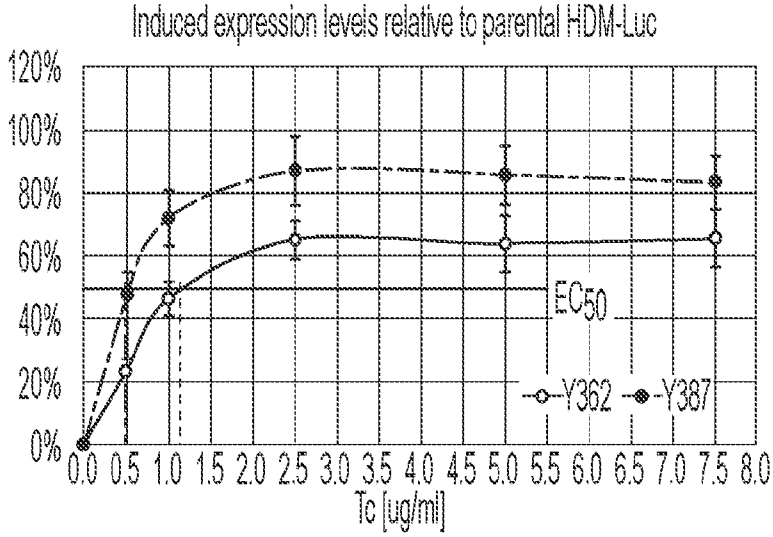


Figure 23B